Opérateurs et systèmes intelligents : se comprendre pour décider. Application à la supervision de drones.

Adrien Metge

▶ **To cite this version:**

Adrien Metge. Opérateurs et systèmes intelligents : se comprendre pour décider. Application à la supervision de drones.. Sciences cognitives. Université de Bordeaux, 2022. Français. NNT : 2022BORD0342 . tel-04033938

HAL Id: tel-04033938

https://theses.hal.science/tel-04033938

Submitted on 17 Mar 2023

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

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

université de BORDEAUX

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DES SCIENCES PHYSIQUES ET DE L'INGÉNIEUR

SPÉCIALITÉ INGÉNIERIE COGNITIQUE

Par Adrien METGE

Opérateurs et systèmes intelligents : se comprendre pour décider. Application à la supervision de drones.

Sous la direction conjointe de : Benoît LE BLANC & Nicolas MAILLE

Soutenue le 2 décembre 2022

Composition du jury :

M. Norbou BUCHLER, Chief networked systems branch – DEVCOM Analysis Center	Rapporteur
M. Julien CEGARRA, Professeur – Institut National Universitaire Champollion	Rapporteur
Mme Nelly CHOUVY, Responsable métier science de l'homme – DGA	Invitée
M. Bernard CLAVERIE, Professeur – Institut Polytechnique de Bordeaux	Président du jury
M. Benoît LE BLANC, Professeur – Institut Polytechnique de Bordeaux	Directeur
M. Nicolas MAILLE, Ingénieur de recherche – Onera	Examinateur
Mme Anne-Lise MARCHAND, Enseignante chercheuse – CREA	Examinatrice

Titre : Opérateurs et systèmes intelligents : se comprendre pour décider.

Application à la supervision de drones.

Résumé :

Cette thèse de doctorat présente un travail de recherche transdisciplinaire qui s'inscrit dans une démarche d'ingénierie cognitique. À l'intersection de l'Intelligence Artificielle (IA) et du Facteur Humain (FH), nous explorons les principes d'interaction qui permettent la mise en place d'une collaboration performante entre un utilisateur et un système de recommandation intelligent dans un contexte de prise de décision. Nous nous concentrons sur le cadre applicatif de la planification, pour lequel nous avons développé un environnement de simulation qui met en scène un opérateur aérien chargé de superviser un drone doté d'un haut niveau d'autonomie décisionnelle. Le scénario des missions, joué sur différents terrains, amène l'opérateur à déterminer un nouveau plan de vol assisté par des recommandations intelligentes. Une analyse de la littérature nous permet de caractériser trois problématiques scientifiques que nous investiguons à travers trois études expérimentales dans cet environnement.

La première problématique s'intéresse aux conséquences d'un changement de la participation de l'IA à la prise de décision sur le ressenti de l'opérateur. Nos résultats indiquent que lorsque l'IA réduit sa participation en arrêtant de proposer des suggestions de plan, le sentiment de l'opérateur d'être responsable et à l'origine de la solution augmente significativement. Cet effet se révèle néanmoins asymétrique. Lorsque l'IA accroît sa participation en introduisant des suggestions de plan qui n'étaient pas proposées dans un premier temps, le sentiment de l'opérateur d'être responsable et à l'origine de solution diminue peu alors même que les plans validés deviennent plus homogènes entre opérateurs, signe de la moindre implication des individus.

La deuxième problématique examine les liens potentiels entre les critères de compromis des plans construits avec l'IA par l'opérateur et les éléments de sa personnalité qui pourraient la prédire, sur des terrains où il n'existe pas de plan concevable qui ait une qualité totalement satisfaisante. Nous mettons en évidence

une préférence individuelle pour la dégradation d'un des trois critères de compromis du plan : certains participants préfèrent plutôt abandonner des cibles, d'autres consommer beaucoup de carburant, d'autres encore prendre un risque plus élevé. Néanmoins, nous n'observons pas de corrélation entre ces profils décisionnels des participants et leurs traits du modèle des Big Five, ni leur confiance en soi.

La troisième problématique porte sur l'élaboration d'une IA qui prenne mieux en compte le processus décisionnel humain. Des entretiens d'autoconfrontation avec des participants nous ont permis d'établir un modèle du processus décisionnel de l'opérateur pour la tâche de replanification dans l'environnement. Nous avons analysé comment la manière dont l'opérateur utilise les différents outils permet de révéler dans quelle phase de construction de la solution il se situe, ce qui nous a permis d'automatiser le suivi de ce processus décisionnel. Nous en tirons des exemples de situations dans lesquelles le système d'IA peut adapter différemment ses recommandations, selon que l'opérateur est identifié comme impliqué dans une phase d'exploration ou bien dans une phase d'exploitation des plans de vol.

Ces résultats permettent d'identifier des enjeux de facteur humain liés aux usages futurs de l'IA dans les équipes humain-IA avant même que ces systèmes ne soient disponibles, et de proposer des principes de conception qui s'appuient sur les mécanismes cognitifs qui sous-tendent la prise de décision humaine pour poser les bases d'une interaction avec le système d'IA. Nos travaux ouvrent la voie à une assistance intelligente dont le paramétrage sera construit au fil de l'interaction.

Mots clés : Équipe humain-autonomie, interaction humain-machine, collaboration, prise de décision, planification, conception centrée utilisateur, facteur humain

Title: Operators and intelligent systems: understanding each other to decide. Application to drone supervision.

Abstract:

This PhD thesis presents a transdisciplinary research work that is part of a cognitive engineering approach. At the intersection of Artificial Intelligence (AI) and Human Factor, we explore the interaction principles that allow the implementation of a powerful collaboration between a user and an intelligent recommendation system in a decision-making context. We focus on the application framework of planning, for which we have developed a simulation environment that stages an air operator in charge of supervising a drone with a high level of decision-making autonomy. The mission scenario, played on different terrains, leads the operator to determine a new flight plan assisted by intelligent recommendations. An analysis of the literature allows us to characterize three scientific problems that we investigate through three experimental studies in this environment.

The first problematic focuses on the consequences of a change in the AI's participation in the decision making process on the operator's feeling. Our results indicate that when the AI reduces its participation by stopping to propose plan suggestions, the operator's feeling of being responsible and at the origin of the solution increases significantly. However, this effect is found to be asymmetric. When the AI increases its participation by introducing plan suggestions that were not initially proposed, the operator's feeling of being responsible and at the origin of the solution decreases little even though the validated plans become more homogeneous.

The second problem examines the potential links between the compromise criteria of the plans constructed with the AI by the operator and the elements of his personality that could predict it, in terrains where there is no conceivable plan that has an acceptable quality. We highlight an individual preference for degrading one of the three plan trade-off criteria: some participants prefer to abandon targets instead, others consume a lot of fuel, and others take a high risk. Nevertheless, we do not observe a correlation between these participants' decision profiles and their Big Five model traits or their self-confidence.

The third issue concerns the development of AI that better accounts for human decision making. Self-confrontation interviews with participants allowed us to build a model of the operator's decision making process for the replanning task in the environment. We analyzed how the way the operator uses the different tools reveals in which phase of the solution construction he is in, which allowed us to automate the monitoring of this decision-making process. We derive examples of situations in which the AI system can adapt its recommendations differently, depending on whether the operator is identified as being involved in an exploration phase or a flight plan exploitation phase.

These results allow us to identify human factor issues related to future uses of AI in human-IA teams even before these systems are available, and to propose design principles that rely on the cognitive mechanisms underlying human decision making to lay the foundations for interaction with the AI system.

Keywords: Human-autonomy teaming, human-machine interaction, collaboration, decision making, planning, user-centered design, human factor

Remerciements

Je souhaite commencer par remercier mes directeurs de thèse, Nicolas Maille et Benoît Le Blanc. Je vous remercie pour la confiance que vous m'avez accordée, et pour vous être montrés si disponibles. Je vous remercie pour votre enthousiasme sans faille et vos précieux conseils tout au long des années tumultueuses qui se sont écoulées.

Je tiens à exprimer ma reconnaissance aux membres du jury pour les échanges scientifiques que vous avez motivés à la lecture du présent document et lors de la soutenance : Dr Buchler, Prof. Cegarra, Mme Chouvy, Prof. Claverie et Dr. Marchand.

Je remercie l'Agence Innovation de Défense et l'Office National d'Études et Recherches Aérospatiales pour avoir financé cette thèse.

Je remercie tous les participants qui ont accepté de participer aux études que j'ai menées.

Aux co-doctorants du laboratoire passés et présents, Florent, Mathieu, Quentin, Kelly, Anne-Laure, Raphaël, Nicolas, Marine, Marin, Yale, Claire, François merci pour cette camaraderie qui était sans aucun doute le sel de cette aventure. Aux membres permanents également, Bruno, Ugo, Jean-Christophe, Francine.

À ma famille, pour votre soutien, pour m'avoir toujours poussé à donner le meilleur de moi-même.

À mes amis de toujours, Elie, Guilhem. Je cite : « Bac + 8 ? C'est ce qui arrive quand personne n'a osé te dire qu'il faut arrêter les études au bout d'un moment ».

Pour finir, à toi, cher lecteur, merci de prendre le temps d'ouvrir ce document, même si ce n'est que pour y récupérer une source (Prébot, 2020)…

Équipe Cognitique & Ingénierie Humaine

Laboratoire de l'Intégration du Matériau au Système

IMS – UMR CNRS 5218

Université de Bordeaux

109 Avenue Roul, 33400 Talence

« Et nous sommes en présence avec l'ordinateur de la contradiction, déjà annoncée dans tout le mouvement technique, maintenant portée à sa complète rigueur, entre le rationnel (des problèmes posés grâce à l'ordinateur et des réponses données) et l'irrationnel des attitudes et tendances de l'homme : l'ordinateur fait éclater ce qu'il y a d'irrationnel dans une décision humaine, montre que tel choix considéré comme raisonnable est passionnel. Il n'est pas dit que cela se traduise par une rationalité absolue, mais il est évident que ce conflit introduit l'homme dans un univers culturel différent de tout ce qu'il avait connu avant. »

JACQUES ELLUL, *Le système technicien*

Table des matières

Résumé : 3
Abstract: 5
Table des matières 10
Table des illustrations 12
Introduction 14
Chapitre 1 : État de l'art et problématiques de recherche 17
I. L'enjeu éthique de la responsabilité humaine devant l'IA 17
II. L'explicabilité de l'IA et ses limites 20
III. Repenser la répartition des rôles entre opérateur humain et IA 24
IV. Les mécanismes de la prise de décision humaine individuelle 26
V. Les mécanismes de la prise de décision humaine en équipe 28
VI. Problématiques de recherche 32
Chapitre 2 : Environnement de simulation développé 34
I. Description de la tâche simulée 35
II. Déroulement d'un essai dans l'interface 38
III. Outils de replanification collaborative 43
IV. Version simplifiée en condition coopérative ou collaborative 47
Chapitre 3 : Étude de l'impact des suggestions de l'IA sur la prise de décision et le ressenti de l'opérateur 51
I. Méthode 51
1. Hypothèses 51
2. Plan expérimental 51
3. Métriques 53
II. Résultats 55
1. Impact de l'évolution du mode d'interaction sur le ressenti de l'opérateur 55
2. Impact du mode d'interaction sur la prise de décision de l'opérateur. 59
III. Conclusion sur l'expérimentation 61
Chapitre 4 : Étude du lien entre la prise de décision de l'opérateur avec l'IA et ses éléments de personnalité 63
I. Méthode 63
1. Hypothèses 63
2. Plan expérimental 63
3. Métriques 64

II. Résultats....64
1. Préférences décisionnelles individuelles....64
2. Profils décisionnels et caractéristiques individuelles....67
III. Conclusion sur l'expérimentation....70
Chapitre 5 : Étude de l'adaptation de l'IA au processus décisionnel de l'opérateur....72
I. Méthode....72
1. Hypothèses....72
2. Méthode de construction pour chaque étape....73
II. Résultats....74
1. Construction d'un modèle du processus décisionnel des opérateurs....74
2. Automatisation du suivi de l'évolution du modèle décisionnel pendant son déploiement....76
3. Identification d'opportunités d'assistance adaptées à l'évolution du processus décisionnel suivi....80
III. Conclusion sur l'expérimentation....84
Chapitre 6 : Discussion générale....85
I. Apports méthodologiques sur la simulation de la prise de décision en équipe « opérateur-IA »....85
II. Apports théoriques sur l'impact de l'évolution de la participation de l'IA sur la prise de décision de l'opérateur....88
III. Apports théoriques sur la variabilité humaine dans la prise de décision avec l'IA....90
IV. Apports théoriques sur la conception d'une IA adaptée au processus décisionnel de l'opérateur....92
Conclusion générale....96
Publications issues des travaux de cette thèse....98
Bibliographie....99
Annexes....108
Annexe 1 : Superposition des plans validés selon le mode d'interaction dans l'étude 1....108
Annexe 2 : Questionnaire NASA-TLX....115
Annexe 3 : Questionnaire Big Five Inventory français....116
Annexe 4 : Échelle d'estime de soi de Rosenberg....118
Annexe 5 : L'échelle de stabilité de la confiance en soi....119
Annexe 6 : Traits de personnalité des participants de l'étude 2....120

Table des illustrations

Figure 1 : Espaces des erreurs humaines et de l'IA.....18
Figure 2 : Compromis des termes coopération et collaboration.....31
Figure 3 : Une vue catégorielle des différents types d'agents.....32
Figure 4 : Sémantique associée à la couleur des cases du terrain dans l'environnement.....35
Figure 5 : Scénario de la tâche......36
Figure 6 : Jauges des critères du compromis de plan dans l'interface......37
Figure 7 : Algorithme de recherche de plans à suggérer à l'opérateur implémenté dans l'environnement......39
Figure 9 : Au début d'un essai, les outils de replanification ne sont pas disponibles.....40
Figure 10 : Au bout de 5 secondes, les nouvelles menaces sont détectées et affichées......40
Figure 11 : Au bout de 5 secondes de plus, les outils de replanification sont affichés......41
Figure 12 : Exemple d'alternatives tactiques proposées.....42
Figure 13 : Une fois l'étape de replanification terminée......42
Figure 15 : L'essai est terminé et un bouton permet de lancer l'essai suivant.....43
Figure 16 : Une représentation des liens entre les niveaux technique, tactique, opérationnel, stratégique de théâtre, stratégique, de grande stratégie, et institutionnel de la guerre......44
Figure 17 : Répartition des rôles entre l'opérateur et l'IA selon les différents outils.....45
Figure 18 : Évolution de la qualité du plan au fil du déroulement d'un essai en condition coopérative.....48
Figure 19 : Fonctionnalité de vérification de la classification d'un ennemi par l'opérateur.....48
Figure 20 : Évolution de la qualité du plan au fil du déroulement d'un essai en condition collaborative.....49
Figure 21 : Évolution de l'assistance à la replanification fournie par l'IA selon le changement du mode d'interaction.....52
Figure 22 : Questionnaire posé dans l'interface après chaque replanification.....53
Figure 23 : Variation observée sur les sentiments de responsabilité, de paternité, de confiance et de charge de travail des participants selon l'évolution du mode d'interaction.....56
Figure 24 : Résultat des ANOVA à mesures répétées mesurant la significativité de l'effet de l'évolution du mode d'interaction sur le sentiment de responsabilité, de paternité, de confiance et de charge de travail des participants......57
Figure 25 : Cercles des corrélations correspondants aux ACP de la confiance, responsabilité, paternité et charge de travail moyenne de chaque participant.....58
Figure 26 : P-value des corrélations de Pearson des variables pour les deux conditions......59
Figure 27 : Séquences des cibles des plans validés par les 20 opérateurs......60
Figure 28 : Superposition des plans validés par les participants sur un terrain.....61
Figure 29 : Dendrogramme de la classification hiérarchique des participants en fonction de leurs critères de compromis moyen......65
Figure 30 : Cercle des corrélations de l'ACP des participants selon leurs critères de compromis moyen.....66
Figure 31 : Score des participants sur une échelle de 0 à 100 pour leur ouverture d'esprit, consciencieusité, extraversion, agréabilité, tendance à la névrose, confiance en soi et stabilité de la confiance en soi......67
Figure 32 : Cercle des corrélations de l'ACP des participants selon leur score dans les sept métriques de caractérisation individuelles mesurées......68
Figure 33 : Caractéristiques individuelles moyennes des participants de chaque groupe décisionnel identifié......69

Figure 34 : Caractérisation individuelle des participants selon leur groupe décisionnel (indication des p-values issues des tests d'analyse de variable). 70
Figure 35 : Plan expérimental de l'étude qualitative.....74
Figure 36 : Verbatim de l'explicitation de la méthodologie de replanification fournie par les participants lors des entretiens.. 75
Figure 37 : Processus décisionnel type des opérateurs pour la tâche de replanificationl. 76
Figure 38 : Exemple théorique de la phase décisionnelle liée à plusieurs possibilités de compromis à l'étape n+1 de modification de plan selon une étape n donnée. 78
Figure 39 : Pseudocode de l'algorithme de suivi du processus décisionnel. 79
Figure 40 : Inférence de la phase décisionnelle de l'opérateur à chaque modification du plan de vol pour un essai de l'étude. 80
Figure 41 : Deux modifications de plans successives lors de la tâche de replanification sur un terrain. 83
Figure 42 : Deux suggestions de plans sont proposées en tant qu'alternative 2 et alternative 3, l'alternative 1 étant le plan à la modification n.....83
Figure 43 : Modèle de la prise de décision collaborative entre coéquipiers humains. 93
Figure 44 : Modèle de la prise de décision assistée par une IA collaborative qui s'adapte au processus décisionnel de l'opérateur.....95

Introduction

Au cours des années 2010, les drones ont pris une place de plus en plus importante pour la conduite des opérations extérieures et la lutte contre le terrorisme dans le cadre de conflits asymétriques. Lors de l'opération Barkhane menée au Sahel par l'armée française, sur 98 frappes effectuées par l'armée de l'air et de l'espace en 2021, 55% ont été réalisées par des drones MQ-9 Reaper de fabrication américaine (Tanguy, 2022). En 2020, la guerre du Haut-Karabagh qui a opposé l'Azerbaïdjan à l'Arménie a marqué un tournant et consacré les drones comme acteurs incontournables dans la troisième dimension également pour les conflits conventionnels. Cet affrontement a été caractérisé par une faible utilisation de l'aviation, par crainte des défenses aériennes, et par l'emploi massif des drones par les forces azéries (Cigolotti et Carlotti, 2021). Le grand raid aérien initial classique, par avions ou missiles de croisière, détruisant ou neutralisant les capacités de défense antiaérienne ennemies, a été remplacé par un raid de drones Bayraktar TB2 de fabrication turque et de drones-kamikaze IAI Harop de fabrication israélienne (Goya, 2020). Les drones y ont ainsi rempli les fonctions complètes, classiques, de l'arme aérienne (renseignement, surveillance, reconnaissance, appui-feu, frappes) et à un coût bien moindre.

Ces évolutions techniques mais aussi de doctrines d'emploi ont été préfigurées en Europe dès 2006 avec le lancement du programme nEURON afin de développer un démonstrateur de drone de combat furtif, pour préparer la densification et les performances des systèmes anti-aériens qui pourraient empêcher des avions de combat conventionnels d'exécuter leurs missions au-dessus d'un territoire contesté. Ce projet, résultat d'une coopération à six pays européens et piloté par Dassault Aviation, a réussi un premier vol de démonstrateur technologique en 2012 et le largage d'une arme de la soute interne en 2015 (Trappier, 2017). En parallèle, le programme de développement d'un RPAS MALE (Medium Altitude Long Endurance Remotly Piloted Aircraft System) européen, ou programme Eurodrone, a été initié pour renforcer la souveraineté dans ce domaine et favoriser la base industrielle et technologique de défense européenne (Groupe de travail ADER, 2016). Ce programme entre l'Allemagne, la France, l'Espagne et l'Italie ambitionne de fabriquer un premier prototype 2024 et la première livraison de série est prévue pour la fin de la décennie (Dupuy, 2013). Enfin, le programme SCAF (Système de Combat Aérien

du Futur) prévoit de répondre à un besoin capacitaire des armées de l'air française, allemande et espagnole à l'horizon 2040, en succédant au Rafale et à l'Eurofighter Typhoon. Selon Le Gleut et Conway-Mouret (2022) ce projet de « système de systèmes » comportera plusieurs composants :

- Un avion de combat omnirôle NGF (Next Generation Fighter), capable de mener à bien les missions d'interception, la défense air/air et la dissuasion.
- Un « Air combat cloud » connectant l'ensemble des plateformes et permettant de mener le combat collaboratif.
- Des « remote carriers » (effecteurs déportés), machines non habitées ayant des capacités de saturation, de leurrage, de renseignement, ou de frappe. Certains d'entre eux seront récupérables par retour direct ou par récupération sur le terrain, d'autres seront consommables à l'instar d'une munition. Ils seront dotés dans une certaine mesure de capacités autonomes, notamment afin de faire face aux menaces qu'ils pourraient rencontrer en avance de phase par rapport aux avions de combat.

Ces remote carriers sont destinés à être subordonnés au NGF et pour les plus évolués de ces drones, le modèle est celui du « Loyal Wingman » qui doit rester contrôlé par les appareils pilotés. Toutefois, l'intégration de fonctionnalités avancées d'Intelligence Artificielle (IA) est pleinement requise, afin de décharger les pilotes des tâches les plus simples, d'aider à la prise de décision, ou encore de ne pas perdre les drones en cas de coupure des liaisons de données. L'IA sera ainsi essentielle à la performance du SCAF et un moyen d'augmenter les capacités de l'homme, qui resterait au cœur du système, plutôt que de le remplacer (Huderbeau, 2022). C'est dans cet esprit qu'a été lancé en 2018 par le ministère des armées le projet « Man Machine Teaming », qui a pour but de préparer les technologies d'IA nécessaires à l'aviation de combat du futur en dotant les différents systèmes de d'avantage d'autonomie « au service d'une relation Homme-Machine élargie et repensée ». Dans cette perspective, ces systèmes intelligents ne se limiteraient plus seulement à la simple exécution des actions demandées par un opérateur mais permettraient un travail collaboratif qui rendrait les actions et décisions des opérateurs plus efficaces et performantes tout en économisant les ressources mentales et physiques de ces derniers. Cette capacité permettrait aux systèmes d'apprendre des situations rencontrées, de s'adapter en conséquence et de partager les informations pertinentes afin d'apporter une aide à la prise de décision et à la planification des opérateurs. Une des principales problématiques du SCAF est donc l'intégration entre

eux de systèmes pilotés par des humains à bord des aéronefs, pilotés à distance, et partiellement autonomes. Si l'industrie de la défense est à l'avant-garde de ces enjeux, les applications civiles sont également multiples. Les drones de petite dimension connaissent un succès grandissant dans le monde civil et devraient avoir des rôles de plus en plus importants nécessitant davantage de coordination avec des opérateurs. Pour anticiper ces évolutions de l'IA du rôle de robot à celui de partenaire artificiel de l'humain, l'objectif de cette thèse est d'investiguer comment opérateur et système intelligent peuvent se comprendre pour décider dans le cadre applicatif de la supervision de drones.

Chapitre 1 : État de l’art et problématiques de recherche

I. L’enjeu éthique de la responsabilité humaine devant l’IA

La robotisation et numérisation croissante du champ de bataille montre les limites de l'approche techno-centrée héritée des paradigmes historiques. La saturation informationnelle et la surcharge cognitive induites par la multiplication des interfaces obligent désormais à reconsidérer la place de l’humain, et ce dès les phases initiales de conception des systèmes d’armes (Briant, 2021). Les opérations actuelles et futures se dérouleront de plus en plus dans des environnements de mission hyperconnectés qui exigeront que les processus de commandement et de contrôle soient rapides et agiles, et où le rôle des opérateurs devrait évoluer vers une fonction de supervision de systèmes intelligents dont les capacités de calcul dépassent déjà les capacités humaines en ampleur et en vitesse (Lyons et al., 2021). Si cette évolution ouvre de nouveaux contextes opérationnels permettant une meilleure efficacité des interventions militaires, il est essentiel que les décisions d'action prises restent entièrement sous la responsabilité des opérateurs humains, car ce sont eux qui porteront la responsabilité politique et juridique devant la société. Il existe actuellement un consensus international sur le principe du contrôle humain des systèmes intelligents, qui se traduit dans les textes par « des garanties et des mécanismes, tels que l'attribution de la capacité de décision finale aux humains, qui sont appropriés au contexte et à l'état de l'art », la possibilité de décider de ne pas utiliser un système intelligent afin de conserver des niveaux de jugement humain, ou encore la possibilité que la décision de l'humain prime sur celle calculée par le système (Tessier, 2021). Pour le ministère des armées, si les systèmes comportant de l’IA sont appelés à fonctionner avec une certaine autonomie, il est alors indispensable que ceux-ci soient robustes, sécurisés et permettent d’assister le soldat et le commandement en confiance, sans effet « boîte noire », tout en conservant la responsabilité humaine de l’action (Task Force IA, 2019). Cette IA de confiance repose notamment sur la rigueur de conception des systèmes qui doit garantir le total respect du cadre fixé par l’homme, ainsi que sur la capacité du ministère à évaluer et certifier ces systèmes. L’homme pourra alors tirer le meilleur parti de son système et obtenir ainsi un réel facteur de supériorité opérationnelle. L’objectif est de combiner jugement humain et puissance des algorithmes pour

décider et agir avec clairvoyance dans des tempos opérationnels toujours plus élevés (Figure 1). Cependant, cette notion d'IA de confiance est l'objet de critiques dans la mesure où la confiance ne se décrète pas et qu'une machine ou un système ne peut pas porter, en soi, la confiance. Ainsi, selon Tessier (2021) : « *C'est bien l'expérience d'une personne qui utilise un système, l'examen de la manière dont il a été conçu et les garanties démontrées de conformité technique qui sont fournies qui vont amener cette personne à avoir confiance, ou non, dans ce système pour répondre à ses besoins* ». En outre, il ne faut pas négliger la contribution positive des opérateurs humains à la résilience des systèmes. Lorsque des systèmes hautement automatisés fonctionnent dans des environnements complexes, les opérateurs humains contribuent à la résilience globale du système par leur capacité à voir et à agir en dehors du champ des capacités de l'automatisation (Prinzel et al., 2021).

Figure 1 : Espaces des erreurs humaines et de l'IA. La forme bleue en pointillés indique l'espace des erreurs possibles que les humains peuvent commettre. La forme rouge dénote les erreurs de l'IA ; sa taille plus petite indique une réduction nette du nombre d'erreurs. La région grise indique les erreurs spécifiques de l'IA qu'un humain ne ferait jamais. Malgré la réduction du nombre total d'erreurs, un modèle de déploiement peut créer de nouvelles zones de responsabilité (en gris), nécessitant des explications. Extrait de Weld et Basal, 2019.

Cependant, les automatismes peuvent altérer les mécanismes de contrôle classiquement utilisés par l'humain : moindre engagement dans la tâche, augmentation de la divagation attentionnelle, moindre aptitude à détecter des erreurs. En particulier s'il est novice, fatigué ou stressé, l'humain est susceptible de se reposer sur ce que préconise la machine et d'être ainsi enfermé dans des choix restreints ou erronés. Enfin, le manque d'informations ou au contraire un flot trop abondant d'informations, ou encore les schémas que l'humain a en tête, peuvent entraîner une mauvaise compréhension du comportement de la machine ou des résultats qu'elle propose, et entraîner des décisions humaines inadaptées (Demiaux, 2017). Par conséquent, bien que l'humain ait formellement le droit de décider

(d'accepter ou de rejeter les conseils donnés), on peut se demander si ce droit est toujours applicable lorsque les implications des suggestions du système sont difficiles à comprendre (Kox et al., 2021). Il est également connu que les humains perdent leur vigilance avec le temps, ce qui entraîne des erreurs de détection des signaux, subissent une atrophie des compétences en raison d'une dépendance excessive, et ne connaissent pas les caractéristiques essentielles des systèmes qu'ils supervisent, ce qui entraîne un manque de confiance et une sous-utilisation des ressources critiques du système (Lachter et al., 2021). Ces dérives sont susceptibles de provoquer une confiance aveugle dans le système et donc une perte d'aptitude du militaire à en remettre en cause les suggestions. La tendance humaine à développer une addiction aux automatismes peut aboutir au fait que l'humain ne veuille plus reprendre le contrôle ou bien perde confiance dans sa capacité à appréhender un niveau élevé de complexité (Comité d'éthique de la défense, 2021). Un système d'aide à la décision ou un système « autonome », conçu pour aider l'être humain ou le remplacer dans certaines tâches, est ainsi susceptible de porter atteinte à l'autonomie humaine, en influençant la décision de la personne, voire en s'y substituant. Il y a un paradoxe entre les raisons qui motivent l'automatisation et le fait d'exiger un contrôle humain des fonctions automatisées : la question se pose de la capacité de l'humain à exercer effectivement ce contrôle.

Pour permettre aux opérateurs de rester maîtres des décisions, il est alors nécessaire d'améliorer leur compréhension du fonctionnement du système qu'ils supervisent, et de mettre en place des outils d'aide à la décision, qui sont construits conjointement par le système intelligent et l'opérateur humain, et qui devront être pleinement assumés par les personnes impliquées dans ce processus. L'homme doit ainsi être dans la boucle globale : une machine peut être autonome mais elle ne peut s'inventer une mission ou modifier celle-ci sans demander l'autorisation à un être humain. Il apparaît important de travailler à la fois sur la manière dont les systèmes intelligents présentent et justifient les éléments qu'ils apportent dans le processus de décision, et sur le processus interactionnel homme-IA qui doit être mis en place (Pagliari et al., 2022). Plus les systèmes deviennent intelligents, par exemple en utilisant des techniques d'apprentissage profond, moins leurs résultats seront compréhensibles pour un décideur humain. Pour maintenir les opérateurs dans la boucle de décision, la prise de décision humaine repose de plus en plus sur un

prétraitement de l'information, qui doit améliorer la compréhension de l'opérateur de la situation autour du dispositif et de son état de fonctionnement. Il s'agit de promouvoir un design propre à renforcer l'autonomie et la réflexivité des personnes, à remédier aux situations d'asymétrie que peuvent établir les algorithmes à leur détriment, à leur permettre de prendre des décisions informées et de manière lucide. Pour cela, des outils de visualisation peuvent permettre à des individus de comprendre pourquoi des recommandations leur ont été proposées voire, encore mieux, de générer en retour des recommandations plus appropriées. Les individus se trouvent par-là même placés dans une posture active. Il s'agit de donner à l'individu la main sur une partie au moins des critères qui déterminent la réponse fournie par l'algorithme, lui permettant éventuellement de tester différentes réponses en fonction de paramétrages différents. À travers le design, c'est toute la relation entre l'homme et la machine qui peut être modifiée, dans le sens d'une responsabilisation de l'homme et d'une augmentation de sa capacité à prendre des décisions éclairées, au lieu d'une confiscation au profit de la machine de sa capacité à faire des choix.

II. L'explicabilité de l'IA et ses limites

Ceci amène à des réflexions sur la compréhension que l'opérateur peut avoir du fonctionnement du système qu'il utilise. L'amélioration de l'«explicabilité» de l'Intelligence Artificielle est devenue en quelques années une priorité majeure pour les états (Gunning et Aha., 2019) et un domaine de recherche à part entière (Alonso et al., 2018) qui se fédère autour du terme XAI (eXplainable Artificial Intelligence). Commençons par souligner le fait que pour la question du XAI, comme pour toutes les problématiques relatives à l'Intelligence Artificielle, une des difficultés réside dans la définition même de l'IA. Deux grands courants de pensée, étroitement liés aux sciences cognitives, structurent la recherche en Intelligence Artificielle (Cardon et al., 2018) : le cognitivisme, et le connexionnisme. La manière d'aborder la problématique du XAI sera ainsi très différente selon le type d'IA considéré. L'Intelligence Artificielle cognitiviste, également dite symbolique, sert à répondre à des questions de logique formelle pour résoudre des problèmes associant des règles, faits et raisonnements connus du savoir humain. Son point fort réside dans son caractère explicite et décorticable, et son point faible dans sa difficulté à gérer de situations imprévues.

L’Intelligence Artficielle connexionniste, elle, décompose les problèmes en sous-éléments connectés entre eux. Ses modèles émergent depuis un réseau d’interconnections d’une multitude d’unités simples. Ses points forts résident dans sa haute adaptabilité et capacité d’apprentissage, et sa robustesse au bruit, tandis que ses points faibles se nichent dans son manque de représentation explicite des connaissances. Les réseaux de neurones artificiels, inspirés du fonctionnement des neurones biologiques, se basent sur ce paradigme. Ils sont généralement optimisés par des méthodes d’apprentissage de type probabiliste. L’augmentation des puissances de calcul ainsi que des quantités de données disponibles a permis aux réseaux de neurones de faire d’importants progrès dans les années 2010, notamment dans la reconnaissance d’images ou le traitement automatique du langage (Le Cun et al., 2015). Cependant, ces outils ont un fonctionnement qui peut être considéré comme particulièrement opaque. En effet, les règles d’un réseau de neurones sont déduites empiriquement des données d’entraînement, et enfouies dans ses paramètres, et il n'existe à l'heure actuelle aucune technique systématique permettant de demander à un tel modèle de fournir une explication claire pour une décision donnée (Pégny, 2018). La recherche pour créer des réseaux de neurones plus interprétables est active. Krause (2018) indique par exemple qu’inspecter les relations entrées-sorties de ces modèles en utilisant des méthodes d’analyse visuelle, peut aider à en saisir le raisonnement.

Cet enjeu d’explicabilité est particulièrement prégnant dans le cadre applicatif de la planification, où l’assistance fournie par l’IA va servir de support à une réflexion humaine qui doit s’en approprier la responsabilité. Barnes et al. (2019) proposent un modèle d’informations à fournir aux utilisateurs pour rendre plus intelligibles des plans de vol. Afin d’améliorer la conscience humaine de la situation, des agents artificiels intelligents devraient indiquer trois niveaux d’information. Premièrement, sur le choix des objectifs déterminés par l’agent et les chemins qui les réalisent. Deuxièmement, sur le raisonnement de l’agent, en détaillant les contraintes environnementales prises en compte et les algorithmes employés pour la recherche de chemin. Troisièmement, sur les projections réalisées par l’agent sur les résultats espérés des solutions, ainsi que les incertitudes qui pourraient conduire à des échecs. Cependant, l’efficacité de cet ajout d’informations sur la compréhension humaine n’a pas été vérifiée expérimentalement. Chakraborti et al. (2019) ont établi

une méthodologie pour rendre des plans plus explicables à des conducteurs de voitures autonomes. Le besoin d'explicabilité d'une solution est défini comme l'écart entre celle-ci et les solutions anticipées par le conducteur. Ils détaillent un modèle de régression permettant d'apprendre à mesurer cet écart de représentation entre conducteur et système de recommandation grâce au feedback humain, et qui peut être intégré au processus de génération de plan pour en produire des justifications. Les auteurs démontrent à travers une étude sur des conducteurs que ceux-ci comprennent mieux les trajets du véhicule autonome grâce à ce modèle. Cependant, les utilisateurs regrettent de ne pas pouvoir choisir de désactiver ces explications dans les cas où ils ne les jugent pas nécessaires. Ces différents travaux ont caractérisé les informations à renvoyer à des utilisateurs humains pour rendre des suggestions de plan informatiques plus compréhensibles. Ils soulignent notamment l'importance de renseigner les utilisateurs sur le fonctionnement calculatoire des algorithmes de recherche de chemin, et la manière dont ils différent de la méthodologie de planification humaine.

Cependant, Miller (2019) montre que la plupart des modèles actuellement utilisés en recherche en XAI n'est pas basée sur la compréhension scientifique que l'on a du processus d'explication. De plus, les expérimentations comportementales humaines y sont rares, situation qui doit évoluer pour produire des techniques empiriquement utilisables. Ces carences dans la recherche en XAI sont pour cet auteur le fruit d'un biais cognitif des chercheurs, qui provenant essentiellement des mathématiques ou de l'informatique n'ont pas habituellement le souci de l'utilisabilité des outils qu'ils mettent au point : les explications de l'IA qu'ils proposent sont des explications faites pour des chercheurs en IA, bien plus que pour des utilisateurs finaux. La plupart de la recherche sur le XAI utilise de fait l'intuition des chercheurs sur ce qui constitue une « bonne explication », plutôt que de se baser sur une compréhension fine de la manière dont les individus définissent, génèrent, sélectionnent et présentent des explications. Miller en appelle ainsi à adopter une démarche hautement interdisciplinaire, intégrant les travaux de la philosophie, de la psychologie et des sciences cognitives notamment, et sur l'évaluation de ces modèles par des utilisateurs.

Si le système intelligent, formalisant sa connaissance et prenant des décisions grâce à une forme de maximisation de l'utilité, cherche à optimiser des variables, à l'inverse, l'opérateur humain agit plutôt grâce à un ensemble d'heuristiques qui l'amène à satisfaire des contraintes. Afin de faire converger ces deux types de comportements, il s'agit de mettre en place une dynamique de dialogue, au cours de laquelle l'opérateur va questionner le système d'IA. Celui-ci, après avoir déterminé les connaissances a priori et les attentes de l'opérateur, va lui fournir une connaissance adaptée. Chakraborti (2018) précise en ce sens qu'une réflexion autour des interactions homme-IA requiert que l'Intelligence Artificielle prenne en compte le modèle mental de l'homme, ainsi que le modèle de ses capacités physiques. L'agent artificiel se doit d'anticiper comment une planification va être perçue par l'homme, et la façon dont cela va le pousser à agir. Borgo (2018) propose lui un cadre général pour fournir des explications pour la prise de décision à travers une planification conjointe homme-IA. Il théorise des explications appropriées comme autorisant l'utilisateur à explorer des alternatives d'action, et à les comparer pour retrouver celle du planificateur, et ce en utilisant différentes métriques. Étant donné la nature contrastive de l'explication, quand l'utilisateur est confronté à une proposition de plan par un planificateur, la question type qu'il se pose est « Pourquoi ce plan contient-il ceci plutôt que cela que j'attendais ? ». Le support de l'IA est particulièrement pertinent dans le cas où celle-ci suggère une option différente de ce que l'opérateur aurait fait. Pour autant, réécrire les étapes d'un algorithme de planification en langage naturel ne constituerait pas une explication efficace ; ni les heuristiques d'évaluation des états sélectionnés par le planificateur dans sa recherche de plan. Bien plus, Borgo défend l'idée que ce qui peut justifier la sélection d'un ensemble d'actions par un planificateur est que cet ensemble d'action se révèle meilleur, ou au moins aussi bien, que l'ensemble d'actions que l'utilisateur choisi. Ainsi, un modèle qui fournirait des explications devrait autoriser l'utilisateur à explorer des alternatives de plans, et à comparer ces résultats avec celui fourni par le planificateur. Une telle approche devrait augmenter la confiance des utilisateurs dans le planificateur, et lui donner de solides arguments pour accepter ou rejeter le plan proposé. De plus, en tant que l'ensemble d'alternatives possibles peut être potentiellement infini, il est nécessaire de guider l'utilisateur à travers le type de questions qu'il pourrait poser. Notons à ce propos que si ce jeu de questions-

réponses prend la forme d'un échange en langage naturel, nous retrouvons ici les considérations développées sur la théorie de l'explication comprise comme dialogue.

