Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Jonas GOURAUD

le mardi 27 novembre 2018

Titre :

Mind wandering dynamic in automated environments and its influence on out-of-the-loop situations

---

École doctorale et discipline ou spécialité :
ED CLESCO : Neurosciences, comportement et cognition

Unité de recherche :
CERCO - ICNA

Directeur/trice(s) de Thèse :

Arnaud DELORME
Bruno BERBERIAN

Jury :
M. Jérôme SACKUR, Rapporteur
M. Axel CLEEREMANS, Rapporteur
Mme Céline LEMERCIER, Examinateur
M. Frédéric DEHAIS, Examinateur
M. François-Benoît VIALATTE, Examinateur

Jonas GOURAUD, Mind wandering and out of the loop situations, “Apprendre le doute n’a pas de prix.”

Jonas Gouraud

MIND WANDERING DYNAMIC IN AUTOMATED ENVIRONMENTS AND ITS INFLUENCE ON OUT-OF-THE-LOOP SITUATIONS

Directors: Arnaud Delorme, Bruno Berberian

Abstract

Higher levels of automation are progressively integrated in critical environments to satisfy the increasing demand for safer systems. Such philosophy moves operators to a supervisory role, also called out-of-the-loop (OOTL) situations. Unfortunately, OOTL situations also create a new kind of human-machine interaction issues, called OOTL performance problem. The dramatic consequences of OOTL performance problem stress the need to identify which mechanisms could influence their appearance. The emergence of thoughts unrelated to the here and now, labeled mind wandering (MW), could affect operators in OOTL situations through the perceptual decoupling induced. This thesis investigates MW dynamic in OOTL situations and its influence on operators. We firstly reviewed the evidences in the literature underlining a link between OOTL performance problem and MW. We completed theoretical insights by reporting pilots' tendency (collected with a questionnaire) to encounter more problems with autopilots when experiencing more task-unrelated MW. Then, we conducted three experiments in OOTL conditions using an obstacle avoidance task. With non-expert population and sessions longer than 45 minutes, we observed a significant increase of MW in OOTL situations compared to manual conditions, independently of system reliability. MW episodes were also accompanied by a perceptual decoupling from the task induced by task-unrelated MW. This decoupling was visible on reports of mental demand as well as oculometric (pupil size, blinks) and encephalographic (N1 component, alpha activity) signals. Overall, our results demonstrate the possibility to use physiological markers of MW in complex OOTL environments. We discuss new perspectives towards the use of MW markers to characterize the OOTL performance problem. Instead of blindly stopping MW episodes, which could have benefits for operators, future research should focus on designing systems able to cope with MW and identify information needed to facilitate the reentry in the control loop when needed.

Jonas Gouraud

DYNAMIQUE DE LA DIVAGATION ATTENTIONNELLE DANS DES ENVIRONNEMENTS AUTOMATISES ET SON INFLUENCE SUR LES SITUATIONS DE SORTIE DE BOUCLE

Directeurs : Arnaud Delorme, Bruno Berberian

Résumé

Des niveaux d'automatisation élevés sont intégrés dans les environnements critiques pour satisfaire la demande croissante de systèmes plus sûrs. Cette philosophie déplace les opérateurs vers un rôle de supervision et crée de nouveaux problèmes appelés problèmes de performance liés à la sortie de boucle (SDB). L'émergence de pensées sans lien avec ici et maintenant, ou divagation attentionnelle (DA), pourrait affecter les opérateurs dans des situations de SDB par le biais du découplage perceptuel induit. Cette thèse a étudié la dynamique de la DA dans les situations de SDB et son influence sur les opérateurs. Nous avons en premier lieu examiné les preuves dans la littérature pointant vers un lien entre le problème de performance lié à la SDB et la DA. Nous avons complété cette analyse théorique en rapportant la tendance des pilotes (collectée avec un questionnaire) à rencontrer plus de problèmes avec leur pilote automatique pour ceux ayant une plus grande propension au MW non lié à la tâche. Nous avons ensuite mené trois expériences dans des conditions de SDB. Nous avons observé une augmentation significative des épisodes de DA dans les situations de SDB quelle que soit la fiabilité du système, par rapport aux conditions manuelles. Les épisodes de DA étaient également accompagnés d'un découplage perceptuel vis-à-vis de la tâche créé par la DA non lié à la tâche. Ce découplage était visible sur des rapports de demande mentale ainsi que les signaux oculométriques et encéphalographiques. Dans l'ensemble, nos résultats démontrent la possibilité d'utiliser des marqueurs physiologiques de la DA dans des environnements de SDB complexes. Nous discutons de nouvelles perspectives d'utilisation des marqueurs de la DA pour caractériser les problèmes de performance liés à la SDB. Sans vouloir arrêter aveuglément l'émergence de la DA, qui pourrait être bénéfique pour les opérateurs, les recherches futures devraient se concentrer sur la conception de systèmes capables de gérer la DA et d'identifier les informations nécessaires pour faciliter la rentrée de l'opérateur dans la boucle de contrôle.

ACKNOWLEDGMENTS

Il est 23h30, et cette partie est la dernière à écrire avant de rendre mon manuscript. Non pas que ce soit la moins importante, définitivement non. A l'inverse, je l'ai intentionnellement laissée pour la fin car je sais déjà que je pourrais passer des heures et des pages pour remercier du fond du cœur toutes les personnes qui m'ont soutenu jusqu'ici. J'ai appris tellement au court de ces trois années, en grande partie grâce aux interactions avec toutes les personnes ci-dessous (et beaucoup d'autres).

Tout d'abord, ma famille. Bea, danke, ohne dich würde ich nicht hier sein, meine *Tobschen*. Zum Glück bist du hier, und ich warte auf unsere gemeinsame Wohnung mit Spaß! Maman, à chaque fois que je viens chez toi c'est avec la certitude que tout y ira bien, et ceci grâce à toi. Tu as su me passer ta force d'avancer et de toujours tenter. Cassandre, merci pour ton aide et ta joie de vivre à chaque fois qu'on se retrouve, je suis plus que content de t'avoir comme sœur. Mes grands-parents, Clisson ou Bouchemaine, vous êtes et resterez toujours mes piliers dans la vie quand des choix complexes arrivent. Merci à mes oncles et tantes, Corinne, JC, Anne, Jean-Noël, Tony, Bertrand, et aux nombreux cousins et cousines. Et enfin Papa, merci pour toutes les valeurs que tu m'as transmises. Je n'oublierai jamais tes conseils, et feraient en sorte de communiquer ta sagesse et ton humour au plus grand nombre.

Je tiens à remercier mes collègues de Salon, pour beaucoup devenus des amis, qui ont appris à me supporter pendant les trois (trop courtes) années de ma thèse. Bruno, heureusement que tu étais là. Merci d'avoir eu confiance en moi pour démarrer ce doctorat, alors même que je n'avais aucune expérience ni en psychologie ni en neuroscience. Tu as su trouver la méthode pour me pousser à donner le meilleur de moi-même, tout en ne rechignant pas à refaire mon éducation scientifique quand tu en sentais le besoin. Et besoin il y a eu… Arnaud, merci pour ta présence. Aïsha, Bertille, François, merci pour les grandes discussions et les éclats de rire qu'on a eu dans le bureau des magiciens. Je vous souhaite bonne chance de votre côté, pour la fin de thèse et après. Raphaël, j'espère qu'on continuera à se voir autour d'une bière, et peut être bien d'une partie de Heroes III. Christian, grâce à toi j'ai pu utiliser le LIPS pendant 3 ans, désolé pour les cheveux blancs. Greg, Kevin, Nawfel, JC, Alex, Mick, Xavier, Gemma, Jean, Andrea, Jose, Naïs, Quang-Huy, Camille, Florent, Floriane, Fabrice, Simon, Laurent, merci infiniment pour les discussions, les blagues, le foot, et généralement la bonne humeur qui règne au labo. Le CSP ne serait pas le même sans chacun de vous !

Enfin, le labo était une chose, mais les weekends et soirées n'auraient jamais atteint ce niveau sans la présence de Thibault, Paul et Claire. Aixois en force \o/ Mais Aix n'est pas seule, et je peux remercier des amis très chers ailleurs. Merci à la famille Rodriguez-Gouedreau pour votre accueil dans la belle région de Provence. David, merci pour ton soutien dans les moments difficiles et les longues soirées de coop (qui vont continuer longtemps encore, c'est sûr). Sam, Dede, merci à vous pour les vacances superbes qu'on passe à chaque fois. Laura, merci à toi pour nos discussions passionantes sur tous les sujets, et ta présence quand j'en avais besoin. Merci Lydie pour ta joie de vivre, Baptiste pour les soirées jeux, à Corinne, Mathieu, Antoine, Florian, Romain, Juliette, Julie, Jasmin, Lila, JB, Cédric, Chloé. Merci aussi au groupe d'allemand, à Karin la meilleure prof. Enfin merci à mes encadrants avant Salon, qui ont su me donner le goût de la recherche et l'envie d'en savoir toujours plus. Merci Michel pour nos échanges sur les convertisseurs AC/DC, ma première expérience de vraie recherche. Merci Stéphane pour tes conseils et ta vision, l'IHM vaincra !

Bien du temps me serait nécessaire pour remercier chacun d'entre vous dans les formes. Mais minuit approche. Dans les moments légers et les passages difficiles, chacun d'entre vous m'a aidé à être qui je suis aujourd'hui, m'a apporté un morceau de sa personnalité. Je vous remercie mille fois pour ça, et espère que nos échanges continueront bien au-delà de ces quelques 26 années. Un petit pas pour l'homme, un grand pas pour le Jonas.

CONTENTS

Chapter 5: study of mind wandering dynamic in out-of-the-loop situations 113

Chapter 6: study of the attentional decoupling with automation varying in reliability 135

LIST OF FIGURES

LIST OF TABLES

CONTENT OF THE THESIS

The introduction of automation in a wide variety of critical environments has brought new problems related to automation-supervisor interactions. Gathered under the term of out-of-the-loop (OOTL) performance problem, we still poorly understand this issue today. This thesis investigate the dynamic of mind wandering (MW) in automated environments and its influence when operators supervise systems (OOTL situations).

The chapter 1 develops the definitions we adopt throughout this thesis. Although automation helped reduce accidents in safety critical environments, moving operators out of the control loop also creates the OOTL performance problem with dramatic consequences. We develop a model vision of OOTL performance problem as a human-system cooperation problem. On the other hand, the abundant research on MW suffers from a lack of accepted general definition. Although generally defined as thoughts unrelated to the here and now, we detail the recently proposed family-resemblance framework and give the rationales behind it. We present the perceptual decoupling created by MW and its negative impact on short-term performances. Finally, we explain why we focus on the dimensions of task proximity (MW depending on the task or not related to the task) and depth (measured in terms of decoupling of the environment).

In the chapter 2, we review the evidences in the literature supporting a link between OOTL performance problem and MW. MW may be particularly influential in automated environments, generally rated as monotonous (except when a problem suddenly occurs). We point to the similarities observed between the influence of MW and the physiological changes observed during OOTL performance problem. In particular, the concepts of complacency and feeling of agency are developed as possible concepts to bridge OOTL and MW fields. We further describe how such a link would help both theoretical and experimental research on OOTL performance problem characterization.

Chapter 3 presents pilots' answers to a questionnaire about their experience of human-autopilot cooperation. In order to complete the evidences found in the literature, we wanted to assess the operational need of our research. We created a questionnaire enquiring about pilots' experience, their perception of the autopilot usability and the frequency of problems encountered when using automated aids. Moreover, we used a validated task-unrelated MW questionnaire to compare pilots' and non-pilots' MW propensity in their daily life. We only report descriptive statistics, as unbiased inferential analysis would require a more robust questionnaire with more answers.

The chapter 4 presents in details all the material and methods used in the subsequent experiments. Because the three experiments use the same environment (LIPS), we provide here our rationales in order to lighten the description of each task and avoid redundancy. Moreover, we present the probes used for subjective reports of MW, adopting a ternary model of the phenomenon. We present the eye-tracker and the electroencephalogram used for physiological measures, as well as their respective functioning. We also detail the step-by-step analysis used for data pre-processing.

We present in the chapter 5 the first experiment. We aimed at comparing the dynamic of MW in an OOTL situation, compared to the same task performed manually. On top of the probes used randomly throughout the experiment, we recorded behavioral measures using a NASA Task Load Index questionnaire filled by participants at the end of the experiment. Furthermore, we placed an eye-tracker to evaluate the use of oculometric markers in complex environments to characterize the perceptual decoupling induced by MW. We finally detail how our findings integrate the literature and could help OOTL research.

In the light of the results obtained in chapter 5, we performed in chapter 6 another experiment addressing the influence of trust on MW. We used the automated mode of the LIPS already used in chapter 5, but with 2 levels of reliability. Throughout the experiment, we measured trust in the system. Moreover, we also measured perceived mental demand in order to be able to pinpoint more accurately the link between system features, MW and operators' perception. Again, we measured oculometric changes using the eye-tracker. After describing the result analysis, we consider how the conclusions articulate with those from the previous experiment as well as other studies on MW and automation.

The final experiment detailed in chapter 7 explored the possibility to characterize MW in OOTL environments with better accuracy and without disrupting operators. Moreover, we went a step further in ecological validation and created a multimodality sensorial task. We used the electroencephalographic signal and analyzed MW influence on event-related potentials and brain wave activity. Moreover, we added an Auditory Steady-State Response to the paradigm in the form of an aeronautical-like noise. The purpose was to explore opportunities to continuously monitor MW.

The final chapter 8 concludes our work by summarizing the different findings of the three experiments. We discuss how these results add to each other on the way to a better characterization of OOTL performance problem through the MW phenomenon. We discuss the limits of our study, and propose tracks for future research to go further our findings. Finally, we detail the recent advances of two other concepts that tackle other aspects of the OOTL performance problem, namely joint-agency and performance monitoring.

CONTENU DE LA THESE

L’introduction de l’automatisation dans une grande variété d’environnements critiques a engendré de nouveaux problèmes liés aux interactions entre l’automatisation et le superviseur. Réuni sous le terme de problème de performance lié à la sortie de boucle (SDB), ce problème est encore mal compris aujourd’hui. Cette thèse examine la dynamique de la divagation attentionnelle (DA) dans les environnements automatisés et son influence quand les opérateurs supervisent un système (situation de SDB).

Le chapitre 1 développe les définitions que nous adoptons tout au long de cette thèse. Bien que l’automatisation ait permis de réduire les accidents dans les environnements critiques pour la sécurité, le fait de sortir les opérateurs de la boucle de contrôle crée également un problème de performance lié à la SDB avec des conséquences dramatiques. Nous développons une vision modèle du problème de performance lié à la SDB en tant que problème de coopération homme-système. D’autre part, l’abondante recherche sur la DA souffre d’un manque de définition commune. Bien que généralement définie comme des pensées sans rapport avec l’ici et maintenant, nous détaillons une définition récemment proposée en donnant nos justifications. Nous présentons le découplage perceptuel créé par le MW et son impact négatif sur les performances à court terme. Enfin, nous expliquons pourquoi nous nous concentrons sur les dimensions de proximité à la tâche (DA liée ou non liée à la tâche) et de la profondeur (mesurée en termes de découplage à la tâche).

Dans le chapitre 2, nous examinons les preuves dans la littérature soutenant l’existence d’un lien entre le problème de performance lié à la SDB et la DA. La DA peut être particulièrement influente dans les environnements automatisés, généralement considérés comme monotone (sauf lorsqu’un problème survient soudainement). Nous soulignons les similitudes observées entre l’influence du MW et les changements physiologiques observés au cours des problèmes de performance liés à la SDB. En particulier, les concepts de complaisance et de sentiment d’agentivité sont développés en tant que concepts possibles pour relier les champs de la SDB et de la DA. Nous décrivons en outre comment un tel lien aiderait à la fois la recherche théorique et expérimentale sur la caractérisation des problèmes de performance liés à la SDB.

Le chapitre 3 présente les réponses des pilotes à un questionnaire sur leur expérience vis-à-vis de la coopération homme-pilote automatique. Afin de compléter les preuves données par la littérature, nous avons voulu évaluer le besoin opérationnel de nos recherches. Nous avons créé un questionnaire pour connaître l'expérience des pilotes, leur perception de la facilité d'utilisation du pilote automatique et la fréquence des problèmes rencontrés lors de l'utilisation d'aides automatisées. De plus, nous avons utilisé un questionnaire déjà validé portant sur la DA non liée à la tâche pour comparer la propension à la DA des pilotes et des non-pilotes dans leur vie quotidienne. Nous rapportons uniquement des statistiques descriptives, une analyse déductive non biaisée nécessitant un questionnaire plus robuste avec plus de réponses.

Le chapitre 4 présente en détail tout le matériel et les méthodes utilisés dans les expériences suivantes. Les trois expériences utilisant le même environnement (LIPS), nous fournissons ici nos justifications afin d'alléger la description de chaque tâche et d'éviter toute redondance. En outre, nous présentons les sondes attentionnelles utilisées pour les rapports subjectifs de la DA, en utilisant un modèle ternaire du phénomène (les participants pouvaient être concentrés, en DA liée à la tâche ou en DA non liée à la tâche). De plus, nous détaillons l'oculomètre et l'électroencéphalogramme utilisés pour les mesures physiologiques ainsi que leur fonctionnement respectif. Nous détaillons également l'analyse pas à pas utilisée pour le prétraitement des données.

Nous présentons dans le chapitre 5 la première expérience. Notre but était de comparer la dynamique de la DA dans une situation de SDB, par rapport à la même tâche effectuée manuellement. En plus des sondes attentionnelles disposées de manière aléatoire pendant toute l'expérimentation, nous avons enregistré les mesures comportementales avec un questionnaire de la NASA Task Load Index rempli par les participants à la fin de l'expérience. En outre, nous avons placé un oculomètre pour évaluer l'utilisation de marqueurs oculométriques dans des environnements complexes pour caractériser le découplage perceptuel induit par la DA. Nous détaillons enfin comment nos résultats intègrent la littérature et pourraient aider la recherche OOTL.

À la lumière des résultats obtenus au chapitre 5, nous avons effectué au chapitre 6 une autre expérience portant sur l'influence de la confiance sur la DA. Nous avons utilisé le mode automatisé du LIPS déjà utilisé au chapitre 5, mais avec deux niveaux de fiabilité. Tout au long de l'expérience, nous avons mesuré la confiance dans le système. De plus, nous avons également mesuré la demande mentale perçue afin de pouvoir identifier plus précisément le lien entre les caractéristiques du système, les épisodes de DA et la perception des opérateurs. A nouveau, nous avons mesuré les changements oculométriques en utilisant le suivi oculaire. Après avoir décrit l'analyse des résultats, nous examinons comment les conclusions s'articulent avec celles de l'expérience précédente ainsi que d'autres études sur la DA et l'automatisation.

L'expérience finale détaillée au chapitre 7 a exploré la possibilité de caractériser la DA dans des environnements liés à la SDB avec une meilleure précision et sans perturber les opérateurs. De plus, nous sommes allés une étape plus loin dans l'exploration d'environnements écologiques avec l'utilisation d'une tâche sensorielle multimodale. Nous avons utilisé le signal électroencéphalographique et analysé l'influence de la DA sur les potentiels évoqués et l'activité des ondes cérébrales. Le but était d'explorer les possibilités de surveiller en continu la DA.

Le dernier chapitre 8 conclut notre travail en résumant les différentes conclusions des trois expériences. Nous discutons de la façon dont ces résultats s'ajoutent les uns aux autres sur la voie d'une meilleure caractérisation du problème de performance lié à la SDB au travers du phénomène de DA. Nous discutons des limites de notre étude et proposons des pistes pour de futures recherches afin d'aller plus loin. Enfin, nous détaillons les avancées récentes de deux autres concepts qui abordent d'autres aspects du problème de performance liés à la SDB, à savoir l'agentivité d'équipe et le suivi des performances.

1 OUT-OF-THE-LOOP SITUATIONS AND MIND WANDERING PHENOMENON

- *Designers have used automation to answer modern problems of safety, efficiency and precision in a wide range of industries; however, it creates* ***monotonous environments*** *for operators.*
- *The* ***out-of-the-loop*** *(OOTL)* ***performance problem*** *highlights human-automation interactions issues in out-of-the-loop situations, i.e. when supervising a system.*
- *As humans will remain in control of critical systems for a long time,* ***understanding the psychological mechanisms*** *underlying OOTL performance problem is a key to safer systems.*
- ***Mind wandering*** *(MW) is a collection of experiences along several dimensions that refers to thoughts unrelated to here and now. It also creates a decoupling from the task.*
- *We investigate the mind wandering aspects of* ***task proximity*** *and* ***depth*** *in order to consider mind wandering impact in automated environments.*

1.1 Description of the current chapter

The two main concepts of this thesis (out-of-the-loop performance problem and mind wandering) remain complex while their definitions vary in the literature. We explore here the genesis of each concept, their different aspects and finally the definitions adopted, as well as the corresponding rationales for our decisions. The subsections 1.2- Out-of-the-loop performance problem and 1.3 - Mind wandering phenomenon present independently the definitions for each concept. We provide a summary of the work presented in the thesis so far in the last subsection 1.4 - Thesis progress recap (chapter 1).

1.2 Out-of-the-loop performance problem

1.2.1 Automation helped safety-critical industry

To continuously improve system safety, industry where safety is essential (transports, power plants, medical) makes extensive use of automation (Baxter, Rooksby, Wang, & Khajeh-Hosseini, 2012; Billings, 1991; Degani & Heymann, 2000a; Raja Parasuraman, 1987; Sheridan, 1992). Automation is “the use of various control systems for operating equipment […] with minimal or reduced human intervention” (Rifkin, 1995, pp. 66–75; Wikipedia, 2018b). The initial rationale of introducing automation is to reduce operators’ workload, reduce operational costs and errors, while increasing precision (Sarter, Woods, & Billings, 1997). In cockpit, automation was first introduced during the 1980s. Designers integrated multiple modes of automation, allowing pilots to fly in autopilot mode. Automated modes can now maintain an altitude, fly to a point, or perform a landing, all without any human intervention (Wiener, 1988). Cars are currently going through the same revolution, as engineers deploy autopilots able to manage the car’s trajectory (human driver supervises and is still responsible). At the same time, the automobile industry is conducting studies of fully autonomous cars (no human intervention or supervision required, see Ackerman, 2017).

Automation can be applied to 4 different functions: information acquisition, information analysis, decision and action selection, action implementation (Chialastri, 2012; Parasuraman, Sheridan, & Wickens, 2000). Information acquisition designates the sensing and registration of environmental data (examples of automated information acquisition are cameras or Pitot probes). Information analysis is the ability to review raw input data in order to extract meaning for understanding or predictive purposes (e.g. calculators). Decision selection uses algorithms processing previously acquired and analyzed data to choose the outcome closer to the system's purpose (e.g. processors). Finally, action implementation is about modifying the system or the environment consistently with the chosen outcome (e.g. food processors).