III. Repenser la répartition des rôles entre opérateur humain et IA

Si au départ l'explicabilité de l'IA est principalement vue comme une manière de permettre de comprendre pourquoi le système d'IA arrive à une solution donnée, et rendre la « boite noire » plus transparente, il est maintenant reconnu qu'il est important de mesurer comment ces systèmes intelligents peuvent influencer les décisions humaines (Ferreira, 2020). Dans le domaine du e-commerce des études mesurent l'impact de l'ajout d'explication sur le pouvoir de persuasion, la confiance et la satisfaction que cela amène à l'utilisateur (Verbert et al., 2013). Par ailleurs, dans le contexte des tuteurs intelligents qui aident des élèves à apprendre en gérant par exemple la progression des exercices proposés ou la sélection des supports éducatifs les plus adaptés, il est montré que l'apprenant apprécie d'avoir des informations sur ce qui guide les suggestions du tuteur (Barria-Pineda, 2020). Cependant, dans ces approches, les rôles entre le système d'IA et l'opérateur humain sont très dissociés. Une étape supplémentaire importante pour préparer l'avenir est l'étude de la coopération entre humains et systèmes d'IA. Une collaboration de nombreux chercheurs aborde la question des implications liées au fait qu'un système d'IA devienne réellement un membre à part entière d'une équipe (Seeber et al., 2020). Ils notent dès le départ qu'un des points clés est que ce système intelligent ait plus de capacités que les systèmes actuels et en particulier passe d'une position de support de la collaboration, comme le font les assistants numériques actuels, à un « engagement » dans les étapes de la résolution de problème et de prise de décision. Une démarche similaire émerge dans le domaine de la créativité ou historiquement le système informatique était soit un outil au service de l'agent qui crée une œuvre, soit le créateur qui présente un résultat à l'opérateur qui est un spectateur. Le travail actuel sur les interfaces à initiative mixtes (Deterding et al., 2017) entre dans ce changement de paradigme qui cherche à ce que la machine devienne un collaborateur à part entière dans le processus de création. La notion d'engagement prend ici aussi une part centrale, l'agent artificiel idéal étant supposé pouvoir prendre des initiatives pour suggérer, contraindre,

produire, évaluer, modifier l’œuvre en cours de construction, tout en prenant en compte les initiatives de l’opérateur humain.

Dans ce contexte, le paradigme Human-Autonomy Teaming (HAT) a récemment émergé pour préparer l'évolution de ces équipes de systèmes humains-IA, en concevant de nouvelles façons de penser les mécanismes d'interaction (Flathmann et al., 2020). Le HAT peut être définie comme une interdépendance dans l'activité d’un ou plusieurs humains et un ou plusieurs agents artificiels, dans laquelle chacun de ces membres de l’équipe est reconnu comme occupant un rôle distinct pour atteindre un objectif commun en tant que collectif (O'Neill et al., 2020). Si dans une interaction conventionnelle, une machine est traitée comme un outil, dans des interactions hautement interdépendantes et dans des conditions appropriées, elle pourrait assumer le rôle d'un coéquipier. Dans une perspective centrée sur l'humain, un système intelligent devrait être considéré comme un sujet individuel et associé pour la conception d'équipes de systèmes humains-intelligents, et pas seulement comme un substitut ou un remplacement insatisfaisant des agents humains (Zhang et al., 2021). La justification générale de l'utilisation de la HAT est que ce paradigme pourrait fournir de meilleures performances par rapport aux humains seuls ou aux machines seules et surtout dans des situations de forte incertitude, bien que des technologies et des outils innovants doivent être créés pour aider à la communication et au développement de connaissances partagées (Demir et al., 2020). Shively et al. (2017) ont ainsi proposé trois principes pour la conception de la HAT. Les deux premiers, la communication bidirectionnelle et la transparence, sont utilisés pour permettre l'exécution conjointe des tâches. Le troisième, la délégation dynamique, implique une allocation plus flexible du travail, en tenant compte de facteurs tels que la charge de travail et la pression temporelle, afin de maintenir la conscience situationnelle de l'opérateur et de garantir que l'automatisation se voit attribuer des tâches dans sa boîte de compétences.

Il convient de noter les travaux de Tokadli et al. (2021), qui ont étudié comment, dans le cadre d'une opération de pilotage unique, le pilote et son coéquipier autonome exprimeraient leurs intentions l'un à l'autre tout en répondant à une situation changeante. Ils ont mené une étude sur une interface de délégation de type playbook qui permet au pilote humain de réaffecter dynamiquement des fonctions

pour répondre à une situation et coordonner l'exécution des tâches avec le coéquipier autonome. Leurs résultats indiquent que les pilotes préféraient utiliser une interface de type playbook en mode monopilote, mais préféraient encore avoir un copilote humain en raison de l'interaction intuitive et de la confiance dans le processus de jugement. Gombolay et al. (2015) ont également mené une étude pour examiner la performance des tâches et la préférence des personnes pour l'attribution collaborative des tâches impliquant un humain et un système intelligent. Ils ont constaté qu'un cadre semi-autonome, dans lequel l'humain décide d'abord des tâches qu'il veut effectuer et l'agent décide ensuite du reste de l'attribution des tâches, est plus satisfaisant que les cadres de contrôle manuel et de contrôle autonome dans lesquels l'humain et le robot attribuent entièrement les tâches. Ces travaux soulignent l'importance de prévoir l'arrivée de systèmes de délégation dynamiques pouvant être examinés à la fois par un utilisateur humain et un système intelligent, afin de s'assurer que les principes d'interaction sur lesquels ils reposent sont suffisamment acceptables et efficaces.

La flexibilité des interactions est connue pour être une caractéristique souhaitable des équipes car elle facilite l'adaptation aux perturbations de la tâche (Nalepka et al., 2021), et a été désignée sous différents noms tels qu'«interaction co-participative », « interaction flexible », « allocation dynamique des tâches », « délégation dynamique » ou « travail d'équipe collaboratif ». Si la prise de décision par l'homme est une tâche qui prend du temps, la prise de décision par une IA entièrement autonome présente également plusieurs inconvénients : l'agent humain n'a aucun moyen de corriger les erreurs commises par le système, n'a aucun contrôle sur les sources auxquelles le système a accès et ne peut pas utiliser l'expérience et l'intuition humaines (van Diggelen et al., 2021). Par conséquent, une forme d'interaction entre l'homme et le système intelligent qui tire parti à la fois des capacités de traitement rapide des big data du système et de la connaissance du domaine et du bon sens de l'homme est souhaitée.

IV. Les mécanismes de la prise de décision humaine individuelle

L'ensemble des décisions jugées comme possibles par un individu est intimement lié à la nature des connaissances qu'il va recevoir ou construire, et à la manière donc celles-ci se combinent pour former un cadre de réflexion cohérent. Pour (Baron, 1998), « prendre une décision consiste à choisir ce qu'on veut et ne veut pas faire afin de produire un résultat satisfaisant ». La théorie de la prise de décision individuelle est structurée en plusieurs grands courants, dont les plus importants sont les approches dites normatives et descriptives. La théorie normative de la décision, formalisée par (Von Neumann, 1944), est originaire de la microéconomie, et a pour objet la prise de décision en univers risqué. Cette théorie est basée sur une certaine idée de la rationalité, et repose sur les axiomes de la théorie de l'utilité, tels que l'ordonnance des alternatives, la transitivité, la continuité ou l'invariance. Dans ce cadre, l'utilité est une unité mesurée par la désirabilité, ou le degré de satisfaction des conséquences des actions décidées. Seulement, dès 1953, Allay montre que ces axiomes sont parfois violés. Les individus ont une préférence pour la certitude, face à un choix avec une plus haute utilité (Allay, 1953). En 1997, Simon critique vivement cette théorie à travers sa thèse de la rationalité limitée : « les choix rationnels imposent des standards impossibles aux individus ». De fait, la rationalité parfaite excède les capacités cognitives des individus, il est impossible pour eux de considérer toutes les conséquences des choix possibles. Bien plus qu'un comportement de maximisation de l'utilité espérée, les individus sont guidés par une volonté de satisfaction de contraintes, plus réaliste au vu de leurs limitations. N'étant pas omniscient, l'individu n'a pas à sa disposition l'ensemble des informations lui permettant d'anticiper les conséquences de toutes les options qui s'offrent à lui : il lui est nécessaire d'utiliser des heuristiques de décision. Tversky (1974) en décrit plusieurs, tels que l'heuristique de représentativité, de disponibilité ou d'ancrage du jugement.

Cependant, si cette théorie semble insuffisante pour restituer les mécanismes effectifs de la prise de décision chez l'homme, on peut considérer qu'elle reste une approche pertinente pour formaliser les mécanismes de la décision des systèmes intelligents. Tout système informatique fonctionnant selon une suite finie et non ambiguë d'instructions, il semble raisonnable de considérer qu'une telle formalisation mathématique est équivalente à un processus de minimisation de contraintes. Les limitations cognitives de l'homme invitent néanmoins à admettre que les mécanismes

de sa prise de décision sont d'une autre nature. La théorie descriptive de la prise de décision, partant de ce constat, se base sur une approche bottom-up pour établir une approche empiriquement plus valide. Cette théorie, qui se fonde sur la recherche expérimentale, se nourrit des sciences cognitives. Ross, (2006) explique par exemple que confronté à une situation dynamique, l'individu ne fonde pas son choix sur un calcul rationnel, ou l'analyse exhaustive des utilités, mais fait plutôt appel à ses expériences passées pour créer des simulations mentales. Klein (2003) va plus loin et comprend l'activité décisionnelle comme englobante et complexe, un phénomène « macro-cognitif émergeant » où les processus perceptifs et cognitifs sont inextricablement liés. Pour reconnaître une situation comme typique et y répondre rapidement, les experts disposeraient de structures mnémoniques permettant de conserver des potentiels d'action significatifs et efficaces. Son modèle Recognition-Primed Decision distingue plusieurs éléments-clés pour comprendre ce processus :

- Il n'y a pas de génération aléatoire d'options de décision puis une sélection, et la première option qui apparaît est généralement satisfaisante.
- Le prise de décision se fonde sur une option satisfaisante, non nécessairement optimale.
- L'évaluation des options se déroule à travers des simulations mentales, et non une analyse statistique.
- Il y a une concentration sur l'amélioration des options déjà trouvées, et non le choix entre différentes options.
- Il y a une prime à l'action sur l'analyse exhaustive.

V. Les mécanismes de la prise de décision humaine en équipe

Roy et Bouyssou (1993) notent que « même si la responsabilité finale d'une décision incombe à un individu clairement identifié, elle est souvent le résultat d'interactions entre de multiples acteurs au cours d'un processus de décision ». Même si dans une organisation impliquant des opérateurs humains et de l'IA, il peut être convenu que la décision finale reste la responsabilité de l'homme, résultat du processus d'interaction entre les différents acteurs. Comprendre les éléments de ces interactions et la façon dont ils peuvent influencer la prise de décision finale est une question importante pour mieux développer des systèmes basés sur une

combinaison d'humains et d'IA. Les sciences managériales et de gestion se sont intéressées entre autres, à la manière dont les décisions sont prises au sein des organisations et peuvent éclairer ces questions. Turoff, White et Plotnick (2011) soulignent que dans la littérature, les termes collaboration, coopération et coordination sont trop souvent confondus et proposent une échelle qui différencie cinq niveaux croissants de communication dans la prise de décision en groupe : compétitif : aucune confiance dans les informations transmises ; informatif : échange honnête d'informations sur ce que chaque partie fait ; coordination : planification mutuelle de ce que chaque partie fait et quand ; coopération : accord mutuel sur les tâches que chaque partie effectuera ; collaboration : accord mutuel pour travailler ensemble sur les mêmes tâches. Dans le contexte de la prise de décision concernant des éléments importants qui pourraient mettre en péril les objectifs ou la sécurité de la mission, ce sont essentiellement les niveaux de coopération et de collaboration qui peuvent être en jeu. Nous allons donc maintenant détailler ce qu'ils recouvrent.

Pour Dillenbourg (1996), la coopération et la collaboration ne se différencient pas par la division ou non de la tâche, mais par la manière dont elle est divisée. Dans la coopération, la tâche est divisée hiérarchiquement en sous-tâches indépendantes. Piquet (2009) définit également le travail coopératif comme une organisation collective du travail dans laquelle la tâche à satisfaire est fragmentée en sous-tâches. Chacune de ces sous-tâches est ensuite attribuée à un acteur, soit selon une répartition parfaitement horizontale où tâches et acteurs sont équivalents, soit selon une logique d'affectation en fonction des compétences de chacun. Il s'agit d'une division rationalisée d'une tâche en actions qui seront réparties entre des acteurs agissant de manière autonome. Le travail coopératif est donc un travail de groupe organisé et planifié hiérarchiquement, avec des échéances et un partage des tâches selon une coordination précise (Roschelle et Teasley, 1995). Chaque membre du groupe sait dès le départ ce qu'il doit faire et ne communique, n'échange ou ne partage des éléments que dans le but d'atteindre son objectif individuel (Kozar, 2010). À la fin, le travail de chacun est rassemblé pour créer un seul objet de travail. En d'autres termes, c'est la succession progressive et coordonnée des actions de chacun qui permet d'atteindre l'objectif final. La responsabilité de chacun est engagée dans le seul accomplissement des tâches qui lui sont propres.

Le travail collaboratif, quant à lui, n'implique pas une répartition a priori des rôles, mais une fusion des contributions individuelles dans l'action. Les interactions interpersonnelles sont permanentes pour assurer la cohérence d'ensemble, condition nécessaire à l'efficacité de l'action et à la réalisation de l'objectif global. La collaboration dans le cadre du travail collectif est une modalité d'action qui dépasse l'action individuelle en s'impliquant explicitement dans une dynamique d'action collective. Il s'agit pour chaque acteur d'un projet de nourrir ses contributions individuelles avec celles des autres. Le travail collaboratif implique un engagement mutuel des individus dans un effort coordonné pour réaliser la même tâche et résoudre ensemble le même problème (Barbour, 2018). Il demande aux membres de l'équipe d'être plus interactifs et requiert plus de motivation et de confiance interpersonnelle que les autres méthodes d'organisation du travail. Schöttle et al. (2014), qui ont analysés des articles en science managériale publiés entre 1977 et 2014 contenants entre autres les termes « coopération » ou « collaboration », ont synthétisé comment dans cette littérature ces deux termes se positionnent par rapport à différents facteurs (Figure 2).

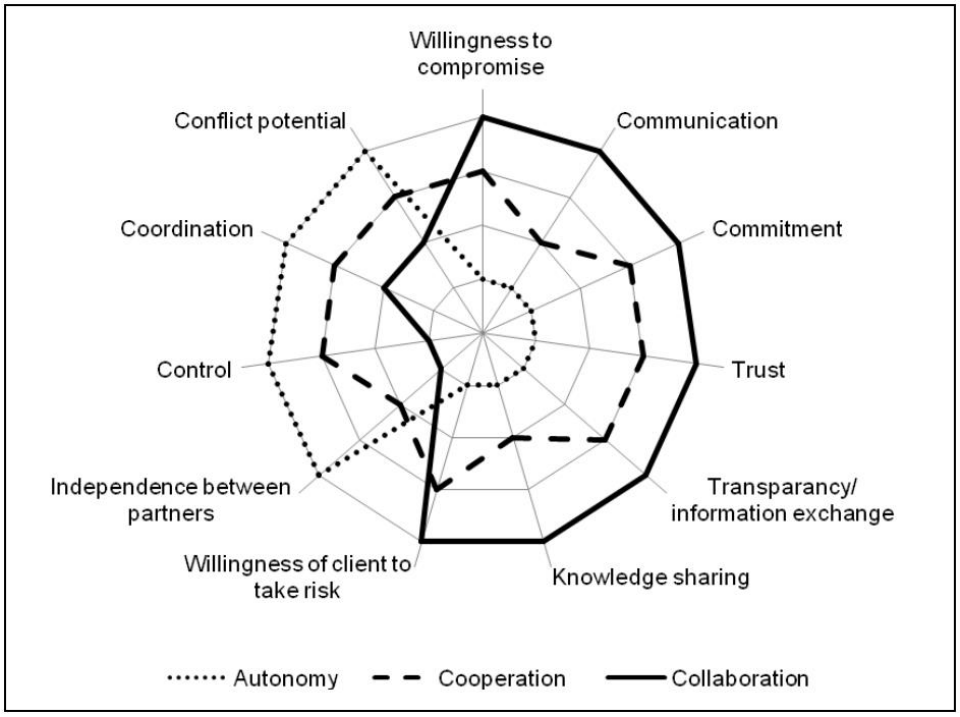


Figure 2 : Compromis des termes coopération et collaboration. Extrait de Schöttle et al., 2014.

Néanmoins, les équipes opérateur-système intelligent sont différentes en nature des équipes exclusivement humaines les concepts de coopération et de collaboration décrivent généralement. Il existe des différences importantes dans la façon dont l'homme et la machine acquièrent et traitent l'information (Kampik, Nieves et Lindgren, 2019), ce qui pourrait rendre la confiance beaucoup plus difficile à établir. Klein (2005) explique que l'une des conditions nécessaires à l'efficacité d'un groupe est l'interprétabilité, c'est-à-dire la capacité de prédire les actions des autres parties avec un degré raisonnable de précision. Chaque membre du groupe doit s'efforcer de rendre ses actions suffisamment prévisibles pour permettre une coordination efficace. Toutefois, les priorités des utilisateurs peuvent changer au fil du temps, ce qui peut obliger le système à gérer et à s'adapter à l'évolution des contraintes et des préférences. Si l'IA nécessite une définition précise pour résoudre mathématiquement le problème, d'un autre côté, les actions humaines sont en fait d'une rationalité limitée et souvent informées par des heuristiques en partie fallacieuses (Tversky et Kahneman, 1992). Il peut y avoir un décalage entre les objectifs mathématiques précis requis parfois par l'IA et les spécifications potentiellement floues qui peuvent être fournies et manipulées par les humains. L'hétérogénéité des équipes peut être une source de difficultés pour établir un modèle mental de l'autre, mais c'est aussi une source d'opportunités car les membres de l'équipe peuvent se compléter en s'apportant mutuellement les compétences qui leur manquent (Nissen, 2014). Ainsi, lorsque les systèmes d'IA doivent interagir avec des utilisateurs humains, il n'est pas tant important de raisonner rationnellement que d'émuler un raisonnement de type humain (Besold et Uckelman, 2018). Un système d'aide à la décision efficace doit non seulement fournir des informations de qualité, mais aussi tenir compte des contraintes de l'utilisateur en termes de ressources cognitives disponibles et de préférences personnelles. Les défis introduits par les humains dans la boucle résident dans cette implémentation d'agents proactifs et sociaux. Selon le modèle de Chakraborti et al. (2017) présenté dans la Figure 3 extraite de leur article, une caractéristique commune à ces deux types d'agents est qu'ils nécessitent tous deux la notion de modélisation mentale des autres coéquipiers, qui ne peut être observée directement et doit être déduite

cognitivement. Une telle modélisation pourrait permettre au robot de déduire les attentes de l'homme concernant son propre comportement et de choisir des comportements conformes à ces attentes.

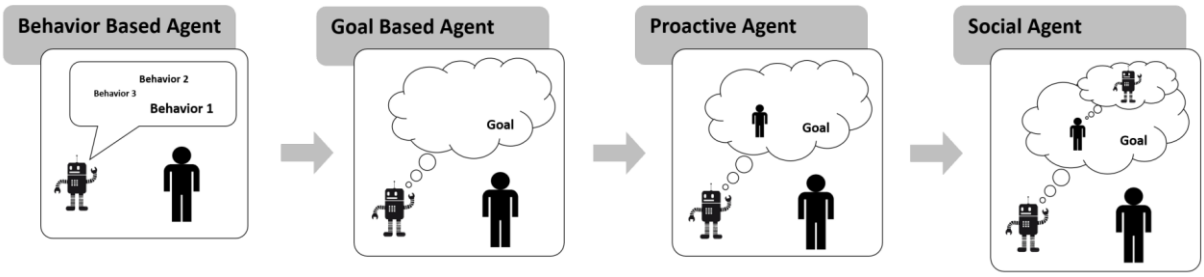


Figure 3 : Une vue catégorielle des différents types d'agents. Chaque type est plus profond que les types précédents en termes de complexité de modélisation (de gauche à droite). Extrait de Chakraborti et al., 2017.

VI. Problématiques de recherche

Pour préparer l'arrivée de drones dotés d'un haut niveau d'autonomie décisionnelle, cet examen de la littérature interdisciplinaire suggère que la condition de l'assurance du contrôle et de la responsabilité de l'opérateur humain réside dans la conception d'un processus d'interaction bidirectionnel et à délégation des tâches dynamique avec le système d'IA, dit collaboratif. Nous identifions trois problématiques relatives à la conception d'un tel processus d'interaction opérateur-IA, et que nous proposons d'investiguer expérimentalement dans cette thèse.

Problématique 1 : Dans quelle mesure l'augmentation de la participation de l'IA à la prise de décision va-t-elle altérer les solutions validées par l'opérateur et son ressenti ?

Problématique 2 : Dans quelle mesure les éléments de personnalité de l'opérateur peuvent-ils être explicatifs de sa prise de décision avec l'IA ?

Problématique 3 : Comment l'IA peut-elle assister l'opérateur de façon adaptée à son processus décisionnel ?

La thèse que nous soutenons est que pour qu'une IA soit une réelle assistance à la décision de l'opérateur il faut que d'une part l'opérateur ait une compréhension de ce que l'IA peut apporter et des limites associées afin de lui faire des requêtes appropriées, et d'autre part que l'IA ait une représentation du processus décisionnel

de l’opérateur pour interagir à bon escient et en respectant l’avancée de son raisonnement.

Chapitre 2 : Environnement de simulation développé

Les études présentées dans cette thèse visent à permettre d'étudier de manière expérimentale plusieurs modalités de travail en équipe entre l'opérateur et le système d'IA afin de mieux cerner l'impact des modes d'interactions, des caractéristiques du système d'IA, des profils d'utilisateurs sur la décision elle-même ainsi que sur la manière dont l'opérateur se sent responsable de cette décision. Le point de vue adopté repose sur l'utilisation d'un environnement de simulation, ou micromonde, où différents outils et différentes dynamiques d'interactions peuvent être testés. Ce micromonde repose sur la planification d'une mission de drone et permet d'avoir une situation à la fois contrôlée qui permette de faire passer différents opérateurs dans des situations comparables, et suffisamment complexe pour que la décision nécessite une aide et soit forcément le résultat d'un compromis. L'interface collaborative développée vise à autoriser une répartition adaptative des rôles entre l'opérateur et l'IA embarquée pour prendre en compte les qualités propres à chaque membre de l'équipe (Schmitt et al., 2017). La conception d'une allocation évolutive des responsabilités nécessite d'élaborer des interfaces non plus centrées sur l'initiative humaine à travers du pilotage, ou de l'initiative du système à travers de la supervision, mais où l'autorité et la prise d'initiative puissent être réparties dynamiquement selon les circonstances (Abbass, 2019). Le système d'assistance développé permet de présenter de sa propre initiative des modifications du plan pour la réalisation de la mission et des outils permettant de l'approuver ou de l'ajuster, comme mis en évidence dans (Szafir et al., 2017).

Dans une première partie nous décrivons la tâche simulée, la nature des décisions à prendre ainsi que la représentation de la situation. Dans une deuxième partie, nous présentons le déroulement d'un essai dans l'interface développée en langage C# avec l'interface graphique Windows Form du framework .NET. Dans une troisième partie, nous détaillons les outils de replanification qui vont permettre la mise en place de séquences d'interactions entre l'opérateur et l'IA afin d'aboutir à une prise de décision collaborative. La quatrième partie présente une version simplifiée de l'environnement où la partie adaptative est bloquée afin de comparer la réalisation de la tâche soit par une interaction coopérative soit par une interaction collaborative entre l'opérateur et le système.

I. Description de la tâche simulée

Le micromonde met en scène un opérateur aérien militaire chargé de superviser un drone pour effectuer des missions d'observation en territoire ennemi. Le but des missions est de survoler plusieurs sites pour les photographier, puis de quitter la zone ennemie tout en minimisant le risque pris et le carburant consommé. L'environnement est représenté par une grille quadrillée aux cases colorées : les cases vertes correspondent aux sites à photographier, les cases grises et noires représentent des zones où le relief est élevé ou très élevé, et les cases rouges des menaces dont la dangerosité estimée est symbolisée par une valeur numérique (Figure 4). Le risque est composé par le fait de passer près des menaces ennemies, mais aussi de passer sur les reliefs (plus on est haut, plus on est détectable par des radars ennemis).

Légende	
	Relief bas
	Relief élevé
	Relief très élevé
X	Ennemi de dangerosité X
	Entrée ou sortie de la zone ennemie
	Cible à photographier
	Etape ajoutée au plan

Figure 4 : Sémantique associée à la couleur des cases du terrain dans l'environnement.

Le même scénario, présenté dans la Figure 5, est réalisé sur différents terrains. Les essais correspondent à des missions qui se déroulent sur des territoires variés mais avec un scénario similaire : 1) le drone se dirige vers la zone ennemie avec un plan de vol initial, 2) des nouvelles menaces sont soudainement détectées, le plan de vol n'est donc plus satisfaisant, 3) l'opérateur interagit avec le système pour définir un nouveau plan de vol, 4) l'opérateur valide un nouveau plan de vol, ce qui termine l'essai. Pendant la phase d'interaction, l'opérateur collabore avec le système pour déterminer un nouveau plan grâce à plusieurs outils à sa disposition.

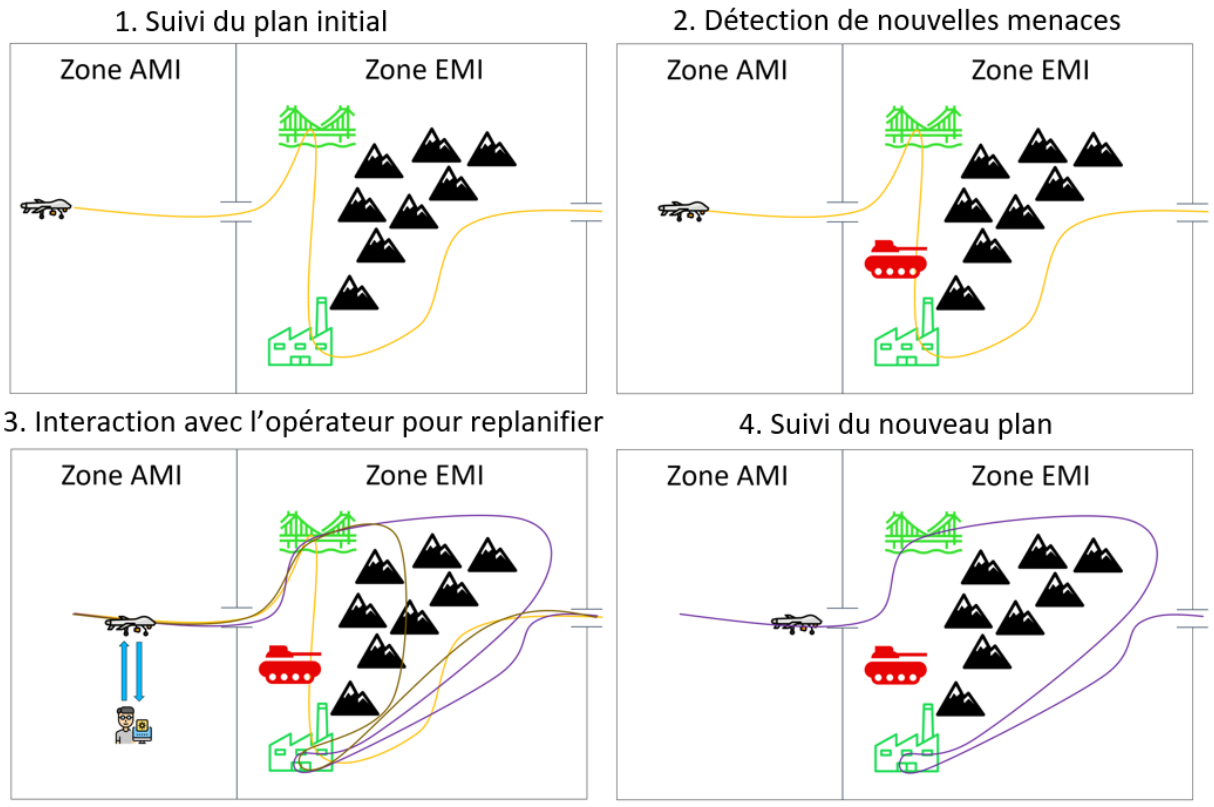


Figure 5 : Scénario de la tâche.

Les terrains représentent des situations à la fois contrôlées, permettant à différents opérateurs de vivre des situations comparables, et suffisamment complexes pour que la décision nécessite une assistance et soit nécessairement le résultat d'un compromis. Les terrains ont été conçus de telle sorte qu'il n'existe pas de plan concevable qui à la fois passe par tous les sites désignés, utilise peu de carburant et prenne peu de risque. L'absence de solution satisfaisante fait rentrer en jeu une dimension subjective de la prise de décision, similaire aux compromis que l'on peut observer dans les situations de crise. Le jugement humain passe par l'expression des préférences individuelles de l'opérateur, qui doit faire des compromis en dégradant au moins un critère.

La qualité du plan de vol courant est représentée selon trois dimensions par trois jauges (Figure 6) : la proportion de sites à photographier traversés, la quantité de carburant consommé et le risque pris. La tâche peut être comprise comme un problème d'optimisation multicritère où ces trois variables doivent être maximisées. L'absence de solution satisfaisante aux problèmes fait intervenir une dimension subjective de la prise de décision, qui se rapproche des arbitrages qui peuvent être

observés en situation de crise (Klein, 2008). Le jugement humain rentre alors en jeu à travers l'expression des préférences individuelles de l'opérateur qui doit réaliser un compromis en dégradant au moins une des variables pour améliorer les autres.

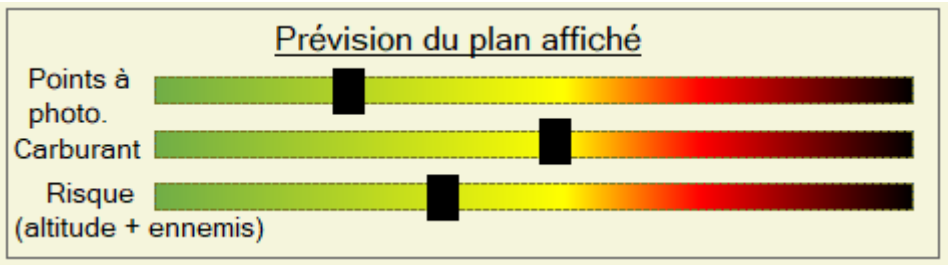


Figure 6 : Jauges des critères du compromis de plan dans l'interface.

Les outils de planification sont basés sur un algorithme de recherche de chemin dans un tableau de 30x20 cases qui permet huit directions de mouvements (Figure 7). À chaque case est associé un coût déterminé par plusieurs paramètres : la distance à la case précédente, le relief de la case et la distance aux menaces pondérée par leur dangerosité. L'algorithme de recherche de chemin dans un graphe A* (Russel et Norvig, 2016), que nous avons implémenté en C#, calcule le segment de plan le moins coûteux entre les sites à photographier deux à deux. Ensuite, l'algorithme combine ces segments pour déterminer le chemin global qui maximise les trois variables du plan. Les valeurs représentées par les jauges correspondent effectivement au résultat objectif de ce calcul qui est trop complexe pour être réalisable par l'opérateur pendant la phase de replanification. La solution calculée est bien optimale mais relativement à la fonction de coût considérée par l'algorithme. La pondération entre les paramètres de cette fonction est un arbitrage qui va être suivi rigoureusement par l'algorithme alors que l'opérateur, en tant qu'humain, n'a pas une rationalité aussi forte et n'appliquera pas forcément le même compromis dans toutes les situations.

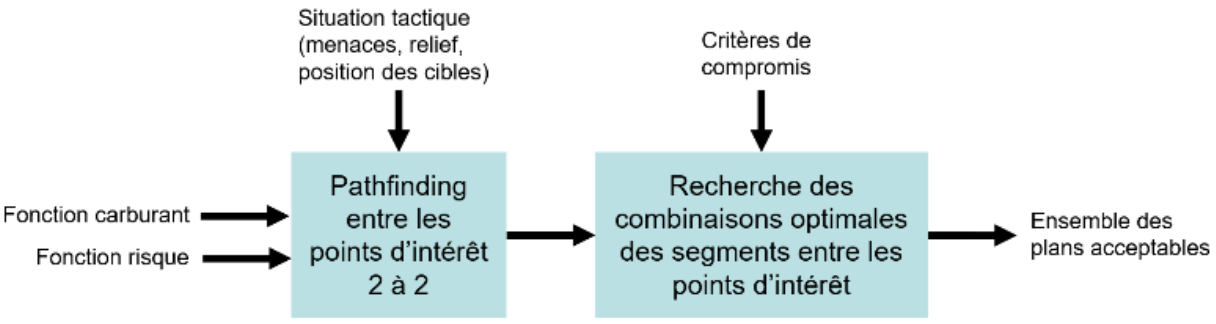


Figure 7 : Algorithme de recherche de plans à suggérer à l'opérateur implémenté dans l'environnement. Premièrement, l'algorithme recherche les segments optimaux reliant les 8 points d'intérêts 2 à 2 (les 6 cibles du terrain, la case d'entrée et la case de sortie de la zone ennemie) grâce à l'algorithme A. À chaque case du terrain est associé un coût prenant en compte des éléments liés à sa situation tactique et le poids en carburant et en risque correspondant et définis par des fonctions carburant et risque. Deuxièmement, l'algorithme recherche les combinaisons optimales de ces segments entre les points d'intérêts par une méthode de séparation et évaluation. Troisièmement, l'application des critères de compromis aux plans construits permet de délimiter l'ensemble de ceux qui sont acceptables.*

II. Déroulement d'un essai dans l'interface

L'interface est développée en langage C# avec l'interface graphique Windows Form du framework .NET. Elle est affichée aux participants sur un écran d'ordinateur avec lequel les participants interagissent grâce à une souris. Les principaux éléments qui la composent sont annotés sur la Figure 8 :

A) L'icône du drone, qui se déplace au fil du temps vers la zone ennemie horizontalement en suivant le tracé du plan.

B) Le tracé orange du plan que le drone va suivre pour effectuer la mission.

C) L'outil de parcours de l'historique des plans.

D) L'outil de replanification manuelle.

E) L'outil de consultation des suggestions de plan de l'IA.

F) L'outil de co-construction de plan.

G) Le bouton de validation d'un nouveau plan, qui termine la tâche de replanification.

H) Les trois variables de qualité du plan (nombre de cibles abandonnées, quantité de carburant consommé, risque pris) représentées par des jauges. Ces jauges sont mises à jour à chaque modification du plan.

I) Description des paramètres du solveur. Si l'opérateur a sélectionné dans l'outil de co-construction le relâchement de la prise en compte d'un critère de consommation de carburant ou de prise de risque dans l'algorithme, l'information y est indiquée.

J) La description de l'étape actuelle de l'interaction.

Les figures 9 à 15 montrent l'évolution de l'interface au fil du déroulement d'un essai.

Figure 8 : Annotation des principaux éléments de l'interface.

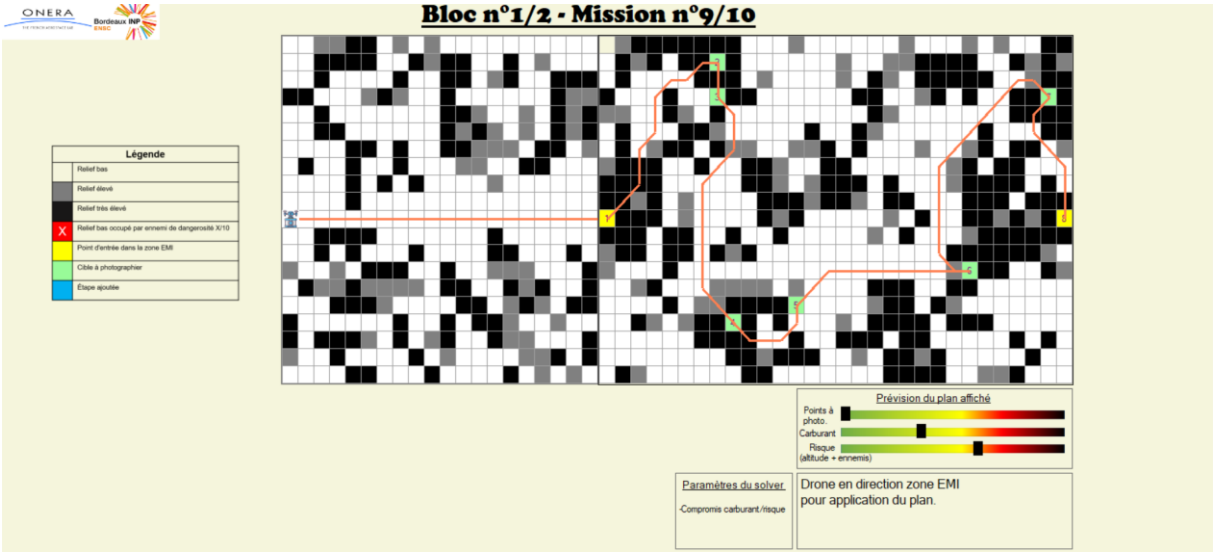


Figure 9 : Au début d'un essai, les outils de replanification ne sont pas disponibles. Les trois jauges de qualité du plan sont à des niveaux acceptables et le plan passe par toutes les cibles. Il n'y a pas encore de menaces détectées.