Table 1. Level of automation taxonomy, adapted from Endsley and Kaber (2004)

LEVELS OF AUTOMATION	FUNCTIONS			
	INFORMATION ACQUISITION	INFORMATION ANALYSIS	DECISION AND ACTION SELECTION	ACTION IMPLEMENTATION
1. Manual control	Human	Human	Human	Human
2. Action support	Human/Computer	Human	Human	Human/Computer
3. Batch processing	Human/Computer	Human	Human	Computer
4. Shared control	Human/Computer	Human/Computer	Human	Human/Computer
5. Decision support	Human/Computer	Human/Computer	Human	Computer
6. Blended decision making	Human/Computer	Human/Computer	Human/Computer	Computer
7. Rigid system	Human/Computer	Computer	Human	Computer
8. Automated decision making	Human/Computer	Human/Computer	Computer	Computer
9. Supervisory control	Human/Computer	Computer	Computer	Computer
10. Full automation	Computer	Computer	Computer	Computer

Each of these classes of function can be automated on different levels, from "human does everything manually" to "machine does everything and ignore human" (see Table 1; for a complete description of each level see Kaber & Endsley, 2004).

For example, teleoperation robots used for remote surgery operations are an example of level 2 “Action support” automation (Sheridan, 1992). They assist the surgeon by presenting important information (electrocardiogram, 3D map, time) and stabilize the tools to allow for precise actions. On the other side, nuclear plants systems operates autonomously. They nevertheless offer consoles for operators to visualize all parameters in order to correct any problem the system could encounter. This is a level 9 “Supervisory control” system.

It is clear today that without automation, safety-critical industries could not achieve such levels of safety (see Figure 1). Between 2012 and 2016, the aeronautics industry achieved a rate between 2.1 and 3.1 fatalities per million departure (International Civil Aviation Organization, 2017), partly thanks to multiple systems offering vital automated aids. For example, the introduction of the Ground Proximity Warning System (triggers an alarm if the aircraft is too close to the ground) answered the necessity to stop “controlled flight into terrain” (functioning airplane under the control of the crew is flown into terrain with no awareness of the crew). Up until 1955, U.S.A. had a rate of 3.5 controlled flight into terrain per year. Since its introduction in 1974, there have not been any case in the U.S. airspace (Sabatini, 2006). In the automobile industry, Tesla cars achieved 130 million miles on highways with their so-called “Autopilot” active, compared to a fatality every 60 million miles for all vehicles on highways (94 million on U.S. roads, see The Tesla Team, 2016).

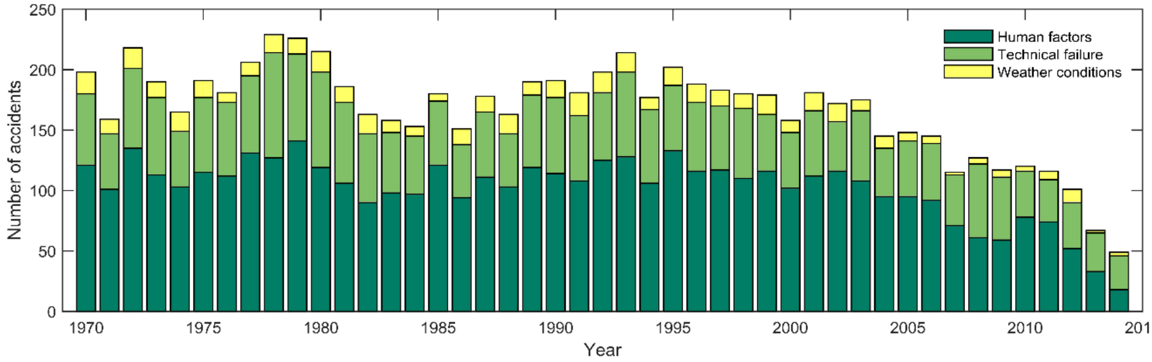


Figure 1. Aeronautical accidents from 1970 to 2014 classified by major cause, from (Peysakhovich, 2016).

It is to be noted that accidents have multiple causes, and because an accident has “weather” as a major cause does not necessarily mean that it does not involve human error also. Figure courtesy of Vsevolod Peysakhovich.

Unfortunately, if implementing higher levels of automation can improve the efficiency and capacity of a system, it also introduces difficulties for human operators.

1.2.2 Automation moves operators out of the control loop

When manually performing a task (e.g. injecting a drug for a nurse), the operator is part of the direct control loop. Even though some intermediate mechanism can modify operators' input (e.g. bikers pedal in circle but the overall system translates forward), he will remain the only initiating agent handling functions in order to fulfill his goal. On the other hand, automation moves operators away from the control loop. Intermediate levels let humans handle some functions, however higher levels put them in a supervisory role. Situations where operators are supervising automated control loop (or parts of it and manually performing others) are called out-of-the-loop (OOTL) situations (Endsley & Kiris, 1995). Unfortunately, operating OOTL can have negative side effects on overall performance, commonly referred to as OOTL performance problem. Therefore, in the definition we adopt, OOTL is a situation that lasts as long as the operator is supervising the system, while OOTL performance problem points to measured performance drops linked to the OOTL situation (see Merat et al., 2018 for another proposed definition). OOTL performance problem is linked to many issues in practice. In accident reports, one can find the terms "total confusion" (Bureau d'Enquête et d'Analyse, 2002, p. 167; National Transport Safety Board, 1975, p. 17), "surprise effect" (Bureau d'Enquête et d'Analyse, 2012a, p. 44, 2016, p. 10) or "no awareness of the current mode of the system" (Bureau d'Enquête et d'Analyse, 2012b, p. 178).

OOTL performance problem is fundamentally an issue of human-automation interaction, ultimately spoiling performances. It is characterized by a degradation of one or more psychological mechanisms necessary for system supervision. OOTL performance problem can arise because of issues of poor performing monitoring (Somon, Campagne, Delorme, & Berberian, 2017), impaired decision-making (Skitka, Mosier, & Burdick, 1999), lower metacognition (Lau, Skraaning Jr, & Jamieson, 2009), reduced perception (Louw & Merat, 2017) or loss manual skills (Cummings, 2004). For example, Endsley and Kiris (1995) designed an automobile navigation task where participants had to choose between trajectories to reach the destination as fast as possible with limited given gas. Each trajectory was given with the estimated gas consumption and time took. Participants were either choosing manually, or had the help of an expert systems giving them the probability that each was the correct solution. Endsley and Kiris (1995) observed that participants performing the task with the expert system were slower to choose if the system broke down, exhibiting a form of manual skill degradation.

Among those problems, the pre-eminence of attention-related problems surprised designers (Mackworth, 1948). Mosier et al. (1994) examined NASA's Aviation Safety Reporting System (ASRS) database and found that 77% of the incidents in which over-reliance on automation was suspected involved a probable "vigilance" failure (even though vigilance is a blurry term that involve many different mechanisms). Similarly, Gerbert and Kemmler (1986) studied German aviators' anonymous responses to questionnaires about automation-related incidents and reported failures of "vigilance" as the largest contributor to human error. Multiple studies addressed the problem and unveiled a progressive degradation of operators' detection ability when interacting with highly automated systems, labeled "vigilance decrement" (see, for example, O'Hanlon, 1981; Strauch, 2002; Wiener, 1987). The attentional part of OOTL performance problems has been reported multiple times within accidents or incidents analysis reports, despite a heterogeneous terminology. For example, the Eastern Airlines L-1011 flight crashed during clear weather because the crew was focused on explaining a red light in the cockpit and didn't notice that the autopilot had disengaged (Federal Aviation Authority, 1972).

As a human-machine interaction issue, OOTL performance problem can originate from both operators' internal states and system properties. Throughout the management of their task, operators can experience many different internal physiological and psychological states, some of them impairing their ability to supervise the system efficiently. Fatigue (Desmond, Hancock, & Monette, 1998), stress (Sauer, Nickel, & Wastell, 2013), loss of agency (Berberian, Sarrazin, Le Blaye, & Haggard, 2012; Sahaï, Pacherie, Grynszpan, & Berberian, 2017), cognitive mismatches (Baxter, Besnard, & Riley, 2007), complacency (Bagheri & Jamieson, 2004), among others, were reported as causes of OOTL problems that can arise at various moments. In 2017, the National Transportation Safety Board (NTSB, American Agency charged with transportation safety accident investigation) considered fatigue in its advocacy priorities for safer transports.

On the other hand, OOTL performance problem can be induced by flaws in system design. Systems are regularly considered opaque, complex and not communicating enough. On the contrary, systems can sometimes issue much information without letting operators time to analyze it, or with no efforts made to prioritize information. In the well-known Rio-Paris flight accident, autopilot issued several messages to the crew (ECAM messages, see Bureau d'Enquête et d'Analyse, 2012a). In a calm environment, with time to analyze the situation, the crew may have read all those messages and understand the problem. Unfortunately, airplanes do not allow much time to think, and so the carefully designed pilot-autopilot communication did not help in this situation. Another problem regularly encountered when automating a task is the redistribution of workload into idle periods interspersed with important workload peaks. Instead of lowering the general level of workload to help operators, some automation only transform things for the worse, as sudden high workload management is difficult to handle after idle times. Called "clumsy automation" by Wiener (1989), this kind of problem is pregnant in many critical systems. Wiener (1989) followed during 3 years Boeing 757 (already equipped at that time with many automated aids) crews to collect their opinion and experiences. When asked if they agreed with the statement "automation does not reduce total workload, since there is more to monitor now", more than half of the pilots answered "Agree" or "Strongly Agree".

Higher levels of automation move operators to a supervisory role, leading them to OOTL situations. Laboratory studies, accident reports and questionnaires point OOTL performance problem as an impairment of psychological mechanisms required to supervise a system caused by both system features and operators' internal states. Researchers already identified multiple consequences OOTL performance problem. However, much work is still needed to understand the sources of human-system interaction difficulties. That can only be done with valid models and concepts.

1.2.3 Out-of-the-loop research needs validated psychological models

Although OOTL situations represent a key challenge for system designers, the performance problems associated remain difficult to characterize and quantify the influence of underlying psychological mechanisms (Bainbridge, 1983; Baxter et al., 2012). Studying supervisors in real conditions is complex because researchers cannot ask to pause the task or disturb the operator too much. Moreover, operators display complex behaviors, which are difficult to translate into models and theories because multiple psychological constructs are intertwined. Even though laboratories allow addressing specific points using controlled experiments, OOTL performance problems only arise after much time using the system. This forces researchers to use long experiments (sometimes several hours) with only a few events, sometimes even only one, to study (Casner & Schooler, 2014, 2015; W. C. Harris, Goernert, Hancock, & Arthur, 1994; Liu, Fuld, & Wickens, 1993; Lorenz, Di Nocera, Röttger, & Parasuraman, 2001; Thackray & Touchstone, 1989).

On top of limiting the pace of scientific discoveries on OOTL problems, such constraints also helped spreading the use of questionable models. Dekker and Hollnagel (2004) denounced the rise of “folk models” in human factors. Folk models are models that substitute one concept for another in their definition (instead of decomposing), overgeneralize to large portions of human behavior and cognition, which make them immune to falsification because of their blurriness (Dekker, 2015; Moray & Inagaki, 2000; Popper, 1972). On the other hand, such models are generally appealing because they intuitively make sense. However, careful examination allows discarding folk models and preferring more rationale and scientific concepts.

A perfect example of folk models is situation awareness. Situation awareness points “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near” (Endsley, 1996; Wickens, 2008). Multiple studies linked a “loss situation awareness” (capture by verbal report about the element perceived in the environment) to performance drops in laboratories (Carmody & Gluckman, 1993; de Winter, Happee, Martens, & Stanton, 2014; Durso, Hackworth, Truitt, Crutchfield, & Nikolic, 1999; Endsley, 1988; Endsley & Rodgers, 1998; Kaber & Endsley, 2004; Matthews & Beal, 2002; McGuinness, 2004). The first problem of situation awareness is that the concept overlooks the human abilities to collect and use information unconsciously to avoid data overload. Multiple studies unveiled effects such as the Unconscious Thought Effect (Dijksterhuis, 2004; Dijksterhuis, Bos, Nordgren, & van Baaren, 2006), implicit learning (Cleeremans, 2006; Cleeremans, Destrebecqz, & Boyer, 1998) and many others (Kouider & Dehaene, 2007; Morris, Öhman, & Dolan, 1998; Mudrik, Breska, Lamy, & Deouell, 2011; van Gaal & Lamme, 2012). We are all able to perform repetitive and well-known tasks while thinking about unrelated matters without being able to report every detail of what we just did (see the next section 1.3 - Mind wandering phenomenon). Worst, this concept of situation awareness masks the most critical issue: what is the role of awareness in the control of complex system?

Second, even though situation awareness can help to understand which information did the operator not consciously perceived, the concept is so wide that it becomes uninformative. What does a drop of situation awareness says? In the light of the definition, a “loss of situation awareness” arises when the operator knows less information than there is in the world around him. Put differently, a loss of situation awareness is “merely the difference between what you knew then versus what I know now” (Dekker, 2015). It is easy to prove that an operator did not know everything when the accident occur. Unfortunately, it is also uninformative regarding the causes of the accident, i.e. *why* the operator did not have all information. Did he not collected relevant information because his attention was directed somewhere else? Was the information pregnant enough to be perceived? Was the issue on consciously accessing to the representation of the environment? Situation awareness can be seen as an abstraction wide enough to cover the whole cognition, but remains too vague to inform automation designers regarding the precise psychological mechanism at hand.

As situation awareness, many concepts of human factors were useful in the past century, when acquisition systems and models were almost impossible to use for ecological situations. Unfortunately, many did not evolve with recent theories and measurement systems. However, some studies did build on these critics to provide better definitions based on actual human mechanisms. For example, Parasuraman et al. (1993b) firstly defined complacency as the “uncritical reliance on the system leading to thinking of it as more competent than it actually is”. However, other studies challenged the very existence of this concept as a way to explain automation-related errors (Moray & Inagaki, 2000). They argued that a participant could miss a target because of many other mechanisms aside complacency. As complacency is a monitoring problem, monitoring should be examined, not detection. Taking their advice into account, Bagheri and Jamieson (2004) investigated complacent behavior by comparing the optimal sample given the dynamic of a source, to the actual sample rate of participants. They also asked participants for their trust regarding automation failure. Linking both measures, they highlighted the complacency phenomenon. Nevertheless, they specified that it could also be a deliberate strategy to optimize mental resources and that there is a need for further studies.

Following on the detailed work of Bagheri and Jamieson (2004), studies investigating the causes of OOTL performance problem should use specific psychological constructs and take care to define it as precisely as possible. Physiological measures can help in this regard, backed by latest technologies. They can provide detailed insight regarding operators' behavior without relying on intermediate questionable cognitive states. This philosophy makes use of the novel field known as neuroergonomics (Gramann, Fairclough, Zander, & Ayaz, 2017; Johnson & Proctor, 2013; Parasuraman, 2011), even though neuroergonomics itself can be subject to "folk models" use. For better accident analysis, it is up to human factor researchers to investigate and propose validated human factors constructs that may explain the performance drops observed.

1.2.4 Operators are still needed in control

Despite recent technology developments and the OOTL performance problem, we still include human operators within automated environments. A natural question would be "if OOTL thwarts human-machine interactions, why not suppress humans and automated at level 10 'Full Automation'?" Multiple reasons back operators' final control up.

The first reason for keeping operators in control is the flexibility of humans to handle multiple situations. Automation is now far from being able to handle all tasks during normal functioning. For example, pilots still deal with ATC communications, weather-related decisions or on-board medical problems. For emergencies, databases include many reports of humans “saving the day” (Degani, Chappell, & Hayes, 1991) when the automation could not. On January 15, 2009, the captain of the US Airways 1549 departing from LaGuardia airport flight saved passengers and avoided a large catastrophe by ditching on the Hudson River after losing thrust in both engines (National Transportation Safety Board, 2010). Safety engineers did not consider the possibility of losing both engines, resulting in an absence of autopilot adequate response as well as adequate checklist. This example illustrates how multiple safety-critical incidents produced an unexpected situation, which was nevertheless handled by the human pilot. Multiple little incidents happen every day while not descending into catastrophes thanks to human intervention.

The second reason for keeping operators in control is the dependability of the automation to its design. Automation can be subject to human errors. Certification processes are here to test every possible state and transition the system can take. However, a bug can always pass the tests. When the system fails, it should provide enough information to help the operator diagnose and correct the problem. On the contrary, bugs happening without proper feedback can have tragic consequences. A perfect example is the accidents linked to the Therac-25, a computerized radiation therapy machine (see Figure 2). During the mid-1980s, this system was involved in several overdose accidents. One of those accidents happened at the Texas Cancer Center in 1986 (Sarter et al., 1997). A technician made a mistake when setting up the device for the treatment. When she tried to correct it, the speed she used when pressing the buttons revealed a software glitch that displayed the corrected information but did not store it for the treatment. She then activated the system but encountered a “Malfunction 54” error. This error being quite common with the system (and the documentation scarcely explaining that this error corresponded to a “dose input 2” error), she decided to activate the machine nonetheless. The patient was hitting with an overdose, and died a few month later from complications (it has to be noted that on this particular day, the technician could not see the patient because both the intercom and the camera feedback were inoperative; for a complete description of the accident, see Leveson & Turner, 1993). This example illustrates how despite extensive testing, designers will never be able to produce programs with zero errors. Systems will always encounter bugs or glitches, which can be handled by human operators if the system is informative enough.

Finally, the last reason for keeping human operators in the loop is that possibilities to make automation more flexible will not be applicable to critical systems before some time. Indeed, many designers try to overcome the problems of traditional automation by developing non-deterministic programs. The goal is to produce systems that could adapt to situations as humans do. Recent examples have demonstrated great promises in the medical domain (NDTV, 2016), automobile (Normand, 2017) or unmanned air vehicles (Ernest & Carroll, 2016). However, an important challenge will be to certify these systems from laboratory to real environments. Current certification processes require deterministic algorithms. They must ensure that the system will always have the same behavior when confronted to the same situation, a condition that non-deterministic algorithms cannot meet (Koopman & Wagner, 2016). Moreover, a trained program can discover solutions that match its own definition of "success" but do not make sense for us. In the video game field, a program trained to play Tetris found that the best way to not lose was to let the game on "pause" indefinitely (Murphy, 2013). Finally, such systems still run on powerful, costly and huge machines. During the famous match of Go against Lee Sedol (one of the strongest players in the history of Go), Google's AlphaGo program used 1920 controller processing units (The Economist, 2016; Wikipedia, 2018a). Such assembly is far from being affordable by many safety-critical industries, and would not necessarily fit in many environments (e.g. aircrafts). While many experts guarantee that such systems will become small and affordable in the years to come, more and more researchers openly show that promises do not meet the facts beyond the hype (Piekniewski, 2018).

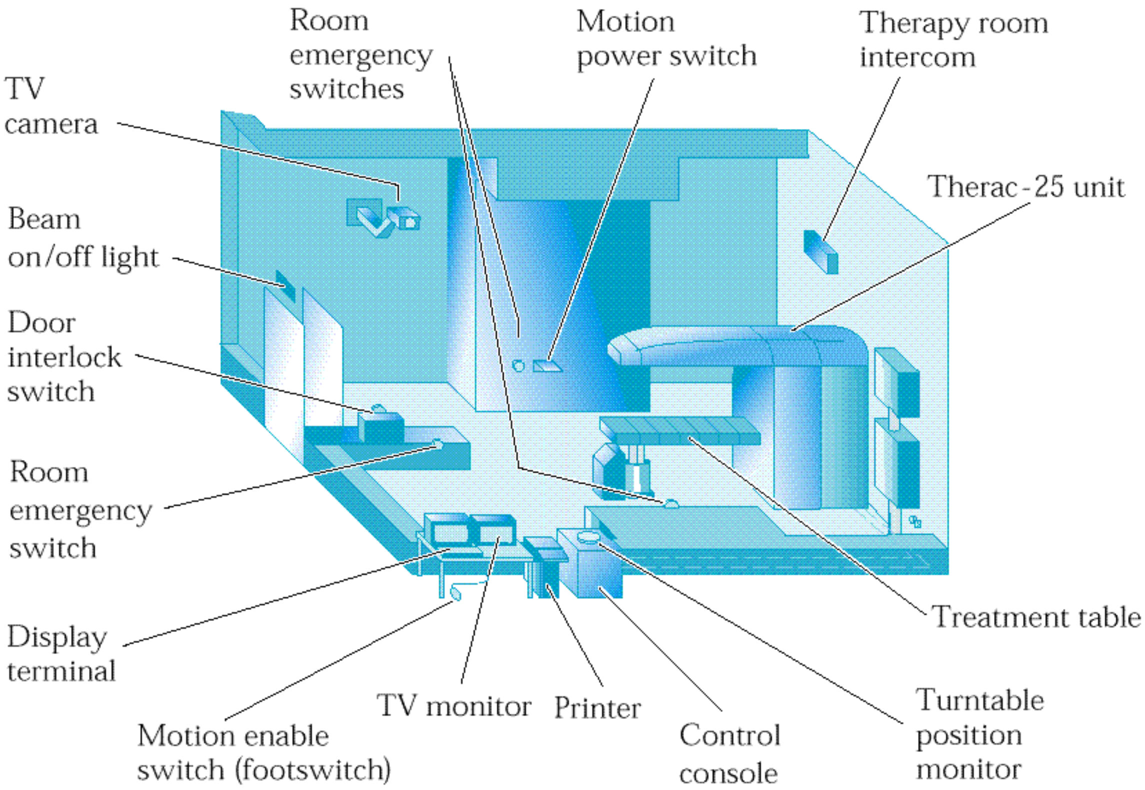


Figure 2. Diagram of the Therac-25 operation room, from (Wang, 2017)

The Therac-25 a computerized radiation therapy machine involved in at least 6 reported overdoses between 1970 and 1990 (Leveson & Turner, 1993)

As shown in this subsection, automation allowed operators to reach unprecedented levels of safety. However, it also brought some new issues by changing the very nature of the tasks to perform when supervising a system, creating OOTL performance problems. Nevertheless, human operators are still needed to handle systems and check critical decisions. We do not want to minimize recent advances in automated commands, nor do we advocate for lower automation. Rather, we underline that despite ground breaking results, critical systems will still need human agents for quite some time. Therefore, we need to pinpoint the psychological mechanisms creating OOTL performance problems and affecting takeover capabilities. The aim is to understand how to facilitate operators' reentry in the loop when an action is needed. An interesting approach would be to address the moments when our attention is internally directed towards either not here or not now (or both) without us being necessarily conscious of it. This is called the mind wandering phenomenon.

1.3 Mind wandering phenomenon

1.3.1 Mind wandering is a family of experiences

Mind wandering (MW) have known a renewed interest since two decades (Callard, Smallwood, Golchert, & Margulies, 2013). Multiple studies have showed that MW affects us in all aspects of our lives (see Figure 3). However, maybe due to the relatively new visibility of this research domain, multiple definition of MW suffers from a lack of common definitions (Christoff et al. in press; Seli, Kane, Metzinger, et al., 2018). In this thesis, we adopt the view of Seli et al. (2018) of a family-resemblance framework. We address the dimensions of task proximity and depth of decoupling. We detail our rationales below.