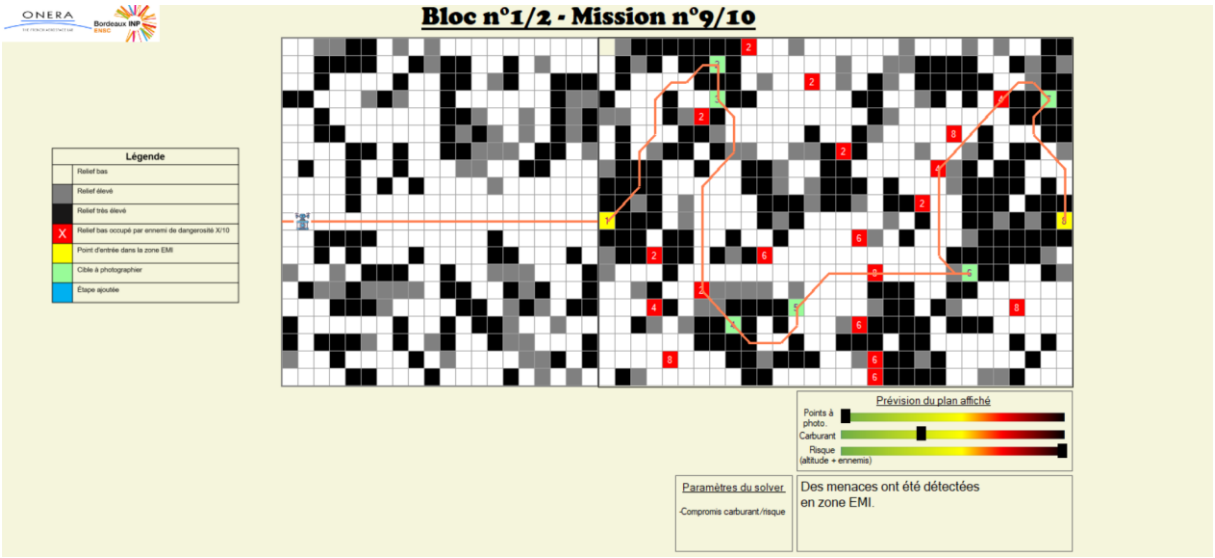


Figure 10 : Au bout de 5 secondes, les nouvelles menaces sont détectées et affichées (cases rouges). La jauge de risque est fortement dégradée car ces menaces augmentent le danger lié à la réalisation du plan.

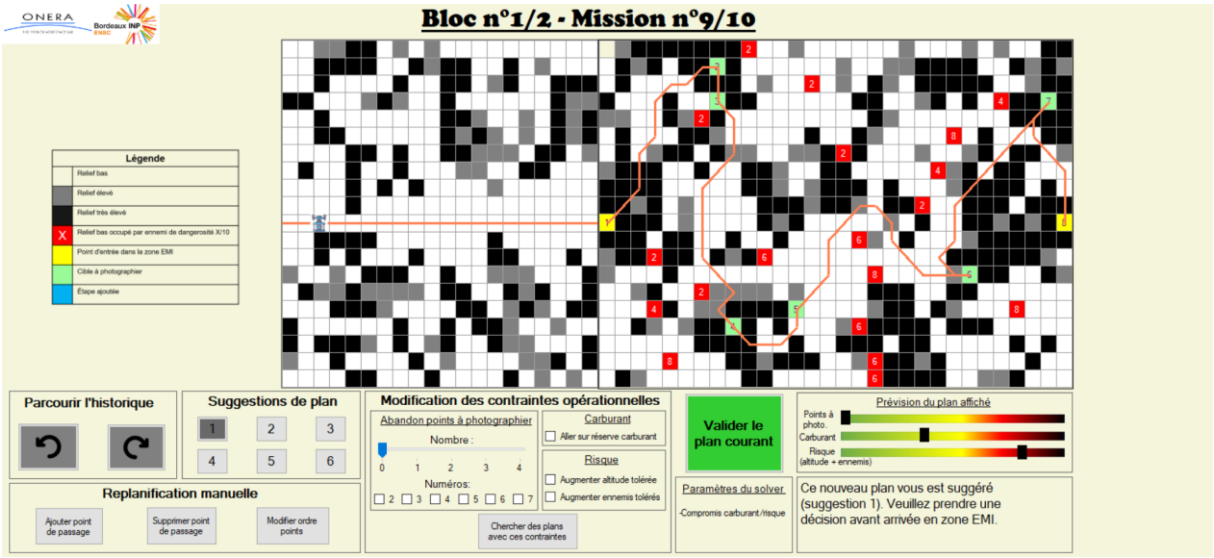


Figure 11 : Au bout de 5 secondes de plus, les outils de replanification sont affichés à l'opérateur. L'IA débute cette phase d'interaction en affichant automatiquement le plan de la première suggestion. L'opérateur peut directement valider la suggestion affichée ou bien utiliser les différents outils de replanification à sa disposition. Cette phase se termine soit en cliquant sur l'outil de validation de plan, soit lorsque le drone arrive à l'entrée de la zone ennemie au bout de 85 secondes supplémentaires.

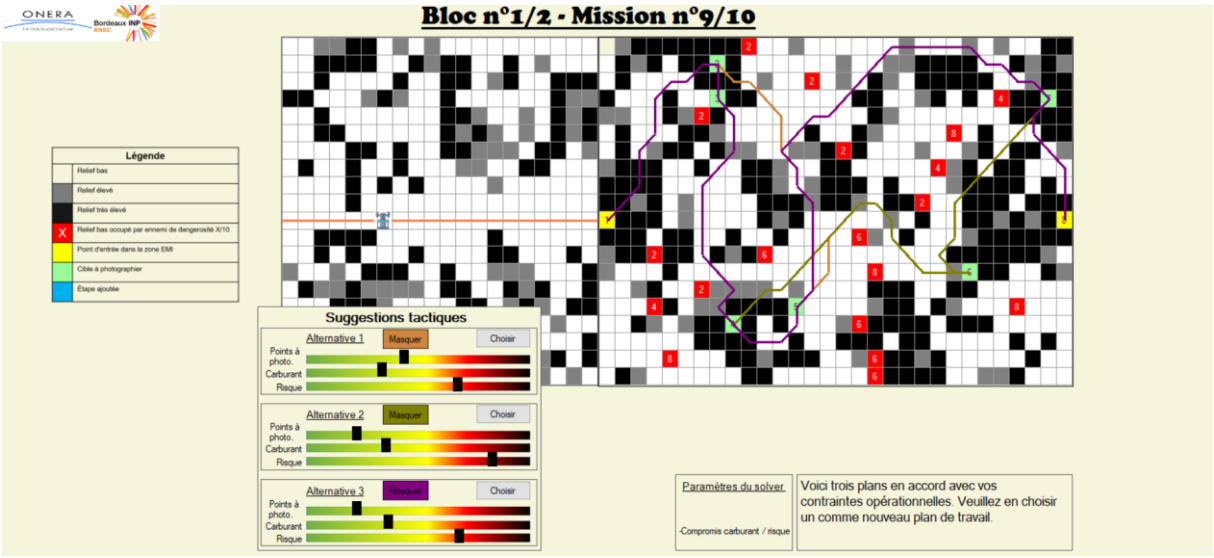


Figure 12 : Exemple d'alternatives tactiques proposées après avoir donné des critères au solveur dans l'outil de co-construction. Trois alternatives tactiques sont superposées avec des couleurs différentes. Elles sont comparables visuellement et grâce aux jauges qui décrivent leur qualité. Il est possible de masquer leur tracé grâce à un bouton.

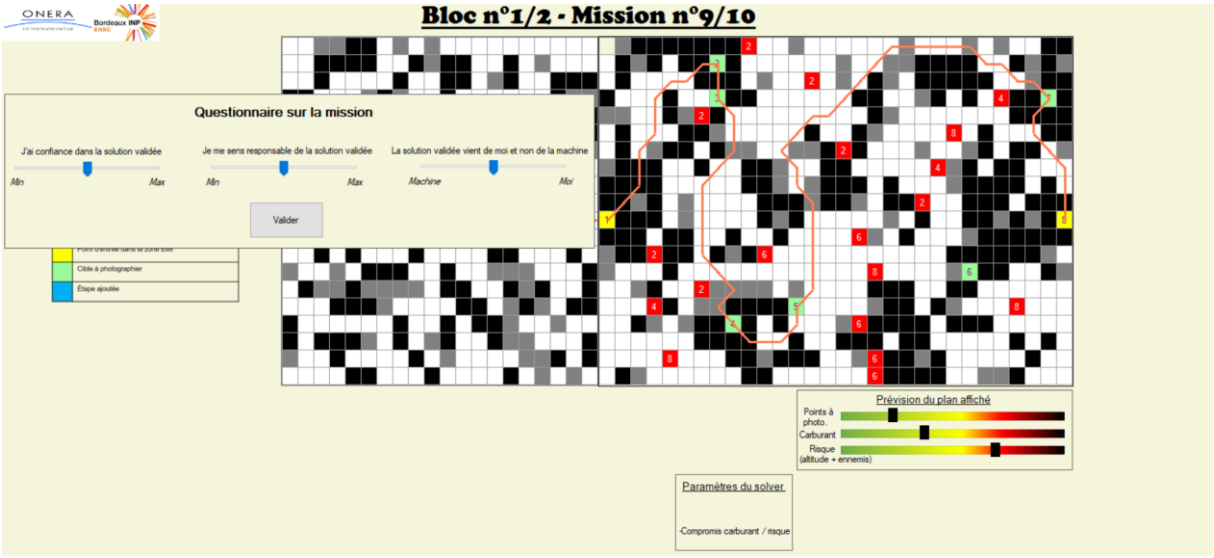


Figure 13 : Une fois l'étape de replanification terminée, l'opérateur doit répondre à trois questions sur son ressenti lié à la tâche de replanification qu'il vient de réaliser en bougeant des curseurs sur des échelles continues. Il s'agit de quantifier sa confiance, son sentiment de responsabilité et la mesure dans laquelle il se sent à l'origine du nouveau plan de vol.

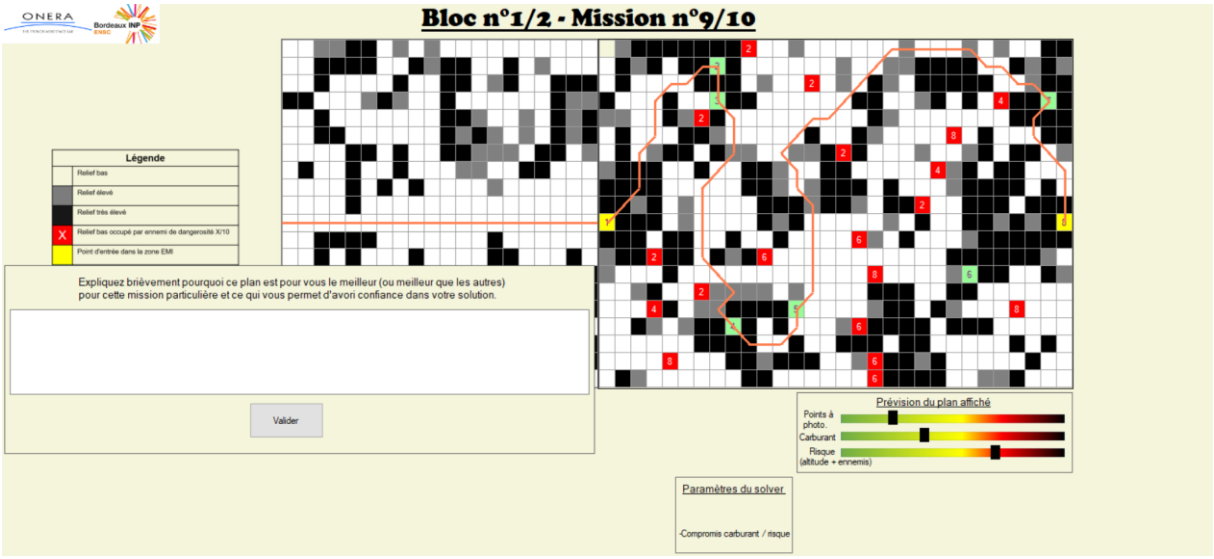


Figure 14 : Une fois les questions répondues, le participant a la possibilité de s'exprimer librement et d'écrire quelques lignes pour expliquer les raisons qui l'ont

amené à valider sa solution en particulier. Cette phase est optionnelle et permet de recueillir des données plus qualitatives qui peuvent aider à comprendre les choix qui sont faits.

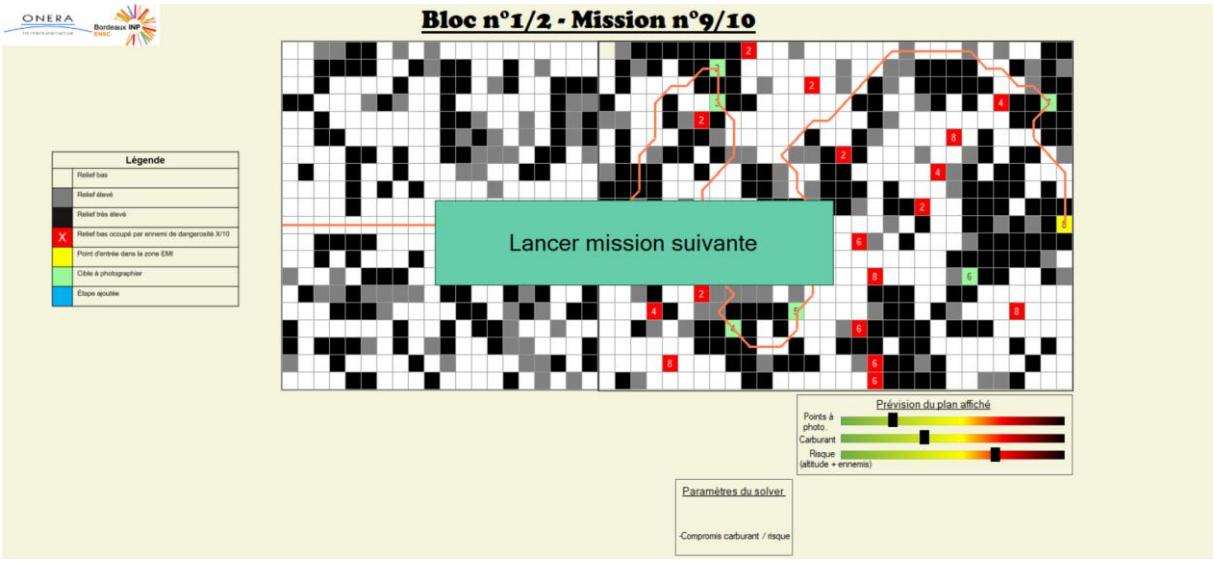


Figure 15 : L'essai est terminé et un bouton permet de lancer l'essai suivant.

III. Outils de replanification collaborative

La guerre est généralement divisée en trois niveaux : stratégique, opérationnel et tactique (Figure 16). Le niveau stratégique est le niveau de guerre auquel une nation détermine les objectifs et les orientations stratégiques de sécurité nationaux, puis développe et utilise les ressources nationales pour atteindre ces objectifs. Ces objectifs nationaux qui servent à leur tour à élaborer les objectifs et la stratégie militaires pour chaque théâtre. En termes généraux, le niveau stratégique de la guerre aborde les questions du « pourquoi » et du « avec quoi » nous allons nous battre et du « pourquoi » l'ennemi se bat contre nous. Ensuite, le niveau opérationnel de la guerre est le niveau de la guerre auquel les campagnes et les opérations majeures sont planifiées, menées et soutenues pour atteindre les objectifs stratégiques au sein des théâtres. La planification au niveau opérationnel de la guerre détermine avec « quels » plans d'action, dans « quel » ordre, pour « quelle » durée, et avec « quelles » ressources. Enfin, le niveau tactique de la guerre est le niveau de la guerre auquel les batailles et les engagements sont planifiés et exécutés pour atteindre les objectifs militaires assignés aux unités tactiques ou aux forces opérationnelles. Si ces concepts sont originaires du militaire, ils ont également

été employés par la science du management, où la stratégie consiste à fixer les priorités, et la tactique à définir comment ces priorités sont mises en place (Mackay and Zundel, 2017). Dans le cadre de cet environnement, nous faisons le choix de considérer comme stratégiques les décisions portant sur le choix des cibles à photographier, la prise de risque, la consommation de carburant et le compromis entre ces critères, et comme tactique celles portant sur l'ordre des cibles et le chemin entre les cibles. Les outils de modification de plan ont été développés pour permettre de manipuler ces deux niveaux.

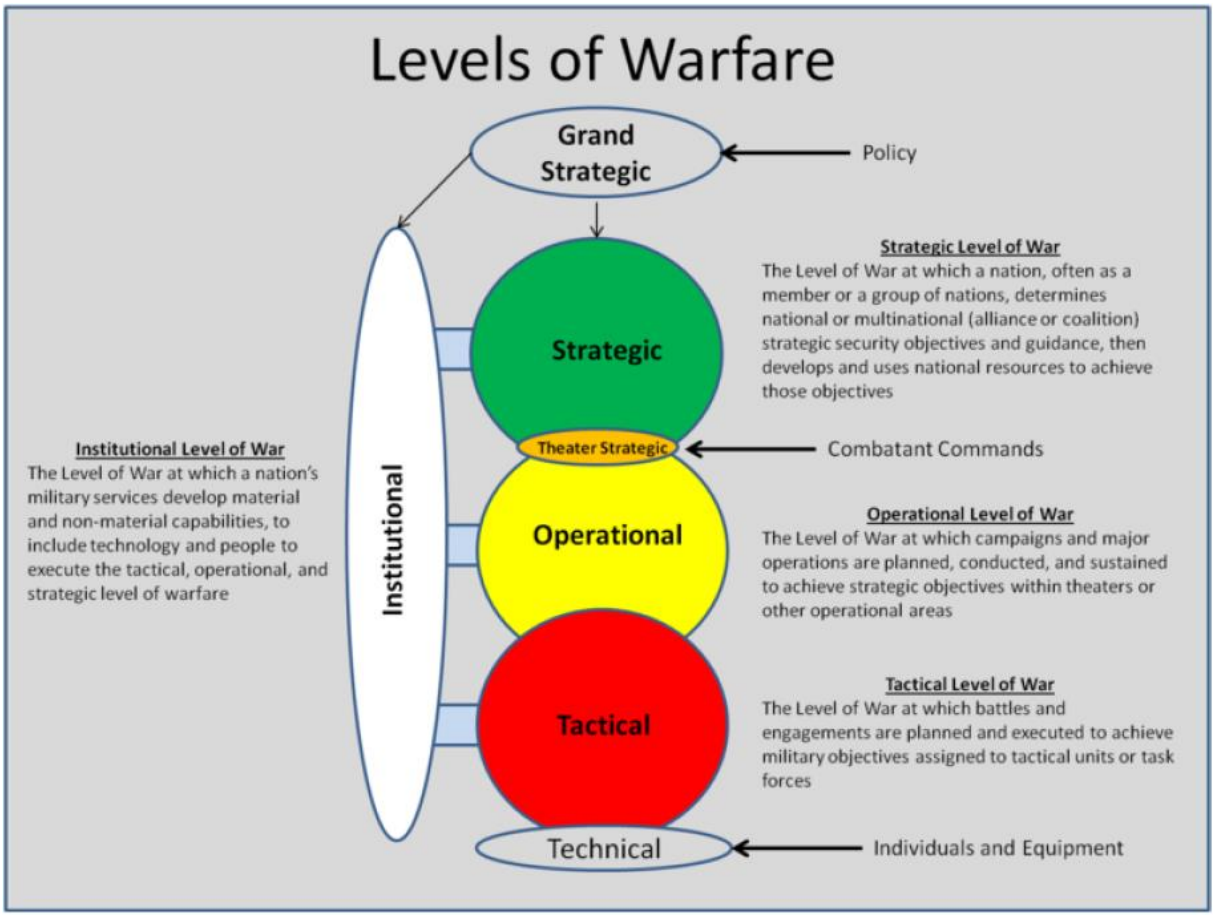


Figure 16 : Une représentation des liens entre les niveaux technique, tactique, opérationnel, stratégique de théâtre, stratégique, de grande stratégie, et institutionnel de la guerre. Extrait de Sukman, 2016.

Les outils de modification de plan implémentés dans l'environnement permettent à l'opérateur de construire le nouveau plan de vol en jouant sur des choix d'ordre stratégique et tactique grâce à différentes répartitions des tâches avec l'IA (Figure 17). La construction d'un nouveau plan nécessite des décisions d'ordre stratégique : quelles cibles conserver, quelle priorité définir entre les différentes contraintes, et des

décisions d'ordre tactique : par quelle cible commencer, quel chemin prendre… Les différents outils sont conçus pour permettre à l'IA et à l'opérateur d'intervenir sur ces deux niveaux de décision. Pour le niveau stratégique, l'opérateur peut être leader (outil C) ou bien l'IA peut être leader (outil S). De même pour le niveau tactique, l'opérateur peut être leader (outil M) ou bien l'IA peut être leader (outil C). L'ensemble de ces outils est ainsi suffisamment complexe pour ne pas limiter un type d'opérateur sur un niveau de décision, ce qui est nécessaire comme nous l'avons vu dans notre revue de littérature pour arriver à une forme de collaboration.

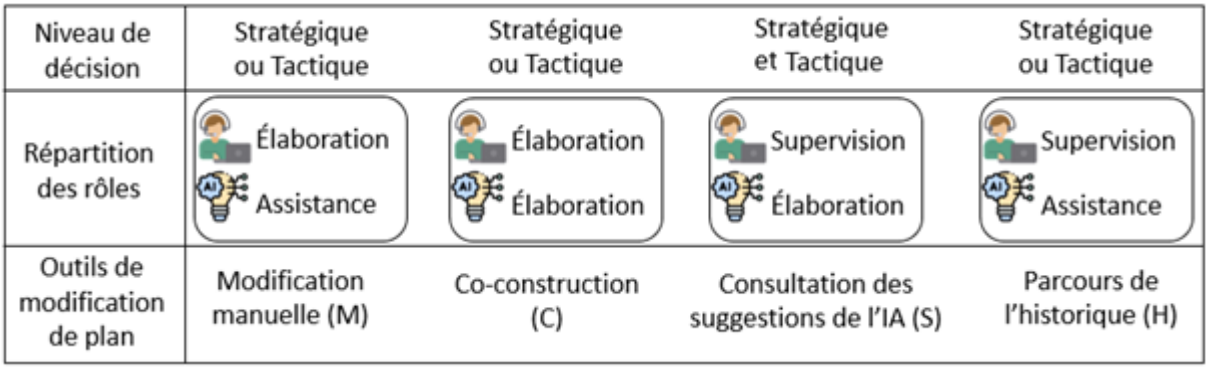


Figure 17 : Répartition des rôles entre l'opérateur et l'IA selon les différents outils de replanification à disposition.

Un outil de modification manuelle du plan (outil M) limite l'IA à une assistance calculatoire ; un outil de co-construction (outil C) repose sur une contribution mixte nécessitant des entrées à la fois de l'opérateur et du système d'IA, notamment pour la sélection des cibles à conserver et leur ordre de passage ; un outil de consultation des suggestions de l'IA (outil S) place le système en position de leader qui va suggérer des solutions globales ; et un outil de parcours de l'historique des plans construits (outil H) donne de la flexibilité à l'interaction en permettant de revenir à des trajets que les membres de l'équipe auraient précédemment construits.

Outil de consultation des suggestions de l'IA. Avec cet outil, l'opérateur se place dans un rôle de superviseur et va choisir parmi plusieurs propositions de plan de vol que le système propose. Pour cela, une fois que des nouvelles menaces sont détectées, le système met à jour la situation tactique en affichant leur emplacement et l'estimation de leur dangerosité, puis calcule leur impact sur le risque prévisionnel. Ensuite, l'IA détermine six suggestions pour réduire ce risque, par exemple en abandonnant un ou deux sites à photographier afin d'utiliser le carburant ainsi

économisé pour contourner les menaces. Dans ce cas, l'IA réalise à la fois des choix stratégiques (quelles cibles conserver, le compromis entre les critères) et des choix tactiques associés (l'ordre de passage des objectifs, et le chemin).

Outil de co-construction. À travers cet outil, l'opérateur et l'IA établissent un dialogue en plusieurs étapes pour déterminer les modifications à effectuer. Tout d'abord, l'opérateur précise les variables du plan qu'il préfère relâcher sous forme de contraintes opérationnelles à ajouter. Ces contraintes peuvent porter sur les sites à photographier, en indiquant un nombre de site à abandonner ou plus précisément le numéro de ceux que l'on souhaite retirer du plan ; sur le carburant consommé, en diminuant son poids relatif dans la valeur de coût associé aux cases ; sur le risque pris, en diminuant le poids relatif de l'altitude et/ou des ennemis. L'utilisateur réalise donc des choix stratégiques, puis, après avoir coché les contraintes qu'il souhaite appliquer, donne la main à l'IA qui détermine des plans qui les prennent en compte. Les trois meilleures alternatives tactiques correspondantes sont affichées dans l'interface ainsi que leur niveau de jauge associé. L'opérateur peut superposer ou masquer les chemins et choisit l'alternative qu'il préfère comme nouveau plan courant. On a ainsi une approche mixte dans laquelle l'opérateur effectue des choix stratégiques et l'IA des mises en œuvre tactiques. Cet outil laisse aussi la possibilité à l'IA de compléter les choix stratégiques de l'opérateur si celui-ci n'a pas spécifié explicitement les cibles à abandonner mais simplement indiqué leur nombre. L'outil offre donc une possibilité de collaboration directe sur un choix stratégique.

Outil de modification manuelle du plan. Grâce à cet outil, l'opérateur modifie manuellement le plan de vol et le système se limite à un rôle d'exécutant en affichant les changements demandés. Trois options sont disponibles : ajouter un point de passage sur le chemin, supprimer un point de passage ajouté ou un site à photographier, modifier l'ordre de passage de deux points de passage ajoutés ou sites à photographier. L'opérateur prend donc les décisions stratégiques et tactiques. Une fois que l'opérateur a indiqué l'évolution à apporter, l'IA recalcule la portion de chemin impactée et se cantonne à une assistance d'optimisation locale.

Parcours de l'historique des plans. À chaque utilisation d'outil, le plan associé est enregistré dans un historique qui peut être consulté avec les boutons d'annulation ou

de rétablissement d'action. Cet outil diffère des autres dans la mesure où il ne vise pas à créer un nouveau plan mais plutôt à donner de la flexibilité à l'interaction en permettant de revenir à des idées que les membres de l'équipe auraient précédemment proposées, à l'image d'un remue-méninge entre personnes humaines.

IV. Version simplifiée en condition coopérative ou collaborative

La première étude expérimentale, décrite dans le chapitre 3, vise à étudier l'impact du passage d'un mode de prise de décision à l'autre, coopératif et collaboratif, sur le ressenti de l'opérateur et les solutions qu'il valide. Une version simplifiée de l'interface y est utilisée, où la tâche de replanification est réalisée soit dans un mode coopératif, soit dans un mode collaboratif selon l'essai. En mode coopératif, la répartition des tâches est fixe et le choix est laissé à l'opérateur de prendre les décisions stratégiques, alors qu'en mode collaboratif l'IA peut également participer à la définition de la stratégie. Alors qu'en mode collaboratif l'opérateur reçoit au moins une suggestion de plan de niveau stratégique de la part de l'IA, dans le mode coopératif cette fonctionnalité n'est pas disponible et l'IA n'est utilisée que pour optimiser au niveau tactique à la demande de l'opérateur.

Dans le mode collaboratif, après l'apparition des menaces, l'IA prend l'initiative d'afficher automatiquement la première suggestion de plan de l'outil de consultation des suggestions. Ensuite, l'opérateur peut utiliser tous les outils de replanification présentés dans la partie 3. Ici, la répartition des rôles n'est pas figée et le système d'assistance va également proposer des changements tactiques pour la réalisation de la mission. Le système le fera de sa propre initiative au début de la phase de replanification, et pourra ensuite suggérer d'autres types de solutions (modification de points de passage, d'objectifs conservés ou supprimés…) à la demande de l'opérateur (Figure 18). Le système peut également demander à l'opérateur de confirmer la classification d'une image pour laquelle pour laquelle l'incertitude serait élevée, et dont le niveau de dangerosité s'il était différent pourrait impliquer des modifications au plan (Figure 19). Le numéro indiquant la dangerosité est supposé être extrait de manière automatique par un système de reconnaissance d'image et

des erreurs de classification sont toujours possibles, d'où l'idée de pouvoir effectuer une vérification croisée par l'opérateur en cas de doute. Les outils présents en mode coopératif sont également présents dans ce mode d'interaction plus avancé, mais les solutions proposées restent les mêmes dans les deux cas, l'algorithme d'optimisation des contraintes étant pour sa part inchangé.

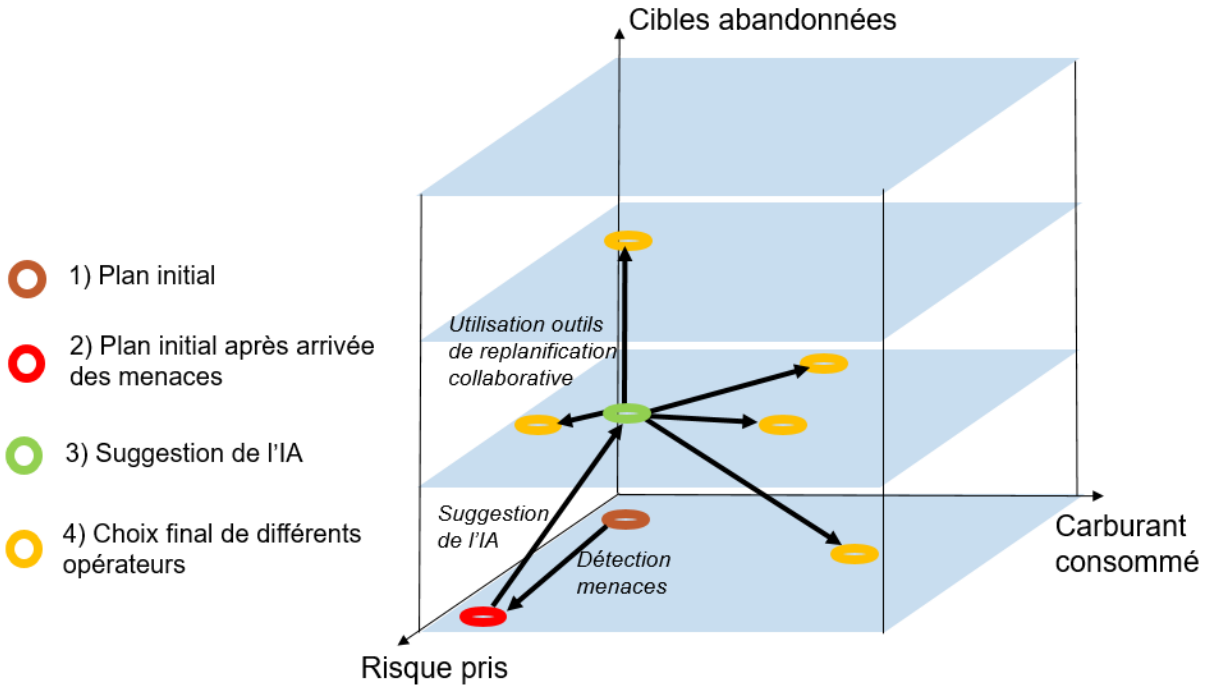


Figure 18 : Évolution de la qualité du plan au fil du déroulement d'un essai en condition collaborative.

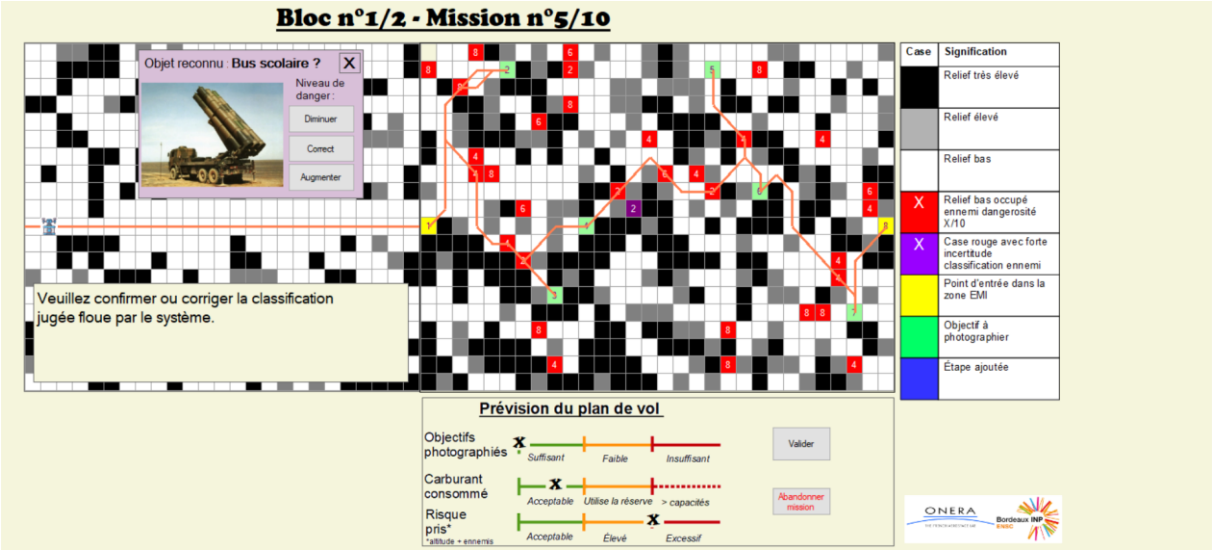


Figure 19 : Fonctionnalité de vérification de la classification d'un ennemi par l'opérateur. Les menaces dont la case est de couleur violette dans l'environnement simplifié ont une classification jugée peu certaine par l'IA. L'opérateur a la possibilité

de vérifier la classification en cliquant sur la case, ce qui affiche une photographie prise de la menace et le nom de l'objet reconnu. L'opérateur peut valider la classification faite s'il la considère juste, ou bien augmenter ou diminuer le niveau de danger lié à la case si celle-ci n'est pas appropriée. Dans ce cas, la valeur de danger inscrite sur la case de la menace est mise à jour et cette évolution de la situation tactique sera prise en compte par l'algorithme lors des prochaines utilisations d'outil de replanification.

Dans le mode coopératif, après l'apparition des menaces, l'IA n'affiche pas de nouveau plan. L'opérateur doit replanifier en utilisant l'outil de modification manuelle ou l'outil de co-construction, et n'a pas accès à l'outil de consultation des suggestions de l'IA (Figure 20). Dans ce mode de fonctionnement, les rôles entre l'opérateur et le système d'assistance sont fixes : l'opérateur définit les éléments stratégiques de haut niveau qui obligent à la modification du plan (points de passage, objectifs retirés de la mission, etc.) puis le système d'assistance produit la trajectoire optimisée en tenant compte de ces éléments.

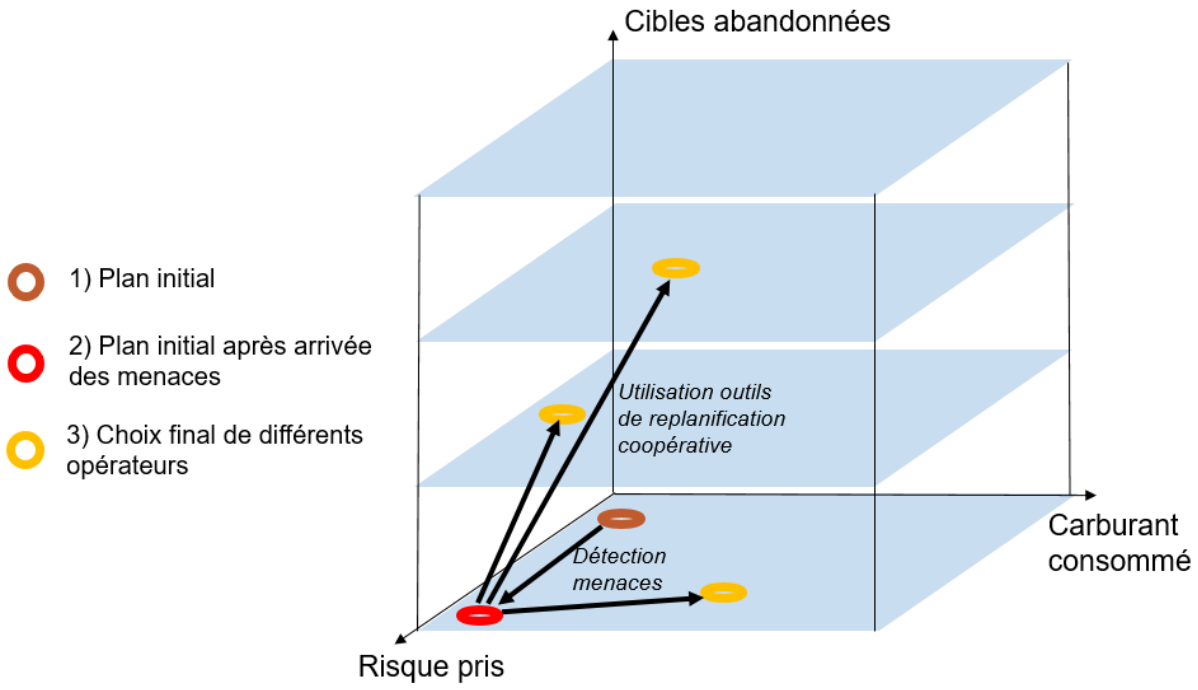


Figure 20 : Évolution de la qualité du plan au fil du déroulement d'un essai en condition coopérative.

Ainsi, nous avons mis en œuvre les modes d'interactions coopératif et collaboratif identifiés dans l'état de l'art à travers deux versions de notre environnement de

simulation. Dans le mode coopératif, la répartition des tâches entre l'opérateur et l'IA est figée. L'opérateur prend les décisions stratégiques et tactiques principales et l'IA est appelée sur la tâche d'optimisation locale du trajet en fonction du réglage des critères de choix. Dans le mode collaboration, l'opérateur mais aussi l'IA ont la possibilité de faire des propositions de choix stratégiques et tactiques. L'optimisation locale reste à la charge de l'IA. L'IA fait toujours une proposition stratégique au départ et en a d'autres que l'opérateur peut consulter s'il le souhaite. Le choix a été fait que l'opérateur garde la main sur le processus d'interaction (il choisit les outils à utiliser) mais le rôle et les tâches attribuées aux deux intervenants varient selon l'outil choisi. Ce mode plus collaboratif mais relativement contrôlé permet d'étudier quel rôle l'opérateur souhaite laisser à l'IA. C'est une étape vers le développement d'une interface de collaboration plus forte où l'IA pourrait prendre plus d'initiatives à tout moment du processus (comme afficher une nouvelle suggestion même si cela n'est pas demandé par l'opérateur).