Studies claiming to address “mind wandering” within the same field will name it daydreaming, task unrelated thoughts, spontaneous cognition, proneness to fantasizing, inner cognition, despite each of these terms pointing a particular reality that does not completely overlap the others. Some studies even include contradiction regarding their own definition. For example, Baird et al. (2012) investigated the impact of MW over the creative process. They introduce MW as “thinking that is unrelated to an overt goal”, i.e. task-unrelated thoughts. They asked participants to perform the Unusual Uses Task (Guilford, 1967). This task requires participants to generate as many unusual uses of a common object in a limited period. Participants performed the Unusual Uses Task during two blocks of 2 minutes each. They were exposed to four different conditions: (a) performing a demanding task between the two blocks, (b) performing a less-demanding task, (c) resting and (d) no break. The issue lies is the (c) condition, resting. Indeed, they themselves reported having troubles interpreting the results for the resting condition: “this comparison is difficult to interpret because the rest condition included no primary task to which internal thoughts could fail to pertain” (this point does not minimize the important results of the study).

Figure 3. The MW at work, from (Grégoire & Lachance, 2013)

Far from only being an epistemological debate, considering these terms as designing the same phenomenon could lead to compile different results on the MW even though experiments do not measure the same entity. A very famous study by Killingsworth and Gilbert (2010) used a smartphone application to measure MW in the daily lives of about 5000. The result was a natural MW rate of 46.9% for people. However, one must note that Killingsworth and Gilbert (2010) defined MW as "stimulus-independent thoughts" happening during a task (question "Are you thinking about something other than what you're currently doing?"). Their definition, although consistent with their experiment, dismissed all experiences of inner cognition happening when at rest, or moments when we willingly evade the moment to fantasize. On the other hand, the Daydreaming Frequency Scale (Giambra, 1993; Singer & Antrobus, 1970) asks questions about the propensity of having thoughts emerging when participants are not performing any task: "When I have time on my hands I daydream…". Despite this definition, some studies used the DDFS to assess the propensity to mind wander while defining MW as task-unrelated thoughts (Stawarczyk, Majerus, Maquet, & D'Argembeau, 2011).

Christoff and colleagues proposed to defined MW in their "dynamic framework" as "mental state, or a sequence of mental states, that arises relatively freely due to an absence of strong constraints on the contents of each state and on the transitions from one mental state to another" (Christoff et al., in press; Christoff, Irving, Fox, Spreng, & Andrews-Hanna, 2016). Unfortunately, MW studies addressed so many dimensions of the phenomenon that such a hard definition approach may not be the more suitable. From physiological, behavioral and epistemological perspective, no rational exists to discriminate between what is MW and what is not as binary. How would one name the thoughts of a driver willingly trying to organize his evening while supervising an autonomous car? They are directed, and so should not be admitted as MW according to the definition of Christoff et al. At the same time, he may have started by unconstrained thoughts, and then realized he was thinking about his evening, for ultimately deciding to continue this train of thoughts. When would MW start? What does "relatively" means in the definition of MW?

As Seli et al. (2018) pointed, if we proposed this driver to define the level of coercion he applies on his thoughts, how would a researcher define what come under MW and what does not? On a scale from 1 to 10, would a 4 count? An what about a 5? What rational would motivate such a binary choice? The driver could experience the same performance drops observed for unconscious MW, which is the inability to react efficiently to sudden events. No physiological, behavioral or epistemological argument can help to draw such a hard line and discard a part of the literature as "not investigating MW".

In order to solve this definition problem, Seli et al. (2018) proposed to adopt a family-resemblance framework to define MW (see Figure 4): instead of trying to frame MW into one hard definition, researchers should see MW as a family of experiences of graded membership along several dimensions. Such approach can be seen for broad terms like "games" or "sport" (Wittgenstein, 1968, p. 33), but also for scientific terms like "cognition" or "conscience". They differ drastically in their definition between fields (as the definition of cognition differs between visual cognition and numerical cognition fields). However, each of these terms is defined probabilistically with examples being more or less prototypical of the whole field. Such family-resemblance framework would suppress the necessity to arbitrarily decide that this research is MW and this one is not. For example, it would consider intentionality in the possible dimensions of MW (Golchert et al., 2016; Seli, Risko, &

Smilek, 2016; Seli, Risko, Smilek, & Schacter, 2016). Other dimensions of MW are awareness (Bastian et al., 2017; Schooler et al., 2011), task proximity (Casner & Schooler, 2014, 2015) or time orientation of the content of the thoughts (Baird, Smallwood, & Schooler, 2011; Stawarczyk, Cassol, & D'Argembeau, 2013).

Nevertheless, the family-resemblance framework of MW should not be seen as the blurriness of several concepts of human factors (e.g. situation awareness). Firstly, this framework is based on decomposition, as MW is broke down along several dimensions. Moreover, adopting such framework requires that studies should clearly state which aspects of the MW family it addresses and how it defined MW for the participants. Unlike "folk models", each prototype of MW defined along dimensions is a precise construct that can be measured and has physiological signature.

For example, the content presence dimensions of MW (i.e. individuals thinking about something or having their mind blank) has demonstrated some influence on behavioral and neural signals (Esterman, Noonan, Rosenberg, & DeGutis, 2013). Even though the family-resemblance framework of MW is still debated, we here adopt this view for the reasons previously mentioned (for a more exhaustive view of all arguments, see Seli et al., 2018). As this thesis is interested in investigating the influence of MW on operators' behavior in operational automated environments, we focused on MW emerging when performing a task. Within this context, we focus on aspects of task proximity (MW when performing a task) and depth (decoupling from the environment). We did not inspect other dimensions because of the difficulty in self-reports and data analysis with too many predictors dimensions. Accordingly, we decided to focus on a limited number of important dimensions consistent with our goal. We detail each of the two chosen dimensions in the next sections.

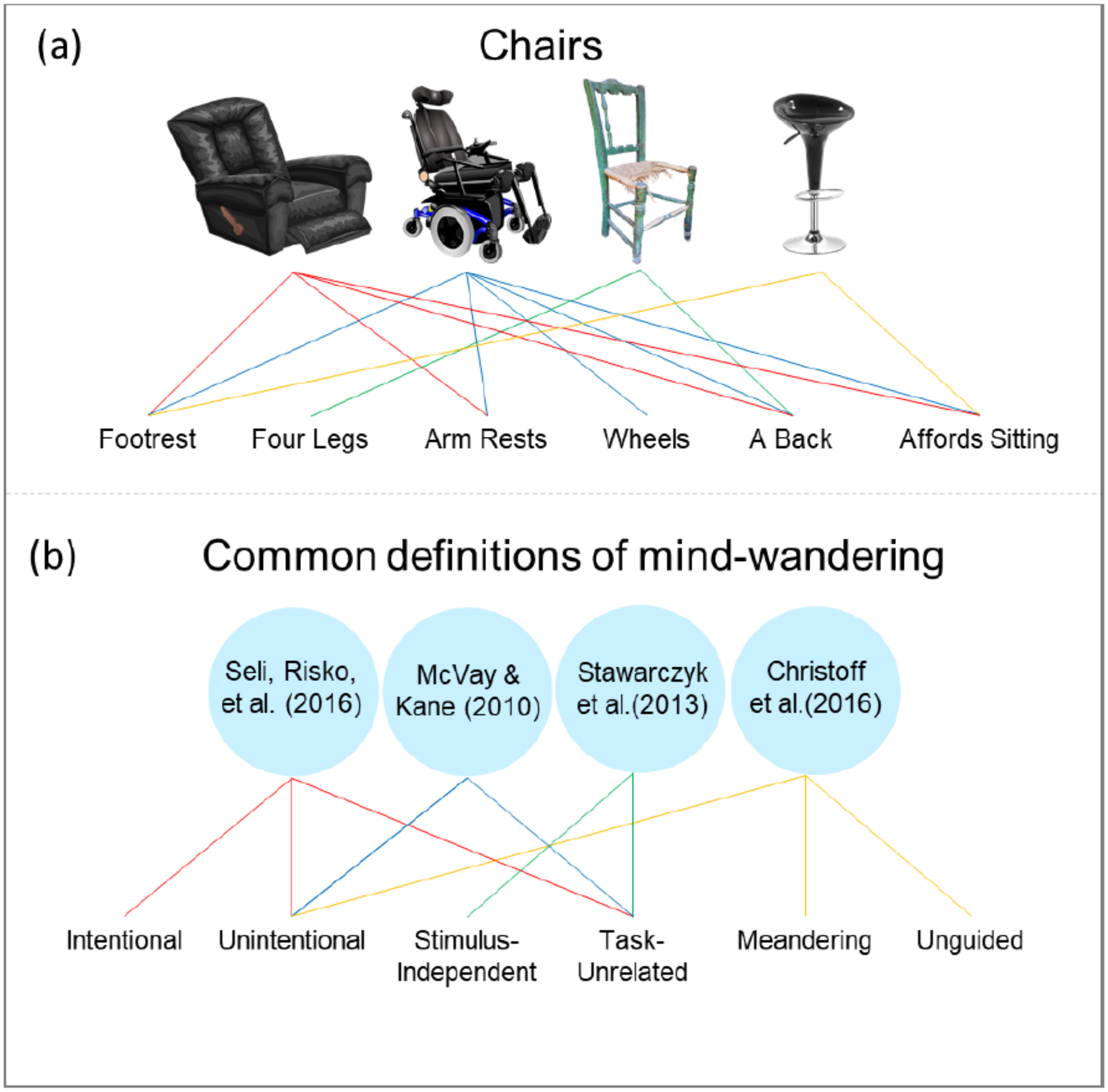


Figure 4. Schematic for a family-resemblances view, from (Seli et al., 2017).

In panel (a) are four chairs that share one or more features with the each other. For example, the first, second, and fourth chairs all afford sitting (the third chair has a broken seat), and the first, second, and third chairs all have backs. However, no single feature runs through all members of the family of chairs. Hence, there is no universal feature that defines membership. The "family" of chairs is held together by overlapping features.

In panel (b) are examples of different definitions of mind-wandering from four articles (Christoff et al., 2016; McVay & Kane, 2010; Seli, Risko, Smilek, et al., 2016; Stawarczyk et al., 2013). Across these articles, mind wandering is defined with reference to specific aspects of conscious experiences. However, the family resemblances-view posits that, just as there is no single feature that defines the chair family, there are no specific features that a thought must have to be granted membership in the mind-wandering family. Instead, by the family-resemblance view, mind wandering is a collection of related experiences that share some, but not all, features.

Figure and description courtesy of Paul Seli and Elsevier

1.3.2 Mind wandering can be task-related or -unrelated

The content of the thoughts during MW involves many parameters of high interest. Among them, the relatedness to the task is of high interest. Task proximity is one of the major aspects of MW episodes. MW propensity increases when the subject performs monotonous tasks (Eastwood, Frischen, Fenske, & Smilek, 2012). Familiar stimuli have been shown to increase MW frequency (Bastian et al., 2017), as do easier or longer tasks (Smallwood & Schooler, 2015; Thomson, Seli, Besner, & Smilek, 2014).

However, where do thoughts stop relating to the task? For example, how do we categorize an operator thinking about how he would have designed the system differently? Obviously, it concerns how he *could do the task*, but not how to *do it right here and now*. The view of participants as being either “on-task” or “off-task” is known as the dichotomy hypothesis (Schad, Nuthmann, & Engbert, 2012). It is a black and white method aimed at simplifying results analysis. However, it implies highly questionable hypothesis on how to consider thoughts at the frontier (e.g. memory of similar tasks, imagining how the task could be, or even previous performances, (Head & Helton, 2016).

To solve this problem, some studies adopted a Likert scale and asked subjects to quantify their focus on the task from “on-task” to “off-task” (Bastian et al., 2017; Christoff, Gordon, Smallwood, Smith, & Schooler, 2009; Farley, Risko, & Kingstone, 2013; Krimsky, Forster, Llabre, & Jha, 2017). Even though this method solves the problem by suppressing the binary view of MW, it postulates that attention exhibits a linear distribution between the task and the rest of the thoughts. Empirically, there exist at least four different types of thoughts regarding task proximity. Thoughts can be either about our immediate environment independent from the task (here and now, e.g. “this CPR screen is old”), about something here but not now (“the last experiment here was with another screen”), about something not here but happening now (“this plane is making too much noise”) or about something neither here nor now (“I cannot wait for the OM football match tonight”). All of these thoughts go along the dimensions of space and time. However, it is not certain that thoughts about here but not now would have the same impact on behavior as thoughts about now but not here. Similarly, how can people experiencing “mind blanking” – not thinking about anything – categorize this in the attentional report? To solve this issue, researchers can ask subjects to speak or write the exact content of their thoughts. A set of raters examine the reports, and agree on several dimensions to categorize thoughts. Complete reports solve the issue of dimensions, as they adapt to each set of participants. However, it becomes extremely difficult for an algorithm to classify the thoughts, preventing efficient automation of the attentional detection. Moreover, reports could be highly disruptive for the natural tendency to mind wandering as subjects must go through an extensive meta-cognitive process.

A third strategy is to propose several categories previously defined to report the attention (Hu, He, & Xu, 2012; McVay & Kane, 2009; Stawarczyk, Majerus, Maquet, et al., 2011). For example, Casner and Schooler (2014, 2015) asked pilots to fly a simulator and report their thoughts on either “on-task thoughts” (e.g. interacting with the system), “task-related thoughts” (e.g. plan the flight ahead) or “task-unrelated thoughts” (e.g. being thirsty and wanting a Radler beer). Such attentional probes allow participants to spend only a few seconds to report their thought category, lowering the impact on the subject, and nevertheless allow a more accurate thought classification. Moreover, their experiment was to our knowledge the only study to investigate MW in an operational setting. For all these reasons, we chose to use the same kind of fixed categorized attentional reports throughout our experiments (see chapter 4).

1.3.3 Mind wandering depth could influence decoupling

1.3.3.1 Mind wandering induces decoupling from the environment

One of the most threatening aspect of MW for safety is the decoupling from the environment (Schooler et al., 2011). Operators engaged in an episode of MW will see their encoding of external information degraded. MW disrupts visual information flow by reducing pupil diameter (K. McIntire, P. McIntire, Mckinley, & Goodyear, 2014; Smallwood et al., 2011) and increasing blink frequency (Smilek, Carriere, & Cheyne, 2010b). Neuronal studies demonstrated during MW an increase of alpha waves power, linked with sensory suppression (Foxe & Snyder, 2011; O’Connell et al., 2009), and a reduction of Event Related Potentials linked to external information perception and processing (Smallwood, Beach, Schooler, & Handy, 2008). At the behavioral level, this decoupling translates into a decrease in performance. Reaction exhibits higher variability (Bastian & Sackur, 2013), while omissions and anticipations are more common (Cheyne, Carriere, Solman, & Smilek, 2011). Accuracy was shown to decrease in both simple paradigms (Kam et al., 2012) and more ecological ones (Yanko & Spalek, 2014).

1.3.3.2 Mind wandering episodes may differ in depth

Some studies investigated the possibility that episodes of MW could integrate an aspect of "depth", or "intensity", directly related to the decoupling from the environment. Cheyne et al. (2009) used a Sustained Attention to Response Task (SART) to investigate the validity of their bi-directional model of inattention. They obtained convergent measures supporting their three postulated states of inattention: level 1 characterized with more erratic reaction time, level 2 with anticipations and level 3 with omissions.

Following the same path, Schad et al. (2012) detailed the "levels of inattention hypothesis" based on the assumption that our mind processes information in a sequential manner involving greater complexity with each step. MW could then thwart information processing at different stages, depending on the depth of the episode. While some MW episodes could be superficial, only impacting higher cognition, some others could completely decouple from the task by blocking external information encoding and "cascade through the cognitive system" to impact more complex processing (Smallwood, 2011). Schad et al. (2012) further strengthened their argument by manipulating a corpus of text by inserting different types of errors, from pseudo-words (lower level errors) to inconsistent statements (higher level errors). During the experiment, those participants who experienced MW exhibited progressive gaze pattern modification depending on error level, supporting a graded nature of the phenomenon.

A graded MW with different decoupling could explain why we are most of the time able to conduct tasks while being in MW, while sometimes we make clear errors that could have been avoided with our full attention.

1.3.3.3 The decoupling hypothesis stands the critics

Nonetheless, the decoupling hypothesis was recently criticized by Helton and colleagues (Head & Helton, 2016). They argued that participant could report MW as a way to rationalize their poor performance. According to this "rationalization" theory, lower performance would cause MW reports, and not the other way around as it is commonly accepted in the MW literature. They tested their hypothesis by conducting a modified GO/NOGO task. Participants were required to click when they saw a number (No-Go stimuli), but to withhold when there was a 3 (Go stimulus). Prior to each number stimulus, a word or a control screen was presented. The last trial (225 trials in total) was always a No-Go trial. Then, participants completed a memory task by reporting if there was a word before the catch trial and if the word was present in a set of words. They also reported if they were on-task or off-task. Participants were dispatched between 5 different conditions: (a) words always precede No-Go stimuli, (b) words always precede Go stimuli, (c) equal likelihood of word presence between No-Go and Go stimuli, (d) only one word before the catch trial and (e) no word at all. Group analysis revealed no significant differences in MW ratings. On the contrary, group (a) achieved for the catch trial lower withhold accuracy than the average of all groups while group (b) achieved higher withhold accuracy (other groups achieved approximately the same accuracy; see Figure 5). At the individual level, they found that participants reporting MW had a lower withhold accuracy. However, when considering the memory task, they did not differ from those reporting being focus. The conclusion of Head and Helton (2016) is that the decoupling hypothesis is wrong because reports of MW did not help discriminate between participants who are aware of the pre-catch trial word. More generally, their hypothesis is that participants report MW to rationalize their poor performance.

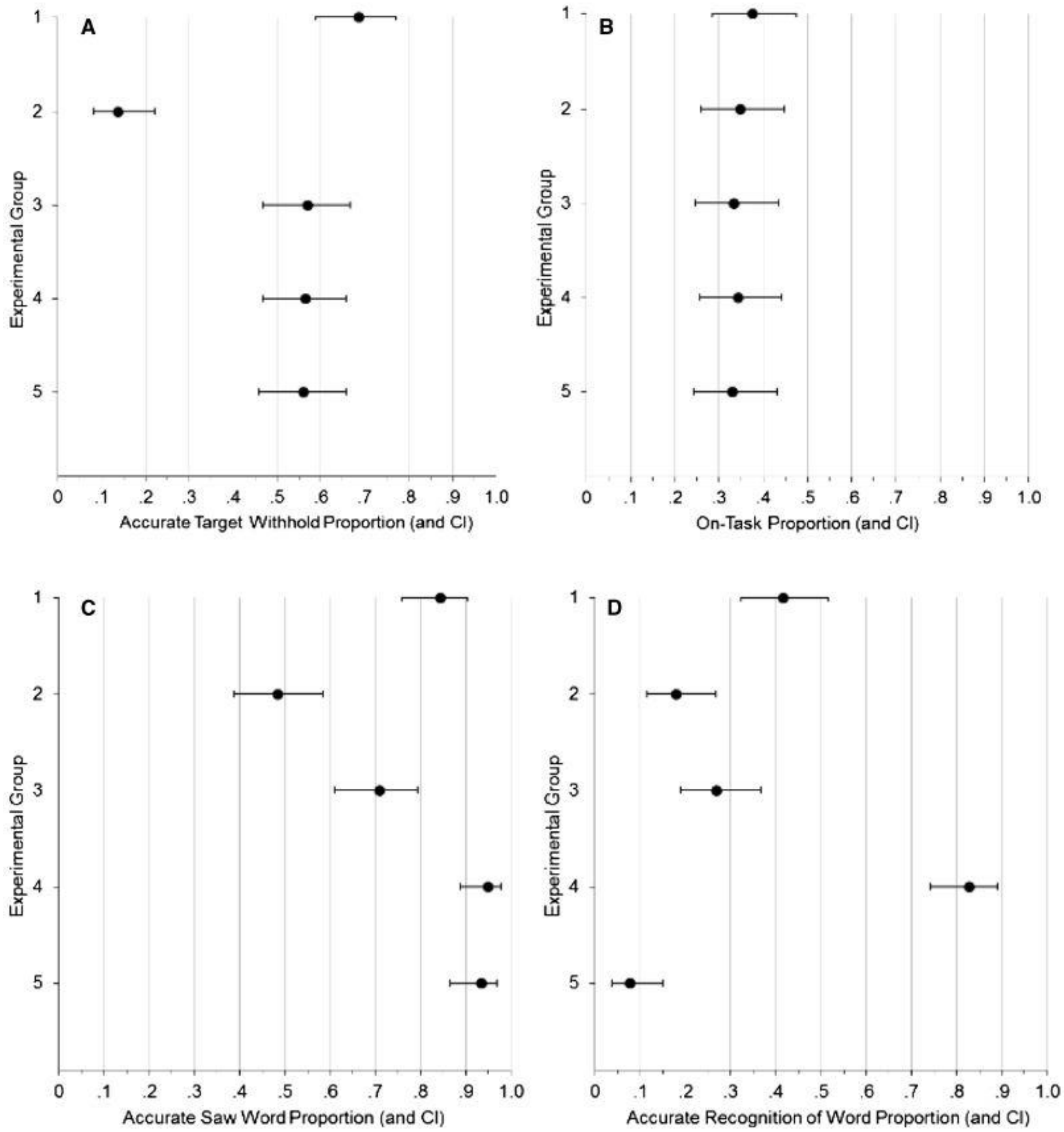


Figure 5. Results of Head and Helton (2016).

Proportions of people who correctly withheld a response to the catch trial (a), reported being on-task to the thought probe (b), reported correctly seeing the pre-catch trial word (for group 5 this is the proportion who correctly reported not seeing a pre-catch trial word) (c), and correctly recognized the pre-catch trial word (d) for each of the experimental groups with 95 % CI. For group 1 the withhold No-Go stimuli was always preceded by a word, for group 2 the Go stimuli were always preceded by a word, for group 3 words were equal-likely to occur before Go and No-Go stimuli, for group 4 a word only occurred before the catch trial and for group 5, no word actually occurred in the task.

Figure courtesy of James Head and Springer.

Although the conclusion of Head and Helton (2016) is supported by a well-thought experiment, it is not exempt from possible flaws that could bias their results. The main task remains the GO/NOGO, while participants are told for approximately 5 minutes that they have to consider words as noise. It is possible then to consider MW and word awareness as dissociated. Indeed participants inhibit the perception of the word in order to focus on the perception of the numbers, except when the words help them regarding the GO/NOGO task. Group (b) was not helped at all by the appearance of the words (which appear before each Go trial). Interestingly, this group has the lowest word awareness and word recognition (group (d) apart). On the contrary, group (a), for which words were the most meaningful (always appearing before a No-Go trial), had both the highest withholding accuracy and the second highest word recognition of all groups. Finally, in their individual level analysis, Head and Helton (2016) mixed all groups to study correlation between MW reports and performance. It is therefore predictable that the influence of word appearance would be lost in the process, yielding biased results. For all these reasons, the study of Head and Helton (2016), although very interesting in its approach, cannot alone rule out the decoupling hypothesis.

Despite the possible flaws in the conclusions of Head and Helton (2016), their warning about phenomenology flaws may indeed concern some paradigms (e.g. oddball or GO/NOGO tasks; Braboszcz & Delorme, 2011; Forster & Lavie, 2014; Robertson, Manly, Andrade, Baddeley, & Yiend, 1997). Nevertheless, many other experiments cannot by criticized with the same arguments. Continuous metrics show similar negative influence of MW on performances (Cowley, 2013; Dündar, 2015; He, Becic, Lee, & McCarley, 2011; Kam et al., 2012; Yanko & Spalek, 2014); if participants were to realize that they did wrong, they would directly correct their behavior. If performance before MW probes were lower, it means that participants were not aware of their poor performance. Similarly, many studies highlighted a link between overall performances and the propensity to MW by measuring MW propensity before the task, for example with questionnaires (Berthié et al., 2015; Galera et al., 2012; Mrazek, Phillips, Franklin, Broadway, & Schooler, 2013). Therefore, participants could not rationalize their poor performance by reporting MW reports.