Chapitre 3 : Étude de l’impact des suggestions de l’IA sur la prise de décision et le ressenti de l’opérateur

I. Méthode

1. Hypothèses

L'objectif de cette première étude est d’explorer la problématique 1 : dans quelle mesure l’augmentation de la participation de l’IA à la prise de décision va-t-elle modifier les solutions validées par l’opérateur et son ressenti ?

Il s’agit de mieux comprendre, de manière expérimentale, comment les modes d'interaction entre un opérateur humain et une IA peuvent impacter la décision résultant du processus interactif et comment cette décision est considérée et acceptée par l'opérateur humain. Nous explorons deux modes d’interaction : un qui se base sur une structure coopérative de l’équipe opérateur-IA, où la répartition des tâches est figée, et l’autre une structure plus collaborative. Nos hypothèses sont :

H1 : L'évolution du mode d'interaction a un impact sur le ressenti de l'opérateur par rapport au processus de décision.

H2 : Le mode d'interaction a un impact sur la variabilité des décisions prises par un ensemble d’opérateurs.

2. Plan expérimental

Un groupe de 20 personnes, tous doctorants ou jeunes ingénieurs en bonne santé (45% de femmes), d'un âge moyen de 27.3 ans (écart-type = 2.9 ans), a participé à cette étude. Tous les participants ont effectué 20 missions sur les mêmes 20 terrains présentés dans le même ordre. La moitié des participants a effectué les 10 premières missions en mode d’interaction coopératif, et les 10 dernières en mode d’interaction collaboratif. Ces participants ont donc subi une augmentation de l’assistance du système pour la tâche de replanification au cours du temps (Figure 21). L'autre moitié des participants a effectué les 10 premières missions en mode d’interaction coopérative, et les 10 dernières en mode d’interaction coopérative. Ces participants ont à l’inverse subi une diminution de l’assistance du système dans la replanification au cours de l’étude. Ce plan expérimental est un plan mixte où on considère deux groupes qui passent les deux modalités de la variable que l’on

contrôle mais en contrebalançant l'ordre de ces modalités. L'analyse des données a consisté à réaliser des analyses de la variance (ANOVA) à mesures répétées pour mesurer l'impact de l'évolution de l'assistance sur des métriques de ressenti de l'opérateur par rapport à sa décision, puis à calculer les corrélations de Pearson entre ces métriques pour déterminer si celles-ci sont corrélées. Ensuite, pour évaluer comment les différents modes d'interactions influencent la prise de décision, nous avons représenté les plans validés sous forme de séquences de cibles, et mesuré leur variabilité grâce à la distance de Damerau-Levenshtein.

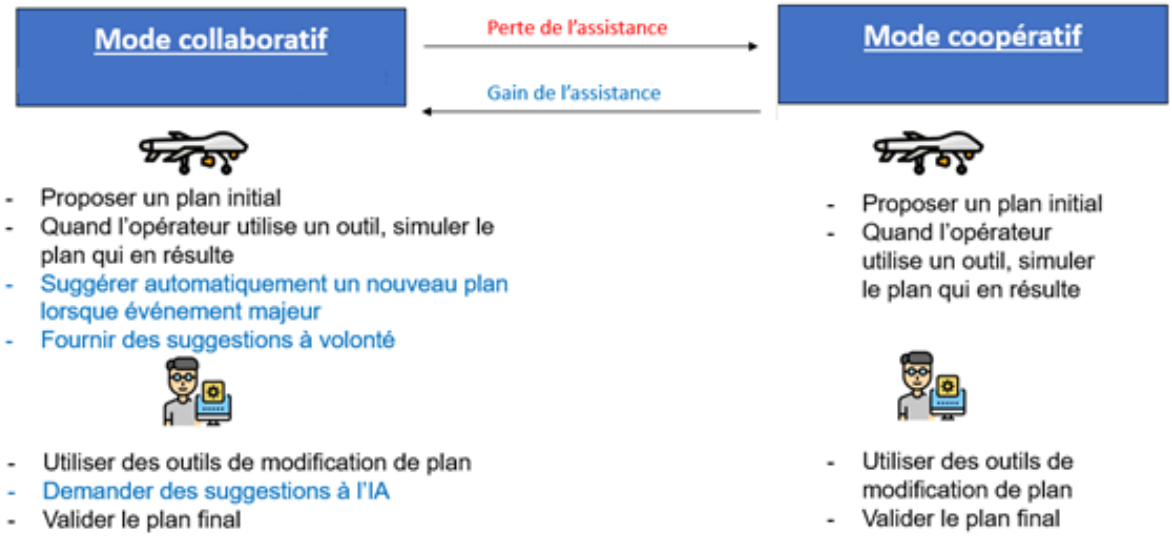


Figure 21 : Évolution de l'assistance à la replanification fournie par l'IA selon le changement du mode d'interaction : perte de l'assistance lors du passage de la collaboration à la coopération, et gain de l'assistance lors du passage de la coopération à la collaboration.

En condition coopérative, la répartition des rôles entre l'opérateur et le système d'assistance est fixée : l'opérateur définit les éléments tactiques de haut niveau qui obligent à modifier le plan (points de passage, objectifs retirés de la mission...) puis le système d'assistance produit la trajectoire optimisée en tenant compte de ces éléments.

En condition collaborative, à l'inverse, la répartition des rôles n'est pas figée et le système d'assistance va également proposer des modifications stratégiques pour la réalisation de la mission. Le système le fera de sa propre initiative au début de la phase de replanification, et pourra ensuite proposer d'autres types de solutions

(modification des points de passage, des objectifs retenus ou supprimés…) à la demande de l'opérateur. Le système peut également demander à l'opérateur de confirmer la classification d'une image pour laquelle l'incertitude serait élevée, et dont le niveau de dangerosité s'il était différent pourrait impliquer des modifications du plan.

3. Métriques

Mesures de l'impact de l'évolution du mode d'interaction humain-IA sur le ressenti de l'opérateur par rapport au processus décisionnel. Afin de répondre à H1, nous avons mesuré le ressenti de l'opérateur par rapport à leur prise de décision à travers quatre dimensions. Lors de chaque mission, après avoir validé un nouveau plan, les participants devaient répondre à trois questions étalonnées sur des échelles de Likert à 7 items :

- sur leur confiance dans la solution validée (Figure 22).
- sur leur sentiment de responsabilité dans la solution validée.
- sur leur sentiment de paternité dans la solution validée, c'est-à-dire selon l'opérateur qui de lui ou de l'IA a le plus participé à sa conception (le score 1 correspondant au sentiment que le plan vient entièrement de l'IA, et le score 7 que le plan vient entièrement de l'opérateur).

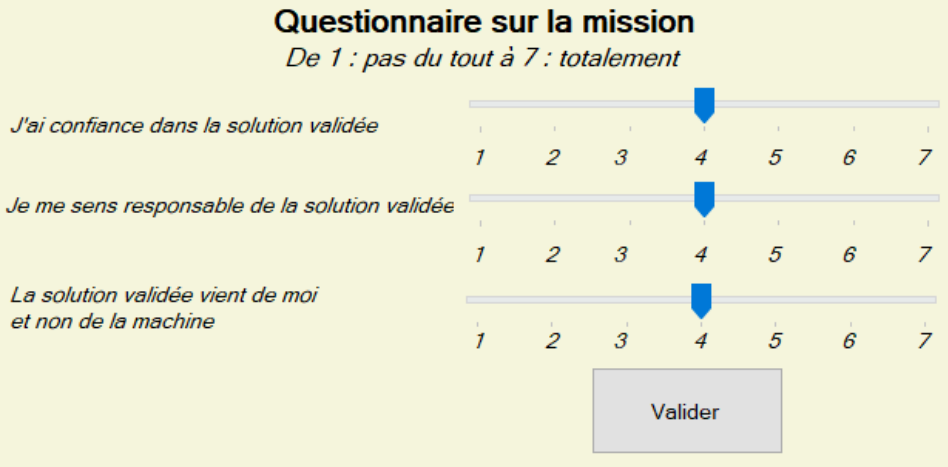


Figure 22 : Questionnaire posé dans l'interface après chaque replanification pour mesurer la confiance, la responsabilité et la paternité ressenties par l'opérateur dans sa prise de décision.

Après avoir effectué toutes les missions dans un mode d'interaction, les participants répondaient ensuite à un questionnaire NASA-TLX destiné à mesurer leur charge de

travail perçue pour la tâche (Hart, 2006 ; Cegarra et Morgado, 2009). Le détail des questions est disponible en Annexe 2.

Mesures de la variabilité des décisions prises selon le mode d'interaction humain-IA. Afin de répondre à H2, nous avons décidé de comparer la variabilité des décisions prises en nous concentrant sur les différences de choix stratégiques, c'est-à-dire sur les cibles par lesquelles les plans validés passent, ainsi que leur ordre. Les missions ont pour objectif de passer par six cibles que nous numérotons de 1 à 6 par commodité. À chaque plan concevable dans les différents terrains peut être associé une séquence de chiffres qui correspond à l'ordre prévisionnel de passage entre ces points d'intérêt. Par exemple, un plan auquel serait associée la séquence de points d'intérêt « 13256 » passerait par le point d'entrée dans la zone ennemie, les cibles 1 puis 3 puis 2 puis 5 puis 6. Cette séquence ne contient pas le numéro 4, ce qui signifie que la cible associée a été abandonnée.

Pour comparer plusieurs séquences entre-elles afin de mesurer la variance entre les plans, nous avons utilisé la distance de Damerau-Levenshtein (Damerau, 1964). Originaire de la théorie de l'information, cette métrique mesure la distance entre deux mots à travers le calcul du nombre minimal d'opérations nécessaires pour transformer un mot en l'autre (Brill and Moore, 2000). L'algorithme de Damerau-Leventhstein trouve ainsi la combinaison la moins couteuse d'insertions, de suppressions, de substitutions, et de transposition qui serait nécessaire pour transformer une séquence en l'autre. Nous avons attribué un poids de 1 aux opérations d'insertion et de suppression, et un poids de 2 à l'opération de substitution, qui peut se comprendre comme une d'insertion puis une suppression ou inversement. Par ailleurs, nous avons généralisé l'opérateur de substitution, qui initialement inverse deux caractères adjacents, pour qu'il puisse inverser deux caractères peu importe leur distance dans la séquence. Nous faisons le choix ici de considérer deux plans qui contiennent les mêmes cibles mais dans un ordre différent comme suffisamment similaires stratégiquement pour attribuer à l'opération de substitution généralisée un coût de 0. Cette simplification, qui consiste à ne considérer comme vraiment différents que des plans qui ne passent pas par les mêmes cibles, permet à l'algorithme de mesurer si deux séquences présentent des différences stratégiques majeures. Ainsi, la version que nous avons implémentée de

l’algorithme calcule par exemple une distance de 1 entre les séquences « 24567 » et « 234567 », les deux étant séparées par une opération d’insertion du caractère « 3 » de coût 1. Si nous prenons un autre exemple, les séquences « 234567 » et « 734562 » ont une distance de 0, n’étant séparées que d’une opération de substitution entre les caractères « 2 » et « 7 » de coût 0.

II. Résultats

1. Impact de l'évolution du mode d'interaction sur le ressenti de l'opérateur.

Nous avons effectué une ANOVA à mesures répétées pour comparer l'effet de l'évolution du mode d'interaction (coopération puis collaboration, ou collaboration puis coopération) sur les métriques du sentiment de l'opérateur vis-à-vis de la solution choisie (sentiment de responsabilité, sentiment de paternité, confiance et charge de travail) (Figure 23). Le résultat formel de l’ANOVA est présenté Figure 24.

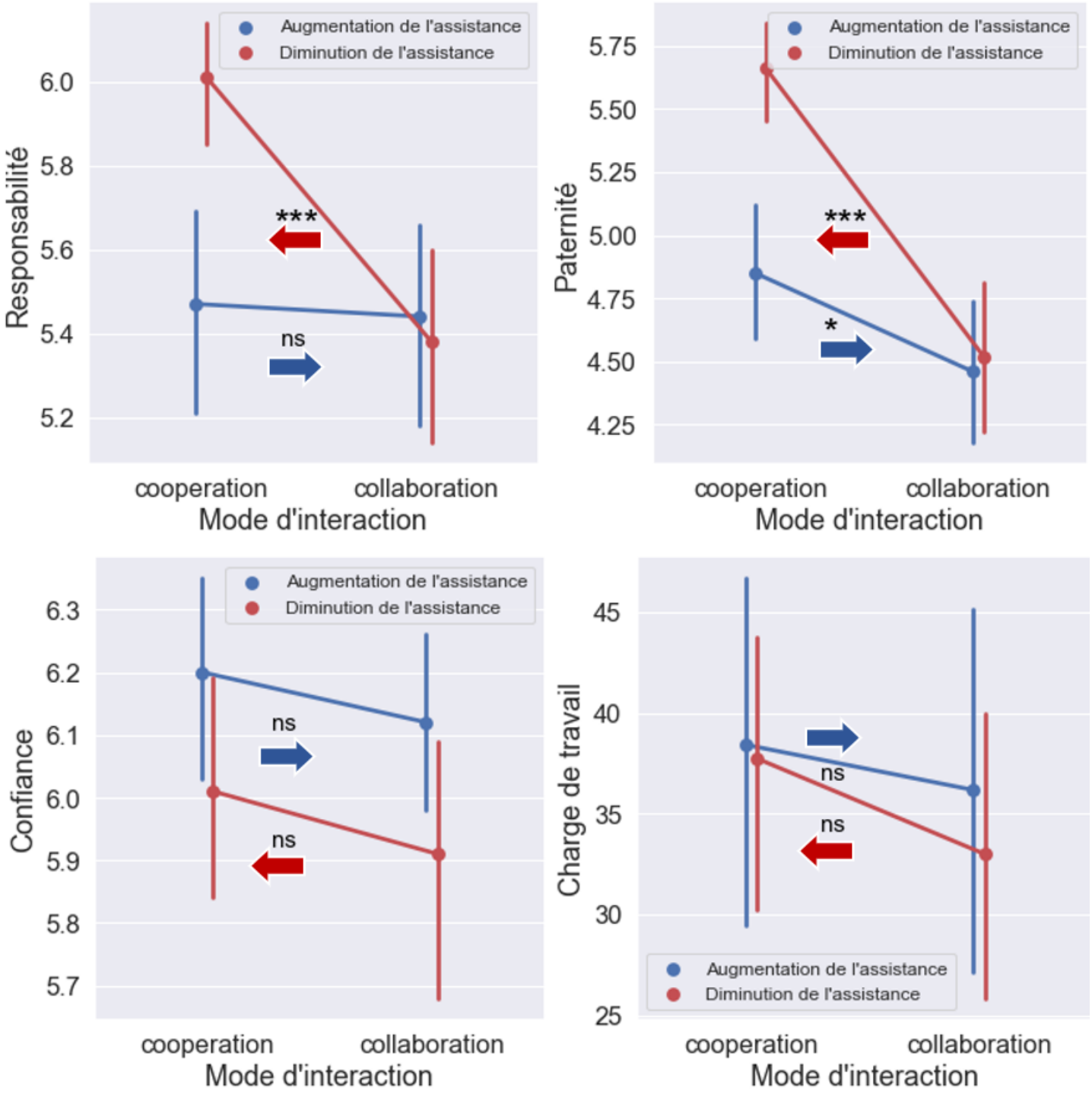


Figure 23 : Variation observée sur les sentiments de responsabilité, de paternité, de confiance et de charge de travail des participants selon l'évolution du mode d'interaction : augmentation de l'assistance lorsque la flèche est rouge, diminution de l'assistance lorsque la flèche est bleue.

Pour les participants qui ont subi une diminution de l'assistance en passant d'une collaboration à une coopération avec l'IA, nous observons une diminution

significative de leur sentiment de responsabilité et ainsi que de leur sentiment de paternité de la décision, mais pas de changement significatif de leur charge de travail ni de confiance dans la décision (Figure 24). Certains de ces résultats sont en accord avec notre hypothèse car ils montrent comme prévu que la transformation de la distribution des tâches avec l'IA influence l'expérience de l'opérateur. Cependant, chez les participants qui ont subi une augmentation de l'assistance en passant d'une coopération à une collaboration avec l'IA, nous observons une diminution significative mais faible de leur sentiment de paternité uniquement, et aucun changement significatif ni de leur sentiment de responsabilité, ni de charge de travail ni de confiance dans la décision. Cette évolution asymétrique n'est pas conforme à notre hypothèse.

	Diminution de l'assistance (collaboration puis coopération)				Augmentation de l'assistance (coopération puis collaboration)			
Evolution de la responsabilité		sum_sq	df	F	PR(>F)	sum_sq	df	F	PR(>F)
	Assistance	19.845000	1.0	23.643282	2.373562e-06	0.020000	1.0	0.013634	0.907167
	Participant	5.832042	1.0	6.948280	9.061781e-03	6.870705	1.0	4.683619	0.031661
	Assistance:Participant	22.205262	1.0	26.455291	6.515735e-07	5.564125	1.0	3.792950	0.052897
	Residual	164.512696	196.0	NaN	NaN	287.525170	196.0	NaN	NaN
Evolution de la paternité		sum_sq	df	F	PR(>F)	sum_sq	df	F	PR(>F)
	Assistance	64.980000	1.0	45.029235	2.035781e-10	3.125000	1.0	1.649793	0.200503
	Participant	3.142991	1.0	2.178001	1.416009e-01	36.516118	1.0	19.278095	0.000018
	Assistance:Participant	31.416760	1.0	21.770894	5.682330e-06	7.975251	1.0	4.210405	0.041506
	Residual	282.840249	196.0	NaN	NaN	371.258631	196.0	NaN	NaN
Evolution de la confiance		sum_sq	df	F	PR(>F)	sum_sq	df	F	PR(>F)
	Assistance	0.500000	1.0	0.518237	0.472452	0.020000	1.0	0.037127	0.847407
	Participant	0.600844	1.0	0.622760	0.430977	6.556011	1.0	12.170142	0.000600
	Assistance:Participant	1.476603	1.0	1.530462	0.217523	0.619510	1.0	1.150017	0.284864
	Residual	189.102552	196.0	NaN	NaN	105.584479	196.0	NaN	NaN
Evolution de la charge de travail		sum_sq	df	F	PR(>F)	sum_sq	df	F	PR(>F)
	Assistance	112.812500	1.0	0.714718	0.410347	25.312500	1.0	0.102849	0.752588
	Participant	17.696335	1.0	0.112114	0.742099	10.909924	1.0	0.044329	0.835901
	Assistance:Participant	0.235602	1.0	0.001493	0.969660	29.940498	1.0	0.121653	0.731796
	Residual	2525.470841	16.0	NaN	NaN	3937.830133	16.0	NaN	NaN

Figure 24 : Résultat des ANOVA à mesures répétées mesurant la significativité de l'effet de l'évolution du mode d'interaction (augmentation ou diminution de l'assistance) sur le sentiment de responsabilité, de paternité, de confiance et de charge de travail des participants.

Nous nous sommes ensuite intéressés aux liens que peuvent entretenir ces différentes variables qui mesurent le rapport des opérateurs à la décision validée selon l'évolution de l'IA. Nous avons moyenné pour chaque participant son score sur chacune de ces variables dans chaque condition. Ensuite nous avons réalisé une Analyse en Composante Principale (ACP) pour ces variables en condition collaborative, et une autre pour les mêmes variables en condition coopérative. L'ACP est une méthode d'analyse des données qui consiste à transformer des variables liées entre elles en nouvelles variables décorrélées les unes des autres afin de résumer l'information en réduisant le nombre de variables.

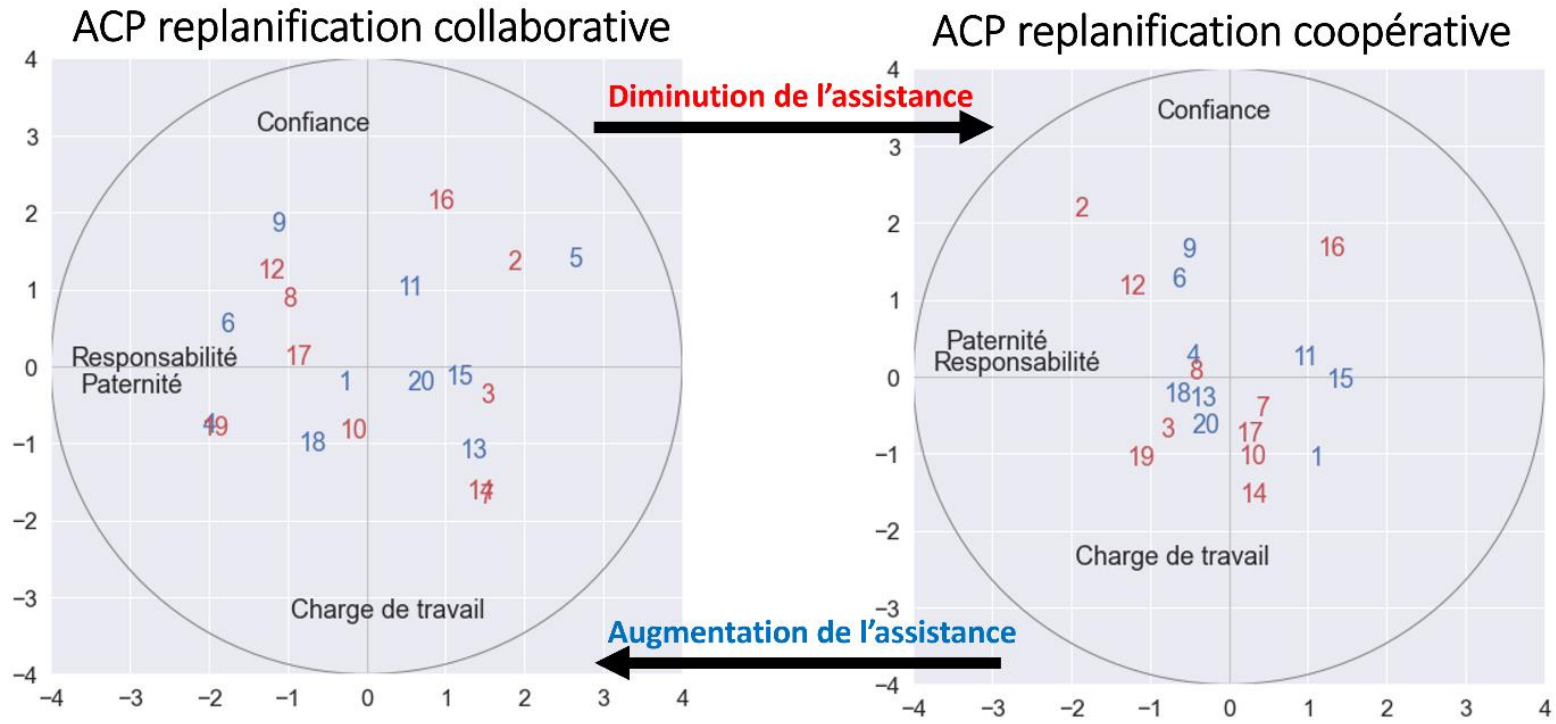


Figure 25 : Cercles des corrélations correspondants aux ACP de la confiance, responsabilité, paternité et charge de travail moyenne de chaque participant en condition collaborative (à gauche) et coopérative (à droite). Chaque numéro identifie un participant.

Les cercles des corrélations tracent le premier plan factoriel de ces deux ACP (Figure 25). Chaque numéro renvoie à un participant et à son niveau moyen pour ces variables dans la condition. Les numéros en rouge correspondent aux participants qui ont commencé en condition collaborative puis qui ont terminé l'étude en condition coopérative. Les numéros en bleus correspondent aux participants qui ont commencé en condition coopérative puis qui ont terminé l'étude en condition

collaborative. Nous observons tout d'abord que les deux cercles des corrélations présentent visuellement dans la même échelle une composition des axes qui est similaire. En condition coopérative comme en condition collaborative, le premier axe du cercle des corrélations montre une corrélation positive importante en le sentiment de responsabilité et le sentiment de paternité des participants. Les participants se sentent responsables de la décision prise dans la mesure où ils considèrent en être à l'origine. Le second axe montre une corrélation négative entre confiance et charge de travail. Plus les participants ont fourni un effort cognitif pour l'élaboration du plan avec l'IA, moins ils ont confiance en ce plan. Les deux axes étant visuellement orthogonaux, ces deux couples de variables ne sont pas corrélés entre eux. Néanmoins, une analyse statistique plus poussée de la significativité des corrélations de Pearson entre ces variables précise que seule la corrélation positive négative entre les variables de paternité et de responsabilité est significative (Figure 26).

Replanification collaborative

	Responsabilité	Confiance	Paternité	Charge de travail
Responsabilité	2.433532e-142	2.579878e-01	0.000104	0.469981
Confiance	2.579878e-01	2.433532e-142	0.565666	0.276855
Paternité	1.037718e-04	5.656659e-01	0.000000	0.510448
Charge de travail	4.699805e-01	2.768555e-01	0.510448	0.000000

Replanification coopérative

	Responsabilité	Confiance	Paternité	Charge de travail
Responsabilité	2.433532e-142	0.505027	0.000032	0.149882
Confiance	5.050269e-01	0.000000	0.574935	0.699977
Paternité	3.245463e-05	0.574935	0.000000	0.399657
Charge de travail	1.498819e-01	0.699977	0.399657	0.000000

Figure 26 : P-value des corrélations de Pearson des variables pour les deux conditions.

2. Impact du mode d'interaction sur la prise de décision de l'opérateur.

La Figure 27 représente pour les 20 terrains la séquence de plan validée par chacun des 20 opérateurs. Par exemple, le participant 1 a validé la séquence « 142635 » pour le terrain 1, ce qui signifie que le plan qu'il a choisi pour cette mission passe successivement par les cibles numérotées 1 puis 4 puis 2 puis 6 puis 3 puis 5. Le code couleur des participants est similaire à celui des figures précédentes : les participants en bleu ont réalisé la tâche de replanification sur les terrains 1 à 10 en condition coopérative puis sur les terrains 10 à 20 en condition collaborative, et réciproquement pour les participants en rouge. Pour mesurer si la variabilité des séquences est différente selon la condition, nous avons calculé pour chaque terrain la moyenne de la distance de Levenshtein entre les 10 séquences qui correspondent

à chaque condition. Pour vérifier si cette tendance est statistiquement significative sur l'ensemble des 20 terrains, nous avons réalisé un test de la somme des rangs de Wilcoxon sur la série des 20 « distances moyennes » en condition coopérative et la série des 20 « distances moyennes » en condition collaborative, avec l'hypothèse unilatérale que la série en condition coopérative est supérieure à la série en condition collaborative. Le test rejette l'hypothèse nulle (F = 265, p-value = 0.04), ce qui indique que les plans validés en condition collaborative sont plus homogènes que les plans validés en condition coopérative. En condition collaborative, lorsque l'IA fait des suggestions de plans, les plans validés par les opérateurs sont plus similaires entre eux qu'en condition coopérative, lorsque l'IA ne fait pas de suggestion de plan. La participation accrue de l'IA d'assistance à la prise de décision semble donc avoir un effet de réduction de la variabilité des réponses humaines.

	Ter1	Ter2	Ter3	Ter4	Ter5	Ter6	Ter7	Ter8	Ter9	Ter10	Ter11	Ter12	Ter13	Ter14	Ter15	Ter16	Ter17	Ter18	Ter19	Ter20
Part1	142635	64523	146253	431562	54236	565213	234154	425136	412536	53621	135624	513462	145236	42531	513642	425136	165243	415632	513462	356412
Part2	14235	16452	612453	456213	14236	2536	53164	24136	145236	53162	563214	54362	45236	243561	153642	423516	165243	415362	153462	356142
Part3	14235	136452	16453	43162	15436	42156	52416	24536	145236	5312	53624	51362	14236	243561	153642	43516	165243	415632	13462	356412
Part4	142635	163452	612453	431562	125436	421536	523146	245136	124536	53162	135624	513462	145236	234561	135642	452316	165234	415632	13462	356412
Part5	142635	163452	614253	431562	154236	421536	523146	245136	145236	53162	513624	514362	145236	243561	153642	423156	165234	415632	153462	356412
Part6	142635	63452	14253	453162	154236	421536	523416	245136	142536	53162	156324	514362	145236	243561	153642	423516	165243	415632	154362	356412
Part7	14235	16452	61253	43562	15436	41536	52316	42536	14236	53162	53624	513642	415236	23561	53642	43516	65243	41362	13462	35614
Part8	146235	136425	612453	453126	154236	421536	523416	245316	145236	53162	513264	154362	415236	243156	513642	423516	651234	413265	513462	315642
Part9	142635	16345	614253	431652	154236	421536	523146	245136	142536	53162	135624	514362	145236	243561	153642	423516	165243	45362	153462	356142
Part10	142635	136452	12453	453162	154236	42136	523146	24563	145236	53162	513246	514632	145236	243561	153642	423516	165234	41532	13462	356412
Part11	42635	163452	612453	43152	54236	421536	23461	425136	145236	53162	15324	41362	14526	24316	15642	423156	165243	413265	15326	315642
Part12	14235	136452	612453	456312	154236	421536	23146	245136	145236	53126	563124	513462	145236	243561	135642	423156	165243	456132	13462	356421
Part13	142635	163452	612453	453162	154236	421563	523416	425136	142536	53162	15624	514362	145236	243561	13642	42516	16543	415632	13462	356142
Part14	142365	13452	612453	431562	15436	42136	23146	21356	145236	53162	56324	54362	45236	243561	13642	42356	65243	41326	13462	356412
Part15	142635	163452	641253	431562	154236	421536	254316	245136	142536	53612	156324	514362	451236	234561	153642	423561	165243	415623	541362	356142
Part16	124635	136452	126453	453126	154236	421356	523146	42563	145236	53612	513624	514362	145236	243561	135642	423561	165234	413256	513462	356412
Part17	14635	13452	126453	453162	145326	421536	523146	245136	145236	53162	513624	514362	145236	243561	153642	423516	165243	413562	153462	356412
Part18	124635	136452	612453	45312	154236	421536	523146	24563	14523	53164	56324	51362	145236	24561	15642	45316	15243	412356	13462	356412
Part19	412635	13452	12453	431562	541236	421536	52346	452136	145236	53162	513624	514362	145236	245631	135642	423561	165234	45632	56432	356412
Part20	142365	163452	612453	431562	145623	421536	523146	245316	145236	53162	135624	514362	145236	243561	15642	423156	165243	415632	153462	312564
Distance Coopération	1.0	1.47	2.13	0.8	1.07	2.0	1.0	2.0	0.8	0.0	0.8	0.2	0.2	1.0	0.8	0.8	0.8	1.73	1.53	0.8
Distance Collaboration	0.8	1.4	0.8	0.8	1.0	1.4	1.0	1.0	0.2	0.6	0.8	0.8	0.8	0.2	0.2	0.8	0.2	1.2	1.0	0.2

Légende : Coopération | Collaboration

Figure 27 : Séquences des cibles des plans validés par les 20 opérateurs sur les 20 terrains, et distance de Levenshtein moyenne entre les séquences deux à deux.

Pour observer plus visuellement ce résultat, nous avons superposé en niveau de gris pour les plans validés par les participants dans chaque condition sur un des terrains

(Figure 28). Le niveau de gris associé à chaque case représente la proportion de plans validés qui la traverse (superposition de 10 plans validés pour chaque case). On observe visuellement qu'il y a plus de cases traversées par au moins un plan en condition coopérative qu'en condition collaborative, ce qui conforte par une autre méthode le résultat précédent. La superposition des plans dans les deux conditions pour les vingt terrains est consultable en Annexe 1. On peut y observer que selon le terrain la variabilité des plans est soit plus importante en mode coopératif, soit égale dans les deux modes, mais jamais plus importante en mode collaboratif.

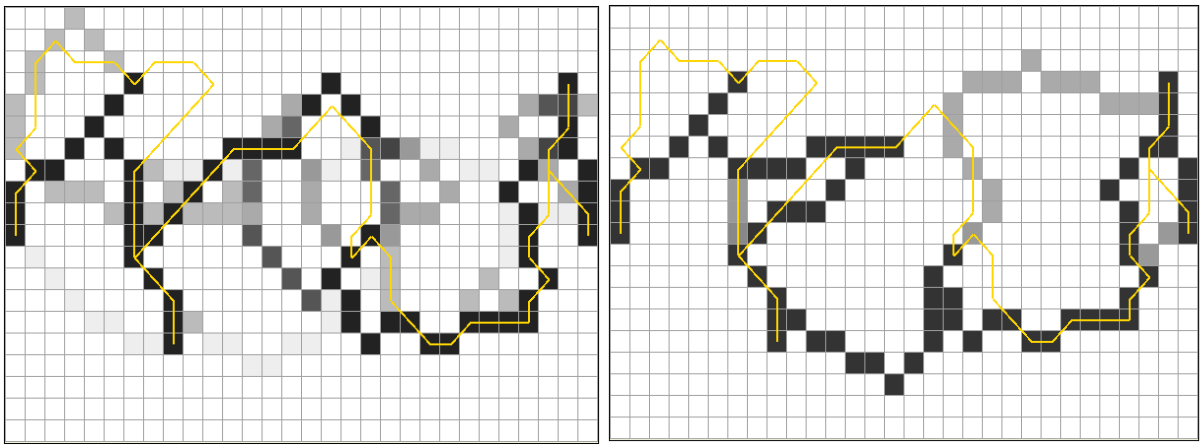


Figure 28 : Superposition des plans validés par les participants sur un terrain selon la condition de travail en équipe avec l'IA (coopération à gauche, et collaboration à droite) pour un des terrains. Le niveau de gris associé à chaque case représente la proportion de plans validés qui la traverse. Le chemin jaune correspond à la première suggestion proposée par l'IA en mode collaboratif.

III. Conclusion sur l'expérimentation

Ces résultats indiquent que faire évoluer la répartition des tâches entre l'opérateur humain et l'IA va influencer l'expérience de l'opérateur. Lorsque l'équipe passe de la collaboration à la coopération, où le système prend progressivement moins d'initiative dans l'action, l'opérateur se sent davantage à l'origine et responsable des décisions prises. A l'inverse, le retrait de l'assistance à laquelle l'opérateur est habitué crée un manque. Cependant, lorsque l'IA introduit des suggestions de plan qui n'étaient pas présentes auparavant, l'opérateur ne se sent pas dessaisi de son rôle pour autant. L'aide est acceptée et ne désengage pas l'opérateur. Cette asymétrie au changement peut être problématique dans la mesure où l'on observe

que la méthode de travail en équipe (coopération ou collaboration) va néanmoins influencer les décisions validées par l'opérateur. Les plans de vol réalisés avec une organisation collaborative sont moins variés que ceux réalisés dans une organisation coopérative, témoignant ainsi d'une normalisation des comportements humains induite par l'IA. Lorsque des suggestions de plans sont proposées par l'IA, l'opérateur aura tendance à valider les plans qui s'en rapprochent, mais sans avoir conscience de cette influence sur son comportement car son sentiment de paternité n'a pas évolué et il se sent autant à l'origine de la solution.

Ainsi, un système dédié à assister un opérateur humain dans une tâche de décision peut insidieusement modifier la solution qu'il valide. Il s'agit d'un problème critique dans le développement de systèmes d'assistance sophistiqués lorsque l'utilisateur n'a pas conscience d'être influencé par le système intelligent. Ce comportement est en quelque sorte normalisé par les suggestions de l'IA et il ne remet plus en question la solution potentielle de la même manière. La réduction de la variabilité humaine peut être considérée comme souhaitable car la prévisibilité du résultat peut conduire à une meilleure fiabilité globale, mais d'un autre côté, l'intérêt de conserver des opérateurs humains réside précisément dans leur capacité à être critiques et subjectifs dans le processus de décision. On peut donc identifier la dégradation de l'esprit critique de l'utilisateur comme un risque lié à la mise en place de ces systèmes d'aide plus adaptatifs. Les utilisateurs peuvent avoir tendance à s'approprier les idées introduites par le système lorsqu'elles sont pertinentes, mais sans avoir nécessairement conscience de cette influence sur leur propre prise de décision. En outre, l'absence de variation de la charge de travail pourrait être due au fait que la tâche est relativement courte car elle n'implique qu'une seule décision. Les résultats seraient peut-être différents sur des tâches qui nécessitent des interactions pendant plus de temps.

Chapitre 4 : Étude du lien entre la prise de décision de l'opérateur avec l'IA et ses éléments de personnalité

Au cours de la première étude, nous avons observé qu'alors que les différents participants ont répondu au même problème avec la même IA et sur les mêmes terrains, ils ont validé des solutions différentes. On peut se demander si ces différences ne seraient pas liées à leur personnalité individuelle, ce qui pourrait potentiellement être utilisé par une IA plus adaptative pour aider l'opérateur de manière personnalisée.

I. Méthode

1. Hypothèses

L'objectif de cette deuxième étude est d'explorer la problématique 2 : dans quelle mesure la prise de décision de l'opérateur avec l'IA est-elle liée à des éléments de personnalité ? Nos hypothèses sont :
H1 : Les participants ont des préférences individuelles pour certains compromis de solution.
H2 : Ces préférences décisionnelles sont liées à des caractéristiques individuelles.

2. Plan expérimental

Un groupe de 24 participants a participé à cette étude, étudiants en école d'ingénieurs et enseignants-chercheurs (50% de femmes), d'un âge moyen de 26.8 ans (écart-type = 3.1 ans). Tous les sujets se sont portés volontaires pour participer à l'étude et ont donné leur consentement éclairé avant de prendre part à l'étude. Tous les participants ont effectué 20 missions sur les mêmes 20 terrains dans le même ordre en collaboration avec l'IA. Le plan expérimental est de type intra-groupe où l'analyse va consister à déterminer des classes de compromis que peuvent préférer individuellement les participants grâce à une classification hiérarchique et une analyse en composante principale, puis à chercher grâce à des ANOVA si les participants classés dans ces différentes classes de compromis présentent des différences de personnalité.