Overall, the MW literature exhibits a remarkable homogeneity regarding its results showing that MW disrupts online adjustment of behavior (see Figure 6 and Kam et al., 2012). Nevertheless, while investigating further, all MW researchers should consider in their protocol design Head and Helton's (2016) warning concerning the phenomenology flaw.

1.4 Thesis progress recap (chapter 1)

We built in this chapter the epistemological structure framing the whole thesis. We defined OOTL performance problem as issues related to automation supervision, and MW as a family of experience related to internally directed cognition. The prototype of MW we adopt is the thoughts unrelated to the here and now during a task and only address dimensions of task proximity and depth of decoupling.

The following chapter will complete the theoretical aspects of this work. We will develop the evidences in the literature for a link between MW and OOTL performance problem and compare the two concepts to look for similarities.

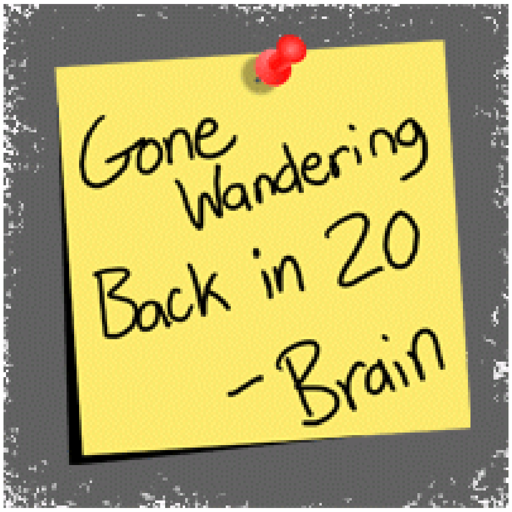


Figure 6. Our brain during MW.

From ("What is MBCT? Do You Want To Know More About Mindfulness?", 2015).

2 REVIEW OF EVIDENCES FOR A LINK BETWEEN MIND-WANDERING AND OUT-OF-THE-LOOP

- *Both OOTL performance problem and MW emergences involve sensory attenuation and lower external stimuli processing*
- *Complacency and loss of agency are two phenomena (possibly complementary) that could link OOTL performance problem and MW.*
- *OOTL performance problem remains a human-machine interface problem that could be influenced by MW through gradual external stimuli processing impairment.*
- *Aside from theoretical contribution, MW markers could be useful to study OOTL situations.*

2.1 Description of the current chapter

This chapter explores the literature seeking for similarities between MW and OOTL. The subsection 2.2 - Mind wandering to complete out-of-the-loop theories presents the theoretical evidences supporting a link between the two phenomena. We analyze aspects such as human-automation cooperation, complacency, feeling of agency and sensory attenuation, as well as the corresponding models. The following subsection 2.3 - Mind wandering markers to study out-of-the-loop introduces the experimental view and present how MW markers could allow studying OOTL performance problem more easily and in a wider range of environments compared to research on the subject nowadays. We provide a summary of the work presented in the thesis so far in the last subsection 2.4 - Thesis progress recap (chapter 2).

2.2 Mind wandering to complete out-of-the-loop theories

As pointed in the last chapter, OOTL situations continue to be the root of various safety issues. The complexity of such automated environments and the difficulty to reproduce OOTL performance problem in laboratories explain this lack of understanding. Nevertheless, a possible linked between MW and OOTL performance problem could help study OTL situations. OOTL research would benefit from such knowledge both experimentally and theoretically.

2.2.1 Interactions between mind wandering and system monitoring

MW seems closely related to boredom. Interestingly, Cummings et al. (2015) recently warned about a possible increase in boredom when integrating higher levels of automation. Moreover, researchers recently observed MW related to automation in automated systems. Casner and Schooler (2015) conducted a study where pilots were instructed to handle the approach (flight phase before landing) in a simulator by following beacons at altitudes given by the air traffic controller (ATC) officer. Probes inquired about their state of mind at predetermined times while the pilots had to report their position to the ATC officer. They observed that pilots were more prone to MW for higher levels of automation when they did not interact with the system. Instead of planning the flight ahead, the pilots were inclined to think about unrelated matters. Although multiple studies have shown that monitoring is stressful and requires high levels of cognitive resources (Helton & Warm, 2008; Warm, Dember, & Hancock, 1996; Warm et al., 2008), vigilance theories do not explain such an increase in MW. On the contrary, could MW theories give a rational explanation in a monitoring environment?

2.2.2 Complacency as a possible explanatory mechanism

Automation technology has changed the very nature of operators' work. Pilots are now required to monitor systems for possible failures. Monitoring tasks request a constant attention from the subject in order to detect seldom and unpredictable events over prolonged periods. This fundamental function is called the sustained attention (Manly, Robertson, Galloway, & Hawkins, 1999). Interestingly, several studies show that efficient sustained attention over hours cannot be achieved (e.g. Methot and Huitema, 1998). If literature suggests that time on task significantly decreases our ability to discriminate infrequent and unpredictable signals (Mackworth, 1948; Raja Parasuraman, 1979; Teichner, 1974; Warm, 1984), then attentional failures also encompass another reality when dealing with automation – that is, the complacency experienced by operators dealing with highly reliable automated systems (Cummings, 2004; Raja Parasuraman et al., 1993b).

Complacency is the adoption of a non-optimal information sampling behavior based on overtrust regarding system's capabilities due to a minimization of automation failure probability (Innes-Jones & Scandpower, 2012; Moray & Inagaki, 2000). Complacency could be a strategy to optimize performances Operators working with systems that fail once every ten million hours of use tend to underestimate the possibility of automation errors and overtrust the system (Amalberti, 2001; Raja Parasuraman & Wickens, 2008). Because they have the feeling that the system does not require them to work efficiently, they instinctively lower cognitive resources allocated to monitoring (Morrison, Cohen, & Gluckman, 1993; Thackray & Touchstone, 1989). The first empirical evidence was the study by Parasuraman et al. (1993b). They tested non-pilot participants on a flight simulation task made of 2D compensatory tracking, fuel management, and system monitoring. In the multiple-task condition, the participants performed the tracking and fuel management tasks manually while the automation handled the system monitoring. In the single-task condition, the participants only had to supervise the automation in the system-monitoring task. In both conditions, automation reliability was variable. The participants were responsible for detecting these failures, and they had to take over when there was a failure. Parasuraman et al. (1993b) observed that participants had a detection rate of over 70% when performing the engine status task manually (baseline condition). Their detection rate substantially declined when performing the task in the multitask condition. Interestingly, the effect was absent when they were in the single task condition, suggesting that the allocation of cognitive resources plays a role in the complacency effect (Bailey & Scerbo, 2007; Moray & Inagaki, 2000). Congruently, operators make fewer eye movements to the raw information sources when using automation than under manual control (Bagheri & Jamieson, 2004; Metzger & Parasuraman, 2001; C. Wickens, Dixon, Goh, & Hammer, 2005), reflecting an allocation of attention to other concurrent tasks. Furthermore, operators tend to less frequently visualize parameters in automation mode than under manual mode, thus blindly trusting the automation diagnosis (Lorenz, Di Nocera, Roettger, & Parasuraman, 2002; Manzey, Bahner, & Hüper, 2006). In a low probability signal context, Manly et al. (1999) used a sustained attention to response task (SART, a GO/NOGO task), to demonstrate a striking positive correlation between signal probability and detection rate.

These results indicate that complacency may relate to MW, as both complacency and MW divert cognitive resources away from the task. Supervising ultra-reliable systems seems to encourage a decrease in cognitive resources allocated to the monitoring task. In this context, resources saved by automation, which the operator should normally use to plan operations ahead, would instead maintain task-unrelated thoughts. Therefore, complacency might lead operators to free cognitive resources and reallocate them to unrelated thoughts. This assertion is supported by an observed increase in MW in a low probability signal environment (Berthié et al., 2015; Casner & Schooler, 2015; Galera et al., 2012) and as one has been on task for a longer period of time (McVay & Kane, 2009; Smallwood, Baracaia, Lowe, & Obonsawin, 2003; Smallwood et al., 2006; Teasdale et al., 1995; Thomson et al., 2014). Nevertheless, the exact direction of this link needs further data. MW could also occur prior to complacency and modify its emergence, for example by lowering the level of confidence needed for the operator to become complacent.

2.2.3 Out-of-the-loop problems and mind wandering: some similarities

2.2.3.1 Agency and the decoupling of human observer from the task

When designers integrate automation in systems, they often believe that it will only be a substitute to the human operator (i.e. substitution myth, see Woods and Tinapple, 1999). However, an important part of the literature has accumulated evidence against this view. Automation does not only simply perform tasks that humans previously handled. It also changes the complexity of the task and creates new issues, thus transforming the nature of human work. Operators give up their direct control over the system for a monitoring role in the supervisory control loop (Moray, 1986a; Sheridan, 1992). These changes are far from trivial – direct control involves manual functions including process planning, decision-making, selecting responses, and implementing strategies.

At the same time, passive information monitoring only requires operators to scan information sources and compare it to previously learned references. In an automated environment, operators can experience loss of manual skills (Baxter et al., 2012) and a feeling of distance from the system (Bainbridge, 1983). In particular, the sense of agency (i.e. feeling of control of effects in the environment) is also lower (Berberian et al., 2012; Obhi & Hall, 2011). Obhi and Hall (2011) observed that the ability to experience a feeling of shared agency during human-human interactions is at least partly inhibited when interacting with a machine. They suggested that unlike human partnership, the discrepancies that are observed during human-machine interactions could arise from the absence of action/observation matching abilities when interacting with automation. Their results were confirmed by Berberian et al. (2012) during a more complex obstacle avoidance task. Participants performed the task with different levels of automation: Full Operator Control (FOC), Automatic Decision - Operator Implementation and Engagement (AD-OIE), Automatic Decision and Implementation - Operator Engagement (ADI-OE) and Full Automatic Control (FAC). They used explicit measures (reports of feeling of agency) and implicit measures (intentional binding, i.e. the perceived interval between voluntarily action and outcome seems shortened) to log agency. Both measures were strongly correlated. At the same time, both measures decreased when automation levels increased, showing a convincing link between agency and system autonomy. Such results are extended by the correlation between the feelings of agency and responsibility (Frith, 2014). The outcome is that automation, by replacing human operators from their actions, also strip them from their sense of responsibility for any outcome, despite them being explicitly designated as supervisors.

Interestingly, MW induces the same phenomenon of decoupling from the task. The operators' attention during MW is shifted from the immediate task toward unrelated concerns (Schooler et al., 2011). A large-scale study conducted by Killingsworth and Gilbert (2010) used smartphone applications to measure MW rates in the daily life. People reported MW for 46.9% of probes. More importantly, MW occurred in at least 30% of the probes for all activities (except making love). In other words, both MW and OOTL performance problem start with a decoupling from the task. Moreover, both involve equally concerning threats for safety in critical systems. For example, MW leads operators to forget to report as instructed (Casner & Schooler, 2015) and slows their adaptation to original tasks (Mooneyham & Schooler, 2013), whereas OOTL performance problem see operators less responsive (Endsley & Kiris, 1995) and with low failure detection rates (Parasuraman & Riley, 1997).

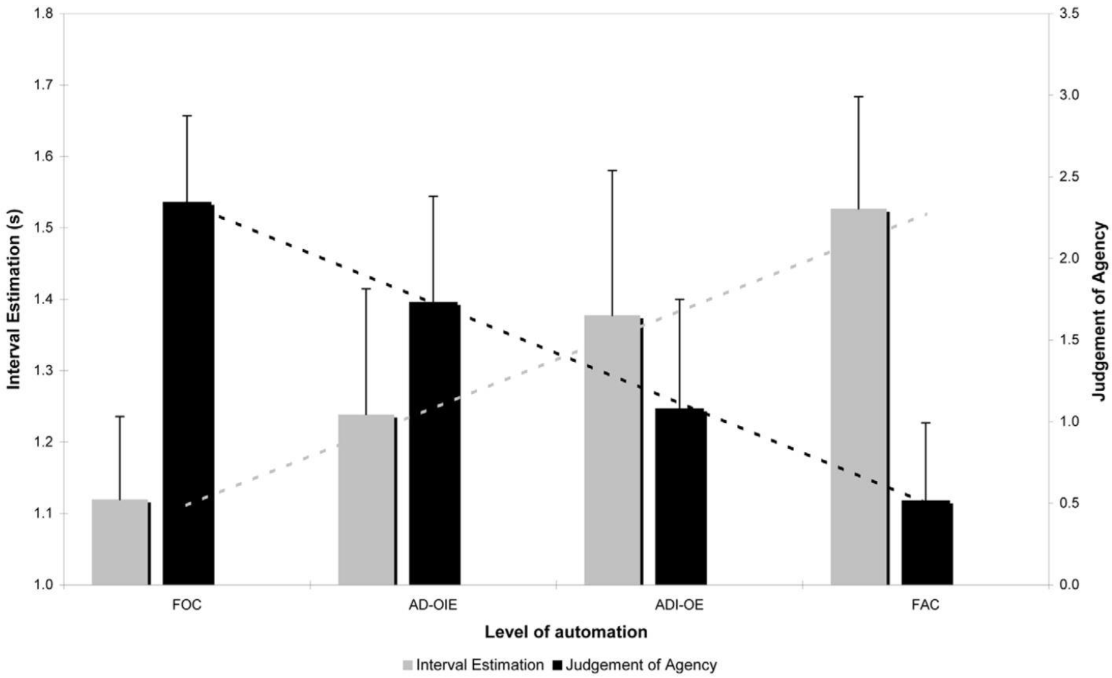


Figure 7. Implicit and explicit measures of agency for each level of automation, from (Berberian et al., 2012).

Automation increases from left to right (FOC < AD-OIE < ADI-OE < FAC). "Interval estimation" measured the effect of intentional binding, while the "judgment of agency" required participants to report explicitly the feeling of agency they had.

2.2.3.2 Sensory attenuation problem

OOTL performance problems often involve problems related to perception and monitoring (Federal Aviation Authority, 1972; Gerbert & Kemmler, 1986; Mosier et al., 1994).

Detection and monitoring of external important information are critical for safety. Failure at these levels influences the whole cognition. Several studies have shown a longer reaction time and lower detection rate following long automated periods. Endsley and Rodgers (1998) found that ATC officers showed poor performance in detecting conflicts when they were passively monitoring the traffic. Willems and Truitt (1999) exposed that, in the same condition, ATC officers were slower to answer questions regarding traffic awareness and they recalled less information as traffic load increased. Interestingly, Caspar et al. (2016)

highlighted a lower sensitivity to external stimuli when operators felt less in control. In their experiment, participants as agents could take money to the “victim” (group 1) or deliver electric shocks to the “victim” to earn money (group 2). Both actions caused a tone to occur after a random delay (between 200 and 800ms). In the free-choice condition, they could freely choose to do the action or not do anything. In the coercive condition, an experimenter stood next to the participants and ordered them to either do the action or not. Caspar et al. (2016) reported lower N1 evoked potential (associated with auditory perception) when people obeyed the experimenter order than when they freely chose to act. Completed by the negative correlation between agency and automation levels (Berberian et al., 2012; Obhi & Hall, 2011), we can conclude that disengagement from the task due to a loss of agency lower external stimuli sensitivity.

Similarly, MW involves a similar reduction in perceptual awareness of the task-relevant environment that lowers the subjects’ ability to detect signals (Blanchard, Bixler, Joyce, & D’Mello, 2014; He et al., 2011; Merat & Jamson, 2008), particularly when dealing with automation (Thackray & Touchstone, 1989). O’Connell and colleagues (2009) used a SART to demonstrate that alpha waves amplitude were more important during MW episodes in occipital scalp sites. Interestingly, tasks analyzing selective attention, where one has to inhibit attention to parts of the environment in order to efficiently perform a task, suggest the involvement of alpha activity as a sensory suppression mechanism (Foxe, Simpson, & Ahlfors, 1998; Foxe & Snyder, 2011), or similarly as reflecting pulsed-inhibition of ongoing cortical processing (Mathewson, 2011). Recently, both electroencephalography (EEG) and magnetic resonance imagery (MRI) have found alpha wave increases in supposedly deactivated regions by manipulating both the level of internally directed attention and the level of self-generated thought (Benedek et al., 2016; Benedek, Schickel, Jauk, Fink, & Neubauer, 2014), thus supporting the idea of alpha waves being a marker of inhibition. Taken together, these findings rule out the possibility that these effects could rely on sensory (bottom-up) processing of the cue and they suggest an endogenous inhibitory effect (top-down). During this time, the system and environment may change, hence increasing risks to the operator of having an out-of-date model of the situation.

2.2.3.3 Understanding perceived information

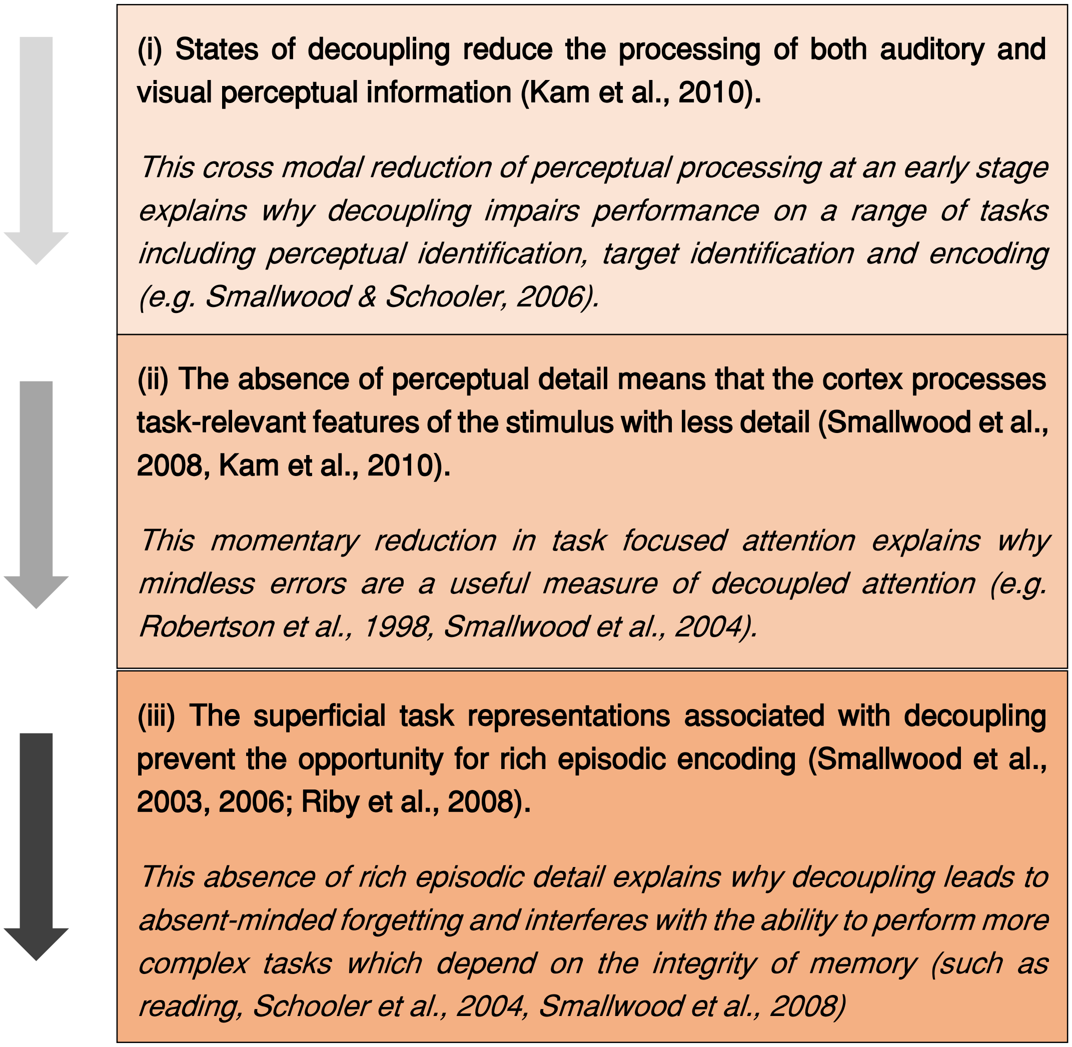
In addition to perception, MW and OOTL performance problem may thwart comprehension. When automation fails or behaves abnormally, operators are required to handle the situation alone. These cases have been well documented in various domains, most notably flight deck and operating room automation (e.g. Degani and Heymann, 2000; Sarter and Woods, 1995a, 1995b). Several fatal crashes and other incidents have been attributed to problems in the flight crew–automation interface (see for example Federal Aviation Authority, 1995). Sarter et al. (1997) referred to this as “automation surprises”, a point where the system behaves differently from what the operator expects causing a longer reaction time and a lower accuracy. In laboratories, Wickens and Kessel (1977) demonstrated that operators removed from the system control show slower reactions and poor response accuracy. Carmody and Gluckman (1993) highlighted that for complex task models, higher level of automation induced heavy losses of understanding. Taken together, these findings demonstrate that automation failures lead to a critical situation where the operator is OOTL and cannot initiate proper recovery actions.

Interestingly, studies reported similar understanding issues for MW. The subjects experience unconscious working memory transfer from the task toward unrelated thoughts. Participants reading a text exhibited comprehension drops (Schad et al., 2012; Smallwood, McSpadden, & Schooler, 2008) and less reactions to text difficulties (Feng, D’Mello, & Graesser, 2013) during MW. Brain studies have shown activity uncorrelated to the environment during the same periods (Konishi, McLaren, Engen, & Smallwood, 2015). A decrement in external stimuli processing is particularly true within monotonous and uninteresting environments (Mosier et al., 1994). In the operational context, studies point to MW as a possible cause of many driving accidents (Galera et al., 2012), plane crashes (Casner & Schooler, 2014), and medical errors (van Charante, Cook, Woods, Yue, & Howie, 1993), maybe due to a lack of a proper model of the situation in critical moments.

2.2.3.4 The cascading model of inattention

Smallwood et al. (2011; 2007) developed the cascading model of inattention (see Figure 8). They suggest that the superficial deficit in information processing induced by MW would cascade, impair a deeper level of understanding and negatively affect the construction of an accurate situation model. The poor-quality model would then decrease the ability of the environment to hold the operator's attention, which in turn would decrease the quality of the model, and so on.

They paralleled their model with the three levels of reading generally used in models: lexical, propositional and situational (Balota, Yap, & Cortese, 2006). Lexical level deals with the meaning of the letters and words, which would to perception decoupling for MW. Next, words are organized into propositions to form sentences, which constitute propositional level. For operators, this level could be the organization of the current external stimuli. Finally, the situational level gathers the extended narrative and provides a context for the reader. This level makes an extensive use of both working memory and long-term memory. Smallwood et al. (2007) proposed that MW could block the information flow from lower levels to higher ones. This would explain why people experiencing MW can still perform well-known tasks (which rely heavily on long-term memory) without being able to handle seldom events (which would require an up-to-date lexical level).