3. Métriques

Métrique sur les critères de compromis de plan. Pour répondre à H1, nous avons enregistré pour chaque plan prévisionnel validé la quantité de carburant dépensé (variable de consommation de carburant), la quantité de risque pris (variable de la prise de risque), et le nombre de cibles photographiées (variable d'abandon d'objectifs). Ces trois variables ont été exprimées sur une échelle de 0 à 100.

Métrique sur les caractéristiques individuelles des opérateurs. Pour répondre à H2, nous avons utilisé sept métriques pour caractériser les opérateurs. Plusieurs semaines après l'étude, pour mesurer leur score sur ces métriques, les participants ont rempli un questionnaire composé de trois questionnaires psychométriques juxtaposés avec un total de 58 questions :

- L'inventaire des Big-Five (BFI-Fr), composé de 45 questions qui décrivent la personnalité en cinq traits centraux : ouverture d'esprit, conscienciosité, extraversion, agréabilité, tendance à la névrose (Plaisant et al., 2010). Le détail des questions est disponible en Annexe 3.
- L'échelle d'estime de soi de Rosenberg (RSES), composée de 10 questions qui mesurent l'estime de soi des individus (Vallieres et Vallerand, 1990). Le détail des questions est disponible en Annexe 4.
- L'échelle de stabilité de la confiance en soi (SESS), composée de 3 questions qui mesurent la variabilité de l'estime de soi des individus dans le temps (Altmann et Roth, 2018). Le détail des questions est disponible en Annexe 5.

II. Résultats

1. Préférences décisionnelles individuelles

Afin de mesurer la nature individuelle des décisions prises par les différents participants, nous avons moyenné pour chaque participant les critères de compromis des 20 plans qu'il a validés. Pour observer la variabilité de ces profils décisionnels, nous avons réalisé un groupement des participants selon leur critère de compromis moyen à travers une classification ascendante hiérarchique (Figure 29). On peut discerner trois groupes ainsi qu'un participant au comportement singulier.

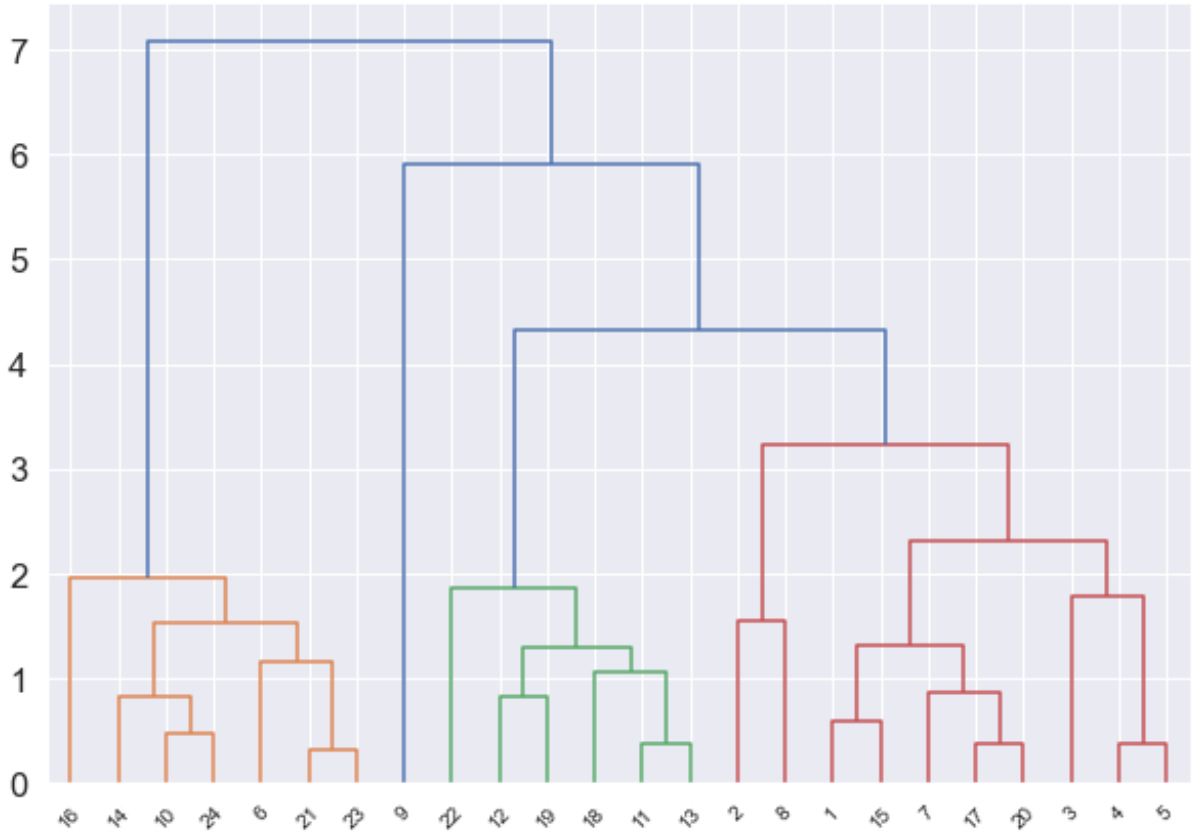


Figure 29 : Dendrogramme de la classification hiérarchique des participants en fonction de leurs critères de compromis moyen.

Pour visualiser la nature de ces groupes, nous avons réalisé une ACP des participants selon leur critère de compromis moyen (Figure 30). Le premier axe du cercle des corrélations montre une opposition entre la variable d'abandon d'objectifs d'un côté, et les variables de consommation de carburant et de prise de risque de l'autre. On peut l'interpréter comme la possibilité d'abandonner des objectifs en échange d'une consommation moindre de carburant et d'une prise de risque plus faible, et réciproquement. Sur le second axe, les variables de consommation de carburant et de risque de risque s'opposent modérément, ce qui signifie qu'il y a également un arbitrage à effectuer entre les deux. Prendre du risque en passant proche de menaces ou au contact du relief pourrait aider dans certaines conditions à réduire la longueur du plan et donc les dépenses de carburant. De même, utiliser beaucoup de carburant en rajoutant des étapes pourrait permettre d'éviter les menaces et donc de réduire le risque pris. Chaque participant se voit affecté une couleur en fonction du groupe que lui a assigné la classification ascendante hiérarchique. On observe que chaque groupe est plus proche que les autres d'une

des trois variables. Les participants du groupe vert ont tendance à valider des plans qui mettent l'accent sur l'abandon d'objectif, les participants du groupe rouge sur la surconsommation de carburant et les participants bleus sur la prise de risque. Ce résultat va dans le sens de H1 et suggère que nos participants ont des profils décisionnels différents malgré leur similarité socio-professionnelle. Notons que le participant numéro 9 n'est pas visible sur le schéma car il présente un comportement aberrant qui consiste à systématiquement valider des plans passant par tous les objectifs, au prix d'une consommation de carburant et d'une prise de risque extrêmement élevées. Son comportement est très distant de celui de tous les autres participants, comme on peut l'observer Figure 29 où il est l'unique membre de son propre groupe de comportement.

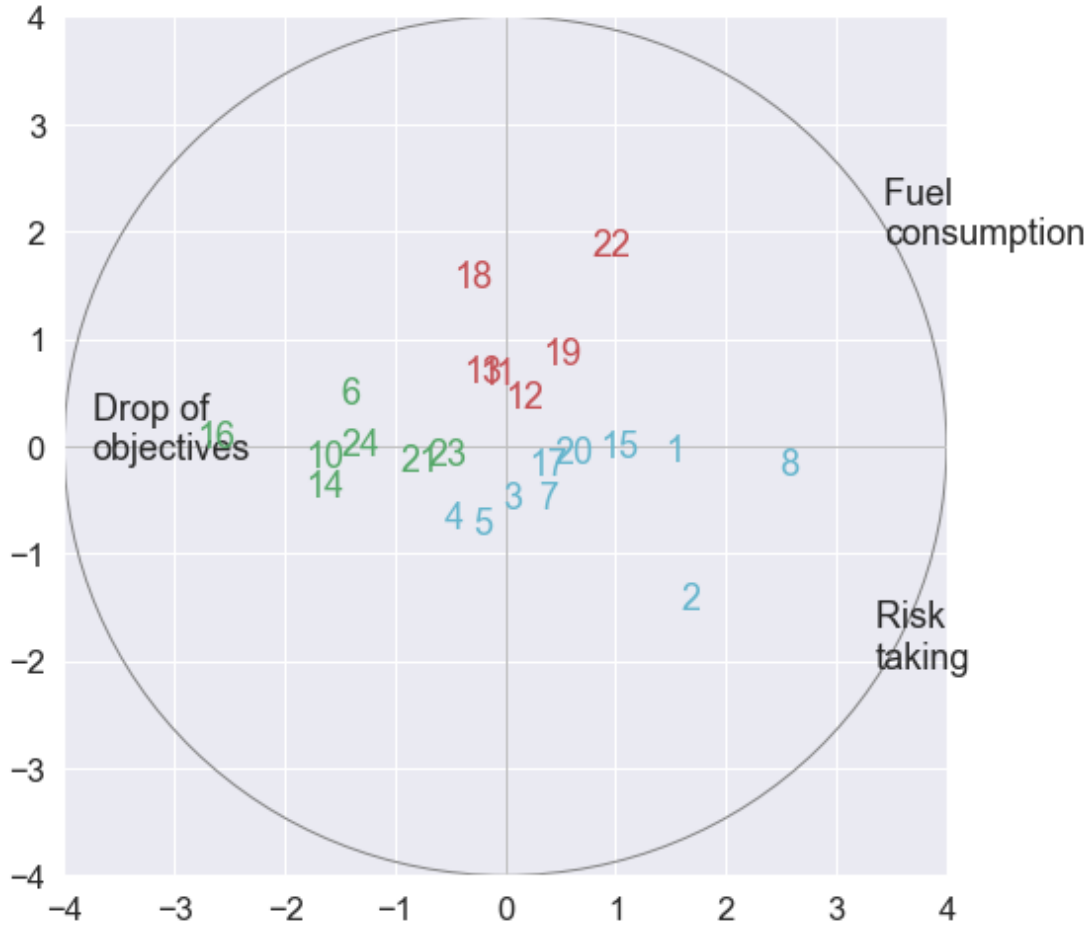


Figure 30 : Cercle des corrélations de l'ACP des participants selon leurs critères de compromis moyen (quantité d'objectifs abandonnées, de carburant consommé et de risque pris).

2. Profils décisionnels et caractéristiques individuelles

20 des 24 participants ont accepté de répondre au questionnaire destiné à quantifier les sept caractéristiques individuelles définies. Pour la suite de cette analyse, nous allons donc exclure les 4 participants pour lequel nous n'avons pas leur score sur ces métriques. La Figure 31 affiche le score des 20 participants restants selon ces sept métriques. On peut constater que les participants exhibent des niveaux différents de ces caractéristiques selon les individus. La Figure est visible en plus grand format en Annexe 6.

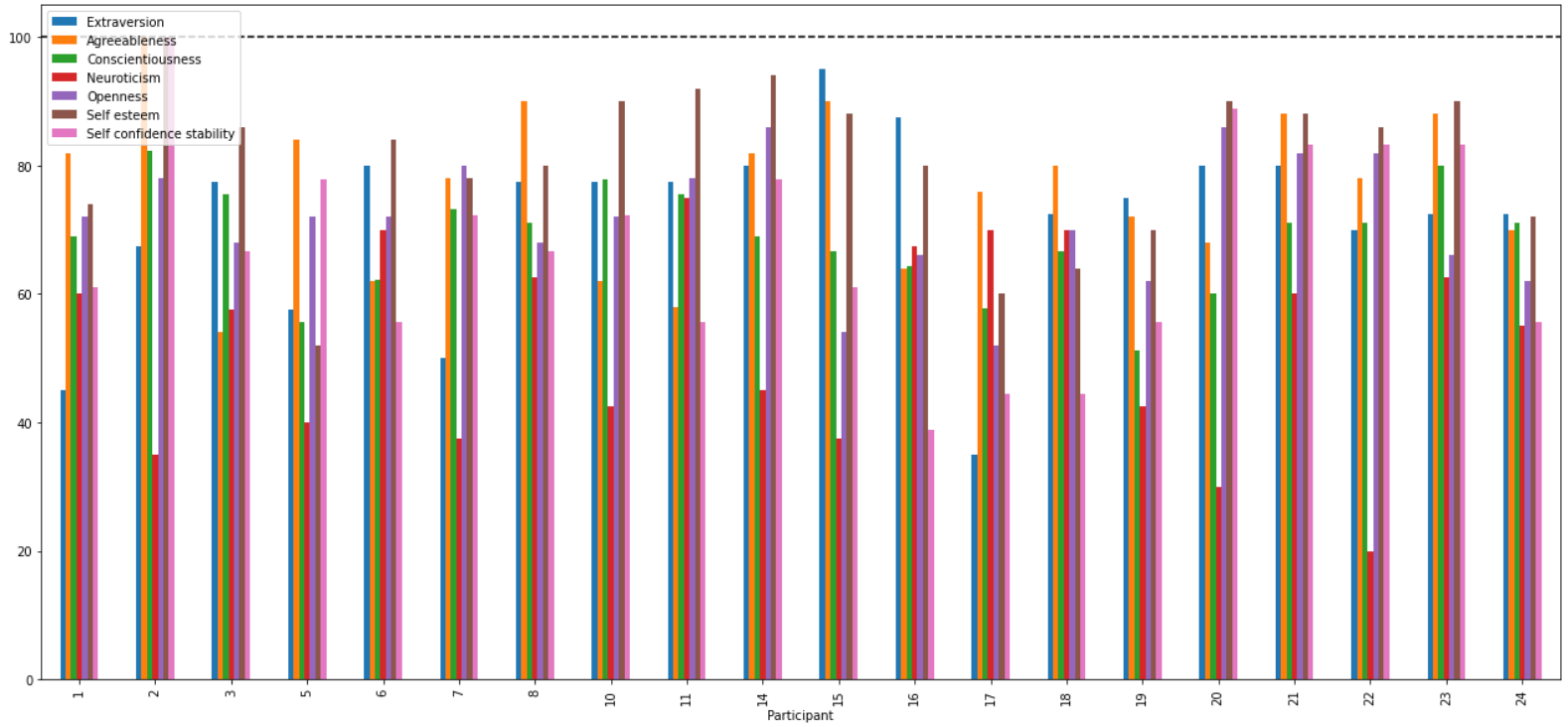


Figure 31 : Score des participants sur une échelle de 0 à 100 pour leur ouverture d'esprit, conscienciosité, extraversion, agréabilité, tendance à la névrose, confiance en soi et stabilité de la confiance en soi.

Afin de mesurer le lien entre les caractéristiques individuelles des participants et leur profil décisionnel, nous avons réalisé une ACP des participants selon leur score dans les sept métriques de caractérisation individuelle (Figure 32). Chaque numéro correspond à un participant, coloré selon son groupe décisionnel. Visuellement, les différents groupes décisionnels sont mélangés et il semble difficile de déduire que les participants aux profils vert (qui préfèrent abandonner des objectifs), rouge (qui préfèrent consommer beaucoup de carburant) et bleu (qui préfèrent prendre

beaucoup de risque) exhibent des ensembles de caractéristiques significativement différents.

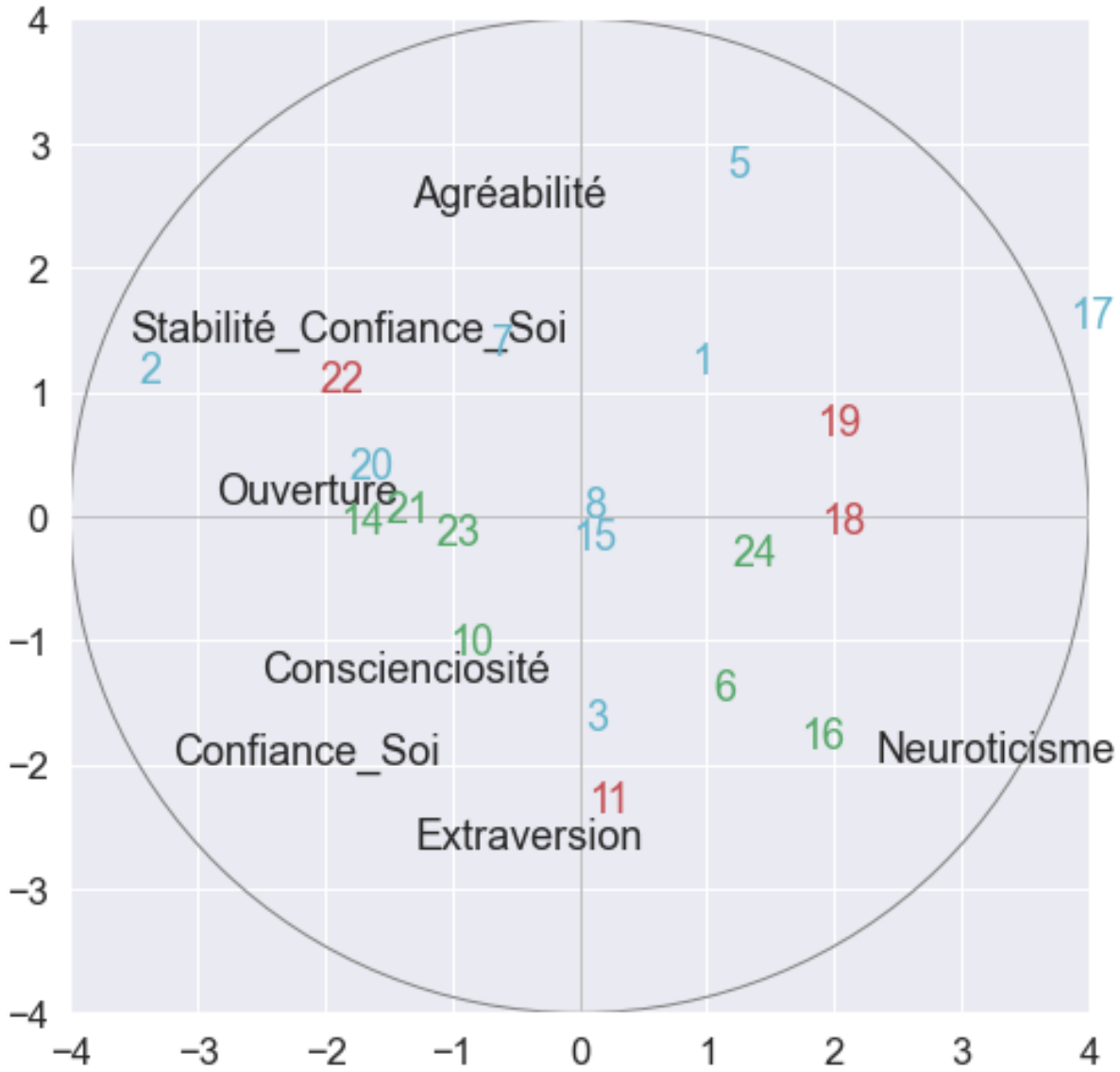


Figure 32 : Cercle des corrélations de l'ACP des participants selon leur score dans les sept métriques de caractérisation individuelles mesurées.

Pour vérifier statistiquement ce résultat, nous avons tout d'abord calculé les caractéristiques individuelles moyennes des participants selon leur groupe décisionnel (Figure 33). Nous avons ensuite réalisé des ANOVA pour tester la significativité de la différence de chacune des sept caractéristiques selon le groupe décisionnel (Figure 34). Aucune p-value n'est inférieure au seuil de significativité retenu ($p < .05$), ce qui signifie que nous ne pouvons pas rejeter l'hypothèse nulle selon laquelle toutes ces caractéristiques individuelles n'ont pas des niveaux significativement différents selon le groupe décisionnel des participants. En d'autres

termes, les données ne montrent pas des profils différents entre les personnes des trois groupes de comportement. Néanmoins, le faible nombre de participants dans chaque groupe pour lesquels nous avons mesuré leurs caractéristiques individuelles (respectivement 9 pour le groupe « prise de risque », 4 pour le groupe « consommation de carburant », et 7 pour le groupe « abandon d'objectif ») incite à nuancer ces résultats.

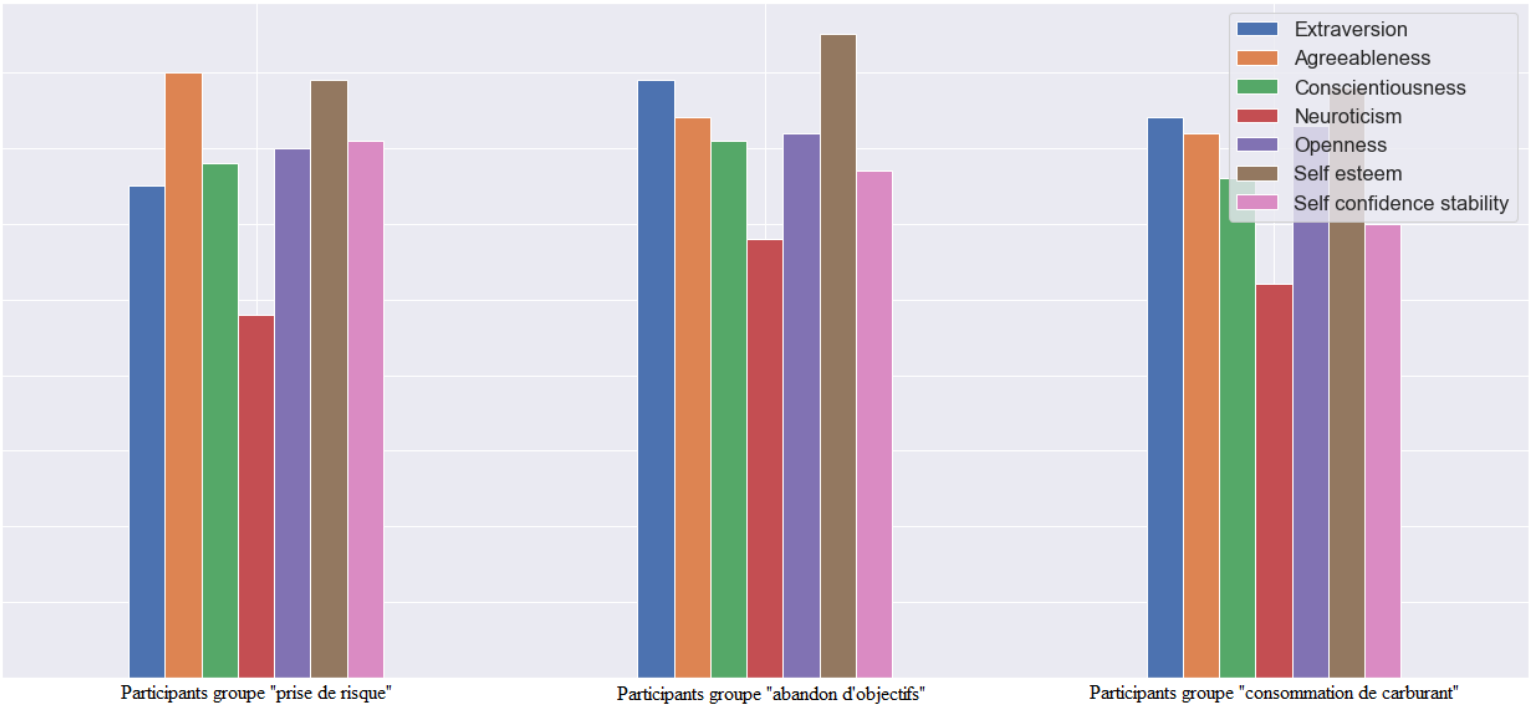


Figure 33 : Caractéristiques individuelles moyennes des participants de chaque groupe décisionnel identifié.

Extraversion et groupe de compromis

	sum_sq	df	F	PR(>F)
Groupe	750.535714	2.0	1.983097	0.168228
Residual	3216.964286	17.0	NaN	NaN

Agréabilité et groupe de compromis

	sum_sq	df	F	PR(>F)
Groupe	259.215873	2.0	0.842581	0.447809
Residual	2614.984127	17.0	NaN	NaN

Conscienciosité et groupe de compromis

	sum_sq	df	F	PR(>F)
Groupe	62.18304	2.0	0.435282	0.654095
Residual	1214.28496	17.0	NaN	NaN

Neuroticisme et groupe de compromis

	sum_sq	df	F	PR(>F)
Groupe	372.256944	2.0	0.727776	0.497435
Residual	4347.743056	17.0	NaN	NaN

Ouverture et groupe de compromis

	sum_sq	df	F	PR(>F)
Groupe	33.371429	2.0	0.164972	0.849259
Residual	1719.428571	17.0	NaN	NaN

Confiance en soi et groupe de compromis

	sum_sq	df	F	PR(>F)
Groupe	222.085714	2.0	0.709215	0.506021
Residual	2661.714286	17.0	NaN	NaN

Stabilité de la confiance en soi et groupe de compromis

	sum_sq	df	F	PR(>F)
Groupe	354.646825	2.0	0.641635	0.538714
Residual	4698.150675	17.0	NaN	NaN

Figure 34 : Caractérisation individuelle des participants selon leur groupe décisionnel (indication des p-values issues des tests d'analyse de variable).

III. Conclusion sur l'expérimentation

Ces résultats indiquent que dans les situations où des compromis doivent être choisis en équipe avec une IA, les individus peuvent avoir des préférences individuelles de critères différentes. Cependant, dans notre étude, ce choix n'est pas directement lié aux traits de personnalité des individus, mais reflète plutôt des éléments liés à leur façon de penser la situation. Un système de recommandation basé sur des personae trop rigides dans un tel contexte pourrait risquer d'enfermer les utilisateurs dans des schémas comportementaux trop simplistes alors que les attentes envers l'IA semblent varier fortement selon l'individu, et davantage liées à la

représentation du problème à un moment donné qu'à des caractéristiques personnelles particulières.

Chapitre 5 : Étude de l’adaptation de l’IA au processus décisionnel de l’opérateur

Lors du chapitre précédent, nous avons déterminé que la personnalité des utilisateurs ne permet pas d’anticiper leur profil décisionnel dans le cadre de la tâche de replanification que nous simulons. Pour que l’IA fournisse une assistance plus personnalisée, s’appuyer sur les caractéristiques individuelles des utilisateurs ne semble pas une approche probante à ce stade. Nous proposons d’explorer une autre voie, qui implique d’aller vers une vraie collaboration avec l’IA dans le processus décisionnel. Une telle IA nécessiterait de pouvoir se représenter le processus décisionnel de l’opérateur et d’en suivre l’évolution pour pouvoir prendre des initiatives contextualisées au fur et à mesure de l’interaction.

I. Méthode

1. Hypothèses

Dans ce chapitre, nous explorons la problématique 3 : Comment l’IA peut-elle prendre l’initiative lors de l’interaction pour assister la prise de décision de l’opérateur de façon adaptée à son processus décisionnel ?

Nous faisons l’hypothèse que la conception d’une IA d’aide à la décision humaine réellement au service du processus décisionnel des opérateurs peut se faire en suivant trois étapes. La première étape est la construction d’un modèle du processus décisionnel des opérateurs pour un cas d’usage, c’est-à-dire l’identification des grandes étapes d’élaboration d’une solution à ce type de problème donné. La seconde étape consiste à automatiser le suivi d’une instance de ce modèle décisionnel. La caractérisation du lien entre l'utilisation faite de l'IA et les intentions de l'opérateur pourrait aider le système à construire une représentation de ce processus de décision au fur et à mesure de son évolution. La troisième étape vise à identifier des opportunités d’assistance adaptées au processus décisionnel de l’opérateur tel que le système se le représente. Dans ce chapitre, nous implémentons les trois étapes de ce modèle afin d’en montrer la pertinence pour concevoir un système d’aide à la décision humaine plus collaboratif.

2. Méthode de construction pour chaque étape

Afin de construire une méthodologie d'assistance à la décision humaine adaptée à son processus décisionnel, la première étape a consisté à déterminer la forme générale de ce processus. Dans ce cadre, nous avons mené une étude expérimentale qualitative dont l'objectif était de déterminer quelles méthodologies de replanification des opérateurs peuvent déployer empiriquement dans notre environnement de simulation. Le plan expérimental est schématisé en Figure 35. Tout d'abord, nous avons réutilisé les données du chapitre précédent, en considérant cela comme un pré-test pour ce chapitre. Les 24 participants de l'étude précédente qui ont réalisé la tâche de replanification sur 20 terrains différents nous ont permis de déterminer les deux terrains présentant le plus de variabilité de plans validés. Ces deux terrains présentent une configuration pour laquelle il n'y a pas de solution qui semble plus optimale de manière évidente pour des utilisateurs. Nous avons conservé ces deux terrains pour la suite de l'étude afin de maximiser la complexité de la décision à prendre. Ensuite, nous avons réalisé la phase d'expérimentation proprement dite avec 8 participants (4 femmes, 4 hommes), jeunes chercheurs et ingénieurs de recherche en aéronautique, d'un âge moyen de 28.1 ans (écart type = 5.2 ans). Les participants ont réalisé la tâche de replanification sur les deux terrains sélectionnés, pendant que leur utilisation des outils de planification était filmée. Ensuite, un entretien d'autoconfrontation individuel était conduit avec chacun d'entre eux. L'entretien d'autoconfrontation est une méthode d'analyse de l'activité humaine qui consiste à confronter un individu aux traces de son activité et à l'inciter à expliquer sa production par rapport à la réalité de sa pratique (Theureau, 2003). Au cours de ces entretiens, l'expérimentateur a visionné en leur compagnie les enregistrements vidéo de leurs réalisations de la tâche afin de recueillir les intentions et les stratégies décisionnelles liées à chaque utilisation des outils de replanification.

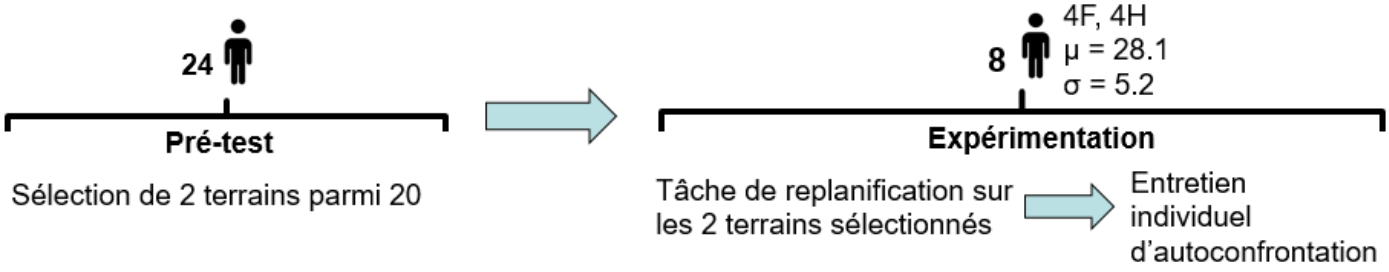


Figure 35 : Plan expérimental de l'étude qualitative.

La deuxième étape a été réalisée par évaluation experte avec les données des sujets utilisés pour la construction du modèle, et consistait à étudier les liens entre l'utilisation des outils et les cases du modèle issu de la première étape. La troisième étape était une analyse qualitative de l'utilisabilité des résultats de la deuxième étape, à travers l'analyse experte du processus décisionnel utilisé à partir de sa reconstruction. L'objectif était de trouver des séquences qui font sens et qu'un collaborateur humain pourrait utiliser pour aider son partenaire.

II. Résultats

1. Construction d'un modèle du processus décisionnel des opérateurs

La synthèse des entretiens a révélé une grande variabilité à la fois des outils de replanification utilisés et des compromis de plans validés par les huit participants sur les deux terrains replanifiés. Malgré cette variabilité inter-individuelle, des récurrences dans les stratégies de planification déployées sont observables. La Figure 36 rassemble certains fragments de phrases prononcées par les participants lors des entretiens et qui renseignent sur leur processus décisionnel. Les intentions des opérateurs peuvent être divisées en deux catégories générales qui semblent correspondre à deux types de phases décisionnelles distinctes. D'une part, les intentions qui se concentrent sur les cibles, à travers des réflexions sur leurs localisations relatives et leur proximité aux menaces afin de déterminer lesquelles pourraient être abandonnées, donc il s'agit là d'intentions d'ordre stratégique (Figure 36 a). D'autre part, les intentions qui portent sur les segments de plan qui relient les cibles entre elles, ce qui consiste à localement prendre plus de risque pour diminuer la consommation de carburant ou l'inverse, donc il s'agit là d'intentions d'ordre tactique (Figure 36 b).

a)
« je vais d'abord regarder **les cibles**, pour savoir **lesquelles virer** »
« tout d'abord regarder le **positionnement des cibles** vis-à-vis des risques, la **distance entre les cibles** »
« au début, chercher **quels objectifs** je devrais enlever si je dois en enlever »
« au début je regardais quelles étaient les **mauvaises cibles**, c'est-à-dire difficile d'accès »
« il fallait regarder **quel point** coûte cher, c'est-à-dire qui est hyper entouré, difficile d'accès »

b)
« ensuite je me demandais si je ne pouvais pas **améliorer** »
« ensuite je me suis demandé si je pouvais encore **améliorer un peu le chemin** ou pas, **à la marge** »
« ensuite j'ai cherché si je pouvais pas encore **optimiser des portions**, **échanger du carburant contre du risque** »
« j'ai ajouté **une déviation** que j'avais identifiée plus tôt **entre C3 et C4** »
« puis j'ai ajouté manuellement la **déviation identifiée avant C7** »

c)
« je me suis demandé si finalement je ne pouvais pas **ajouter manuellement C6** »
« je suis retourné **sur C6** qui correspond à ça »
« donc je me suis dit qu'il fallait forcément **abandonner C5**, et revenir à l'idée de toute à l'heure »
« mais finalement, je me suis demandé si on ne pouvait pas **ajouter C7**, mais différemment, en faisant un détour plus long »

Figure 36 : Verbatim de l'explicitation de la méthodologie de replanification fournie par les participants lors des entretiens. Les termes « objectif », « point » ou « Cx » désignent les cibles liées aux terrains. Les mots mis en gras caractérisent le type de modification de plan réalisée ou considérée sur le plan par l'opérateur, et les mots soulignés l'étape temporelle du raisonnement à laquelle la modification intervient.

Ce dilemme entre optimiser l'utilisation des ressources existantes (dans notre cadre, chercher à améliorer les choix tactiques pour une stratégie donnée) et trouver un moyen d'en créer de nouvelles (dans notre cadre, chercher d'autres choix stratégiques) correspond à un modèle descriptif de la science managériale connu sous le nom de « conflit exploration-exploitation dans le processus cognitif humain » (Berger-Tal et al., 2014 ; Sinha, 2015). Ce modèle a également fait l'objet de recherches en neurosciences, où des corrélats neurophysiologiques à ces deux états ont pu être observés, l'exploitation activant les régions du cerveau associées à la

recherche de récompense pour évaluer la valeur des choix actuels, et l'exploration faisant appel aux régions associées au contrôle attentionnel pour investiguer des choix alternatifs (Laureiro-Martinez et al., 2015). Par ailleurs, les marqueurs temporels de raisonnement que nous avons notés tendent à indiquer que ces cycles décisionnels d'exploration et d'exploitation peuvent se succéder plusieurs fois au cours de la replanification, et en débutant systématiquement par une phase d'exploration (Figure 36c). Tous les participants ont de fait commencé les interactions par une phase d'exploration, puis se sont engagés dans une alternance de phases d'exploitation et d'exploration jusqu'à la validation du suivi d'un plan (Figure 37). Les deux mots « exploration » et « exploitation » ayant une calligraphie voisine, nous proposons de les distinguer par leur lettre médiane : (R) pour exploration et (T) pour exploitation.

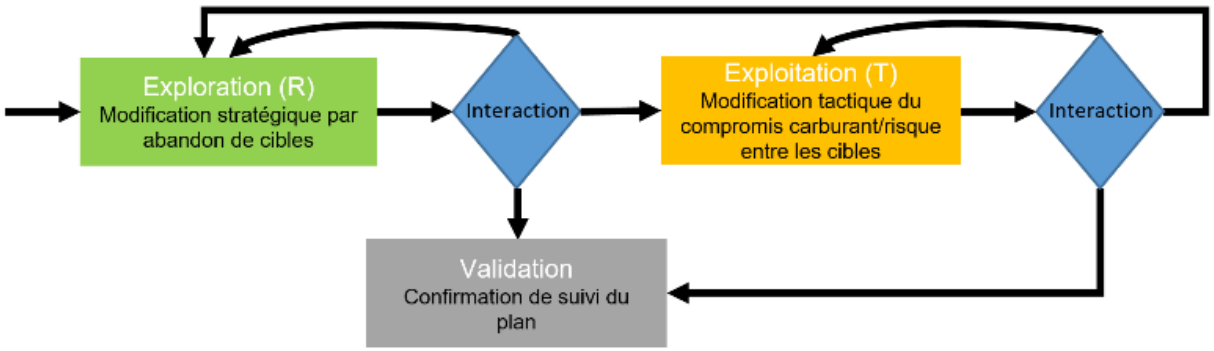


Figure 37 : Processus décisionnel type des opérateurs pour la tâche de replanification de plan de vol.

2. Automatisation du suivi de l'évolution du modèle décisionnel pendant son déploiement

Nous avons observé que l'outil de parcours de l'historique des plans relève systématiquement de l'exploitation et est utilisé soit pour comparer les propriétés tactiques de plusieurs plans, soit pour retourner à un plan précédent afin qu'il redevienne le plan de travail. A l'inverse, le rôle d'assistance ou d'élaboration de l'IA compris dans les autres outils a pu être utilisé pour de l'exploration ou bien pour de l'exploitation. La consultation d'une suggestion de plan de l'IA est une action d'exploration lorsque le plan proposé par le système est vu pour la première fois ou lorsque l'opérateur est dans une phase de recherche initiale, mais est une action

d’exploitation lorsque l'opérateur y revient après en avoir déjà pris connaissance. Pour les outils de modification manuelle et de co-construction, où l’opérateur participe à l’élaboration de la solution, les critères de recherche de plan qu’il fournit à l'IA contiennent des indices sur son but sous-jacent. Lorsque ces critères de recherche comprennent notamment l’abandon des mêmes cibles que le plan précédent, l'opérateur est dans une phase d'exploitation et emploie l’IA pour de l’optimisation tactique. Lorsque ces critères de recherche portent sur une famille de plans qui n'a jamais été examinée, alors l'IA est employée à des fins d'exploration pour envisager de nouvelles stratégies. Au-delà du type d’outil de replanification utilisé, c’est le ou les critères de compromis du plan que l’opérateur cherche à modifier qui semblent renseigner sur sa volonté d’explorer ou d’exploiter des solutions. Les modifications changeant les cibles à abandonner caractérisent les phases d’exploration, tandis que les modifications conservant les mêmes cibles et portant donc exclusivement sur les quantités de risque pris et de carburant consommé caractérisent les phases l’exploitation. Si une modification du plan implique un changement des cibles abandonnées, alors l’opérateur travaille à un niveau stratégique et se trouve à cette étape dans une phase décisionnelle d’exploration (R). Si la modification du plan n’implique pas de changement des cibles abandonnées, alors cette modification porte exclusivement sur les critères de carburant consommé et de risque pris, ce qui correspond à un travail tactique et une phase décisionnelle d’exploitation (T). La Figure 38 représente l’espace théorique des compromis de plan selon les trois dimensions que nous considérons : cibles abandonnées, carburant consommé et risque pris. Cet exemple théorique montre la phase décisionnelle dans laquelle se trouverait un opérateur selon différentes possibilités de compromis de plan à l’étape n+1 étant donné une étape n : exploration si le compromis se déplace entre autres sur l’axe lié aux cibles abandonnées, exploitation si le compromis se déplace seulement sur les axes de risque pris et de carburant consommé.