(i) States of decoupling reduce the processing of both auditory and visual perceptual information (Kam et al., 2010).

This cross modal reduction of perceptual processing at an early stage explains why decoupling impairs performance on a range of tasks including perceptual identification, target identification and encoding (e.g. Smallwood & Schooler, 2006).

(ii) The absence of perceptual detail means that the cortex processes task-relevant features of the stimulus with less detail (Smallwood et al., 2008, Kam et al., 2010).

This momentary reduction in task focused attention explains why mindless errors are a useful measure of decoupled attention (e.g. Robertson et al., 1998, Smallwood et al., 2004).

(iii) The superficial task representations associated with decoupling prevent the opportunity for rich episodic encoding (Smallwood et al., 2003, 2006; Riby et al., 2008).

This absence of rich episodic detail explains why decoupling leads to absent-minded forgetting and interferes with the ability to perform more complex tasks which depend on the integrity of memory (such as reading, Schooler et al., 2004, Smallwood et al., 2008)

Figure 8. The cascade model of inattention, from (Smallwood, 2011).

2.2.3.5 The exact nature of the link mind wandering – out-of-the-loop

After comparing MW and OOTL on multiple aspects, a question arises: how can they be linked? Casner and Schooler (2014) highlighted the blurry situation of pilots left with spare time and no guidance about how to actively monitor the automation. This spare time could encourage the operators to think about unrelated concerns and this would drive them away from important matters, such as their current position or the mode of the system. Without knowledge of the situation, OOTL risk rises and threatens operations.

A way MW and OOTL could interact could be through working memory. When experiencing MW, task-unrelated thoughts flood working memory (McVay & Kane, 2010). Depending on the individual's working memory capacity, MW thoughts might fully occupy working memory capacity, preventing new resources from allocation to the ongoing task. As the observed vigilance decrement will lower available working memory, operators' capacity may be full even more quickly within highly automated environments. At the same time, complacency could drive the operator to lower the amount of working memory capacity allocated to the task. Unrelated thoughts would promptly use the working memory capacity freed by complacency. Various results examining the relations between MW/working memory and OOTL/working memory support our framework. Examining the trial-by-trial co-occurrence of MW and performance declines during a working memory span task, Schooler et al. (2014) found that MW precedes poor performance. Our framework states that when filled with task-unrelated thoughts, working memory capacity cannot cope with new cognitive needs. Then operators experience a drop in performance. Similarly, maintaining a good situation awareness—closely linked to whether one is OOTL (Kaber, Onal, & Endsley, 2000)—requires working memory capacity through the active manipulation and use of information (Durso et al., 1999). When MW uses executive resources, the individual will see her situation awareness decrease, leading to a higher risk of being OOTL.

Nevertheless, the link between MW and OOTL remains unclear. Characterizing its features could help to both better define OOTL and understand some of the situations that have led to tragic accidents. To achieve this goal, MW markers could help study OOTL situations. We highlight some possible directions for research in the following sections.

2.3 Mind wandering markers to study out-of-the-loop

OOTL performance problem can arise from many different aspects of cognition and human-system interactions. If a link was assessed between such issues and MW, recently unveiled physiological markers of MW could help track OOTL. Contrary to many concepts in human factors, research on MW has already identified its physiological signature, while the experience of MW is easier to report than hypothetical vigilance levels or debated situation awareness (see subsection 1.2.3 - Out-of-the-loop research needs validated psychological models). Unfortunately, the different dimensions of MW make it difficult to characterize the phenomenon itself. Nevertheless, MW possess an objective reality and has a quantified impact on behavioral and physiological signal. This may allow OOTL research to discover new aspects of the performance problem using real-time internal state monitoring in a near future.

2.3.1 Self-report of MW are more robust than vigilance reports

MW markers are sorted using the triangulation classification among self-reports, physiological, and behavioral measures (Smallwood & Schooler, 2015). Self-reports remains the ground to conclude that a participant is MW or not. Most experiments use probes to determine where is directed participants' attention (Braboszcz & Delorme, 2011; Feng et al., 2013; Gilbert, Dumontheil, Simons, Frith, & Burgess, 2007; Smallwood et al., 2004; Uzzaman & Joordens, 2011). Although subjective reports have their limitation (Overgaard & Fazekas, 2016; Tsuchiya, Frässle, Wilke, & Lamme, 2016), they demonstrated results more robust than any other measure at date. MW reports have demonstrated a high correlation with neurophysiological measures (Cowley, 2013; Smallwood, McSpadden, et al., 2008; Smilek, Carriere, & Cheyne, 2010a). This robustness could prove to be useful when studying OOTL situations. Whereas it may be challenging for someone to report their vigilance level (partly because the term remains blurry), reporting the content of the thoughts could be useful in evaluating the risk of OOTL performance problem.

However, self-report markers of MW bring us back to the problem of disrupting measures in operational environments. Nevertheless, other markers have demonstrated promising results and yield satisfying detection rates in the near future.

2.3.2 Behavioral markers are fundamental for OOTL performance problem

Behavioral markers of MW come in a wide variety. Within this category, reaction time measurements take an important place. Multiple studies highlighted the progressively faster reaction time during MW during repetitive high-frequency tasks (such as the Sustained Attention to Response Task, SART), linking it to impulsive behavior (Cheyne et al., 2011; Smallwood et al., 2003, 2011, 2004, however see Dang, Figueroa, & Helton, 2018 for the limitations of the SART). Reaction time allows us to track the subject's attention without disturbing them. It contains much information, such as omissions (subject does not react to a stimulus although they were instructed to, see Bastian and Sackur, 2013) and anticipations (reaction lower than 100ms, see Hu et al., 2012). Cheyne et al. (2009) proved the robustness of the coefficient of variability (on a given interval, mean reaction time divided by its variability) to study MW (Bastian & Sackur, 2013; Esterman et al., 2013).

Parallel to those results, subject accuracy is extensively used, whether it is during trial to trial tasks (Braem, Coenen, Bombeke, van Bochove, & Notebaert, 2015; Durantin, Dehais, & Delorme, 2015; Konishi et al., 2015) or during continuous monitoring, such as in a car simulators (He et al., 2011; Yanko & Spalek, 2014). Cowley (2013) used a simulated driving route in mixed world (urban and rural) to study how drivers behaved in normal conditions. Driving accuracy was measured throughout the task. She also probed participants 5 times and asked them to report if their thought content and awareness of their own thoughts. Interestingly, the probes where participants reported the highest rate of MW were also the probes with the highest number of lane deviation and seconds speeding.

Overall, behavioral markers can highlight performance decrements induced by MW in many different tasks. They can also be used for OOTL characterization; for example, reaction time to take manual control over a system (de Winter, Happee, Martens, & Stanton, 2014) or accuracy to detect automation failures (Metzger & Parasuraman, 2001). Unfortunately, these measures are also of limited use outside the laboratory. Reaction time is useful when the participants have to perform actions regularly, whereas OOTL situations offer environments where actions are seldom required. Considering that accuracy measures the participants' shift from the goal, it is also limited to situations where the operator is already experiencing OOTL performance problem. A situation that is hardly acceptable in many automated environments where any error can cost much (nuclear plants, airplanes, military operations…).

In order to overcome the problems of disruption and error consequences, physiological measures could be useful to detect the dynamics of the problem.

2.3.3 Oculometric markers promise much

Oculometric signal allow us to derive different markers for potential use in detecting attentional lapses occurring during both MW and OOTL problems (for a description of the measurement tool, see subsection 4.4.1 – Measuring attention through oculometry).

The pupillometric signal is strongly modified by the content of thoughts. Whether the focused on object is interesting or boring, physical or conceptual, exterior to our mind or just and idea, all those parameters will influence pupil size. Researchers demonstrated that pupil size changes when participants experience MW (Lowenstein & Loewenfeld, 1962; Mittner et al., 2014; Yoss, Moyer, & Hollenhorst, 1970). More generally, the literature on vigilance already linked a lower pupil baseline to periods of lower sensibility to external stimuli (McIntire et al., 2014; Nishiyama, Tanida, Kusumi, & Hirata, 2007). It should be noted that some studies highlighted a higher pupil baseline during task-unrelated MW (Franklin, Broadway, Mrazek, Smallwood, & Schooler, 2013; Smallwood et al., 2011).

Nevertheless, Konishi et al. (2017a) recently observed an inverse U-curve relationship between the pupil diameter and performance. They used a paradigm of continuously switching between a 0-back task and a 1-back task (see Figure 9). Non-target stimuli were pairs of geometrical shapes (square, triangle or disk) presented on the left and on the right of a vertical line. Pairs changed automatically while participants waited for the target screen. At the end of 0-back task blocks, the target screen displayed the geometrical shapes, the line in the middle and the target shape on the line. Participants had to indicate whether the target shape in the middle was the same as the shape on the right or on the left. In the 1-back task, the target screen consisted of two question marks replacing the shapes. Participants had to indicate whether the target shape in the middle was present on the left or on the right during the last trial.

Results linked a smaller pupil diameter with a decrease in performances and MW episodes as internally directed cognition. In contrast, a larger pupil diameter was correlated with external distractions (e.g., conversation, noise, or itching) and was accompanied with a decrease in performance. These results corroborate the experiment of Unsworth and Robison (2016; however this experiment could be biased by the rationalization hypothesis, see Head & Helton, 2016). This behavior is correlated with norepinephrine activity in the locus coerulus (i.e. the LCNE system) and is thought to be linked with the role of surprise (Aston-Jones & Cohen, 2005; Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Jepma & Nieuwenhuis, 2011).

Apart from pupillometry, MW also impacts gaze position (Grandchamp, Braboszcz, & Delorme, 2014), eye movement pattern (He et al., 2011; Smilek et al., 2010a), blink count (Uzzaman & Joordens, 2011), and saccades (Bixler & D'Mello, 2014). Reading tasks highlighted differences in on and off-text fixations (Bixler & D'Mello, 2015; Reichle, Reineberg, & Schooler, 2010), reading speed (Feng et al., 2013; Franklin, Smallwood, & Schooler, 2011), especially related to text difficulty (Schad et al., 2012), within-word fixations, and reading regression (or going back a few words if one did not understand the sentence; Uzzaman & Joordens, 2011). Even though oculometry and MW were studied mainly during laboratory tasks, other research has been conducted in more ecological conditions. He et al. (2011) used a driving simulator to exhibit the narrowing influence of MW on the visual field of drivers.

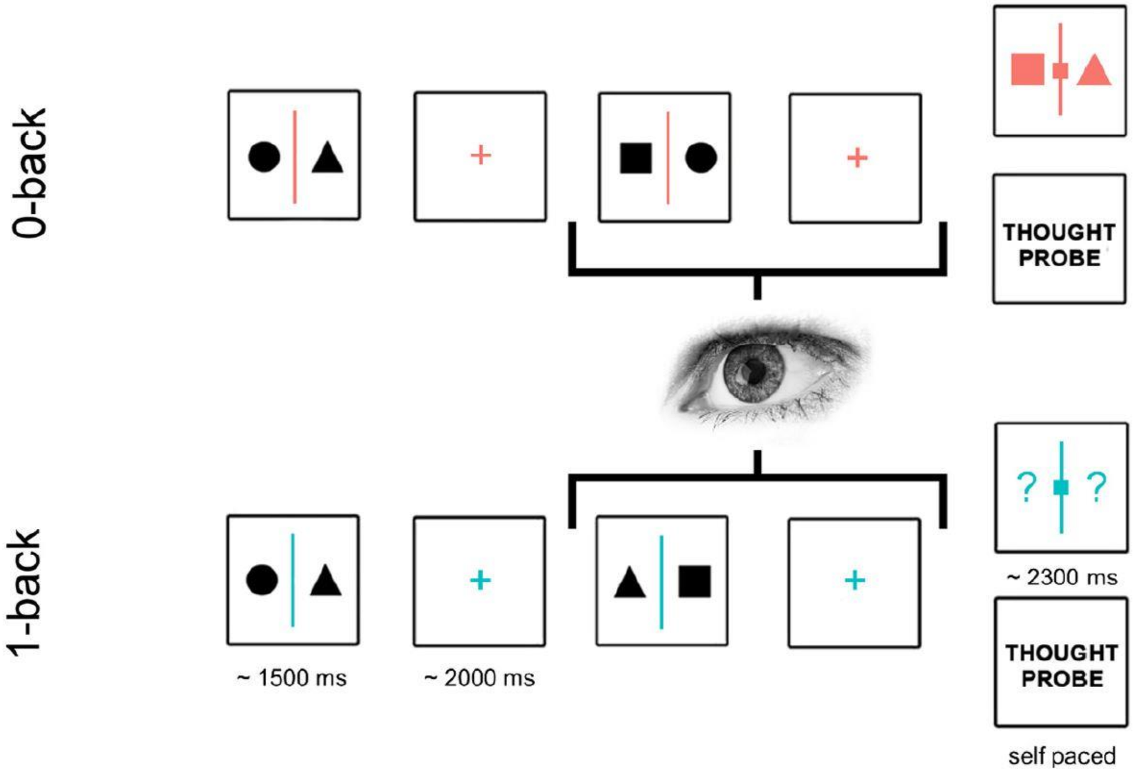


Figure 9. The paradigm used in (Konishi, Brown, Battaglini, & Smallwood, 2017b).

In both conditions, after a certain number of Non-Targets (NTs) participants were faced with a target decision and a thought probe. In the 0-back condition, the decision is based on the presently perceived stimulus (is the square on the left or the right?); the NTs are thus irrelevant to the task, allowing for long periods in the 0-back condition when attention is unconstrained by the ongoing task. Conversely, in the 1-back condition the target decision is based on the previously attended NT (was the square on the left or the right?). Under these conditions, participants must maintain external attention on the NTs in order to perform accurately in the task. We selected a time window of ~3.5 s, corresponding to the NT and fixation cross immediately preceding a target or a thought probe, to analyze the effects of average pupil size on behavior and internal reports.

Figure courtesy of Mahiko Konishi and Elsevier.

Already driven by the promises of oculometry, some projects investigated the use of oculometry within operational conditions. Peysakhovich et al. (2016) demonstrated the usability of oculometric signal during real flights despite changes in light exposition. On the automobile field, the adaptIVe project (funded by the European Commission) investigated the influence of driving conditions on the ability to take back manual control for autonomous vehicles. They used a driving simulator while recording oculometric signal (Louw, Kountouriotis, Carsten, & Merat, 2015; Louw & Merat, 2017).

The possibilities offered by oculometry (non-invasive measures, sensibility to attentional, robustness in operational conditions) already made it an unavoidable measure in the quest to understand OOTL performance problem. However, much is still to discover in order to fully understand the variations of the oculometric signal.

2.3.4 Electrocardiogram and skin conductance markers

Heart rate and skin conductance have been used for a long time to detect periods of boredom (Smith, 1981) and they continue to be part of the latest developments. Their robustness allowed Pham and Wang (2015) to create a classifier which accurately identified lapses of attention during learning. They designed the AttentiveLearner, an intelligent mobile learning system. The system is able to implicitly extract heart rate by analyzing fingertip transparency changes during learning and infer the attention of the learner. They achieved an accuracy of 71.22% (*kappa* = 0.22) on students learning by watching videos on the smartphone.

Heart rate and skin conductance have also shown promising results when used to determine pilots' arousal in real-time (Boucsein, Haarmann, & Schaefer, 2007). Researchers found the signature of the effects of boredom on amplitude and variability on both markers. Interestingly, Smallwood et al. (2004) reported similar effects when studying MW.

Since MW may favor OOTL performance problem, heart rate and skin conductance could also be useful to study OOTL. Regrettably, it is possible that MW influence on the signal would be lost within operational environment because of stress, movement, and temperature influence. Consequently, more studies are required in this field.

2.3.5 Neuronal markers

Neural markers of attention lapses betray MW episodes and its dynamics (for a description of the physical principle, see 4.5.1 – Electroencephalography). Researchers have mostly used EEG or functional MRI (fMRI) to study those markers, with the notable exception of the HbO2 concentration using functional near-infrared spectroscopy (fNIRS; Durantin et al., 2015). EEG activity has a high temporal resolution and a relatively low cost (Luck, 2014), allowing its extensive use for MW research. MW influence on brain waves was suggested by EEG data with an accent on the alpha band (8–14Hz), although the direction of the influence is still debated (Braboszcz & Delorme, 2011; O'Connell et al., 2009), and event related potentials (ERPs). Sensory attenuation has been observed on the visual component P1 and the auditory component N1 (Kam et al., 2011), while the lack of stimulus processing was shown using P3 (Schooler et al., 2011), N400 (O'Connell et al., 2009) and fERN (Kam et al., 2012).

By contrast, fMRI has a fine spatial resolution but a poor temporal resolution. It highlights neuronal networks involved in MW in order to build a map of the wandering mind. Several studies have highlighted brain regions differently involved in the phenomenon, such as the Default Mode Network (Mason et al., 2007; van den Heuvel & Hulshoff Pol, 2010), the executive network (Christoff et al., 2009; Christoff, Ream, & Gabrieli, 2004) and the task-positive network (Mittner et al., 2014). Compared to other markers, neural markers of MW could not only answer the question of "when" OOTL occurs, but also the "why" and "how". This could provide the OOTL performance problem with the physiological definition that it lacks (or at least for one aspect of it).

MW research has identified an important set of markers to detect its occurrence. Due to the proximity with psychophysiological measures recently used in automation studies, these markers may also prove to be useful for OOTL research. However, many unknowns remain regarding some aspects of both MW and OOTL performance problem.

2.4 Thesis progress recap (chapter 2)

In this chapter, we compared the results of multiple studies, implicitly underlying some degree of link between MW and OOTL performance problem. Completing chapter 1, our theoretical review allowed us to explain why such work stem from the modern literature, while the possible benefits advocate for further studies. In the next chapter, we will switch to ecological considerations and explore the perception of pilots regarding problems associated to human-automation cooperation (see Figure 10).

Figure 10. Abstract representation of the reading of this thesis.

From (Zgur_, 2013).

3 STUDY OF PILOTS' PERCEPTION OF AUTOMATION AND PROBLEMS ASSOCIATED

- *We explored with a questionnaire pilots' use of automated aids, their **perception of autopilot usability**, problems associated to **human-automation cooperation** and **task-unrelated MW rates** in daily life.*

- *Pilots reported **favorably perceiving autopilot** and rating it as "Usually" usable on average, a rate in line with the high frequency of use during monotonous flight phases.*

- *Scores on **perceived usability** and **task-unrelated MW** seemed to influence scores on vigilance problems in the cockpit.*

- *When aggregating answers by a non-pilot population on the task-unrelated MW questionnaire, we did **not find any difference** in reported rates between pilots and non-pilots.*

3.1 Description of the current chapter

Before immersing ourselves in experimental tasks in laboratory, we wanted to assess that the operational need for our work existed. We questioned how operators perceive automated aids and if they were aware of attentional issues when supervising automation. We chose to use a questionnaire to collect pilots' feedback, as a population interacting much with automation and for several decades. This chapter details the genesis of the questionnaire in the subsection 3.3 - Questionnaire creation. We then examine the results in both individual- and part-level analysis in the subsection 3.4 - Results and discuss implications after in the subsection 3.5 - Discussions on the result. We provide a summary of the work presented in the thesis so far in the last subsection 3.6 - Thesis progress recap (chapter 3).

3.2 Gather users' feeling

3.2.1 Pilots' feeling towards automation

As for any major change in technologies, the transition from traditional cockpits to the so-called "glass cockpits" necessitated a phase of adaptation. It also encountered some resistance; indeed, pilot were initially reluctant to let a system handle their lives, and lose their sense of control (or agency, Haggard, 2017). Many pilots' unions refused to transition to glass cockpits for as long as their company allowed it (for a complete discussion on the subject, see Amalberti, 1999). It must be noted that the transition is very demanding, and requires pilots to completely modify their work. English is the only language used by automated aids, contrary to personal written notes. Pilots new to the possibilities offered by automation may rely too much on aids like FMS (Flight Management System) and its automated modes (LNAV, HNAV, auto-thrust…) to control many aspect of the flight (Bureau d'Enquête et d'Analyse, 1992).

Nevertheless, decades later, we can say that automation greatly helped reach an unprecedented level of safety for aviation. This may have helped pilot accepting automation as a sort of new crewmember. Moreover, many pilots trained from start with some automation. However, to our knowledge, no recent data exist of pilots' acceptability of their autopilot, decades after their introduction. On the other side, the 'ironies of automation' still exists and failures regularly occur in the pilot - system relation (Bainbridge, 1983; Baxter et al., 2012). Among them, target detection is regularly at fault and greatly influences safety. Several studies have addressed the problem in laboratories (Durantin, 2015; See, Howe, Warm, & Dember, 1995; Thackray, Bailey, & Touchstone, 1975; Thackray & Touchstone, 1989; Warm, 1984). However, only seldom studies investigated how pilots reflected on this problem (see 1.2.2 - Automation moves operators out of the control loop). Gerbert and Kemmler (1986) studied German aviators' anonymous responses to questionnaires about automation-related incidents. Pilots reported failures of vigilance as the largest contributor to human error, while associating these errors with background conditions such as nervousness, high tension or oversaturation of information.

To our knowledge, no study has investigated pilots' perception of MW. Casner and Schooler (2014, 2015) observed task-unrelated MW rates for pilots in simulators similar to other studies for non-pilot people. However, they proposed the possibility that in real conditions, pilots may be able to control the occurrence of MW. This hypothesis was supported by their results: when pilots did not have pauses to mind wander, they would experience MW with negative consequences on performance.

3.2.2 Automated modes

The Federal Aviation Administration (FAA) describes two components of the automated avionics system: the autopilot is the set of servo actuators and circuits that actually do the control movement, and the Flight Director handles calculations required for higher automated modes (Federal Aviation Administration, 2009). It is possible to use the Flight Director without the autopilot; in that case, the Flight Director only issues present results of computation in the form of command bar cues. When the Flight Director is used with the autopilot, the first issues commands which the latter simply follows. In order to simplify this thesis, we adopt the common denomination "autopilot" for all parts related to automated commands in the aircraft.

The autopilot is a state machine, i.e. a deterministic successions of modes (or machine internal states) linked by transitions. Modes correspond to behaviors of the system. Pilots can select different modes to let the automation handle tasks ranging from very simple ones (e.g. "keep this altitude") to more complex ones (e.g. go-around when the landing is aborted). Transitions between modes are based on a "if… then…" model, e.g. "if the system is in mode 1 and condition A is true, then the system goes on mode 2" (Degani & Heymann, 2000b). These conditions can become true either from pilot action or from situation events. For example, we can have "if the autopilot is in CRUISE mode and the pilot pull the button X, then the autopilot enters DESCENT mode".

On the contrary, situation-triggered transitions are hard to anticipate. Nowadays autopilot support a proliferation of modes and transitions that hardly support the "explainability" of the system for pilots (Sarter & Woods, 1995b; Swartout, Paris, & Moore, 1991). Multiple studies described several examples of transitions not even documented in the training manual which resulted in dramatic consequences (2000b; Sheridan & Parasuraman, 2005). For example, pilots of an A300 were trying to override the go-around mode they had inadvertently engaged by acting on the joystick; unfortunately, in this this particular aircraft, the go-around mode could not be disconnected by manual inputs and had to manually deselected. This resulted in a power struggle between pilot and autopilot that eventually led to the crash (Sheridan & Parasuraman, 2005).