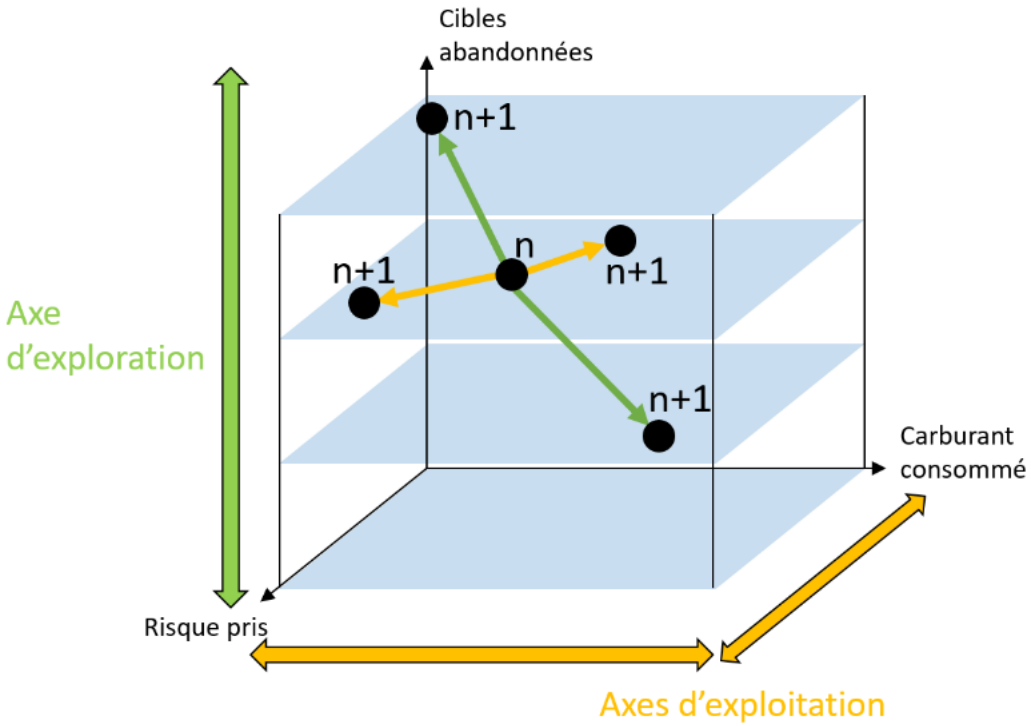


Figure 38 : Exemple théorique de la phase décisionnelle liée à plusieurs possibilités de compromis à l'étape n+1 de modification de plan selon une étape n donnée.

Ces éléments issus de l'analyse experte nous ont permis de concevoir une méthode systématique pour suivre l'évolution du processus décisionnel de l'opérateur au fur et à mesure de son évolution, c'est-à-dire la succession des phases d'exploration et d'exploitation par lesquelles il passe. Cette méthode est automatisable et nous en proposons un algorithme en pseudocode (Figure 39).

```
ALGORITHME Suivi_Processus_Decisionnel

ENUMERATION
        Phase_Decision {exploration, exploitation}

VARIABLES
// contient les cibles abandonnées actuellement
        cibles_Abandonnees[6] : TABLEAU BOOL

// contient les cibles abandonnées à chaque utilisation d'outil
        evolution_Cibles_Abandonnees : LISTE TABLEAU ENTIER

// contient la phase du processus décisionnel correspondant à chaque utilisation d'outil
        evolution_Phase_Decision : LIST Phase_Decision

DEBUT
// lorsque l'opérateur a utilisé un outil de modification de plan on vérifie si les cibles
// abandonnées ont changé
        QUAND Outil_Utilise
                AJOUTER COPIE de cibles_Abandonnees DANS
                evolution_Cibles_Abandonnees

// si les cibles abandonnées n'ont pas changé alors la modification est d'ordre tactique et
// donc l'opérateur est dans une phase d'exploitation
                SI evolution_Cibles_Abandonnees(taille(evolution_Cibles_Abandonnees) ==
                evolution_Cibles_Abandonnees(taille(evolution_Cibles_Abandonnees-1)
                        AJOUTER exploitation DANS evolution_Phase_Decision

// si les cibles abandonnées ont changé alors la modification est d'ordre stratégique et donc
// l'opérateur est dans une phase d'exploration
                SINON
                        AJOUTER exploration DANS evolution_Phase_Decision
FIN
```

Figure 39 : Pseudocode de l'algorithme de suivi du processus décisionnel.

C'est à partir de ce pseudocode que nous avons conduit manuellement les analyses en appliquant ce processus de traitement sur les données d'interaction issues de

l'étude. La figure 40 fournit un exemple de suivi du processus décisionnel issu du fonctionnement de l'application de l'algorithme. Nous réussissons à reconstruire automatiquement les phases décisionnelles par lesquelles passe cet opérateur et en cohérence avec ce qu'il a rapporté au cours de l'entretien individuel. Chaque colonne du tableau correspond à une utilisation d'un outil de replanification pendant cet essai. La première ligne indique les numéros des cibles qui sont abandonnées par le plan prévisionnel à cette étape. La seconde ligne indique le résultat du critère d'inférence des phases décisionnelles, qui consiste à examiner s'il y a un changement de cibles abandonnées entre la modification de plan précédente et la modification de plan actuelle, c'est-à-dire entre une colonne et la colonne précédente. La troisième ligne en déduit la phase décisionnelle dans laquelle se trouve l'opérateur après chaque interaction.

Cible(s) abandonnée(s)	3,7	3,4	4	4,5	5	5,7	5	3	3	3,4	3,5	5	5	5	5,7	5	5
Changement de cibles abandonnées ?	oui	oui	oui	oui	oui	oui	oui	oui	non	oui	oui	oui	non	non	oui	oui	non
Phase de décision inférée	R	R	R	R	R	R	R	R	T	R	R	R	T	T	R	R	T

Figure 40 : Inférence de la phase décisionnelle de l'opérateur à chaque modification du plan de vol pour un essai de l'étude.

3. Identification d'opportunités d'assistance adaptées à l'évolution du processus décisionnel suivi

Ces liens entre les modifications des critères de compromis de plan et le processus décisionnel humain nous ont permis de développer une méthodologie de reconnaissance des phases d'exploration et d'exploitation par lesquelles passent les opérateurs au cours de la replanification. Ce processus peut être reconstruit a posteriori une fois la replanification achevée et un nouveau plan de vol validé, mais aussi dynamiquement pendant la réalisation de la tâche. Cette capacité du système à bâtir une représentation décisionnelle de son coéquipier préfigure des modalités d'assistance à la replanification qui s'appuient sur les mécanismes cognitifs humains. Cette représentation des intentions humaines permet au système de concevoir l'opérateur comme un agent rationnel et d'en éclairer les croyances, c'est-à-dire l'espace des solutions considéré, et les désirs, c'est-à-dire ses préférences

individuelles en termes de critères de compromis de plan. Si actuellement dans l'environnement de simulation le système ne participe à l'élaboration de la décision qu'en réponse à des demandes de l'opérateur soit d'idées ou d'évaluation de choix stratégiques, soit d'évaluation d'optimisation tactiques, cela permet d'envisager une interaction plus symétrique à travers la proposition contextualisée d'idées de choix stratégiques ou d'optimisations tactiques. Nous envisageons ces prises d'initiative du système comme des assistances à l'exploration et à l'exploitation, à travers l'ajout sur le terrain dans l'interface de deux plans alternatifs que l'opérateur pourrait choisir comme nouveau plan de travail. Afin d'aider l'opérateur à les juger, leurs critères de compromis respectifs seraient également affichés.

L'assistance à l'exploration pourrait être mise en œuvre quand deux modifications de plan n et n-1 sont classifiées comme correspondant à des phases d'exploration. Afin d'aider l'opérateur dans ses recherches stratégiques, les propositions de plans alternatifs consisteraient à abandonner des cibles que l'opérateur n'avait pas envisagé d'abandonner. L'assistance à l'exploitation, elle, pourrait être mise en œuvre quand une modification n classifiée comme phase d'exploitation succède à une autre n-1 classifiée comme phase d'exploration. L'IA participerait à l'optimisation tactique du plan en proposant soit d'autres ordres de passage entre les cibles, soit la modification des segments entre les cibles pour diminuer le risque pris ou le carburant consommé. Un exemple de l'assistance à l'exploration tel qu'il pourrait être implémenté de cette manière est présenté dans les Figures 41 et 42. Ces exemples viennent des séquences d'action des participants de l'étude. La Figure 41 présente deux modifications de plans successives lors de la tâche de replanification sur un terrain. Ces deux actions sont classifiées comme des phases d'exploration, ce qui active l'assistance à l'exploration. Les deux modifications abandonnent une cible commune, la cible 3 (située en haut à gauche du terrain), et une autre cible, qui diffère. L'IA va proposer des alternatives qui correspondent à ce schéma d'exploration sur la Figure 42. Deux suggestions de plans sont proposées en tant qu'alternative 2 et alternative 3, l'alternative 1 étant le plan à la modification n. Leurs critères de compromis et le masquage des tracés permettent de les comparer. Le plan de l'alternative 1 abandonne les cibles 3 et 4, le plan de l'alternative 2 les cibles 3 et 5, le plan de l'alternative 3 les cibles 3 et 6. On remarque que les alternatives 2 et 3 ont des critères de compromis similaires, avec une dépense de carburant

inférieure mais un risque supérieur à l'alternative 1. Ces informations apportées à l'opérateur en réaction à ses modifications de plan précédentes pourraient nourrir sa réflexion stratégique et lui faire gagner du temps dans l'exploration de solutions.

Ainsi, l'analyse qualitative permet de trouver sur les situations effectivement faites des séquences qu'un opérateur humain pourrait utiliser pour aider son coéquipier. L'analyse de l'ensemble des séquences possibles et de leur indication sur le processus décisionnel n'a pas été faite à ce stade mais est techniquement réalisable automatiquement grâce à la démarche que nous avons présentée.

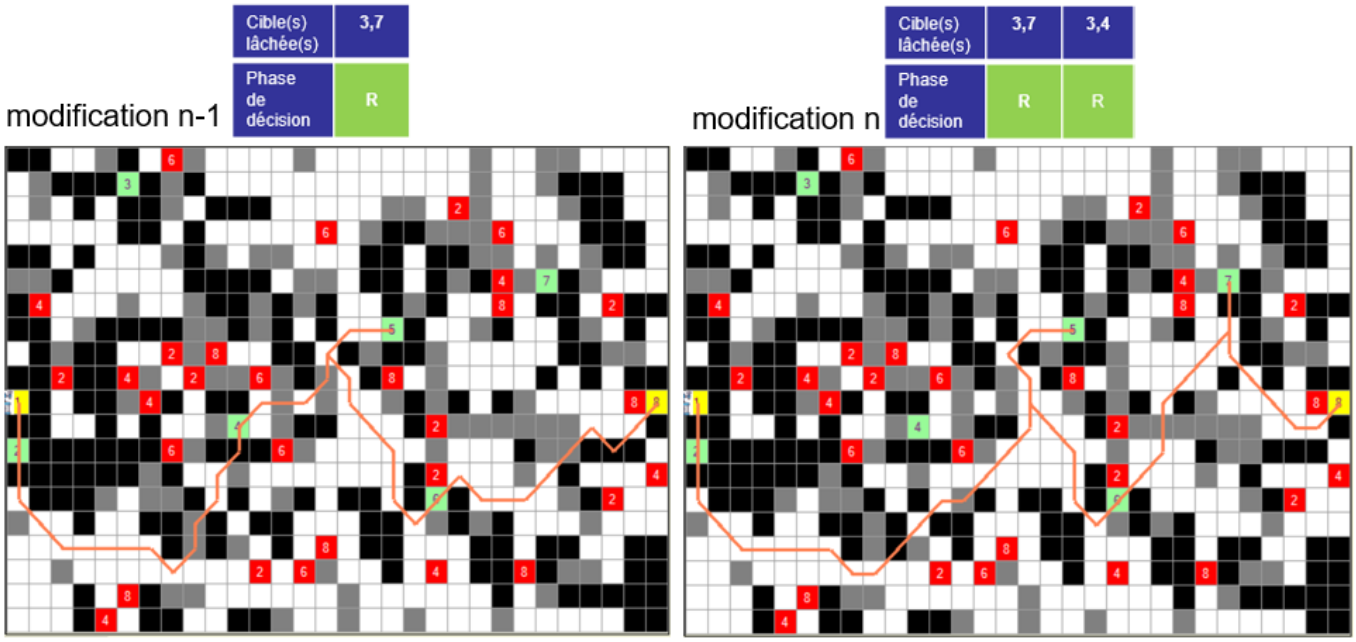


Figure 41 : Deux modifications de plans successives lors de la tâche de replanification sur un terrain.

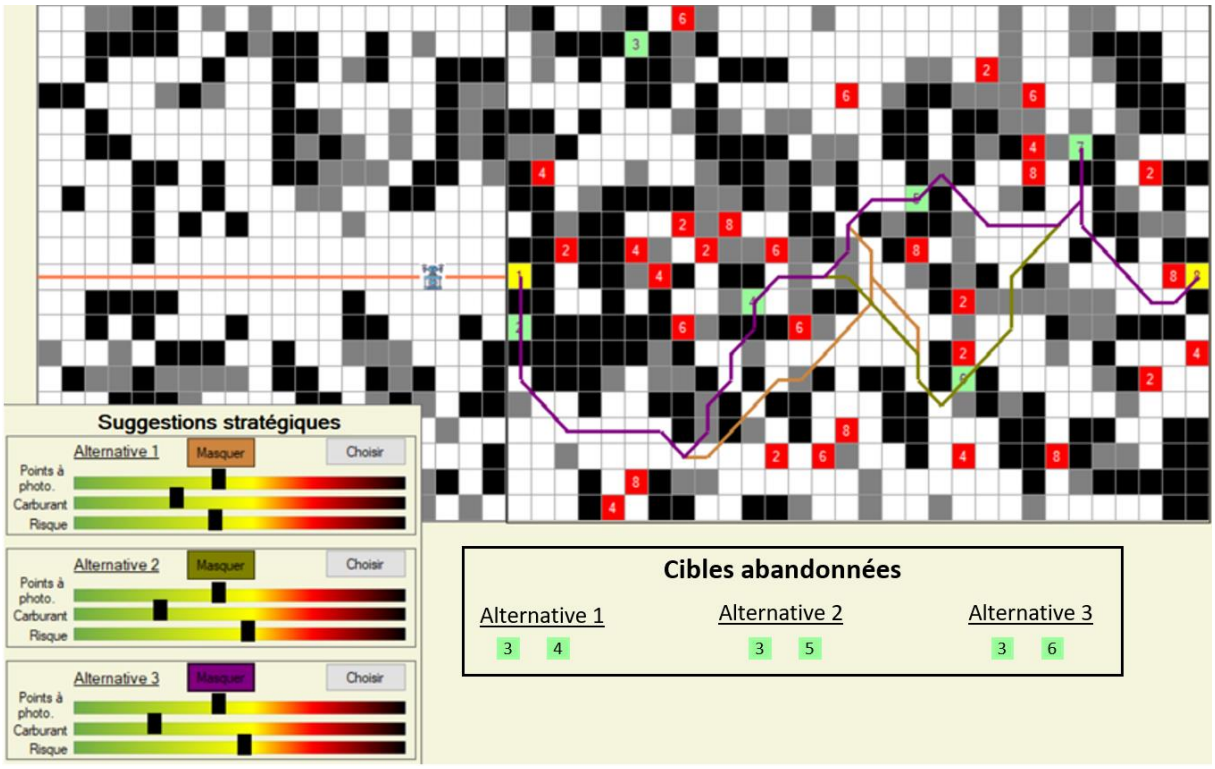


Figure 42 : Deux suggestions de plans sont proposées en tant qu'alternative 2 et alternative 3, l'alternative 1 étant le plan à la modification n.

III. Conclusion sur l'expérimentation

Dans ce chapitre, nous avons présenté une implémentation des trois étapes que nous avons identifiées pour la conception d'une IA d'assistance qui s'appuie sur le processus décisionnel humain. Nous avons établi un modèle du processus décisionnel de l'opérateur pour une tâche de replanification d'un plan de vol de drone dans un environnement de simulation à travers des entretiens d'autoconfrontation. Ensuite, nous avons analysé comment la manière dont l'opérateur utilise les différents outils permet de révéler dans quelle phase de construction de la solution il se situe, ce qui nous a permis d'automatiser grâce à un algorithme le suivi de ce processus décisionnel. Enfin, nous avons proposé deux opportunités d'assistance adaptées à l'évolution de ce processus : une aide à l'exploration, et une aide à l'exploitation des plans de vol. Ces résultats montrent qu'il est possible de concevoir des outils d'aide à la décision qui laissent une part de créativité à l'opérateur et permettent ainsi d'avoir des indices sur ce qu'il cherche à réaliser et la manière dont il cherche à atteindre son but. Ceci ouvre la possibilité à l'outil d'assistance d'avoir un comportement plus adaptatif qui prenne mieux en compte les besoins actuels de l'opérateur en se basant sur les processus cognitifs mis en œuvre plutôt que par la construction d'un profil basé par exemple sur sa personnalité.

Chapitre 6 : Discussion générale

I. Apports méthodologiques sur la simulation de la prise de décision en équipe « opérateur-IA »

Cette thèse propose tout d'abord un apport d'ordre méthodologique à travers la conception d'un environnement de simulation qui permette d'étudier l'impact de diverses modalités d'interaction humain-IA, alors même que le système d'IA n'est pas encore complètement spécifié, ni même développé. Cette méthodologie permet très en amont dans le processus de développement d'argumenter des choix sur les rôles que devraient prendre les différents agents, sur les informations échangées, sur les manières d'interagir. Dans le contexte de la replanification de plan de vol, nous avons développé un micromonde qui représente de manière simplifiée le problème auquel sont confrontés les opérateurs. Cela permet de trouver les solutions optimales avec un algorithme de recherche opérationnelle (qui va émuler ce que ferait l'algorithme d'IA dans le monde réel), tout en restant suffisamment complexe pour que la décision nécessite l'assistance de cette IA. Le calcul se fait en temps réel et prend en compte les modifications dont l'utilisateur est à l'initiative. La décision prise repose sur des choix stratégiques et tactiques qui se basent sur la construction d'un plan de vol à travers une exploration des solutions possibles. Cette tâche de prise de décision est suffisamment difficile pour que l'assistance de l'IA soit nécessaire et que l'opérateur s'engage dans une interaction complexe à travers l'interface. La cinquantaine de personnes ayant participé aux différentes expérimentations a rapporté avoir trouvé l'IA performante, ce qui indique que l'utilisation d'algorithmes de recherche opérationnelle peut permettre de simuler, au regard d'utilisateurs humains, des applications reposant sur des algorithmes d'IA plus complexes, comme de l'apprentissage automatique lorsque l'espace des solutions est restreint.

L'utilisation de cette approche peut être un réel atout pour la conception et la définition détaillée d'un nouveau système. Dans notre cas, même si notre finalité n'est pas de développer un système de replanification pour des opérateurs de drones, l'utilisation de l'interface a permis d'identifier plusieurs pistes d'améliorations des outils proposés qui relèvent de l'ergonomie et amélioreraient l'expérience

utilisateur. L'outil de consultation de l'historique des plans, qui a une structure de pile, se révèle insatisfaisant lorsque les participants cherchent à effectuer un retour arrière. Cela rend compliqué la comparaison de plusieurs plans créés à partir d'une même étape, et d'autant plus si plusieurs étapes de modification les séparent de cette étape. Une solution pourrait être d'ajouter la possibilité d'enregistrer un plan comme favori, et qu'il soit possible de l'afficher à tout moment en cliquant sur un bouton d'affichage du plan favori. Il serait de cette manière possible en un clic de comparer le plan courant et un plan issu d'une branche alternative. Une autre solution, plus ambitieuse, pourrait être de placer l'IA à l'initiative de ce type de comparaison en laissant à l'IA le soin d'identifier les possibles plans favoris en fonctions de la recherche de l'opérateur, puis en laissant l'IA déterminer les moments où proposer ces comparaisons avec un des plans mis en favoris. Une autre suggestion des participants a porté sur l'outil de modification manuelle du plan, et plus particulièrement sa fonction d'ajout d'étape sur le plan. Actuellement, l'opérateur doit dans un premier temps sélectionner la cible précédant l'étape qu'il veut ajouter, et dans un deuxième temps indiquer où il veut ajouter l'étape, pour que dans un troisième temps l'IA replanifie entre la cible précédant l'étape et la cible la suivant. L'ordre des deux premières étapes est peu intuitif bien que les utilisateurs utilisent judicieusement cet outil après une ou de deux utilisations, une fois qu'ils en ont intégré le fonctionnement. Une solution pourrait être de supprimer cette première étape. L'utilisateur indiquerait simplement sur le terrain où il veut ajouter une étape, et l'IA déterminerait entre quelles cibles par lesquelles passent le plan celle-ci s'intégrerait de façon optimale. Cependant, ce fonctionnement qui augmente la participation de l'IA va à l'encontre de la philosophie de cet outil qui est utilité au contraire pour avoir un contrôle fin sur le plan et y appliquer des modifications d'ordre tactique. Lorsque les participants ajoutent une étape au plan, ils ont généralement une idée précise de la déviation qui va être impliquée et qui est même souvent recherchée. Plusieurs participants ont proposé une autre solution, qui consiste à implémenter un fonctionnement de glisser-déposer pour ajouter des étapes au plan. L'opérateur cliquerait sur le plan et maintiendrait le bouton de la souris enfoncé. La cible précédant et la cible suivant le point du plan sélectionné seraient détectées. Ensuite, à chaque fois que le curseur serait déplacé sur le terrain et tant que le bouton ne serait pas relâché, le système pourrait afficher à la volée le segment de plan entre les deux cibles et passant par la case survolée. Cette fonctionnalité

permettrait également de favoriser la comparaison de l'ajout de différentes étapes, qu'il suffirait alors de survoler sans avoir relâché le bouton de la souris pour observer comment ils s'intègrent au plan.

Cette démarche de construction d'un micromonde dédié, représentatif de ce que l'opérateur aura à faire, pourrait être utilisée très en amont dans des processus de conception de nouveaux services à base d'IA et amène des éléments pertinents pour spécifier le système cible. Son utilisation a permis de mettre en évidence l'intérêt de repenser les interactions afin de faciliter le travail en équipe opérateur-IA : plusieurs approches sont envisageables, de la mise en mémoire (en réserve) par l'opérateur de plans particuliers pour pouvoir y revenir de manière simple par un outil dédié, à l'identification par l'algorithme d'IA des plans ayant retenu l'attention et devant être gardés pour une comparaison probable. L'émulation de ce que fera le système d'aide à la décision et son utilisation par un opérateur dans une démarche expérimentale est ainsi une approche viable et pertinente pour orienter la conception du système final. Ces résultats peuvent ensuite guider la conception de systèmes plus appliqués et être précisés à travers l'utilisation d'environnements de test plus écologiques.

Par ailleurs, l'utilisation de ce micromonde offre un cadre contrôlé qui permet de tester rapidement des hypothèses scientifiques plus générales sur les interactions humain-IA, en répliquant les mêmes conditions expérimentales pour un même participant ou entre plusieurs participants. Son utilisation nous a permis de mieux comprendre certains principes généraux qui sous-tendent les interactions des équipes humain-IA, sur leur manière d'interagir, sur les invariants de ces processus décisionnels et les facteurs qui peuvent être pris en compte. Nous avons conçu plusieurs types d'outils pour la replanification qui exploitent les possibilités qu'offrirait l'algorithme d'IA implémenté : aider l'opérateur à faire des choix tactiques, des choix stratégiques, à optimiser, ou encore proposer des solutions complètes. Le fait de laisser un opérateur utiliser ces différents outils soit de manière contrainte, soit de manière libre, permet d'étudier de manière expérimentale d'une part l'impact du rôle assigné à l'IA sur les choix et le ressenti de l'utilisateur, d'autre part la façon dont l'utilisateur souhaiterait interagir avec l'IA. Cette méthode d'émulation de l'IA, de développement d'outils et d'étude expérimentale des interactions se révèle être

réalisable et intéressante pour anticiper et étudier la manière dont un opérateur interagit avec une IA et l'impact du type d'interaction sur la décision retenue. Ce type d'expérimentation a permis de valider que la construction conjointe passe bien par une succession de phases d'exploration et de phases d'exploitation des solutions. L'opérateur utilise aussi une stratégie de retour arrière où il compare plusieurs plans précédemment explorés ou exploités qui ont suscité son intérêt.

Ainsi, les différents outils de replanification implémentés ont été appréciés et facilement pris en main. Cette souplesse d'utilisation, qui était initialement destinée à comparer plusieurs fonctionnalités laissées à la libre disposition des participants, s'est finalement révélée être un concept d'interaction en soi qui favorise l'appropriation de l'IA par des individus différents. Permettre à l'opérateur de choisir dynamiquement le rôle donné à l'IA, de la proposition décisionnelle à la supervision, est une première piste pour mieux adapter l'IA à l'évolution des ressources cognitives et temporelles de l'opérateur au fil de la mission.

II. Apports théoriques sur l'impact de l'évolution de la participation de l'IA sur la prise de décision de l'opérateur

Le premier apport théorique de cette thèse porte sur l'impact de l'évolution de la participation de l'IA sur la prise de décision de l'opérateur. L'augmentation croissante du niveau d'autonomie décisionnelle des systèmes aéronautiques laisse entrevoir l'arrivée de drones qui vont interagir sur des fenêtres temporelles limitées avec les opérateurs en charge de leur supervision. La responsabilité des opérateurs restera cependant pleinement engagée malgré cette évolution d'une fonction de pilotage à distance ou de suivi en permanence des informations en provenance du drone, à une fonction de supervision du drone concentrée sur des phases précises et peut-être éphémères qui amèneront à ce que les informations disponibles soient déjà prétraitées, voire synthétisées par les assistants intelligents. C'est pour cela que nous avons voulu investiguer expérimentalement comment cet accroissement de la participation de l'IA à la planification va impacter la prise de décision de l'opérateur et sa responsabilité ressentie. Pour cela, nous avons défini deux modes d'interaction avec l'IA : coopératif, où l'opérateur utilise les outils de replanification, et collaboratif,

où l’IA prend plus d’initiative et replanifie automatiquement lorsque les menaces sont détectées, l’opérateur gardant la possibilité de demander à l’IA de proposer d’autres suggestions de plan. L’évolution du mode d’interaction de la coopération à la collaboration correspond à une augmentation de l’assistance de l’IA, et l’évolution du mode d’interaction de la collaboration à la coopération correspond à une diminution de l’assistance de l’IA. L’enjeu de notre étude était de mieux cerner les effets de ces changements d’interaction sur l’expérience utilisateur.

Lorsque l’assistance de l’IA diminue, nous observons que les utilisateurs humains ressentent un manque des suggestions de plans qui leur étaient fournies et qui était le point de départ de leur réflexion pour définir une nouvelle solution. Ils doivent compenser la perte de cette aide par une implication accrue dans la tâche qui les fait se sentir plus investis dans leur rôle que précédemment, et donc plus responsables. Néanmoins, lorsque la participation de l’IA augmente, les utilisateurs rapportent se sentir toujours autant à l’origine des plans construits alors que des suggestions ont été introduites et vont influencer leur prise de décision. Les plans de vol réalisés avec une organisation coopérative sont plus variés que ceux réalisés dans une organisation collaborative. Lorsque des suggestions de plans sont proposées par le système, l'opérateur aura tendance à valider les plans qui s'en rapprochent, mais sans être conscient de cette influence sur son comportement. Un système dédié à assister un opérateur humain dans une tâche de décision peut ainsi insidieusement modifier la solution validée par cet opérateur. Il s'agit d'un problème critique lors du développement de systèmes d'assistance sophistiqués alors que l'utilisateur n'a pas conscience d'être influencé par le système d'IA. Ce comportement est en quelque sorte normalisé par les suggestions de l'IA et il ne remet plus en question la solution potentielle de la même manière. D'une part, la réduction de la variabilité humaine peut être considérée comme souhaitable car la prévisibilité du système peut conduire à une meilleure fiabilité globale. D'autre part, l'intérêt de conserver des opérateurs humains réside précisément dans leur capacité à faire preuve d'un esprit critique et subjectif dans le processus de décision. Une piste d’amélioration pourrait être l'ajout d'explications sur les forces et faiblesses des solutions proposées, ou la mise en évidence des choix sous-jacents qui conduisent à ces solutions, pour stimuler la vigilance de l'opérateur.

Ces résultats montrent que la manière dont les tâches sont réparties entre l'opérateur et le système intelligent a un impact sur l'expérience de l'opérateur, et que le passage à une répartition plus dynamique des rôles nécessitera de repenser le processus de prise de décision afin de garantir à l'opérateur un niveau adéquat de conscience de la situation. La principale limite de cet apport théorique réside dans le fait que dans le mode d'interaction collaboratif, les propositions de plan émises à l'initiative de l'IA ne le sont qu'après l'événement qui modifie la situation tactique. Lors de la suite de l'interaction, l'IA ne fait que calculer et afficher les plans résultant de l'utilisation des outils de replanification. Or, nous avons vu dans notre analyse de la littérature que dans une équipe de personnes humaines, la collaboration est caractérisée par une possibilité de spontanéité dans la gestion du travail à accomplir et en réaction aux actions des autres membres. Pour que l'interaction opérateur-IA que nous avons implémentée soit plus collaborative, l'IA devrait avoir la capacité de commenter et de réagir aux modifications de plans effectuées par l'opérateur. Seulement, fournir des suggestions de plan à l'opérateur pendant qu'il utilise une succession d'outil de replanification nécessite de définir lorsque celui-ci en a besoin et la manière dont elles peuvent être pertinentes. La difficulté réside dans la prise en compte du processus décisionnel de l'opérateur et dans la manière dont de telles propositions de l'IA pourraient l'assister, ce qui est l'objet des deux autres apports théoriques. Enfin, la question de la reproductibilité de ce résultat pour d'autres types de décision reste ouverte. Il est difficile à ce stade d'estimer la robustesse de ce résultat, et de savoir dans quelle mesure il est spécifique à l'environnement de simulation utilisé et donc lié à son contexte et ses outils particuliers.

III. Apports théoriques sur la variabilité humaine dans la prise de décision avec l'IA

Le second apport théorique de cette thèse porte sur l'étude des liens entre les traits de personnalité des utilisateurs et les types de décision auxquels ils aboutissent lorsqu'ils collaborent avec une IA. La tâche de replanification que les utilisateurs devaient réaliser n'avait aucune solution satisfaisante et il fallait nécessairement réaliser un compromis entre plusieurs mesures de performances (la réalisation de tous les objectifs, le risque pris et le carburant utilisé). L'analyse du compromis validé indique que les utilisateurs avaient des préférences individuelles

concernant la variable à dégrader. Nous avons effectué une classification hiérarchique sur le compromis moyen des utilisateurs en fonction de ces trois variables, et avons observé trois groupes distincts de décisions prises. L'analyse en composantes principales des participants en fonction de leur compromis moyen montre que certains participants préfèrent consommer beaucoup de carburant, d'autres préfèrent abandonner certains objectifs et d'autres encore préfèrent prendre des risques élevés pour mener à bien la mission. Bien que notre groupe de participants soit assez homogène, nous nous sommes demandé si ces différences pouvaient être expliquées par des types de personnalités différentes. Nous avons donc cherché si certains traits de personnalité permettaient de différencier ces trois groupes. L'analyse ne permet pas d'établir un lien entre ces traits mesurés et le type de compromis que les opérateurs préfèrent valider. Cependant, si le type de décision validé ne semble pas se réduire au caractère individuel, nous avons observé un lien entre les groupes de compromis de plan réalisés et le rôle attribué à l'IA lors de la phase de construction du plan. Les opérateurs qui préfèrent abandonner beaucoup d'objectifs commencent généralement la phase d'interaction en consultant les suggestions de l'IA, puis co-construisent une solution avec le système en donnant des critères à optimiser. Les opérateurs qui valident des solutions à haut risque utilisent aussi initialement l'outil de consultation des suggestions, mais préfèrent ensuite modifier manuellement le plan pour introduire des contraintes tactiques. Les opérateurs dont les solutions consomment beaucoup de carburant élaborent eux-mêmes de nouveaux plans et demandent ensuite au système de participer à l'élaboration. Ces préférences individuelles dans le type de critère de compromis de plan à privilégier ne sont pas liées à la personnalité des participants. Cette variabilité humaine dans la prise de décision réalisée avec l'IA ne peut être réduite à des éléments de caractérisation personnelle qui l'expliqueraient. Ainsi, il semble dans le cadre de notre étude que ce ne soient pas les traits de caractère qui guident le type de choix fait, mais que celui-ci soit corrélé à la séquence des outils utilisés, c'est-à-dire à la façon dont l'opérateur choisit d'interagir avec l'IA.

Or, on peut remarquer que dans le cadre des interactions sur le web par exemple, beaucoup de sites essaient actuellement de suivre des informations sur les personnes (choix musicaux, achats réalisés, centres d'intérêts, lieux regardés…) pour proposer du contenu lié à cet historique ou à leur profil démographique. On y

retrouve la démarche qui consiste à proposer du contenu cohérent avec le groupe associé à la personne et donc à son type de compromis préféré ou à ses choix préalables. Si cette approche est généralement satisfaisante pour la musique et la vidéo où le consommateur a tendance à rester sur les choix proposés, on peut faire l'expérience que pour les achats l'IA de recommandation a souvent un temps de retard et propose ce qu'on a déjà acheté et qui ne nous intéresse donc plus. Nos résultats expérimentaux peuvent contribuer à orienter le développement de ce type d'IA. En particulier ils mettent en évidence que c'est plutôt le comportement actuel de l'utilisateur qui est à exploiter pour guider les propositions de l'IA que des éléments liés à sa personnalité. Ceci nous invite à évaluer comment les interactions et les rôles choisis au sein de l'équipe peuvent être utilisés pour guider le fonctionnement des systèmes intelligents afin qu'ils contribuent efficacement à aider les utilisateurs humains dans leur représentation de la situation et dans leur prise de décision. La population de nos études étant assez homogène, il serait également bon de vérifier si ce résultat se confirme avec des populations plus variées en termes d'âge ou de catégories socioprofessionnelles qui pourraient avoir des attentes et un usage différents de l'IA.

IV. Apports théoriques sur la conception d'une IA adaptée au processus décisionnel de l'opérateur

Nous avons vu que l'opérateur n'attend pas uniquement des solutions globales de l'IA mais est prêt à entrer dans un processus collaboratif pour définir le plan de vol. Un tel processus collaboratif implique que les différents membres de l'équipe, coéquipier humain et coéquipier IA, puissent adapter leur comportement pour réagir de manière dynamique aux actions de l'autre afin d'atteindre le but commun. Dans notre environnement de simulation, si l'opérateur a la capacité grâce aux outils à sa disposition de demander librement à l'IA d'effectuer des modifications stratégiques ou tactiques ou de lui fournir des plans entiers, à l'inverse l'IA ne prend actuellement des initiatives qu'au début de l'interaction. Lorsque les nouvelles menaces sont détectées, l'information est rapportée à l'opérateur et un nouveau plan de vol qui les prend en compte est affiché sans qu'aucune contribution humaine ne soit nécessaire. Mais par la suite lorsque l'opérateur réagit à cette suggestion, l'IA se cantonne à répondre aux demandes de calcul impliquées par l'utilisation des outils.

La collaboration réelle est donc limitée à une certaine phase de l'interaction et peut être qualifiée d'incomplète.

Permettre à l'IA de fournir des suggestions de plan de vol également lorsque l'opérateur est en train d'effectuer des modifications successives va impliquer de pouvoir interférer avec le flux de travail humain en y greffant des opérations non sollicitées. Pour que ces interférences de l'IA soient jugées utiles et donc acceptées par l'opérateur, celles-ci doivent apporter une plus-value à la seule utilisation des outils pour la création du nouveau plan de vol. Une telle plus-value peut être envisagée par exemple comme une réduction du nombre d'étapes nécessaires pour arriver au plan désiré et donc un gain de temps, ou encore comme une assistance à la définition d'un plan acceptable en aidant l'opérateur à parcourir l'espace des solutions. Pour cela, entre coéquipiers humains, chacun va se créer un modèle mental de la manière dont l'autre se représente le problème et la solution à y apporter, et va le comparer à sa propre représentation du problème (Figure 43). Cet écart de représentation va être progressivement comblé par une réaction réciproque aux actions du coéquipier, jusqu'à arriver à un consensus ou bien une sortie de l'échange par l'un ou l'autre (Hoffman et al., 2018).

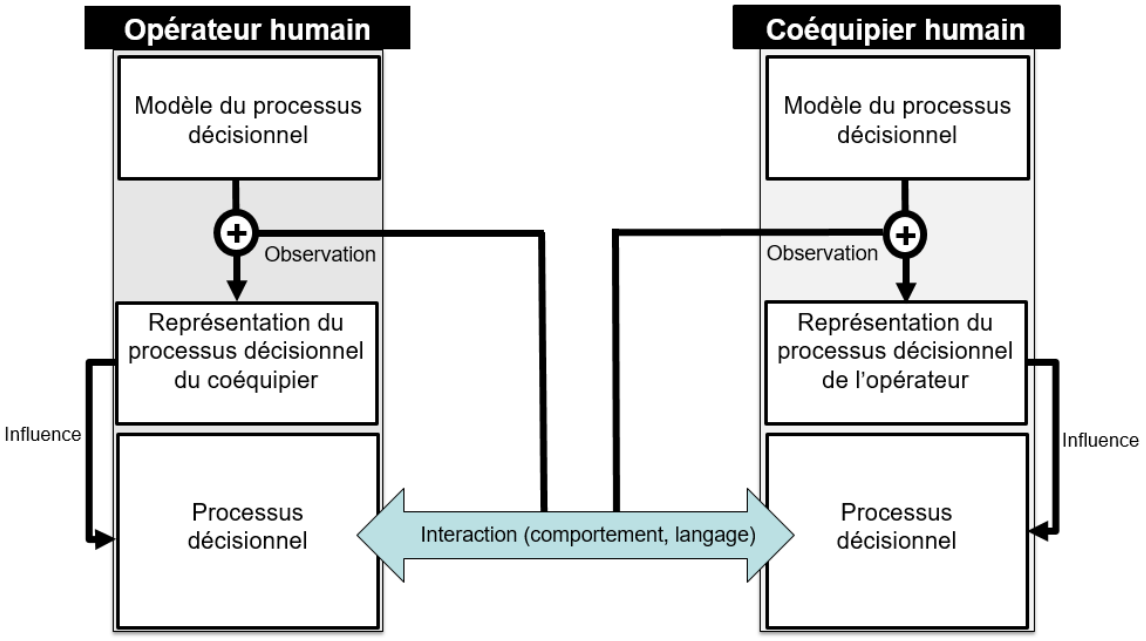


Figure 43 : Modèle de la prise de décision collaborative entre coéquipiers humains.

Dans un collectif humain-IA qui veut tendre vers un niveau de collaboration comparable à celui mis en place par un collectif de personnes humaines, l'IA devrait

de la même manière avoir accès à une forme de représentation de la solution à atteindre que l'opérateur a pour pouvoir aider l'opérateur à y parvenir, ou bien à améliorer cette solution à atteindre à travers une démarche critique. Seulement, nous avons observé dans la deuxième étude expérimentale que le choix des opérateurs et la manière de faire ce choix ne sont pas liés à un profil d'utilisateur général. Un tel profil aurait pu constituer une représentation utilisable par l'IA, mais il pourrait être plus efficace de se baser sur le comportement même de l'opérateur. Dans notre environnement, ce comportement qui puisse être accessible par l'IA se limite aux outils de modifications utilisés : quels outils, dans quel ordre. La dernière étude expérimentale a permis de montrer que ces outils permettent effectivement de modéliser dans une certaine mesure le processus décisionnel de l'opérateur et de s'appuyer sur ce modèle pour que l'IA y réagisse dynamiquement. Nous avons suivi trois étapes pour cela : premièrement, construire un modèle du processus décisionnel des opérateurs pour un cas d'usage, c'est-à-dire l'identification des grandes étapes d'élaboration d'une solution à ce type de problème donné, deuxièmement, automatiser le suivi d'une instance de ce modèle décisionnel, et troisièmement, identifier des opportunités d'assistance adaptées au processus décisionnel de l'opérateur (Figure 44).

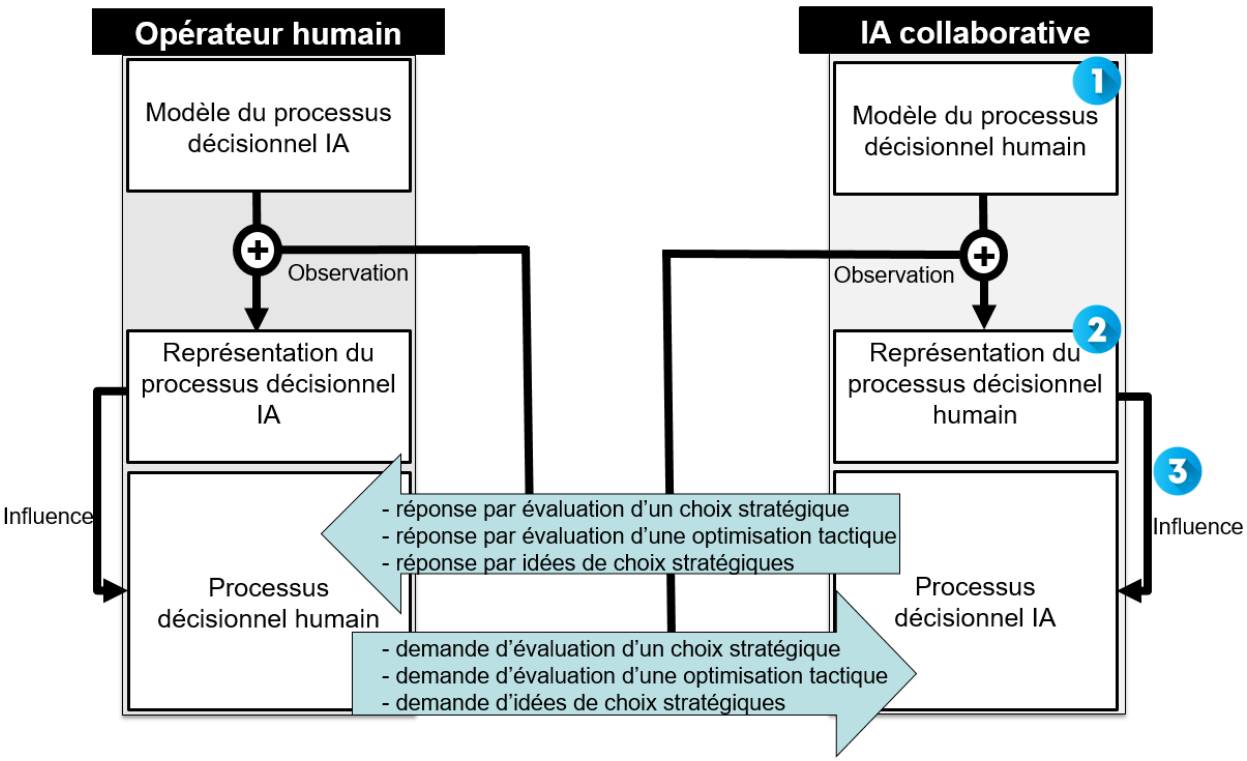


Figure 44 : Modèle de la prise de décision assistée par une IA collaborative qui s'adapte au processus décisionnel de l'opérateur.

Nous avons montré grâce à une analyse experte que la manière dont l'opérateur utilise les outils de modification permet de révéler dans quelle phase de construction de la solution il se situe. Pour les outils de modification manuelle et de co-construction, où l'opérateur participe à l'élaboration de la solution, les critères de recherche de plan qu'il fournit à l'IA contiennent des indices sur sa stratégie décisionnelle sous-jacente. Ces liens entre les modifications des critères de compromis de plan et le processus décisionnel humain nous ont permis de développer un algorithme qui peut automatiser la reconnaissance des phases d'exploration et d'exploitation par lesquelles doivent passer les opérateurs au cours de la replanification. Ce suivi du processus décisionnel permet d'envisager plusieurs opportunités d'assistance adaptées à son évolution : une aide à l'exploration, et une aide à l'exploitation des plans de vol.

Cette démarche montre qu'il est possible de configurer des outils d'aide à la décision qui laissent une partie de la créativité à l'opérateur et permettent ainsi d'avoir des indices sur ce qu'il cherche à réaliser et la manière dont il veut atteindre son but. Ceci ouvre la possibilité à l'outil d'assistance d'avoir un comportement plus adaptatif qui prend mieux en compte les besoins actuels de l'opérateur en se basant sur les processus cognitifs mis en œuvre plutôt que par la construction d'un profil basé par exemple sur sa personnalité. Bien que ce modèle soit réalisé dans le cadre restreint de la décision liée à la replanification d'un trajet, il montre l'intérêt d'étendre le paradigme d'interaction humain/système et de ne pas se limiter à une explicitation pour l'humain de ce que fait le système. Une adaptation réciproque des deux agents est envisageable, du moment que les outils mis à disposition permettent d'accéder au moins partiellement à leurs intentions. La fréquence des interventions de l'IA reste une question ouverte à ce stade, et nécessiterait d'étudier expérimentalement dans quelle mesure elles ne risqueraient pas d'altérer voire de se substituer au processus décisionnel humain en essayant de l'accompagner.

Conclusion générale

Les travaux de recherche menés dans le cadre de cette thèse ont eu pour objectif d'anticiper des enjeux de facteur humain liés à l'évolution de l'intelligence artificielle passant d'un rôle d'outil à celui de partenaire pour la prise de décision humaine. Le cadre applicatif visé est celui de la replanification de plan de vol de drone par une équipe composée d'un opérateur humain et d'un système doté d'IA. Nous avons réalisé une revue de littérature interdisciplinaire qui a mobilisé des travaux en intelligence artificielle pour le développement d'algorithmes d'optimisation de plans de vol, en management pour la prise de décision en mode collaboratif, ou encore en sciences cognitives afin d'identifier comment l'arrivée des systèmes aéronautiques à haut niveau d'autonomie décisionnelle vont nécessiter de repenser les interactions avec l'opérateur pour qu'il puisse exercer pleinement sa responsabilité. Nous avons suivi une démarche expérimentale à travers la réalisation de trois études avec des participants pour investiguer nos problématiques scientifiques : dans quelle mesure l'augmentation de la participation de l'IA à la prise de décision va altérer les solutions validées par l'opérateur ; dans quelle mesure les éléments de personnalité de l'opérateur peuvent être explicatifs de sa prise de décision avec l'IA ; comment l'IA peut venir assister l'opérateur de façon adaptée à son processus décisionnel. Nous avons pour cela développé un environnement de simulation simplifié où des participants incarnent un opérateur de drone chargé d'effectuer des missions qui vont l'amener à replanifier en interagissant avec l'IA, ce qui nous a permis d'éprouver certains usages futurs de l'IA.

Les résultats expérimentaux nous ont appris que l'évolution de l'assistance apportée par l'IA va influencer l'expérience de l'utilisateur humain. Lorsque l'IA augmente sa participation à la prise de décision, l'opérateur n'a pas forcément conscience que sa propre contribution diminue et que les solutions validées tendent à se rapprocher des propositions de l'IA. L'introduction d'une IA plus collaborative qui va proposer des suggestions de solutions ne va pas nécessairement induire de diminution du sentiment de responsabilité humaine, ce qui est problématique dans la mesure où les plans validés par les opérateurs se rapprochent de ces suggestions. Cette observation soulève un risque de sous-estimation des effets de l'IA sur la prise de décision humaine qui pourrait être altérée à son insu. Par ailleurs, nous avons

mesuré une importante variabilité humaine dans l'usage fait de l'IA et dans les solutions validées, mais qui ne peut pas s'expliquer par des éléments de personnalité des différents opérateurs. Classer les opérateurs en se basant exclusivement sur des éléments de caractérisation individuelle risque d'enfermer le l'IA dans des schémas trop rigides, alors que ces différences entre individus semblent plutôt relever d'une représentation dynamique de la situation. Cependant, nous avons montré que le système peut malgré cela interagir différemment selon l'opérateur et accompagner son processus décisionnel individuel dans la mesure où l'IA est capable de construire un modèle du processus décisionnel humain pour la planification et d'en suivre l'évolution. Nous avons observé que ce processus décisionnel est caractérisé par une alternance de phases d'exploration et de phases d'exploitation des solutions, et que l'usage fait des capacités de l'IA diffère selon ces phases. L'opérateur utilise aussi une stratégie de retour arrière où il compare plusieurs plans précédemment explorés ou exploités qui ont suscité son intérêt. De même qu'un équipier humain attribue des états mentaux aux autres membres de l'équipe qui vont guider la manière dont il va interagir avec eux, ces recherches confortent notre thèse selon laquelle la capacité de l'IA à bâtir une théorie de l'esprit pourrait être la clef pour en faire un coéquipier réellement collaboratif. Même si le chemin à parcourir est encore long pour y parvenir, ce travail apporte une pierre à l'édifice et participe à la formalisation des systèmes d'assistance intelligents qui resteront au service de la décision et de la responsabilité humaine.

Publications issues des travaux de cette thèse

Metge, A., Maille, N., & Le Blanc, B. (2022, June). Décision assistée par un système intelligent: de l'utilisation de l'IA à l'intention de l'opérateur dans la replanification. In *CNIA 2022: Conférence Nationale en Intelligence Artificielle*.

Metge, A., Maille, N. & Le Blanc, B. (2022, April). Dynamic replanning assisted by intelligent system: from observed behavior to operator's intent. Poster presented at: Human Factors and Ergonomics Society Europe Chapter 2022 Annual Conference; Turin, Italie.

Metge, A., Maille, N., & Le Blanc, B. (2021, October). Interface de collaboration humain-IA: application au cas de la replanification de vol de drone. In *ERGO'IA 2021*.

Metge, A., Maille, N., & Le Blanc, B. (2021, June). Transition between cooperative and collaborative interaction modes for human-AI teaming. In *CNIA 2021: Conférence Nationale en Intelligence Artificielle* (pp. pp-38).

Metge, A., Maille, N., & Le Blanc, B. (2021, March). Operators and autonomous intelligent agents: human individual characteristics shape the team's efficiency. In *1st International Conference on Autonomous Intelligent Cyber-defense*.

Metge, A., & Maille, N. (2020, June). IA et assistance à la décision humaine: influence des informations transmises sur l'évaluation d'un nouveau plan de vol. In *PFIA/RJCIA 2020 (La Plate-Forme de l'Intelligence Artificielle)*.

Bibliographie

Abbass, H. A. (2019). Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust. *Cognitive Computation*, 11(2), 159-171.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of the Econometric Society*, 503-546.

Alonso, J. M., Castiello, C., & Mencar, C. (2018, June). A bibliometric analysis of the explainable artificial intelligence research field. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems* (pp. 3-15). Springer, Cham.

Altmann, T., & Roth, M. (2018). The Self-esteem Stability Scale (SESS) for cross-sectional direct assessment of self-esteem stability. *Frontiers in psychology*, *9*, 91.

Barbour, Rachael. (2018). Collaboration versus Co-operation Grassroots Activism in Divided Cities and Commu-nication Networks.

Barnes, M. J., Lakhmani, S., Holder, E., & Chen, J. Y. (2019). *Issues in human-agent communication*. Army Research Laboratory.

Baron, R.A. (1998). Cognitive mechanisms in entrepreneurship: Why and when entrepreneurs think differently than other people. *Journal of Business ventuing*, 13(4), 275–294.

Barria-Pineda, J. (2020, July). Exploring the Need for Transparency in Educational Recommender Systems. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 376-379).

Berger-Tal, O., Nathan, J., Meron, E., & Saltz, D. (2014). The exploration-exploitation dilemma: a multidisciplinary framework. *PloS one*, 9(4), e95693.

Besold, Tarek & Uckelman, Sara. (2018). The What, the Why, and the How of Artificial Explanations in Automated Decision-Making. *arXiv preprint arXiv :1808.07074*.

Boquet, J. « L'intelligence artificielle entraînera-t-elle une mutation du domaine militaire ?», *Air&Cosmos*, n°2635, 05.04.2019, pp. 12-14.

Borgo, R., Cashmore, M., & Magazzeni, D. (2018). Towards Providing Explanations for AI Planner Decisions. *ArXiv*, abs/1810.06338.

Briant, R., (2021). La synergie homme-machine et l'avenir des opérations aériennes, *Focus stratégique*, n° 106, Ifri.

Brill, E., & Moore, R. C. (2000, October). An improved error model for noisy channel spelling correction. In *Proceedings of the 38th annual meeting of the association for computational linguistics* (pp. 286-293).

Cardon, D., Cointet, J. & Mazières, A. (2018). La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux*, 211(5), 173-220.

Cegarra, J., & Morgado, N. (2009, September). Étude des propriétés de la version francophone du NASATLX. In *Communication présentée à la cinquième édition du colloque de psychologie ergonomique (Epique)* (pp. 28-30).

Chakraborti, T. (2018). *Foundations of Human-Aware Planning-A Tale of Three Models* [Doctoral dissertation, Arizona State University].

Chakraborti, T., Kambhampati, S., Scheutz, M., & Zhang, Y.L. (2017). AI Challenges in Human-Robot Cognitive Teaming. *ArXiv*, abs/1707.04775.

Chakraborti, T., Sreedharan, S., & Kambhampati, S. (2019). Balancing Explanations and Explicability in Human-Aware Planning. *IJCAI*.

Cigolotti, O., Carlotti, M.-A. (2021). Haut-Karabagh : dix enseignements d'un conflit qui nous concerne. Rapport d'information fait au nom de la commission des affaires étrangères, de la défense et des forces armées, *rapport d'information n° 754* (2020-2021).

Comité d'éthique de la défense (2021). Avis sur l'intégration de l'autonomie dans les systèmes d'armes létaux. *Rapport du ministère des armées*. https://www.defense.gouv.fr/sites/default/files/ministere-

armees/20210429_Comit%C3%A9%20d%27%C3%A9thique%20de%20la%20d%C3%A9fense%20-%20Avis%20int%C3%A9gration%20autonomie%20syst%C3%A8mes%20armes%20l%C3%A9taux.pdf

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, *7*(3), 171-176.

Demiaux, V. (2017). Comment permettre à l'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle, *rapport du CNIL*. http://www.senat.fr/rap/r19-642/r19-642.html

Demir, M., McNeese, N. J., & Cooke, N. J. (2020). Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies*, 140, 102436.

Deterding, S., Hook, J., Fiebrink, R., Gillies, M., Gow, J., Akten, M., Smith, G., Liapis, A., & Compton, K. (2017). Mixed-initiative creative interfaces. In *CHI 2017 Extended Abstracts - Proceedings of the 2017 ACM SIGCHI Conference on Human Factors in Computing Systems: Explore, Innovate, Inspire* (pp. 628-635).

Dillenbourg, P., Baker, M., Blaye, A. & O'Malley, C. (1996) The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds) *Learning in Humans and Machine: Towards an interdisciplinary learning science*. (Pp. 189- 211). Oxford: Elsevier.

Dupuy, R. (2013). L'industrie européenne de défense : changements institutionnels et stratégies de coopétition des firmes. *Innovations*, 42, 85-107. https://doi.org/10.3917/inno.042.0085

Ferreira, J. J., & Monteiro, M. S. (2020, July). What Are People Doing About XAI User Experience? A Survey on AI Explainability Research and Practice. In *International Conference on Human-Computer Interaction* (pp. 56-73). Springer, Cham.

Flathmann, C., Schelble, B., Tubre, B., McNeese, N., & Rodeghero, P. (2020, November). Invoking Principles of Groupware to Develop and Evaluate Present and

Future Human-Agent Teams. In *Proceedings of the 8th International Conference on Human-Agent Interaction* (pp. 15-24).

Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. (2015). Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots*, 39(3), 293-312.

Goya, M. (2020). Les enseignements opérationnels de la guerre du Haut-Karabakh. *Blog La Voie de l'épée*, 12.

Groupe de travail ADER. (2016). Du Reaper à l'UCAV : de nouveaux enjeux pour l'Armée de l'air. *Revue Défense Nationale*, 786, 99-108. https://doi.org/10.3917/rdna.786.0099

Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44-58.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. *arXiv preprint arXiv:1812.04608*.

Huderbeau E. (2019). « Airbus au cœur du SCAF », *Air&Cosmos*, n°2632, 15.03.2019, pp.26-27.

Klein, G. (1997). The Recognition-Primed Decision (RPD) model: looking back, looking forward. In C.E. Zsambok & G.A. Klein (Eds.), *Naturalistic Decision Making* (pp. 285-292). New Jersey: Lawrence Erlbaum Associates.

Klein, G. (2008). Naturalistic decisionmaking. *Human factors*, 50(3), 456-460.

Klein, G., Ross, K.G., Moon, B.M., Klein, D. E., Hoffman, R.R., & Hollnagel, E. (2003). *Macrocognition*, IEEE Intelligent Systems, 18(3), 81-85.

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. *Organizational simulation*, *53*, 139-184.

Klein, Gary & Hoffman, Robert & Mueller, Shane. (2019). Naturalistic Psychological Model of Explanatory Reasoning: How people explain things to others and to themselves. In *International Conference on Naturalistic Decision Making.*

Kox, E., Siegling, L., & Kerstholt, J. H. (2021). Autonomous systems as intelligent teammates: social psychological implications. In *24th International Command and Control Research & Technology Symposium.*

Kozar, O. (2010). Towards Better Group Work: Seeing the Difference between Cooperation and Collaboration. In *English teaching forum* (Vol. 48, No. 2, pp. 16-23). US Department of State. Bureau of Educational and Cultural Affairs, Office of English Language Programs, SA-5, 2200 C Street NW 4th Floor, Washington, DC 20037.

Krause, J., Perer, A., & Bertini, E. (2016). Using visual analytics to interpret predictive machine learning models. *arXiv preprint arXiv:1606.05685*.

Lachter, J., Hobbs, A., & Holbrook, J. (2021). Thinking Outside the Box: The Human Role in Increasingly Automated Aviation Systems. In *62nd International Symposium on Aviation Psychology* (p. 134).

Laureiro-Martínez, D., Brusoni, S., Canessa, N., & Zollo, M. (2015). Understanding the exploration–exploitation dilemma: An fMRI study of attention control and decision-making performance. *Strategic management journal*, 36(3), 319-338.

Le Gleut, R., Conway-Mouret, H. (2022). 2040, l'odyssée du SCAF – Le système de combat aérien du futur. Rapport d'information fait au nom de la commission des affaires étrangères, de la défense et des forces armées, *rapport d'information* n° 642 (2019-2020).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human-Autonomy Teaming: Definitions, Debates, and Directions. *Frontiers in Psychology*, 12, 1932.

Mackay, D., & Zundel, M. (2017). Recovering the divide: a review of strategy and tactics in business and management. *International Journal of Management Reviews*, *19*(2), 175-194.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38.

Nalepka, P., Gregory-Dunsmore, J. P., Simpson, J., Patil, G., & Richardson, M. J. (2021). Interaction flexibility in artificial agents teaming with humans. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Nissen, H. A., Evald, M. R., & Clarke, A. H. (2014). Knowledge sharing in heterogeneous teams through collaboration and cooperation: Exemplified through Public–Private-Innovation partnerships. *Industrial Marketing Management*, *43*(3), 473-482.

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2020). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors*.

Pagliari, M., Berberian, B., & Chambon, V. (2022). What is new with Artificial Intelligence? Human-agent interactions through the lens of social agency. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2022.954444

Pégny, M. & Ibnoushein, I. (2018). Quelle transparence pour les algorithmes d'apprentissage machine ?. *Revue d'intelligence artificielle*. 32. 447-478. 10.3166/ria.

Piquet, A. (2009). Guide pratique du travail collaboratif: Théories, méthodes et outils au service de la collaboration. *Document destiné au «Groupe Communication» du réseau Isolement Social*, *17*(2), 7-9.

Plaisant, O., Courtois, R., Réveillère, C., Mendelsohn, G. A., & John, O. P. (2010, March). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. In *Annales Médico-psychologiques, revue psychiatrique* (Vol. 168, No. 2, pp. 97-106). Elsevier Masson.

Prinzel, L., Ellis, K., Koelling, J., Krois, P., Davies, M., & Mah, R. (2021). Examining the Changing Roles and Responsibilities of Humans in Envisioned Future in-Time

Aviation Safety Management Systems. In *78th International Symposium on Aviation Psychology* (p. 346).

Roschelle, Jeremy & Teasley, Stephanie. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. *Computer Supported Collaborative Learning*. 10.1007/978-3-642-85098-1_5.

Ross, K. G., Shafer, J. L., & Klein, G. (2006). Professional judgments and "naturalistic decision making". In Ericsson, K. A., Charness, N., Hoffman, R. R. & Feltovich, P. J. (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 403-419). Cambridge: Cambridge University Press.

Roy, B., & Bouyssou, D. (1993). *Aide multicritère à la décision: méthodes et cas* (p. 695). Paris: Economica.

Russell, S.J., & Norvig, P. (2016). Artificial intelligence - a modern approach, 3rd Edition*. Prentice Hall series in artificial intelligence*.

Schmitt, F., Roth, G., & Schulte, A. (2017, July). Design and evaluation of a mixed-initiative planner for multi-vehicle missions. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 375-392). Springer, Cham.

Seeber, I., Bittner, E., & Briggs, R. O. (2020). Triparna de Vreede, Gert-Jan De Vreede, Aaron Elkins, Ronald Maier, Alexander B Merz, Sarah Oeste-Reiß, Nils Randrup, et al. 2020. Machines as teammates: A research agenda on AI in team collaboration. *Information & management*, *57*(2), 103174.

Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., Johnson, W. W. (2017, July). Why Human-Autonomy Teaming?. In *International Conference on Applied Human Factors and Ergonomics* (pp. 3-11). Springer, Cham.

Simon, H. A. (1997). Models of Bounded Rationality. Vol. 3: Empirically Grounded Economic Reason. Cambridge, MA: The MIT Press.

Sinha, S. (2015). The exploration–exploitation dilemma: a review in the context of managing growth of new ventures. *Vikalpa*, 40(3), 313-323.

Sukman, D. (2016). The institutional level of war. *The Strategy Bridge*, *5*.

Sycara, K., Hughes, D., Li, H., Lewis, M., & Lauharatanahirun, N. (2020, September). Adaptation in human-autonomy teamwork. In *2020 IEEE International Conference on Human-Machine Systems* (ICHMS) (pp. 1-4). IEEE.

Szafir, D., Mutlu, B., & Fong, T. (2017). Designing planning and control interfaces to support user collaboration with flying robots. *The International Journal of Robotics Research*, 36(5-7), 514-542.

Tanguy, J.-M. (2022). Exclu défense : Le Reaper toujours superstar de Barkhane. *Air&cosmos*, 07.01.2022. https://air-cosmos.com/article/exclu-dfense-le-reaper-toujours-superstar-de-barkhane-25774

Task Force IA (2019). L'Intelligence Artificielle au service de la défense. *Rapport du ministère des armées*. https://www.defense.gouv.fr/sites/default/files/aid/20200108-NP-Rapport%20de%20la%20Task%20Force%20IA%20Septembre.pdf

Tessier, C. (2021, June). Éthique et IA: analyse et discussion. In *CNIA 2021: Conférence Nationale en Intelligence Artificielle* (pp. pp-22).

Theureau, J. (2003). L'entretien d'autoconfrontation comme composante d'un programme de recherche empirique et technologique. *Cahiers de l'INSEP*, 34(1), 81-86.

Tokadli, G., Dorneich, M. C., & Matessa, M. (2021). Evaluation of Playbook Delegation Approach in Human-Autonomy Teaming for Single Pilot Operations. *International Journal of Human–Computer Interaction*, 37(7), 703-716.

Trappier, É. (2017). Politique de défense : un point de vue industriel. *Revue Défense Nationale*, 800, 44-50. https://doi.org/10.3917/rdna.800.0044

Turing, A. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433-460.

Turoff, M., White, C., & Plotnick, L. (2011). Dynamic emergency response management for large scale decision making in extreme hazardous events. In *Supporting real time decision-making* (pp. 181-202). Springer, Boston, MA.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124-1131.

Tversky, Amos and Kahneman, Daniel, (1992), Advances in Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty*, 5, issue 4, p. 297-323.

Vallieres, E. F., & Vallerand, R. J. (1990). Traduction et validation canadienne-française de l'échelle de l'estime de soi de Rosenberg. *International journal of psychology*, *25*(2), 305-316.

van Diggelen, J., Barnhoorn, J., Post, R., Sijs, J., van der Stap, N., & van der Waa, J. (2021, February). Delegation in Human-machine Teaming: Progress, Challenges and Prospects. In *International Conference on Intelligent Human Systems Integration* (pp. 10-16). Springer, Cham.

Verbert, K., Parra, D., Brusilovsky, P., & Duval, E. (2013, March). Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces* (pp. 351-362).

Von Neumann, J. & Morgenstern, O. (1944) Theory of Games and Economic Behavior, Princeton University Press.

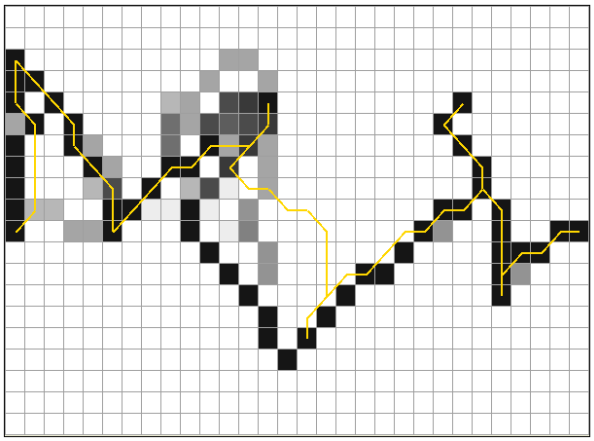
Weld, D. S., & Bansal, G. (2019). The challenge of crafting intelligible intelligence. *Communications of the ACM*, 62(6), 70-79.

Zhang, R., McNeese, N. J., Freeman, G., & Musick, G. (2021). "An Ideal Human" Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3), 1-25.

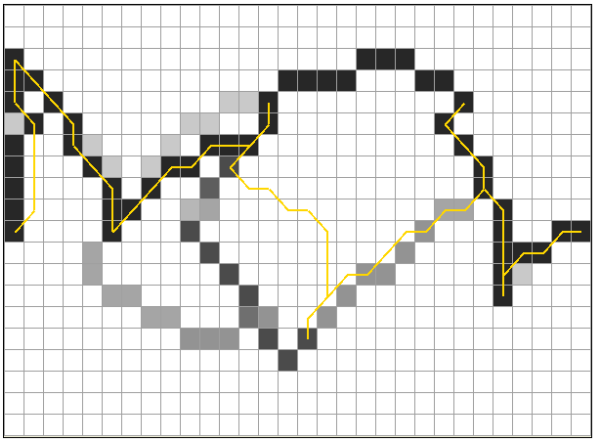
Annexes

Annexe 1 : Superposition des plans validés selon le mode d'interaction dans l'étude 1

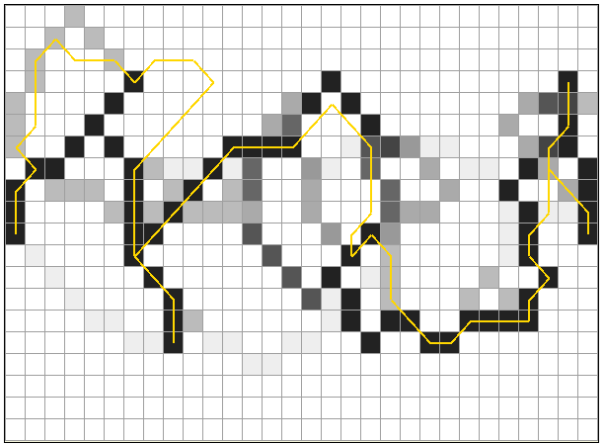
Superposition des plans validés par les participants sur chaque terrain selon la condition de travail en équipe avec l'IA (coopération à gauche, et collaboration à droite). Le niveau de gris associé à chaque case représente la proportion de plans validés qui la traverse. Le chemin jaune correspond à la première suggestion de plan qui est proposée exclusivement en mode collaboratif.