There is no consensus on the literature on how to classify automated modes, partly because modes differ from one aircraft manufacturer to the other. After reviewing the literature, we chose to classify modes into three different categories:

- **Flight envelope modes**: modes ensuring that the aircraft remains in its flight envelope (set of flight and aircraft parameters where it can fly). As the minimum possible automation, pilots cannot disconnect nor override these modes on many aircrafts. We chose to label "manual mode" moments when pilots fly only with envelope protection modes. Examples are protection in pitch, overspeed protection or protection in angle of attack.

- **Selected modes**: pilot select the flight parameter (speed, altitude, heading…) and its value for the autopilot to match. Autopilot will reach the value and maintain it by modifying aircraft parameters (engine, flaps, aileron…). These modes allow pilots to adapt to the situation and remain in control of aircraft parameters changes. Examples are heading capture (HDG), speed capture (A/T) or altitude capture (ALT).
- **Managed modes**: autopilot handles parameters in order to follow the flight path entered by the pilot in the Flight Management Console. Autopilot will modify flight parameters along the flight to follow the flight path. Examples are lateral navigation (LNAV), go around (GA TRK), approach (APP).

3.2.3 Wiener investigation of pilots' perception of B757

Wiener (1989) investigated how pilots perceived the automated features of the Boeing 757. He sent two versions of a questionnaire (in 1986 and 1987) including 5-points Likert scales on the attitude toward automation, open ended questions on various topics and biographical questions on aircraft experience.

The B757, equipped with state-of-the-art automated aids, represented a change in cockpit sophistication: integrated systems, advanced autopilot and autothrottle and the Engine Indicating and Crew Alerting System (EICAS), which role was to monitor other systems to detect any failure. Apart from the interest in knowing how pilots received these new features, this aircraft was chosen partly because of its shorter stage lengths, and hence greater experience of crews operating in terminal areas (take-off, approach, and landing). Finally, 116 models of the B757 were already in use when the questionnaire was sent (Boeing, 2011). This investigation yielded very interesting results and tapped pilots' perception of the new automated aids. The distribution of answers for many questions was almost symmetrically distributed between agreement and disagreement (e.g. "I look forward to more automation" or "Automation does not reduce workload"). As stated by Wiener (1989), “in general the pilots are enthusiastic about flying an advanced technology aircraft, but they express mixed feelings about the impact of automation on workload, crew errors, and ability to manage the flight.”

Almost 30 years later, how has perception of the automation evolved? Do pilots still identify problematic areas for pilot-system interactions?

3.2.4 Questionnaire objective

Our aim is firstly to record pilots’ perception of automation and the human-autopilot problems that might arise in order to assess operational needs. Moreover, we wanted to extend the range of the Mind Wandering Questionnaire (see subsection 3.3.2.5 - Mind wandering questionnaire (Part 5) and Mrazek et al., 2013) already used for children, teenagers and adults to another population (namely pilots). In accordance with this view, results will be mainly descriptive and statistics should be regarded more as pointing tendencies than hard evidences. We measured experience, autopilot acceptability, human-autopilot problems and MW propensity in the daily life.

3.3 Questionnaire creation

3.3.1 Creation process

We chose an online questionnaire to investigate our hypothesis. Online questionnaires are suitable to gather the feedback of a population generally hard to reach (changing hours, worldwide distribution, difficulty to reach) while minimizing resources (no need to travel, low-cost, easy to setup). We used Google Forms, as the service is free, allows multiple types of items (questions) and directly produces figures online to quickly grasp data distribution. All answers were anonymous without us being able to identify respondents.

We firstly conducted a brainstorming with aeronautics experts to gather items in French for each hypothesis to be tested. After narrowing possible items panel to 15 per part (except for part 5, which was based on an already existing questionnaire, see Mrazek et al., 2013), we asked 3 pilots (one commercial pilot, one army pilot and one amateur pilot) to review our items. We performed another brainstorming to eliminate, add or fuse items according to their feedback. Finally, we asked two independent translators to translate items in English and aggregated the resulting translations.

We also conducted validation of population items for the non-pilot questionnaire. This questionnaire was only in French, so we reused the same translation for items of the part 5 MWQ (common with the pilot questionnaire).

3.3.2 Pilot questionnaire

We here describe each item and give the corresponding rationales. The whole questionnaire is available in appendix A and appendix B (English and French versions). The parts are demographic items, pilots' experience of the autopilot, pilots' perception of the autopilot, encountered problems when operating the autopilot and pilots' MW propensity in daily life.

For items proposing different predefined answers, we chose to use a 7-points Likert scale. This type of Likert scale has already been extensively used and validated in psychometrics (Cox & Isham, 1980; Miller, 1956; Nunnally & Bernstein, 1994; Preston & Colman, 2000).

3.3.2.1 Demographic items (Part 1)

The first part of the questionnaire gathered demographic information. This was primarily done to see how our respondent poll was representative of the pilot population. We enquired about respondents' age and gender.

3.3.2.2 Experience of the autopilot (Part 2)

The present section aimed at understanding how much flight experience pilots had, and how do they manage their flight. Pilots had to report information such as their hours flying time, or their number of years as a pilot. Moreover, we asked pilots to report the aircraft integrating automation they considered having the most experienced with. In the parts 3 and 4, they would have to refer to this aircraft to answer this question. Moreover, we enquired about how they used automation, i.e. in which parts of the flight they used it and how often were they required to use autopilot by their airline or the procedures. Depending on the item, items could be either open for writing questions or propose the categories *never, rarely, sometimes, occasionally, frequently, usually* or *always*. Only the item "Time flown on manual mode (percentage of total flight time...)" proposed to answer with the following categories: *Below 10% of flight time*, *10-30% of flight time*, *31-50% of flight time*, *51-70% of flight time*, *71-90% of flight time* or *More than 90% of flight time*. Finally, we asked pilots to report how they divided their time in the cockpit, e.g. what were their non-flight related activities.

3.3.2.3 Autopilot usability (Part 3)

The following section aimed at understand pilots' perception of the autopilot. To our knowledge, there was no questionnaire suitable to evaluate usability of a system that users are already using. We used as a basis the USE questionnaire (Lund, 2001). This questionnaire covers several dimensions: Usefulness, Satisfaction and Ease of use. Moreover, it already uses a 7-points Likert scale. We discarded items that were not suitable for our environment ("I would recommend it to a friend") and created others during the brainstorming sessions. We asked participants to refer to their experience on the aircraft indicated in the first part ("Aircraft with the most experience among the previously given"). Responses could be *Completely disagree*, *Mostly disagree*, *Somewhat disagree*, *Neutral*, *Somewhat agree*, *Mostly agree* and *Completely agree*.

3.3.2.4 Human-autopilot problems (Part 4)

The following section aimed at understanding how often pilots experience various problems when operating autopilot. We wanted to obtain an idea of the frequency of some OOTL performance problems (see subsection 1.2.2 - Automation moves operators out of the control loop). To our knowledge, there is no existing prototype of questionnaire measuring human-autopilot problems in the cockpit. We measured aspects of monitoring, detection, alertness and memory. We created all items during the brainstorming sessions. We asked participants to refer to their experience on the aircraft indicated in the first part ("Aircraft with the most experience among the previously given"). In order to avoid conflicts of terminology, we asked for the frequency specific events in the cockpit, e.g. "When I use the autopilot, I miss some ATC calls because I was thinking about matters unrelated to the flight" or "When I use the autopilot, I check flight parameters as regularly as in manual flying". Responses could be *Never*, *Rarely*, *Sometimes*, *Occasionally*, *Frequently*, *Usually* or *always*.

3.3.2.5 Mind wandering questionnaire (Part 5)

The last part of the questionnaire aimed at measuring the propensity to mind wander in the daily life, outside the cockpit and for both pilots and normal people. To be able to compare it to the activity of piloting (and to the part 4), we needed items measuring MW proximity to the task. We used the Mind Wandering Questionnaire (Mrazek, Phillips, et al., 2013). Unlike other questionnaires (Broadbent, Cooper, FitzGerald, & Parkes, 1982; Carriere et al., 2008; Singer & Antrobus, 1970; Stawarczyk & D'Argembeau, 2016), the MWQ focuses only on task-unrelated MW. This questionnaire has already been validated with children, teenagers and adults (Faber, Bixler, & D'Mello, 2017). The original MWQ displayed a 5-points Likert scale, which we converted to a 7-points Liker scale. Responses could be *Never*, *Rarely*, *Sometimes*, *Occasionally*, *Frequently*, *Usually* or *always*. Items were:

- While reading, I realize I haven't been thinking about the text and must therefore read it again
- I find myself listening with one ear, thinking about something else at the same time
- I do things without paying full attention
- My mind wanders during lectures or presentations
- I have difficulty maintaining focus on simple or repetitive tasks

3.3.3 Non-pilot questionnaire

The non-pilot questionnaire consisted of demographic items and the MWQ described in the previous subsection. Demographic questions were here to verify that our respondent poll were representative of the pilot population. We also discarded all participants with critical professions (where attentional problems could lead to direct and important problems, e.g. taxi driver, surgeon, or air controller). We did so to ensure that we would compare people having to maintain high levels of externally directed attention like pilots, to people with no such requirement in their daily work.

3.4 Results analysis

Results given here are only descriptive. Given our purpose to give us a better idea of the operational need and the small number of answers, we will not conduct inferential statistics.

3.4.1 Population description

46 pilots answered the pilot questionnaire. Age ranged from 24 to 67 years-old (*M* = 37.4, *95% CI* = [34.5; 40.4]). All respondents were male, (compared to 6.7% of female in all pilots in the world according to the International Society of Women Airline Pilots, 2018). 31 pilots answered the French version, 15 pilots answered the English version. There were 16 amateur pilots, 22 professional pilots and 8 military pilots. All operated on airplanes.

For the non-pilot questionnaire, we collected 86 answers. We suppressed answers from safety-critical workers (e.g. surgeon, truck driver…). Moreover, in order to be able to compare the two questionnaires, we semi-randomly dismissed non-pilot answers until we obtained a similar age distribution to pilots. Our final poll included 46 respondents whose age ranged from 20 to 58 years-old (*M* = 33.5%, *95% CI* = [30.5; 36.28]). Six respondents were female while all others were male.

3.4.2 Hours flight time and reference aircraft

There was important differences in the years of experience as a pilot (from 1 to 36 years, *M* = 12.01, *95% CI* = [9.3; 15.3]). Hours flying time ranged from 150 to 14500 hours (*M* = 2083 hours, *95% CI* = [1126; 3038]). Hours flying time for the reference aircraft ranged from 60 to 5500 (*M* = 1111, *95% CI* = [699; 1524]).

Aircrafts reported as reference aircraft are gathered in the appendix D with the number of reports and the field of use. Pilots reported various aircrafts. The most reported being Soccata TB20 and Diamond DA42 for amateur pilots (6 and 5 times, respectively), Transall C160 for military pilots (2 times, all other aircrafts where reported one time each), Airbus 320 and Boeing 737 for professional pilots (6 and 3 times, respectively).

3.4.3 Distribution of time when in the cockpit

Pilots reported fly in manual mode on average 21.6% (*95% CI* = [14.2; 29.1]) of the flight. This rate varied substantially between categories of pilots: amateur and military pilots reported similar rates, *M* = 29.3%, *95% CI* = [17.3; 41.5], and *M* = 28.8%, *95% CI* = [6.1; 51.0], respectively, even though the standard error on the mean highlight a lower agreement in military pilots. Professional pilots reported a lower rate and higher agreement, *M* = 14.1%, *95% CI* = [5.0; 23.2], in accordance with lower variety of missions in commercial aviation.

Reported time spent performing flight-related activities varied substantially between pilots (see Figure 11). By replacing categories around their central value (and therefore responses are 5%, 20%, 40%, 60%, 80% and 95%), we computed the mean and standard error on the mean (also approximate, this technique allows us to derive estimated average). Pilots reported spending on average 47.7% of their flight time performing flight-related activities (*95% CI* = [39.0; 56.5]). However, category of pilots played a minor role in this, even though the variability halts any robust conclusion. Amateur, military and professional reporting 48.7% (*95% CI* = [35.5; 61.7]), 45.0% (*95% CI* = [19.5; 70.2]) and 48.0% (*95% CI* = [35.8; 60.1]), respectively.

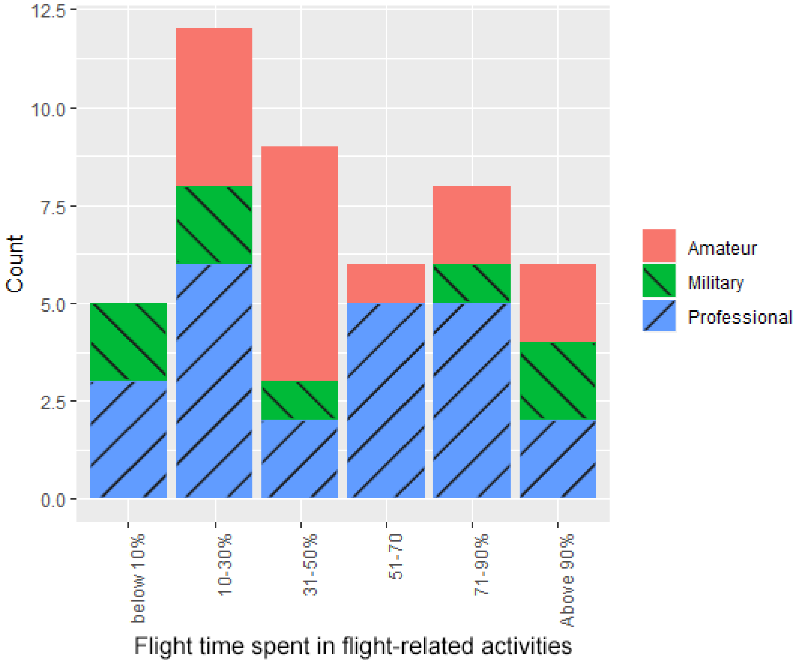


Figure 11. Reported time spent performing flight-related activities, in percentage of all flight, for each category of pilot.

Table 2 presents all non-piloting tasks. Tasks were separated for each category of pilot as it may dramatically differ. Results show that one of the main tasks for all pilots is to discuss with their crew. Amateur pilots regularly fly with passengers they know for leisure. Similarly, military and professional pilots operate with at least a copilot and regularly with a complete crew with whom they can talk about the mission, the flight or unrelated matters (similar discussions were reported for train and truck drivers, see Cummings et al., 2015). The second main task for amateurs is outside surveillance. Indeed, amateur pilots regularly fly in non-controlled airspaces (i.e. no air traffic controller to watch for air traffic separation), which leads them to spend much time looking outside for any incoming flying object. Military and professional pilots, on the contrary, have many more instruments available, which can be seen in the percentage of flight time dedicated to system monitoring. They also reported similar periods of rest, personal reading and eating. Finally, professional pilots reported “knowledge maintaining”. This task seems important to the flight, however no exact description was specified (reading the manual, looking at videos).

Table 2. Non-piloting tasks and reported percentage of flight time for each category of pilot.

“//” indicates categories not reported by pilots. “?” indicates that pilots reported this task without specifying the percentage of flight time. “*” indicates that this percentage is accompanied with “when in multi-crew”

	Amateurs		Military		Professional	
Activity	Percentage of flight time	Number of reports	Percentage of flight time	Number of reports	Percentage of flight time	Number of reports
Outside surveillance	100-50	3	//	//	//	//
System monitoring	?	1	50-10	3	70-10-95	3
Discussions	30-30-50-20-10	7	40-10	3	20-75-5	6
Admiring the view	20-20	3	//	//	?	1
Rest	5	1	15-5	2	40*-5-10	3
Landing preparations	10	1	?	1		
Personal reading	//	//	30	1	5-20-30-5	6
Eating	//	//	5	1	7-10-5	3
Military operations	//	//	20	1	//	//
Bathroom needs	//	//	//	//	2	1
Administrative tasks	//	//	//	//	25-3	2
Knowledge maintaining	//	//	//	//	20	2
Reading NOTAMs	//	//	//	//	1	1
Various	10-10	3	//	//	//	//

3.4.4 Pilots’ use of automation

Items addressing pilots’ preference regarding autopilot use are gathered below. All items specifically concerned experience related to the reference aircraft (see previous subsection 3.4.2 - Hours flight time and reference aircraft). All items could have as answers. For statistical purpose, we replaced Likert steps *never, rarely, sometimes, occasionally, frequently, usually* or *always* to steps number 1 to 7.

Firstly, pilots prefer to use managed modes (higher automation capable of following the flight plan) “frequently” on average, compared to selected modes (lower automation, maintain parameters like altitude or speed, see Figure 12), $M = 5.3$, *95% CI* = [4.8; 5.7] (amateur: $M = 5.0$, *95% CI* = [4.1; 5.9]; military: $M = 5.5$, *95% CI* = [5.1; 5.8]: professional: $M = 5.4$, *95% CI* = [4.8; 6.0]). Similarly, pilots reported preference to use autopilot during cruise “usually” on average (see Figure 13), $M = 6.4$, *95% CI* = [6.1; 6.7] (amateur: $M = 6.3$, *95% CI* = [6.0; 6.7]; military: $M = 6.3$, *95% CI* = [5.7; 6.9]: professional: $M = 6.5$, *95% CI* = [6.0; 7.0]). Both items confirm that pilots often use automation during most of the flight for the cruise (low on events).

However, for more complex and critical situations, autopilot landing is still used mostly on low-visibility conditions (e.g. fog). When they have the possibility, pilots prefer to land manually (letting “Rarely” on average autopilot handle landing, see Figure 14), $M = 2.0$, *95% CI* = [1.6; 2.4] (amateur: $M = 1.6$, *95% CI* = [1.1; 2.0]; military: $M = 3.4$, *95% CI* = [2.1; 4.7]; professional: $M = 1.7$, *95% CI* = [1.3; 2.2]). Military are the only population willing to “occasionally” autoland when given the possibility to do it manually. It may be due to the difficult missions military pilots have to deal with, extensively tiring, and lead them to prefer the relief of autoland. However, these results must be taken with caution, as no possibility to answer “Not Applicable” was provided. This absence could have led pilots (mainly amateur) to report “never” when they did not have autoland available on their aircraft.

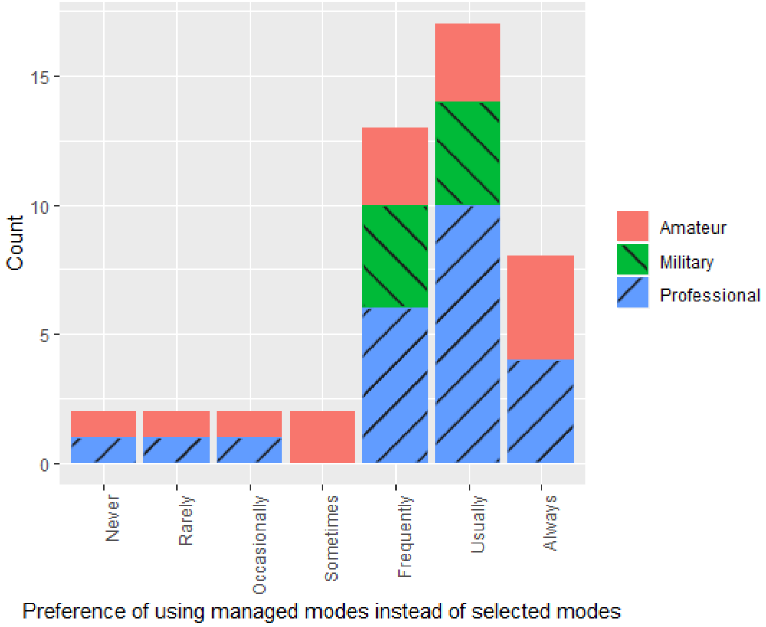


Figure 12. Reported preference of using managed modes (LNAV, VNAV, …) on selected modes (HDG SEL, LVL CHG, …), for each category of pilot.

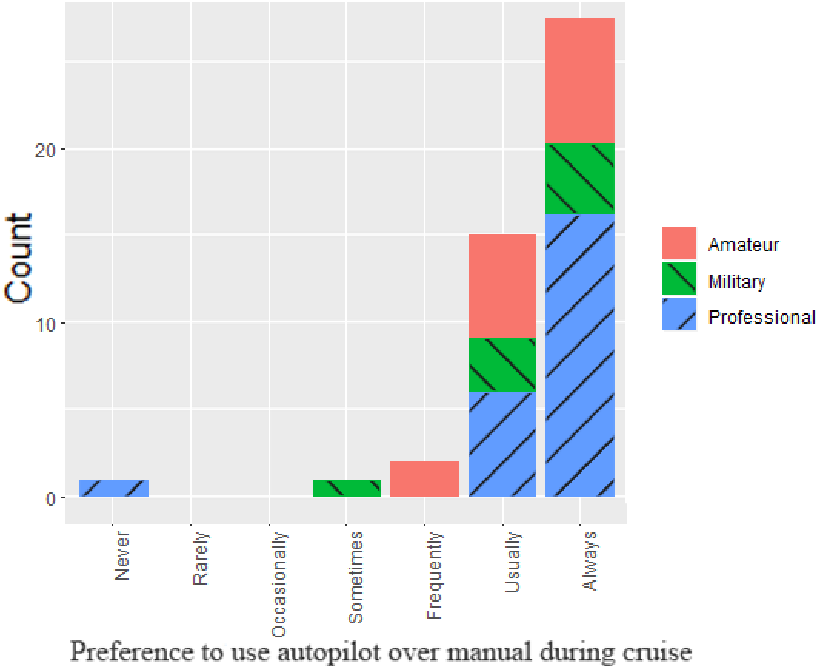


Figure 13. Reported preference to use autopilot on manual during cruise for each category of pilot.

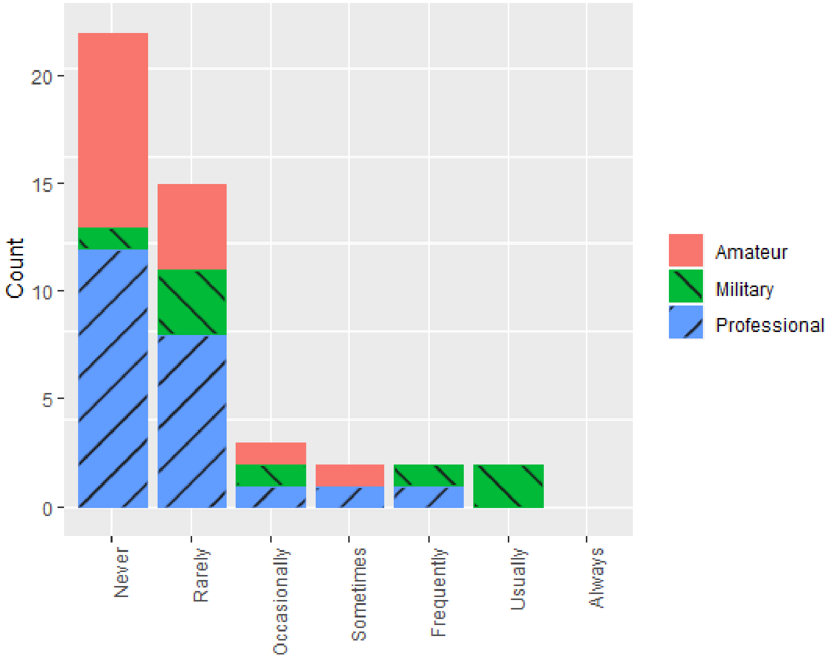


Figure 14. Reported preference to use autopilot on manual modes during landing for each category of pilot.

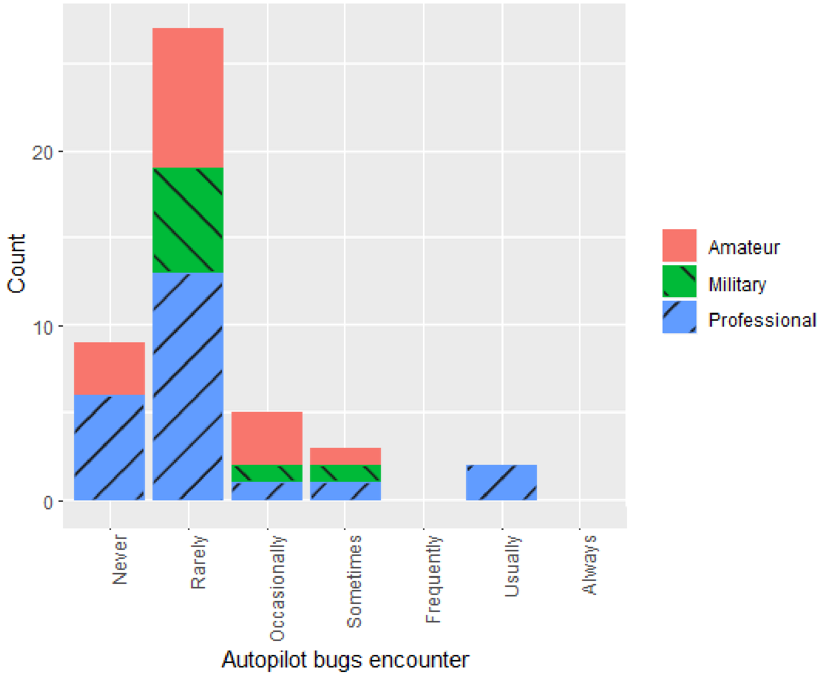


Figure 15. Reported frequency of bugs encountered in the autopilot for each category of pilot.

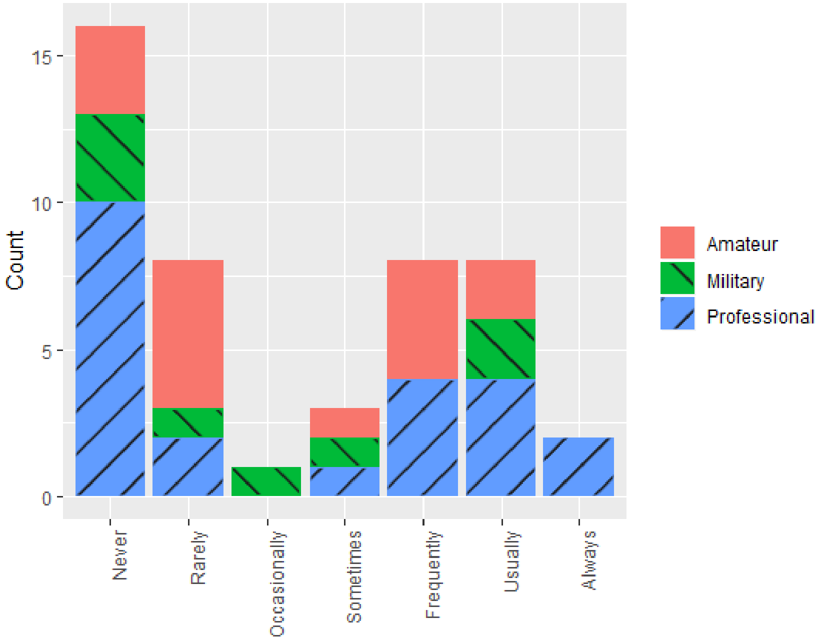


Figure 16. Frequency of orders to use autopilot from higher authorities for each category of pilot.

Finally, pilots reported bugs on average “Rarely” (see Figure 15), *M* = 2.2, *95% CI* = [1.9; 2.5] (amateur: *M* = 2.1, *95% CI* = [1.7; 2.5]; military: *M* = 2.4, *95% CI* = [1.9; 2.9]; professional: *M* = 2.2, *95% CI* = [1.7; 2.8]). Contrary to other items, we here are limited with the interpretation of bugs according to pilots. Are terrible bugs that happen once every one million flying hours considered “rare”? Did pilots answer “rarely” knowing the zero-risk does not exist? Are there small-impact bugs that are completely handled by pilots in a regular basis? More information is needed to exploit these results.

The general feeling of pilots regarding automation could have been biased by external factors. Amon them, professional and military pilots have to obey orders from higher authorities (international regulations, airlines, hierarchy). We asked pilots to report how often was they obeyed orders or procedures to use autopilot. They reported using autopilot “occasionally” on average, even though pilots were strongly divided in their responses regardless of their category (see Figure 16), *M* = 3.2, *95% CI* = [2.6; 3.9] (amateur: *M* = 3.3, *95% CI* = [2.3; 4.2]; military: *M* = 3.0, *95% CI* = [1.6; 4.4]; professional: *M* = 3.3, *95% CI* = [2.4; 4.3]).

Overall, these results confirm that pilots use automation because they appreciate it, even though higher authorities can impose it sometimes. Nevertheless, they still prefer to handle some critical tasks when possible, such as landing.

3.4.5 Relations between parts of both questionnaires

We needed to aggregate answers into part-level to give a general view of our results. This gave us one value for each part and each respondent. We computed Intra-Class Coefficients (ICC; Lebreton & Senter, 2008) in order to validate the inter-rater agreement and inter-rater reliability and verify that in each part items measured similar dimensions. Coefficients are gathered in Table 3. Results provides fair to strong agreement and reliability of items for all parts (Cicchetti, 1994; Glick, 1985; Lebreton & Senter, 2008). These values support the possibility to aggregate answers into part-level.

Table 3. Measures of ICC(3,k) of each part of the questionnaire.

The first result is the mean ICC of all items for each part. Between brackets is the range of results (minimum and maximum of ICC obtained among all items).

Population	Measures	ICC(3,k)
Pilots	Part 3 (perception of the autopilot)	$M = .59$, $F(45,315) = 2.5$, $p < .001$
	Part 4 (human-autopilot problems)	$M = .84$, $F(45,405) = 6.3$, $p < .001$
	Part 5 (MWQ)	$M = .82$, $F(45,180) = 5.5$, $p < .001$
Non-pilots	MWQ	$M = .62$, $F(45,180) = 2.7$, $p < .001$

All individual items descriptive statistics and figures are presented in appendixes E, F, G, and H. Table 4 presents mean and 95% confidence intervals for each part of the two questionnaires. Scores for the question “Autopilot is complex” were reversed, as complexity is negatively correlated with usability. The Figure 17, Figure 18, Figure 19 and Figure 20 show the distribution of means for each category of pilot and for non-pilots. We rounded mean results to obtain these figures.

Table 4. Mean and 95% confidence intervals for each part of the questionnaire.

Item	Item	Mean and 95% confidence interval
Pilots	Part 3 (autopilot usability)	5.6 [5.4; 5.8]
	Part 4 (human-autopilot problems)	2.3 [2.1; 2.5]
	Part 5 (Mind wandering questionnaire)	3.2 [3.0; 3.5]
Non-pilots	Mind wandering questionnaire	3.6 [3.3; 3.8]

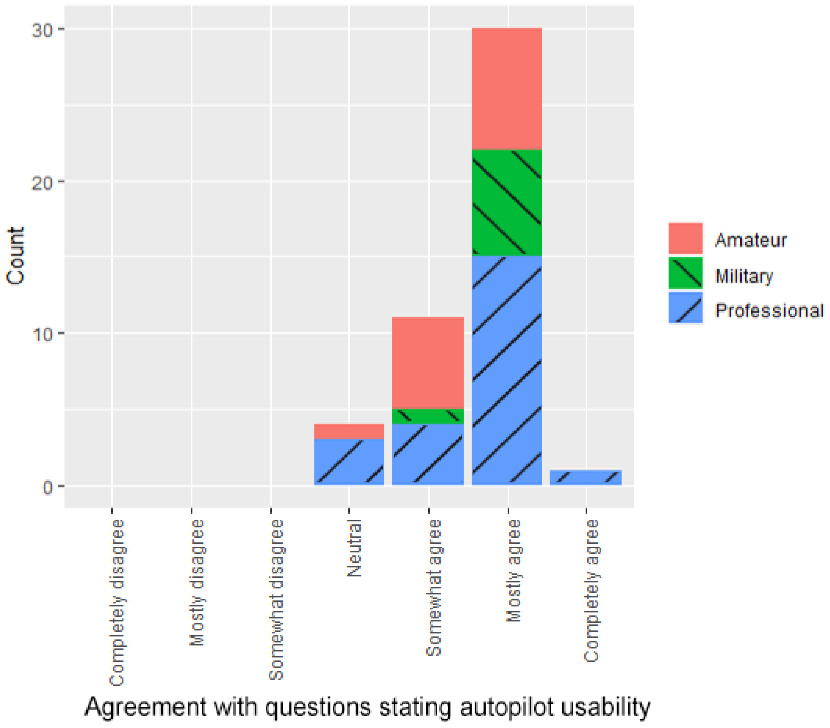


Figure 17. Means of part 3 “Autopilot usability” answers for each pilot.

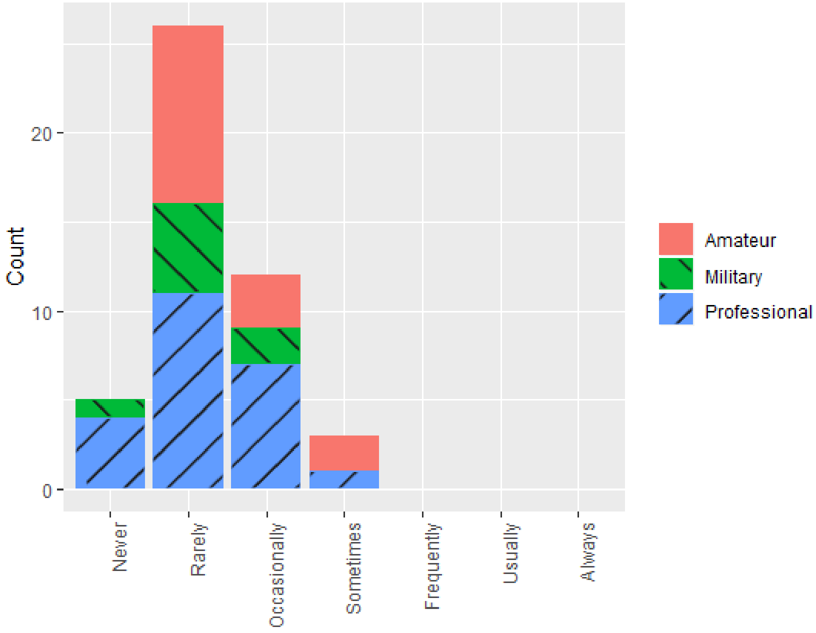


Figure 18. Means of part 4 “Human-autopilot problems” answers for each pilot.

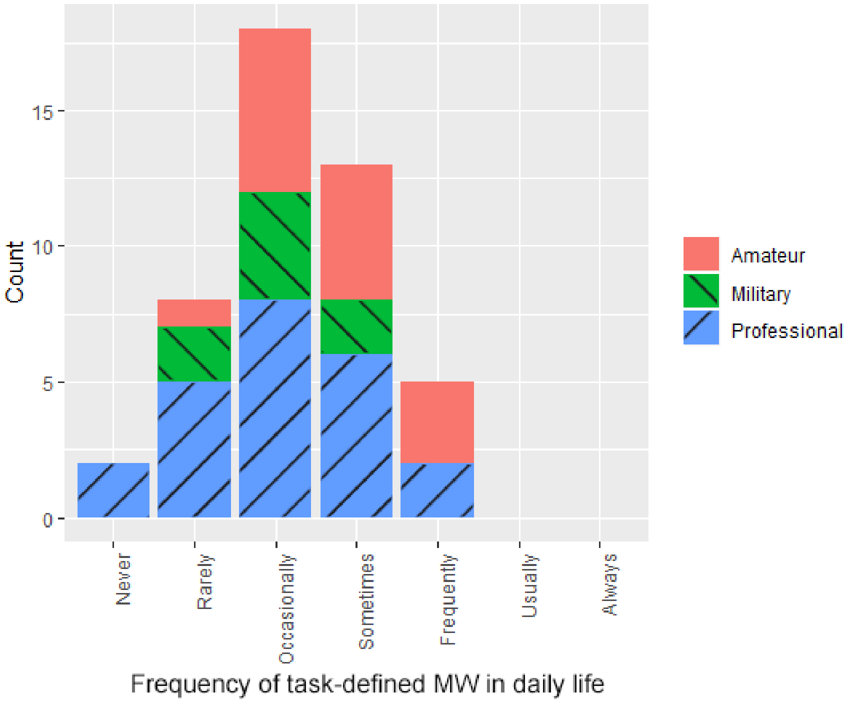


Figure 19. Means of part 5 “Mind wandering questionnaire” answers for each pilot.

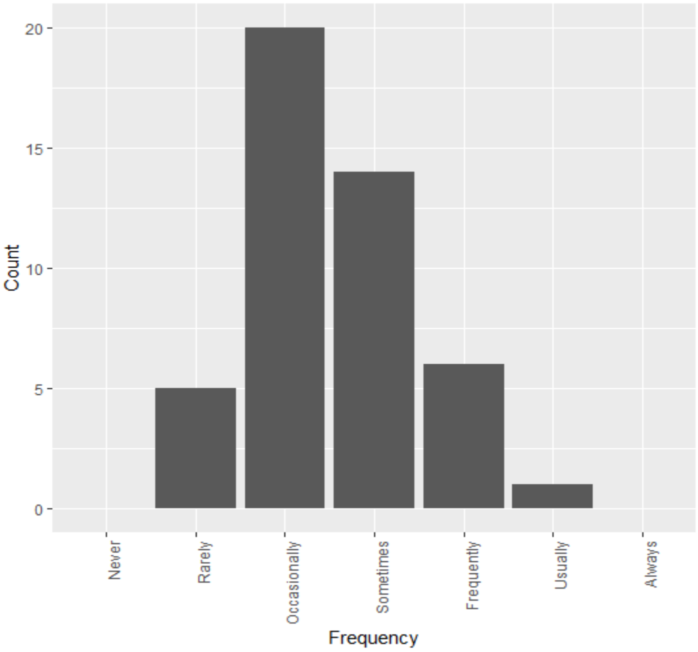


Figure 20. Means of part 5 “Mind wandering questionnaire” answers for non-pilots.

Results show that pilots perceive autopilot as being at least “Usually” usable for two third of them. All means stayed above “Sometimes” usable, confirming that modern pilots appreciate the autopilot. Even though some issues can remain, pilots nowadays are far from being nostalgic of the “manual era” and recognize how autopilot can help them cope with long periods of cruise without much events.

Moreover, pilots reported mostly “Rarely” experiencing problems when operating the autopilot. Although such report should be reassuring, the potential dramatic consequences of any problems in the human-autopilot cooperation should drive us to understand in more details any of these issues. We plotted part 4 “Human-autopilot answers” against both part 3 “Autopilot usability” and part 5 “Mind wandering questionnaire” to see if a tendency could be highlighted in the relation between these scores (see Figure 21 and Figure 22). We observed a negative correlation between scores on part 4 and part 3, and a positive correlation between scores on part 4 and part 5. Those observations support the idea that a better perceived usability may decrease problems related to human-autopilot interactions, while higher rates of task-unrelated MW increased those risks.

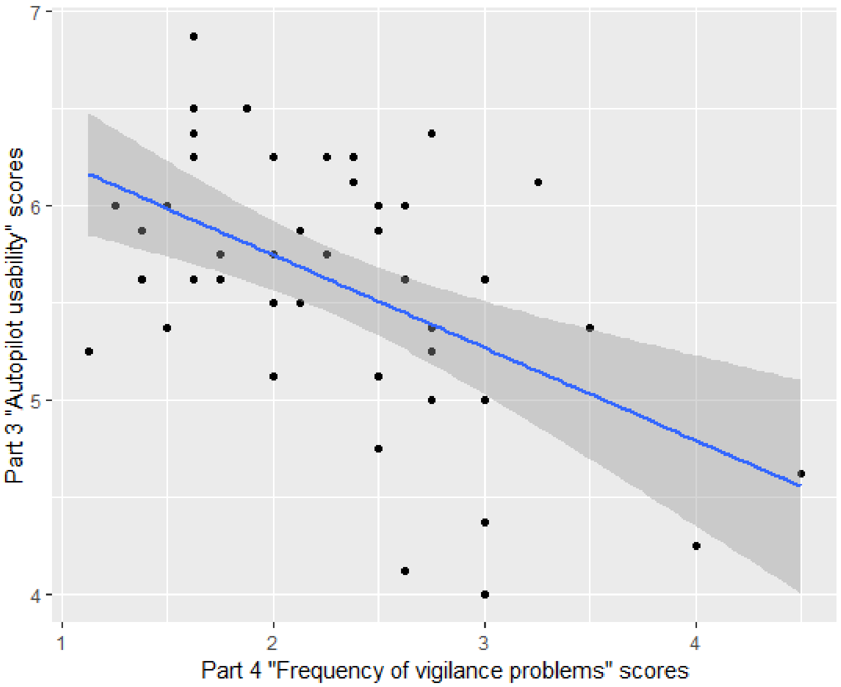


Figure 21. Regression on the scores of part 4 and part 3.

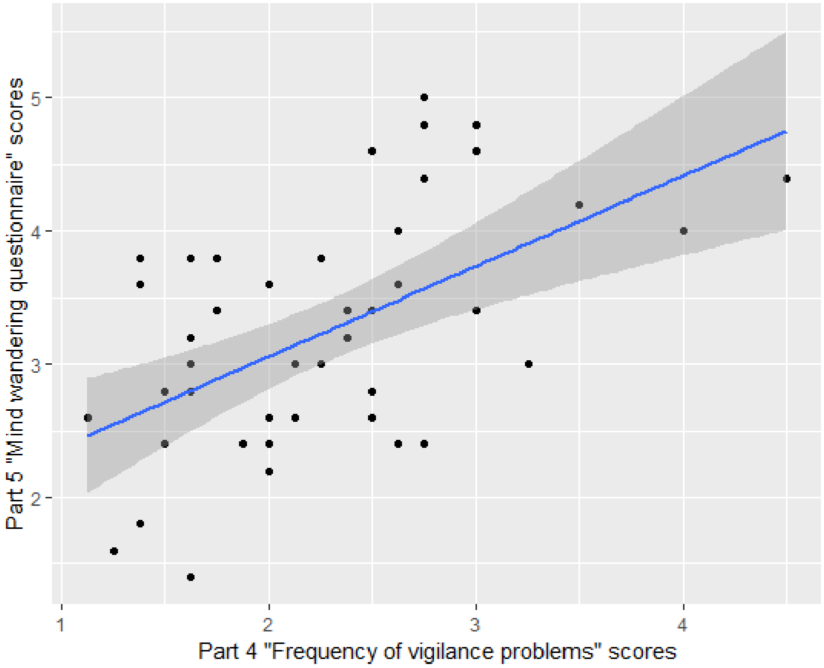


Figure 22. Regression on the scores of part 4 and part 5.

Finally, both pilots and non-pilots reported predominantly "Occasionally" experiencing task-unrelated MW. The bootstrapped 95% confidence interval of both populations show similar reports of task-unrelated MW rates between the two populations. Even though our results cannot definitely demonstrate the absence of difference between pilots and non-pilots, our analysis supports such hypothesis.

3.5 Discussions on the results

3.5.1 Results of the questionnaire

We investigated in parallel pilots' perception of their autopilot and factors influencing attentional capacity in and out of the cockpit. All our results should be considered as given tendencies for further experimental research. On the contrary, no attempt should be made to use these results to advocate for robust significant effects. The small poll of respondents and the variability in the missions between categories of pilots. Nevertheless, we observed that (1) pilots reported favorable perception of their autopilot favorably, (2) perception of autopilot usability and task-unrelated MW rates seem to influence attentional problems in the cockpit and (3) we did not find any difference in task-unrelated MW between pilots and non-pilots. We review each of these results below.

3.5.2 Pilots perceive their autopilot favorably

Answers from pilots on the part 2 "Experience of the autopilot" highlighted pilots' appreciation of the autopilot. Part 3 "Autopilot usability" scores corroborated these results, as pilots reported perceiving the autopilot as being between "Frequently" and "Always" usable. Far from the first decades of automation introduction (Amalberti, 1999), modern pilots seem to understand and appreciate automated aids in order to relieve them from long flight burden. This time save by automation is used to perform other professional activities, such as planning or knowledge maintaining. Pilots also reported non-flight-related activities, such as newspapers reading, discussions or admiring the view. Even though such activities could already exist before automation introduction, it is worth noting that other operators in critical environments reported similar activities to cope with boredom (e.g. truck drivers or nuclear plant operators, see Cummings et al., 2015).

Nevertheless, pilots prefer to remain in manual control for critical phases of the flight, such as landing, when meteorological conditions and procedures allow them. Explanations could either be because of a lack of trust, or because they want to maintain their manual skills.

3.5.3 Part 3 and 5 influence human-automation problems frequency

The regression on part 4 “Human-autopilot problems” against part 3 “Autopilot usability” scores displayed a positive slope. On the contrary, the regression on part 4 “Human-autopilot problems” scores displayed a negative influence on part 5 “Mind wandering questionnaire” scores.

The influence of usability perception on the use of a system and the performances has been widely documented (Jardina & Chaparro, 2013; Nielsen & Levy, 1994; Sonderegger & Sauer, 2010). In particular, the tendency observed in our results is in line with studies investigating the link between usability and user engagement in human-machine cooperation. Poor usability can be a barrier to engagement (O’Brien & Toms, 2008), leading people to reject the cooperation and reduce interactions with the system to a minimum. For example, nuclear plant operators reported to Andersson (2008) that they preferred to use lower levels of automation because of system opacity on the higher levels. Poor usability may contribute to high distractibility and higher engagement in non-flight-related activities, resulting in problems when interacting with the autopilot.

The observed positive slope between part 4 and part 5 is in line with the extensive literature investigating the influence of MW and the perceptual decoupling on performance. This is particularly true in tasks requiring high and prolonged externally directed attention, like OOTL environments (see the evidences reviewed in chapter 2.

3.5.4 Pilots and non-pilots seem to mind wander as much

The last result of our questionnaire is the absence of obvious difference of task-unrelated MW rates between pilots and non-pilots. This would mean that pilot training does not change their propensity to MW wander, or at least that this difference is trivial compared to individual differences (Forster & Lavie, 2014; Golchert et al., 2016; Song & Wang, 2012).