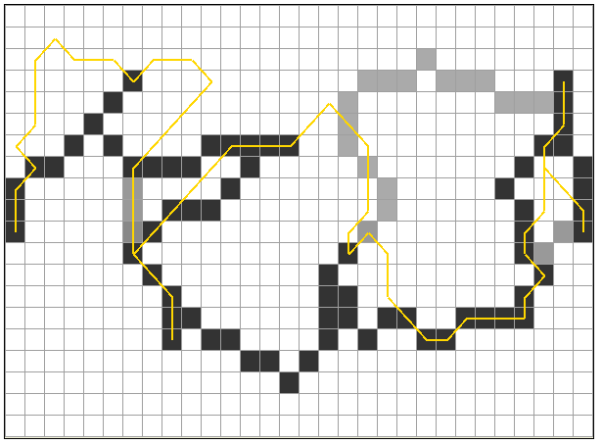
Terrain 1 : Coopération



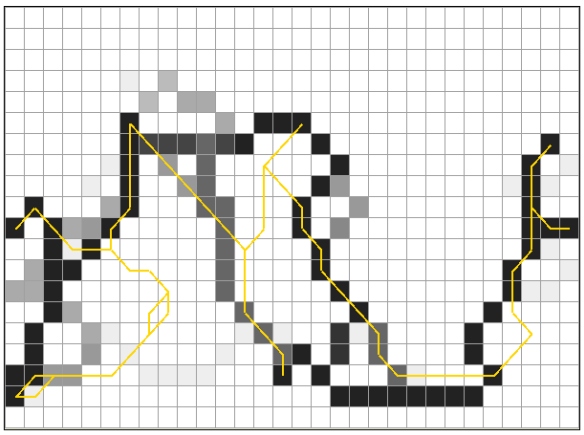
Terrain 1 : Collaboration



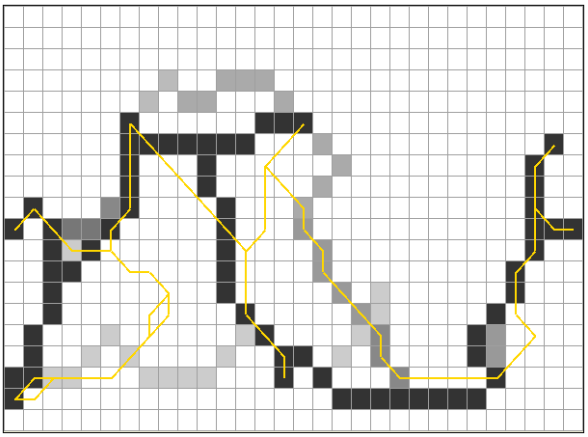
Terrain 2 : Coopération



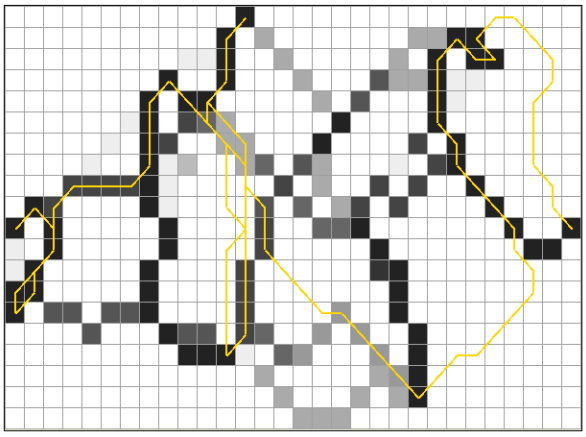
Terrain 2 : Collaboration



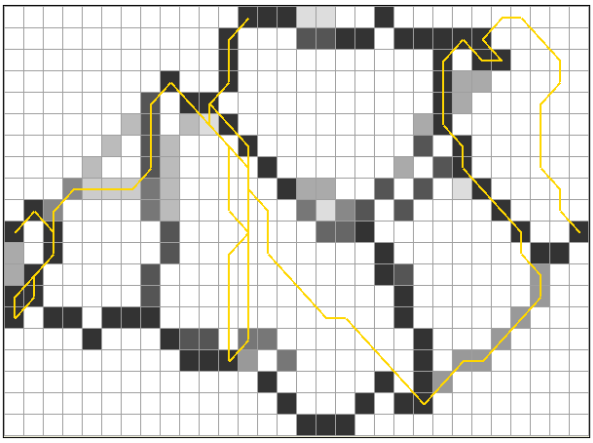
Terrain 3 : Coopération



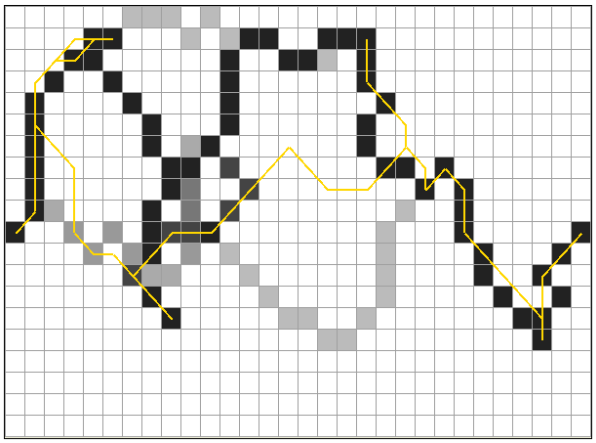
Terrain 3 : Collaboration



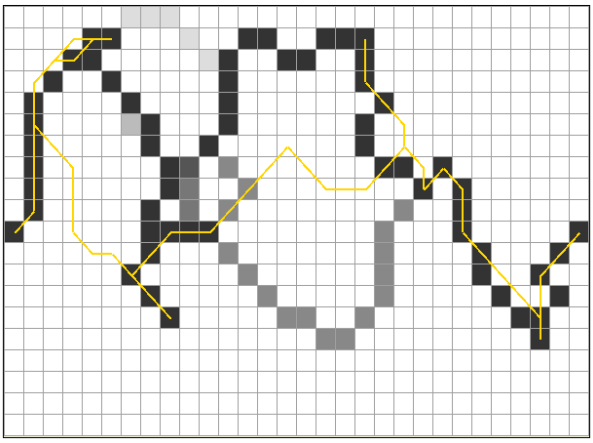
Terrain 4 : Coopération



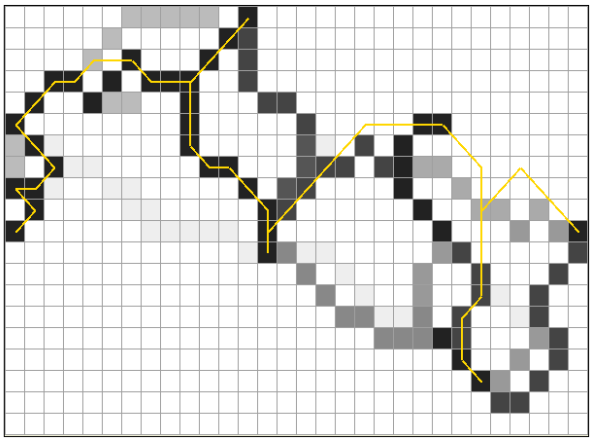
Terrain 4 : Collaboration



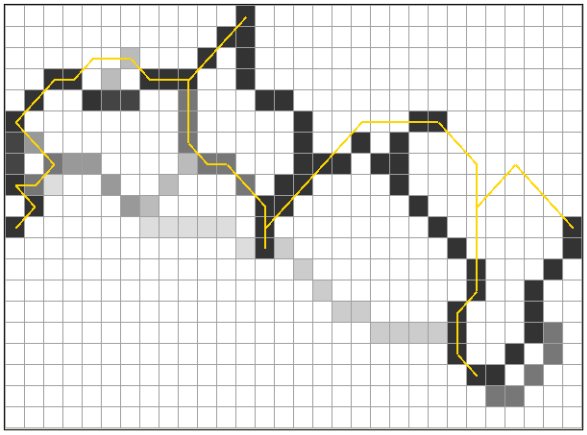
Terrain 5 : Coopération



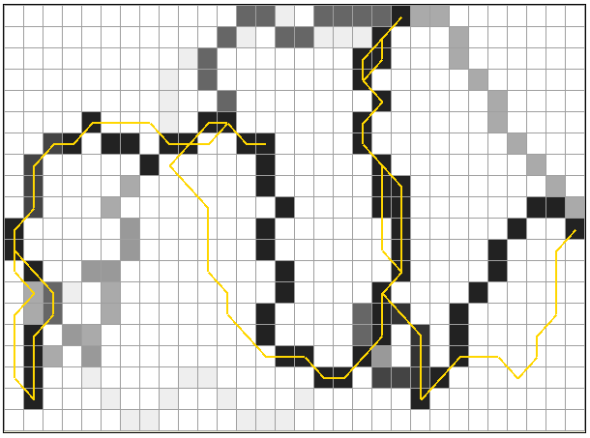
Terrain 5 : Collaboration



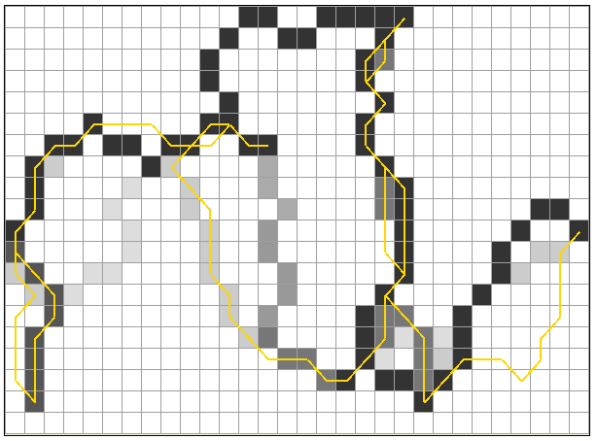
Terrain 6 : Coopération



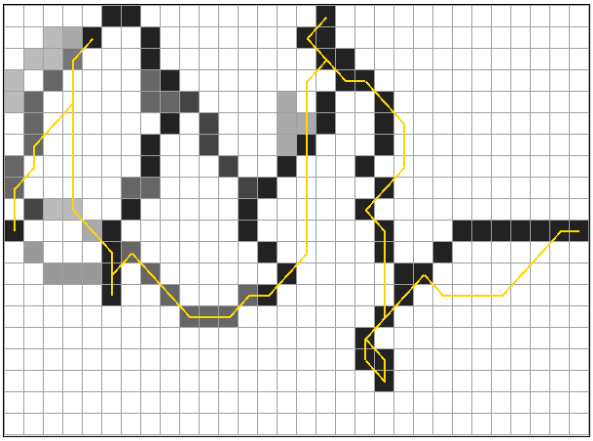
Terrain 6 : Collaboration



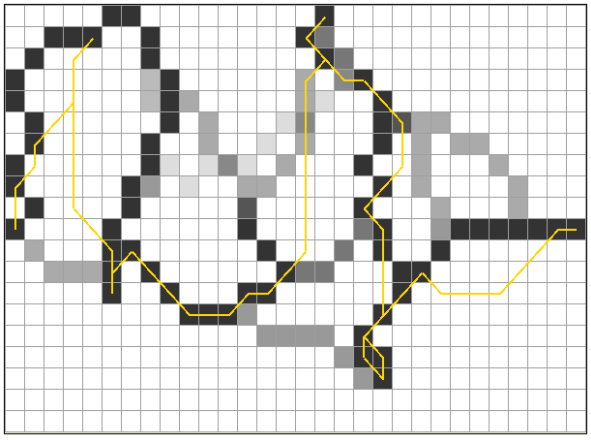
Terrain 7 : Coopération



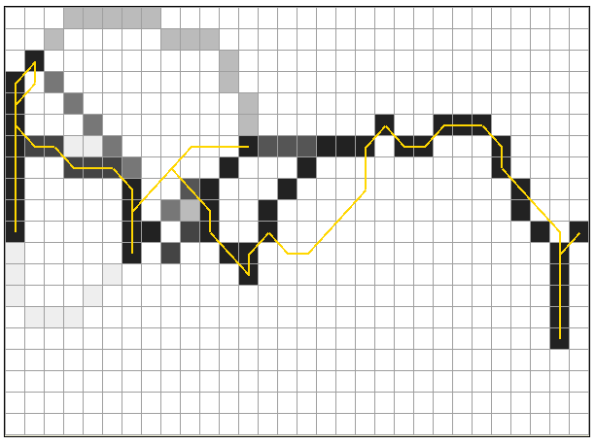
Terrain 7 : Collaboration



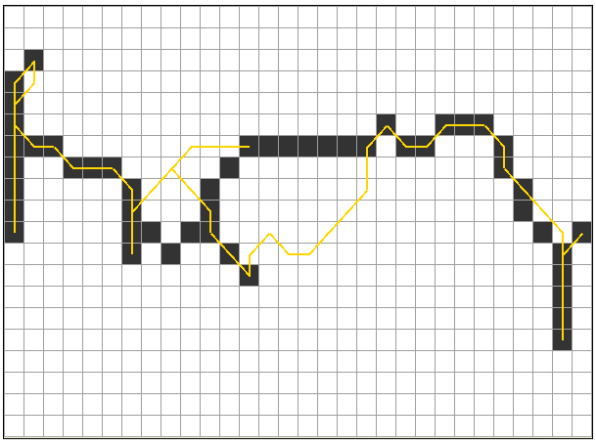
Terrain 8 : Coopération



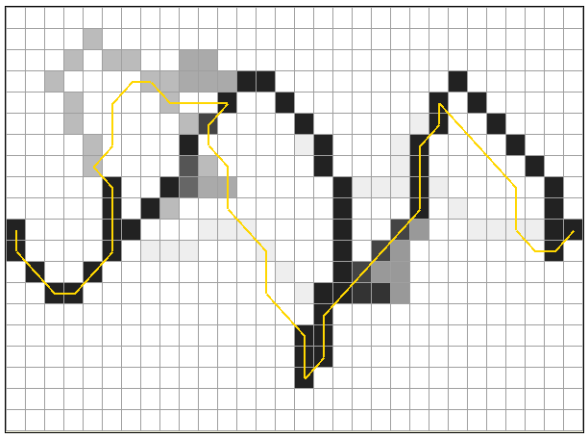
Terrain 8 : Collaboration



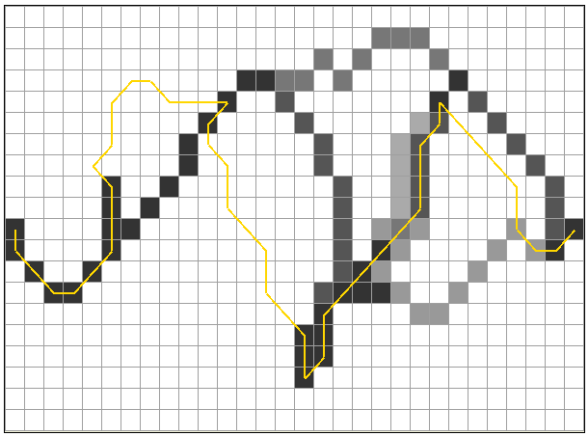
Terrain 9 : Coopération



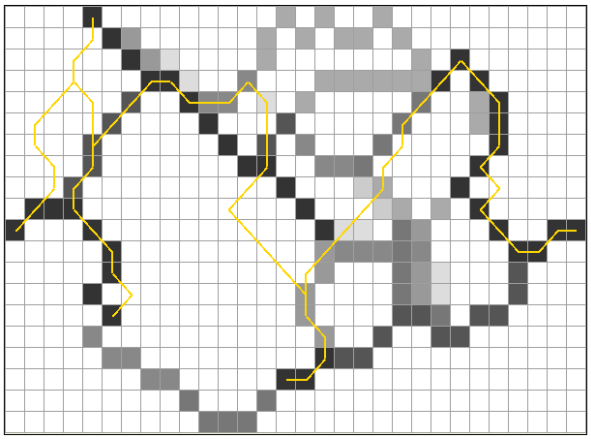
Terrain 9 : Collaboration



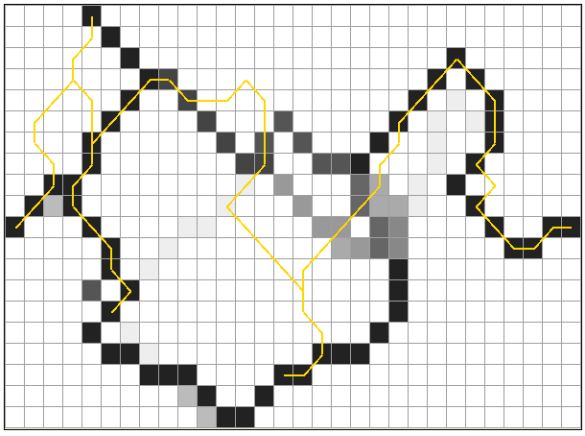
Terrain 10 : Coopération



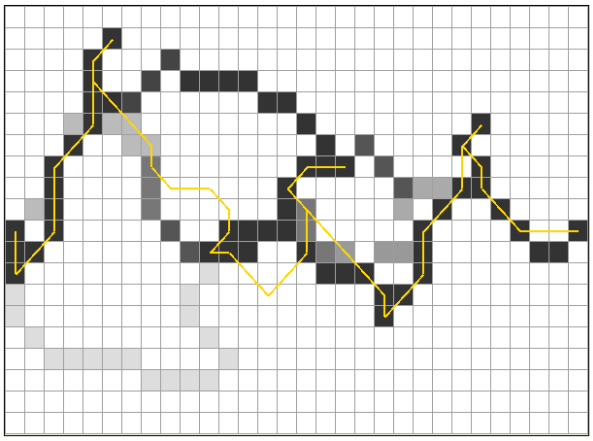
Terrain 10 : Collaboration



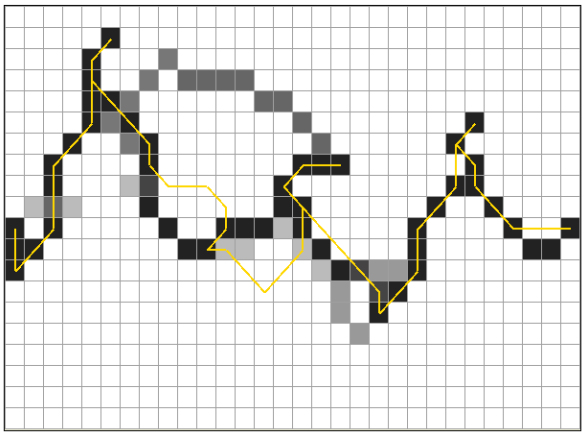
Terrain 11 : Coopération



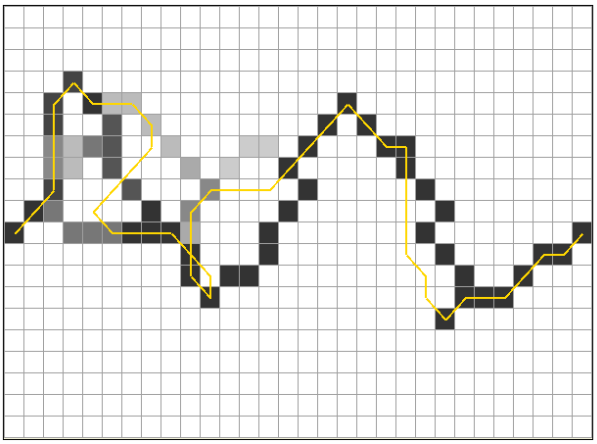
Terrain 11 : Collaboration



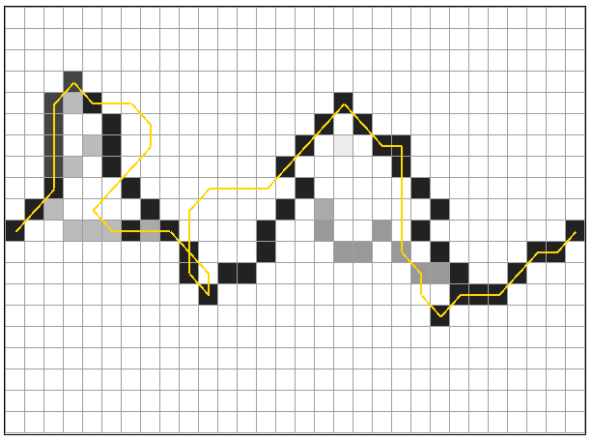
Terrain 12 : Coopération



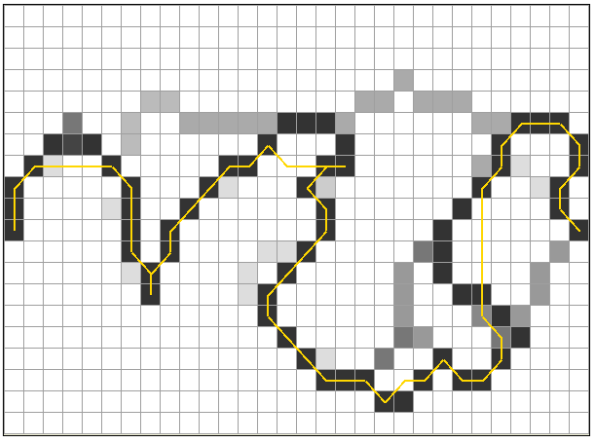
Terrain 12 : Collaboration



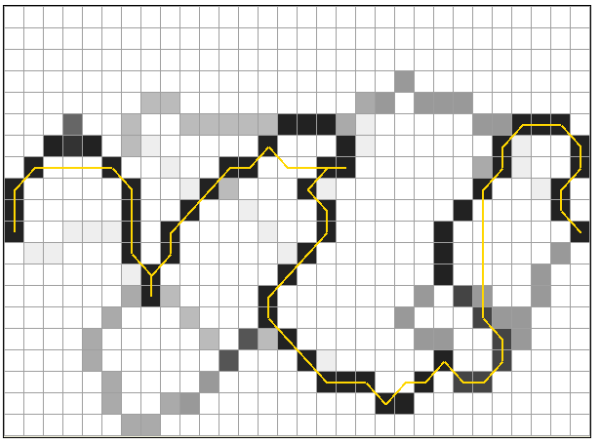
Terrain 13 : Coopération



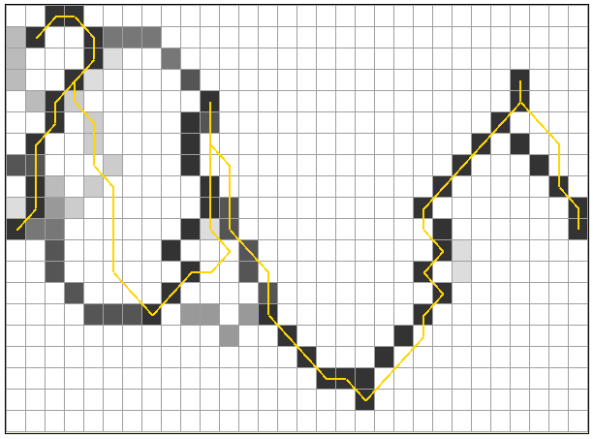
Terrain 13 : Collaboration



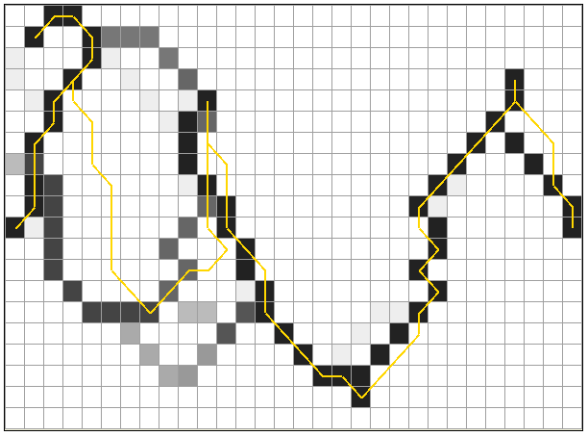
Terrain 14 : Coopération



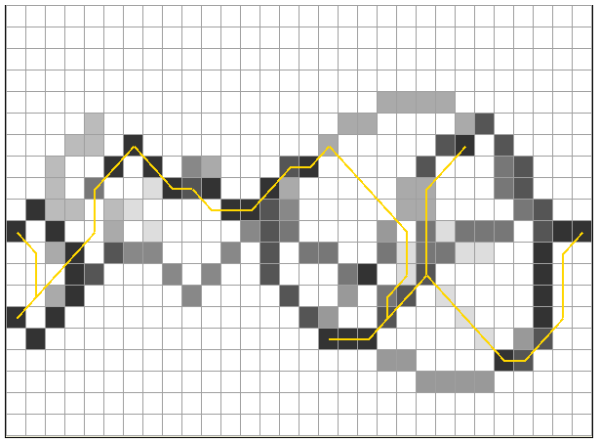
Terrain 14 : Collaboration



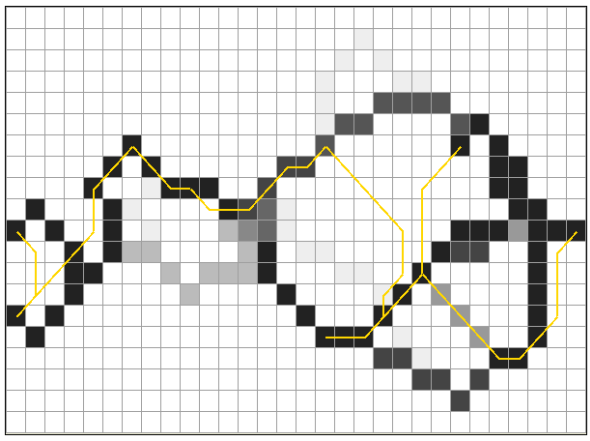
Terrain 15 : Coopération



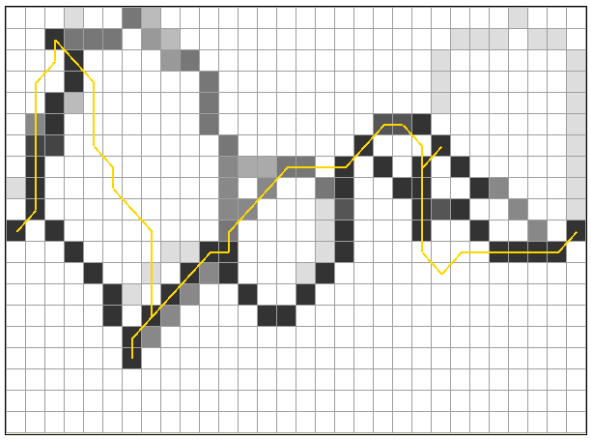
Terrain 15 : Collaboration



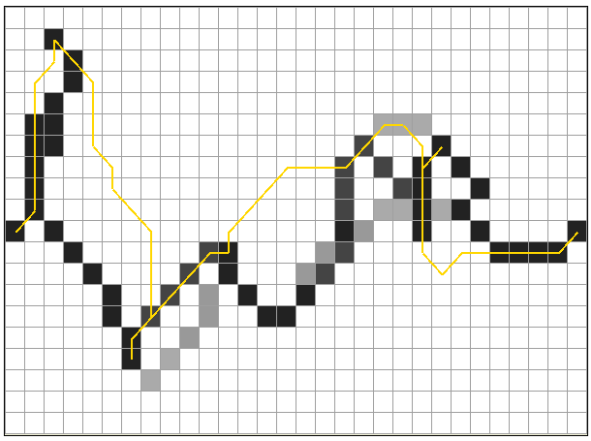
Terrain 16 : Coopération



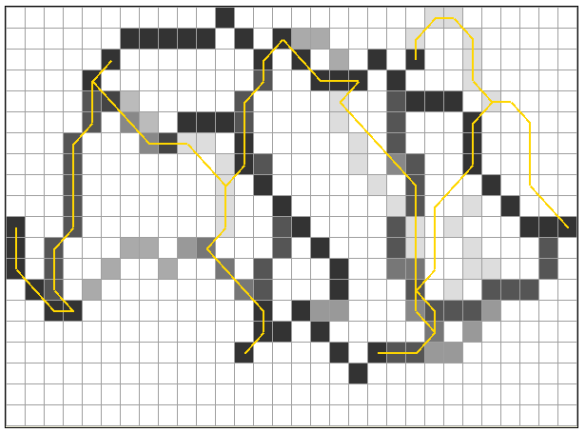
Terrain 16 : Collaboration



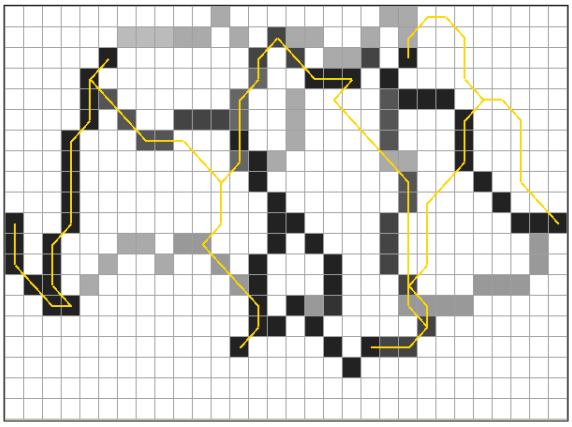
Terrain 17 : Coopération



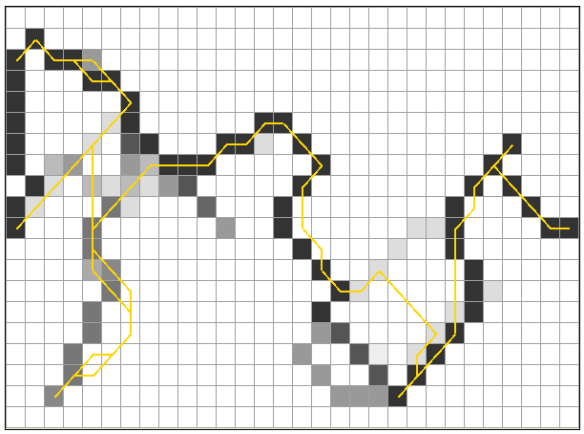
Terrain 17 : Collaboration



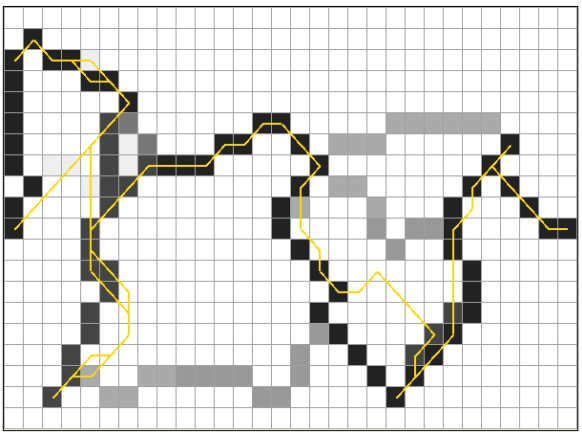
Terrain 18 : Coopération



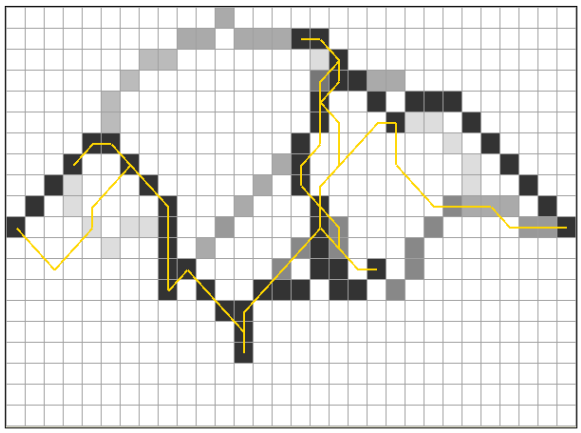
Terrain 18 : Collaboration



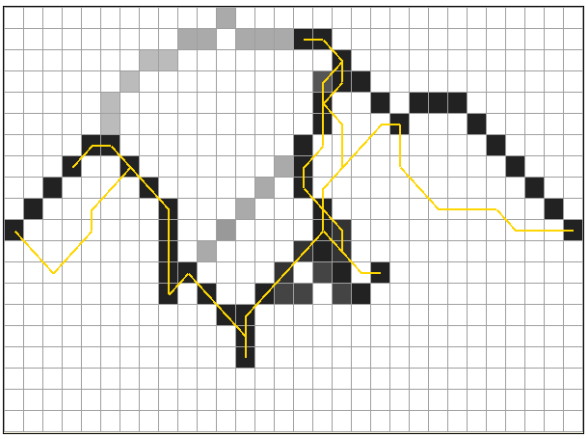
Terrain 19 : Coopération



Terrain 19 : Collaboration



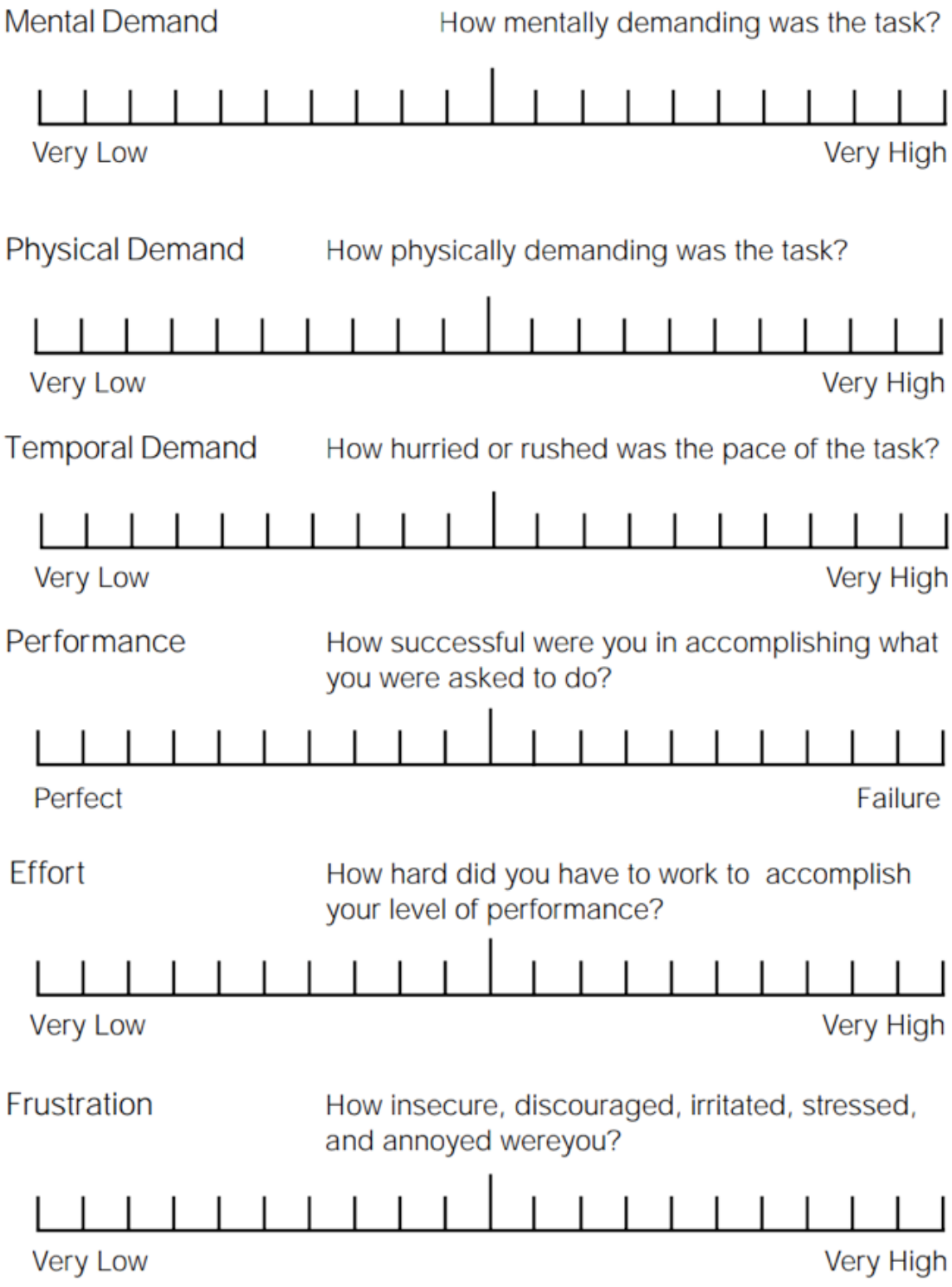
Terrain 20 : Coopération



Terrain 20 : Collaboration

Annexe 2 : Questionnaire NASA-TLX

Le questionnaire NASA-TLX est composé de 6 questions destinées à quantifier la charge de travail perçue pour une tâche (Hart, 2006).



Annexe 3 : Questionnaire Big Five Inventory français

L'inventaire des Big-Five est composé de 45 questions qui décrivent la personnalité en cinq traits centraux : ouverture, conscienciosité, extraversion, agréabilité, névrose (Plaisant et al., 2010).

Instructions:
Vous allez trouver un certain nombre de qualificatifs qui peuvent ou non s'appliquer à vous.
Par exemple, acceptez-vous d'être quelqu'un qui aime passer du temps avec les autres?
Ecrivez devant chaque affirmation le chiffre indiquant combien vous approuvez ou désapprouvez l'affirmation :

1 désapprouve fortement
2 désapprouve un peu
3 n'approuve ni ne désapprouve
4 approuve un peu
5 approuve fortement

Je me vois comme quelqu'un qui ….

1.___ est bavard
2.___ a tendance à critiquer les autres
3.___ travaille consciencieusement
4.___ est déprimé, cafardeux
5.___est créatif, plein d'idées originales
6.___ est réservé
7.___ est serviable et n'est pas égoïste avec les autres
8.___ peut être parfois négligent
9.___ est "relaxe", détendu, gère bien les stress
10.___ s'intéresse à de nombreux sujets
11.___ est plein d'énergie
12.___ commence facilement à se disputer avec les autres
13.___ est fiable dans son travail
14.___ peut être angoissé
15.___ est ingénieux, une grosse tête
16.___ communique beaucoup d'enthousiasme
17.___ est indulgent de nature
18.___ a tendance à être désorganisé
19.___ se tourmente beaucoup
20.___ a une grande imagination
21.___ a tendance à être silencieux
22.___.fait généralement confiance aux autres
23.___.a tendance à être paresseux
24.___ est quelqu'un de tempéré, pas facilement troublé
25.___ est inventif
26.___ a une forte personnalité, s'exprime avec assurance
27.___ est parfois dédaigneux, méprisant
28.___ persévère jusqu'à ce que sa tâche soit finie
29.___ peut être lunatique d'humeur changeante
30.___ apprécie les activités artistiques et esthétiques
31.___ est quelquefois timide, inhibé
32.___ est prévenant et gentil avec presque tout le monde
33.___ est efficace dans son travail
34.___ reste calme dans les situations angoissantes
35.___ préfère un travail simple et routinier
36.___ est sociable, extraverti
37.___ est parfois impoli avec les autres
38.___ fait des projets et les poursuit
39.___ est facilement anxieux

40.___ aime réfléchir et jouer avec des idées
41.___ est peu intéressé par tout ce qui est artistique
42.___ aime coopérer avec les autres
43.___ est facilement distrait
44.___ a de bonnes connaissances en art, musique ou en littérature
45.___ cherche des histoires aux autres

Vérifier que vous avez bien répondu à toutes les questions. Merci

Score= moyenne des items dans la dimension
E (Extraversion, Energie, Enthousiasme)
8 items: 1, 6R 11, 16, 21R, 26, 31R, 36;
A (Agréabilité, Altruisme, Affection)
10 items: 2R, 7, 12R, 17, 22, 27R, 32, 37R, 42, 45R;
C (Conscience, Contrôle, Contrainte)
9 items: 3, 8R, 13, 18R, 23R, 28, 33, 38, 43R;
N (émotions Négatives, Névrosisme, Nervosité)
8 items: 4, 9R, 14, 19, 24R, 29, 34R, 39;
O (Ouverture, Originalité, Ouverture d'esprit)
10 items: 5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44;

Chaque facteur correspond à la moyenne de la somme des items en inversant les items marqués d'un R « reverse » (5 devenant 1 ; 4 : 2 ; 3 : 3, 2 : 4 et 1 : 5).

Note. Copyright © 1991 by Oliver P. John.
Traduction et adaptation française par Odile Plaisant et al., 2005.
Reproduit avec permission.

Annexe 4 : Échelle d’estime de soi de Rosenberg

L'échelle d'estime de soi de Rosenberg, en anglais *Rosenberg’s Self-Esteem Scale* (RSES) est composée de 10 questions qui mesurent l'estime de soi des individus (Vallieres et Vallerand, 1990). Le détail des questions est disponible en Annexe 4.

Pour chacune des caractéristiques ou descriptions suivantes, indiquez à quel point chacune est vraie pour vous en encerclant le chiffre approprié.

Tout à fait en désaccord	Plutôt en désaccord	Plutôt en accord	Tout à fait en accord
1	2	3	4

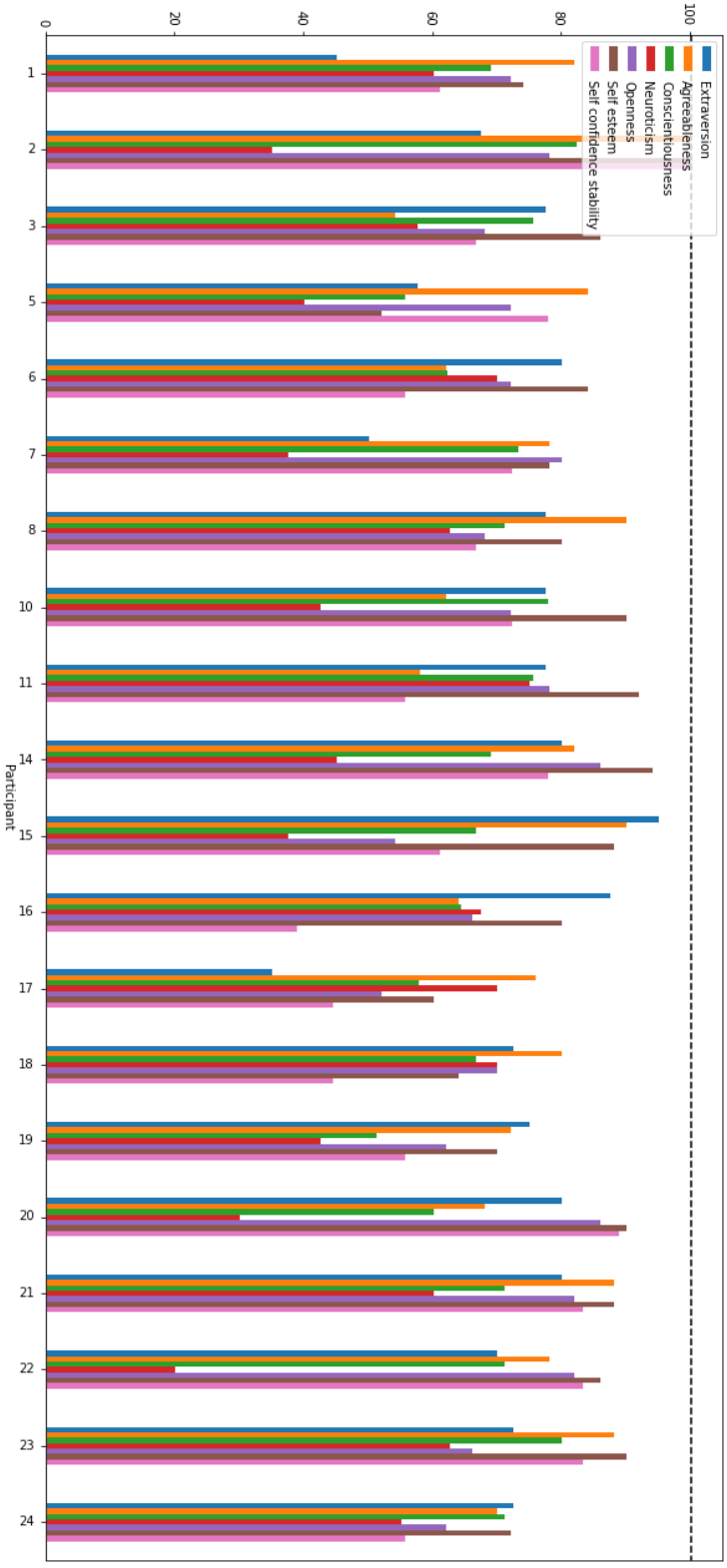
1.	Je pense que je suis une personne de valeur, au moins égale à n'importe qui d'autre	1-2-3-4
2.	Je pense que je possède un certain nombre de belles qualités.	1-2-3-4
3.	Tout bien considéré, je suis porté à me considérer comme un raté	1-2-3-4
4.	Je suis capable de faire les choses aussi bien que la majorité des gens	1-2-3-4
5.	Je sens peu de raisons d'être fier de moi.	1-2-3-4
6.	J'ai une attitude positive vis-à-vis moi-même.	1-2-3-4
7.	Dans l'ensemble, je suis satisfait de moi.	1-2-3-4
8.	J'aimerais avoir plus de respect pour moi-même	1-2-3-4
9.	Parfois je me sens vraiment inutile.	1-2-3-4
10.	Il m'arrive de penser que je suis un bon à rien.	1-2-3-4

Annexe 5 : L'échelle de stabilité de la confiance en soi

L'échelle de stabilité de la confiance en soi, en anglais *Self-Esteem Stability Scale* (SESS) est composée de 3 questions qui mesurent la variabilité de l'estime de soi des individus dans le temps (Altmann et Roth, 2018).

Items*
My attitude toward myself is very stable.
How I estimate my abilities compared with others changes frequently.
My positive and negative feelings toward myself often blend into each other.

Responses were made on a 6-point Likert scale.

Annexe 6 : Traits de personnalité des participants de l'étude 2

Score des participants de l'étude 2 sur une échelle de 0 à 100 pour leur ouverture d'esprit, conscienciosité, extraversion, agréabilité, tendance à la névrose, confiance en soi et stabilité de la confiance en soi.