However, Casner and Schooler (2014) proposed that pilots may exert a certain degree of control over their MW. More precisely, their experiment observed a strong association between monitoring breaks and the propensity to engage in MW. This raises the question on the existence of such ability to control effectively MW independently of the propensity to mind wander. Moreover, if assessed, could such an ability of MW control be selected, acquired, or even taught? Natural selection may favor such ability for the attentional advantage that it creates, but only if the drawbacks do not counterbalance it. Kane and colleagues investigated the relation between working memory (cognitive system responsible for temporarily holding information) and MW. They firstly found that people with wider working memory capacity (WMC) outperformed those with lower WMC on attention tasks (Kane, Bleckley, Conway, & Engle, 2001). They continued by looking at WMC relation with MW rates and concentration (Kane et al., 2007). In their task, subjects had to remember a short list of items. Before reporting this list, participants had to perform an unrelated task. The intermediate task could be verifying either an equation, if a sentence was meaningful, or whether a grid pattern was vertically symmetrical. They observed that higher-WMC participants reported less MW during high concentration periods. However, at the lowest levels of self-reported concentration, higher-WMC individuals were more prone to mind-wander. This important result underlines the complexity of the relation between MW, individual differences and task requirements. More experiments may be able to decide on whether pilot training may influence MW rates.

The relatively small poll of respondents compared to studies we previously cited, i.e. between 77 and 108 respondents for Mrazek et al., 2013; 363 respondents for Smilek et al., 2010a; 201 pilots for Wiener, 1989) may prevent us from unveiling differences between pilot and non-pilot populations. More work comparing pilot to non-pilot populations on traditional tasks (e.g. SART, reading, car simulators) may yield important results.

3.6 Thesis progress recap (chapter 3)

In this chapter, subjective reports highlighted the influence of task-unrelated MW on reported problems in the cockpit. Our questionnaire reveals that pilots mind wander as any other population (even though they may be able to control it). This is in line with the view of MW as a way to cope with boredom in monotonous environments requiring much attentional resources (Cummings et al., 2015). Similar to the theoretical foundation of the first two chapters, we will detail in the next chapter the material and methods chosen for the three chapter-experiments.

4 MATERIAL AND METHODS USED FOR EXPERIMENTATIONS

- *All **material and methods** used in multiple studies or requiring extensive explanations are detailed here (reminders and small customizations are still within each experimental chapter).*
- *Our simulated environment was an **obstacle avoidance task**, which could be manually or automatically operated, implemented as the LIPS (Laboratoire d'Interactions Pilote-Système).*
- *Attentional probes discriminated between attentional states **focused, task-related MW and task-unrelated MW**.*
- *To gather physiological measures, we used the **eye-tracker SmartEye** for oculometric signal and the **EEG ActiCHamp** for neuronal signal.*

4.1 Description of the current chapter

In order to lighten the “Material and methods” parts of the following three experimental chapters, we detail here all experimental means used and the pre-processing procedures adopted. The subsection 4.2 - The obstacle avoidance task describes our need for an automatable and ecological task, which was met with the LIPS environment. The subsection 4.3 - Attentional probes details… well… attentional probes. We also break pre-processing procedures into steps and give the R functions used. Finally, subsections 4.4 - Oculometry and 4.5 - Electroencephalography detail the physical principles behind these measurements as well as the functioning of each tool. We provide a summary of the work presented in the thesis so far in the last subsection 4.6 - Thesis progress recap (chapter 4).

4.2 The obstacle avoidance task

4.2.1 The need for an automatable task

The subject of this thesis was to investigate MW dynamic when interacting with automated systems, compared to when manually operating. Our environments needed to be both manually and automatically operable. Such constraints directly ruled out traditional experimental environments within the MW literature. The Sustained Attention to Response Task (SART) would be difficult to handle. The traditional version of the task requires participants to push a button when a number appears, except for a 3 (they then have to withhold). Supervising the computer performing the TASK would dramatically make the task more complex, thus manual and automated tasks too different to be compared (we tried). Similarly, an automated version of a reading task would be difficult to create. Performance during reading is assessed via with questions on the comprehension at the end. Participants would have to read all the same in order to assess when the automation provides wrong answers.

In order to comply with our aim, we needed a more complex task for which automation would be meaningful (i.e. provide an environment with “minimal or reduced human intervention”; Rifkin, 1995, pp. 66–75; Wikipedia, 2018b). Moreover, we also wanted a task as ecological as possible, i.e. close to what supervisors can encounter.

Obstacle avoidance tasks meet those two constraints. Such tasks can be automated by letting systems detect and avoid obstacles. Moreover, one can implement aeronautical features (plane as the mobile, flying objects as obstacles) in order to increase ecological validity. Finally, such tasks are close to what drone operators, and more generally pilots, are performing.

4.2.2 The Laboratoire d’Interactions Pilote-Système (LIPS)

The LIPS (Laboratoire d’Interactions Pilote-Système, or Pilot-System Interactions Laboratory) is a distributed simulation environment, intended for the simulation of human operators, pilots and controllers, in aeronautical operations scenarios. The ONERA laboratory developed internally this environment. The LIPS handles simulation of the flight of different categories of air vehicles (airplanes, helicopters, drones, military or civil…) using various configurable models (kinematics, pseudo 6DOF…). Operators can order these vehicles with different control devices (joystick, cab….), via different graphic interfaces (pilot view with head-up symbologies, navigation pages with cartography and flight plan…).

The LIPS proposes a generic version of autopilot. Decision functions exists (using Petri net formalism via ProCoSA software) with guiding laws for vertical and horizontal axis and auto thrust. Overall, the autopilot proposes all generic automated flying modes from manual to full automation. Each vehicle can be equipped with a sensor for locating other vehicles or obstacles. The radar or ADS-B functions with their behavior characteristics (ranges, refresh time, errors and possible biases) is simulated.

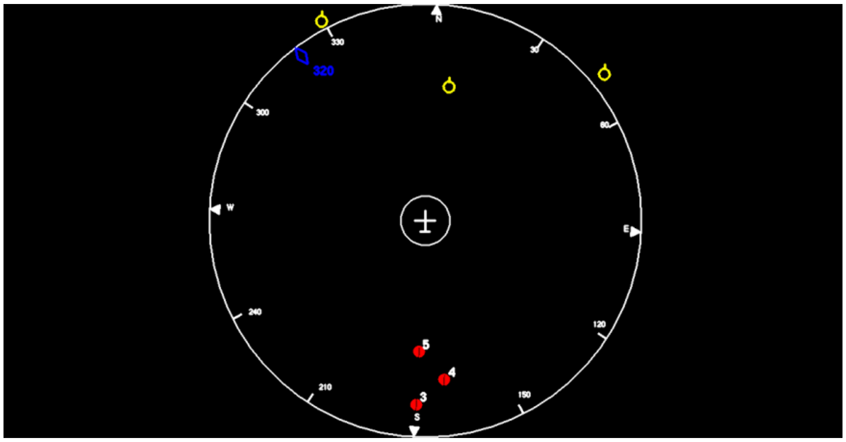


Figure 23. Screenshot of the LIPS interface.

The plane in the center is static and the surround (yellow and red numbered symbols) are moving. During left and right avoidance maneuver, again, the plane remains static and the background rotates.

For these thesis experiments, we only used the navigation and critical situation screen (see Figure 23). An unmanned air vehicle (UAV) depicted as a plane seen from above stayed at the center of a 2D radar screen and moved following waypoints arranged in a semi-straight line with clusters of obstacles along the way (every 45s on average). Each cluster contained between 1 to 5 obstacles, including one on the trajectory. We informed the participants to control the movements of the UAV to avoid obstacles. In the three experiments presented in chapters 5, 6, and 7, we manipulated the level of automation of the task and the reliability of the automated system.

4.3 Attentional probes

4.3.1 Attentional probes in mind wandering literature

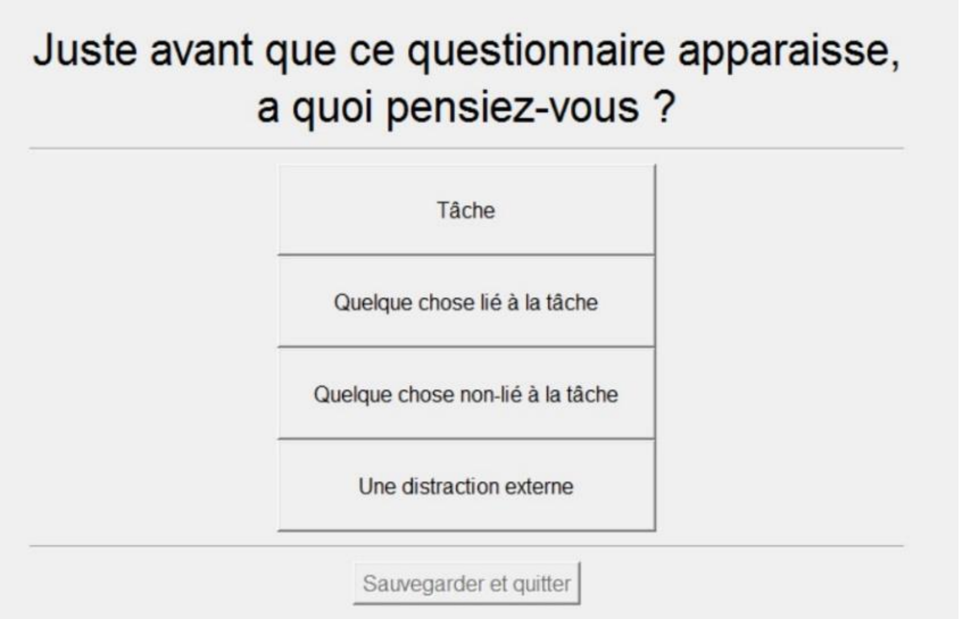
Even though critics exist concerning the validity of real-time subjective reports (Tsuchiya, Wilke, Frässle, & Lamme, 2015), studies showed that no-reports could also present flaws and mistakenly take a psychological construct for another (Overgaard & Fazekas, 2016). This is especially true for MW and the multiple dimensions associated (see 1.3.1 – Mind wandering is a family of experiences). For example, Stawarczyk and D’Argembeau (2016) observed different influences for sleepiness and MW. Although close in their apparent impact, the influence of both phenomena on reaction time variability could be discriminated. Moreover, Smilek et al. (2010a) demonstrated that attentional reports during SART correlated heavily with reports of MW in everyday life. In any case, MW field still relies heavily on these reports to correlates multiple measures with MW/focus periods. Models using physiological measures do not achieve (at the moment) adequate detection rate (Bixler & D’Mello, 2014, 2015; Blanchard et al., 2014; Melinscak, Montesano, & Minguez, 2014; Pham & Wang, 2015).

4.3.2 Attentional probes in our experiments

We used Python 3.6 to program attentional probes. On average every 1’30 or 2 minutes (depending on the experiment), an attentional probe appeared on a secondary screen next to the main screen. For technical reasons, the obstacle avoidance task was not paused when the attentional probes appeared. Participants filled it as soon as it appeared. We informed them that any successful or failed trial during this interval would not affect their overall performances. Moreover, we explained that we used attentional answers for informational purposes, not to assess performance. This limited the possibility that participants would be reluctant to report their distraction.

In all experiments, attentional probes included the following question (originally in French, see Figure 24): “When this questionnaire appeared, where was your attention directed?” Answers could be “On the task” (focused, e.g., thinking about the next obstacle, the decision to make, the incoming waypoint), “Something related to the task” (task-related MW, e.g., thinking about performance, interface items, last trial), “Something unrelated to the task” (task-unrelated MW, e.g., thinking about a memory, their last meal, or a body sensation) or “External distraction” (e.g., conversation, noise). We verbally gave the preceding examples to participants to illustrate each category prior to the experiment. We were primarily interested in reports of being focused or having task-related or task-unrelated MW. The possibility to report “task-related MW” was proposed to avoid participants to report MW when thinking about their performance (Head & Helton, 2016). We chose these three categories after

Casner and Schooler studies (2014, 2015), the first ones to investigate MW in automated conditions. We did not use binary categorization of MW in order to provide a more accurate picture of the different prototypes of MW (Seli, Beaty, Cheyne, et al., 2018). Moreover, the use of a Likert scale based on the proximity to the task would require the critical assumption that task-related MW has effects in between focus periods and task-unrelated MW. “Noise” answer was proposed to avoid participants to report MW if they were focused on any external signal. Experiment of chapter 6 also includes other questions that we detail in the corresponding subsection 6.3.2.3 - Attentional probes.



Juste avant que ce questionnaire apparaisse, a quoi pensiez-vous ?

Figure 24. Screenshot of attentional probes in French

Because of program focus issues, attentional probes could not be displayed on the same screen as the LIPS. Indeed, after participants answered probes, the command window stayed in the center of the screen, in front of the LIPS. This behavior would have force participants to click on the LIPS to bring the focus back on it. This may lead individuals to think about something else than the task, and could offer participants the possibility to click where they should not… In order to avoid such annoyances, we chose to display attentional probes on a different screen for all experiments.

4.4 Oculometry

Even though we still need attentional reports to detect MW episodes, physiological measures may allow use to characterize MW influence on operators more accurately. Ultimately, the opportunity to rely ultimately on physiological measures would offer countless possibilities. Among these, oculometry seems to fit particularly operational needs (see the subsection 2.3.3 – Oculometric markers).

4.4.1 Measuring attention through oculometry

Oculometry is the measurement of the condition (e.g. pupil size, blinks) and movements (e.g. saccades, fixations) of the eye. A remote or head-mounted ‘eye tracker’ connected to a computer allows collecting eye-tracking data. While there are many different types of non-intrusive eye trackers, they generally include two common components: a light source and a camera. The light source (usually infrared) illuminate the eye. The camera tracks the reflection of the light source along with visible ocular features such as the pupil and eyelid. Coupled with a model of the eye, this data allows extrapolating the rotation of the eye to have eye dynamic (direction of gaze, pupil diameter) and eye events (blinks, saccades, fixations).

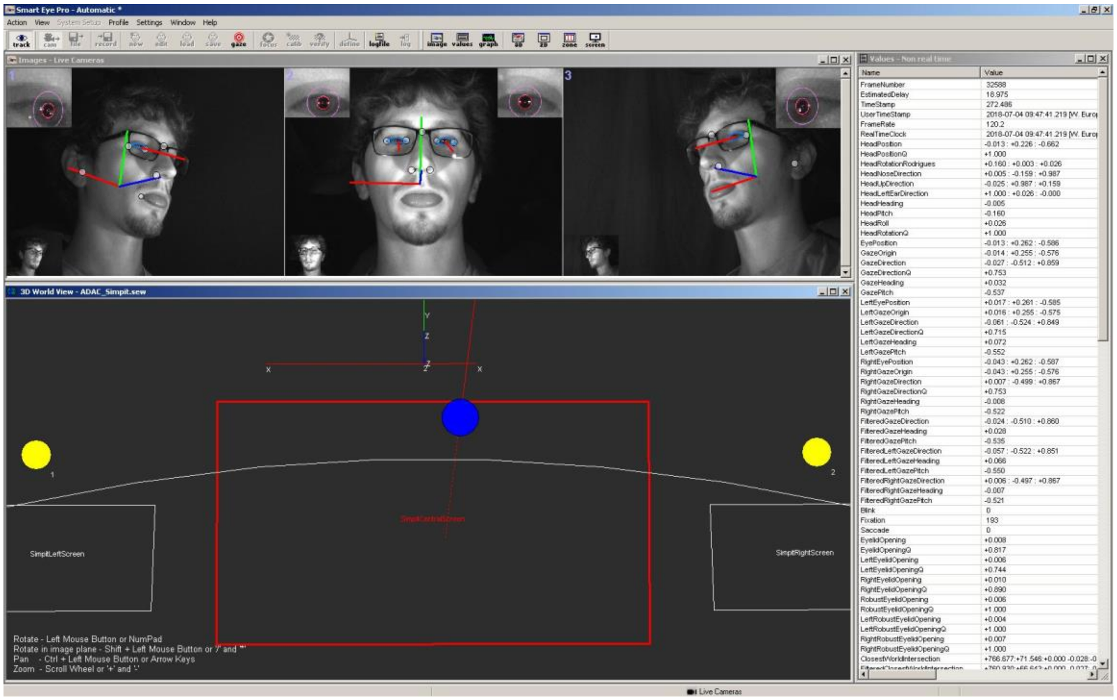
Data yielded by the eye-tracker can help point attentional states. Pupil size variation has been associated with locus coerulus (LC) activity, mainly through the norepinephrine circuit already linked to arousal and external/internal attention (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010). Researchers already linked MW reports with pupil diameter, blink rate, saccades, fixations and gaze position (Franklin et al., 2013; Grandchamp et al., 2014; Reichle et al., 2010; Smilek et al., 2010b).

4.4.2 SmartEye for acquisition

The SmartEyePro system is a deported tracking system well suited for the demanding environment of a vehicle cockpit. The system measures the subject's head pose and gaze direction in 3D. The configurations used three cameras and two infrared illuminators. We used a deported eye tracker because our working position (the LIPS) was static. Moreover, participants with lunettes can experience discomfort with mounted eye trackers. Finally, deported eye-trackers are less invasive which lowers the impact on the natural propensity to mind wander. We used R for filtering and treatment (R Core Team, 2016) because of the flexibility provided by the open-source license.

In all experiments, we used the SmartEye Pro software to compute fixations, saccades and blinks (see Figure 25). Blinks were computed using sliding windows of 700ms. Saccades were defined in SmartEye Pro parameters as gaze velocity over 35 deg/s. Saccades were limited to 200ms. Fixations were frames where the gaze velocity remained below 15 deg/s.

The 10 seconds preceding each probe were extracted from oculometric data. This period length is in line with the literature investigating MW and oculometric markers (Bixler & D'Mello, 2014, 2015; Franklin et al., 2013; He et al., 2011). Extracts before "On the task" and "Something related to the task" were classified as "Focus" to avoid any influence of poor performance on mind wandering related measures (Head & Helton, 2016). Extracts before "Something unrelated to the task" were classified as "MW". Extracts before "External distraction" were discarded as noise.



Name	Value
FrameNumber	32588
EstimatedDelay	16.975
TimeStamp	272.486
UserTimeStamp	2018-07-04 09:47:41.219 [W. Euro
FrameRate	120.2
RealTimeClock	2018-07-04 09:47:41.219 [W. Euro
HeadPosition	-0.013 : +0.226 : -0.662
HeadPositionQ	+1.000
HeadRotationRodrigues	+0.160 : +0.003 : +0.026
HeadNoseDirection	+0.005 : -0.159 : +0.987
HeadUpDirection	-0.025 : +0.987 : +0.159
HeadLeftEarDirection	+1.000 : +0.026 : -0.000
HeadHeading	-0.005
HeadPitch	-0.160
HeadRoll	+0.026
HeadRotationQ	+1.000
EyePosition	-0.013 : +0.262 : -0.586
GazeOrigin	-0.014 : +0.255 : -0.576
GazeDirection	-0.027 : -0.512 : +0.859
GazeDirectionQ	+0.753
GazeHeading	+0.032
GazePitch	-0.537
LeftEyePosition	+0.017 : +0.261 : -0.585
LeftGazeOrigin	+0.016 : +0.255 : -0.575
LeftGazeDirection	-0.061 : -0.524 : +0.849
LeftGazeDirectionQ	+0.715
LeftGazeHeading	+0.072
LeftGazePitch	-0.552
RightEyePosition	-0.043 : +0.262 : -0.587
RightGazeOrigin	-0.043 : +0.255 : -0.576
RightGazeDirection	+0.007 : -0.499 : +0.867
RightGazeDirectionQ	+0.753
RightGazeHeading	-0.008
RightGazePitch	-0.522
FilteredGazeDirection	-0.024 : -0.510 : +0.860
FilteredGazeHeading	+0.028
FilteredGazePitch	-0.535
FilteredLeftGazeDirection	-0.057 : -0.522 : +0.851
FilteredLeftGazeHeading	+0.066
FilteredLeftGazePitch	-0.550
FilteredRightGazeDirection	+0.006 : -0.497 : +0.867
FilteredRightGazeHeading	-0.007
FilteredRightGazePitch	-0.521
Blink	0
Fixation	193
Saccade	0
EyelidOpening	+0.008
EyelidOpeningQ	+0.817
LeftEyelidOpening	+0.006
LeftEyelidOpeningQ	+0.744
RightEyelidOpening	+0.010
RightEyelidOpeningQ	+0.890
RobustEyelidOpening	+0.006
RobustEyelidOpeningQ	+1.000
LeftRobustEyelidOpening	+0.004
LeftRobustEyelidOpeningQ	+1.000
RightRobustEyelidOpening	+0.007
RightRobustEyelidOpeningQ	+1.000
ClosestWorldIntersection	+766.677:+71.546:+0.000 -0.028:-0

Figure 25. Screenshot of the SmartEye software.

Above is the visual flow of the three cameras; below is the world as we defined it in the system; on the right are all captured parameters in real time.

4.4.3 Procedure for oculometric data pre-processing

We performed pupillometry pre-processing using the R packages *reshape* (Wickham, 2007), *psych* (Revelle, 2017), *ggplot2* (Wickham, 2009, p. 2) and *robfilter* (Fried, Schettlinger, & Borowski, 2014). We divided our pre-processing into different steps (between brackets are the R functions and packages used):

- **Epoch extraction**: we defined oculometric epochs as data during the 10-second intervals preceding each questionnaire. This duration is in line with the literature (Bixler & D'Mello, 2014, 2015; Franklin et al., 2013; He et al., 2011).
- **Epoch removing**: we deleted epochs when the subject was looking at the main screen to avoid any luminosity effect (e.g., to avoid reporting effects when it was only a case of the people reporting

MW looking more outside the screen). We also discarded all epochs that included some actions by participants during the interval (i.e., if participants clicked on a button during the 10 seconds). This ensured that all epochs were free of phasic activity linked to decisions (which could mask the MW influence).

- **Conditioned filtering**: pupil diameters smaller than 1 mm and larger than 10 mm were excluded (due to the physical limits of pupil diameter, see Lemercier, 2014). Pupil diameters differing from the preceding value by more than 80% were also excluded (due to pupil dynamic limits). Pupil diameter with a quality metric (computed by the SmartEye software) below 0.01 were excluded, in order to discard tracking losses (given by a quality of 0). 10-second epochs were discarded if their resulting pupil diameter series consisted of more than 30% discarded samples. We excluded 5.5% of all segments, which is in line with the literature (Smallwood et al., 2011).
- **Interpolation**: segments were completed using linear interpolation.
- **Smooth filtering**: after interpolation, a second moving average filter was applied (moving window of 50 frames or 417 ms).
- **Standardization**: data for each participant were standardized by subtracting the mean and dividing by the standard deviation of all retained epochs for this participant.

Aside from pupillometry, the SmartEye Pro software computed fixations, saccades and blinks. Blinks were computed using 700ms sliding windows. Saccades were defined in SmartEye Pro parameters as gaze velocity over 35 deg/s. Saccades were limited to 200ms. Fixations were frames associated with a gaze velocity below 15 deg/s.

4.5 Electroencephalography

Electroencephalography (EEG) is an electrophysiological monitoring method to record electrical activity of the brain. Placing electrodes on the scalp with a conductive gel or paste allows recording.

4.5.1 Electroencephalography physical principle

EEG measures voltage fluctuations resulting from ionic current within the neurons of the brain (Niedermeyer & da Silva, 2005). Neurons are constantly exchanging ions with the extracellular milieu, for example to maintain resting potential and to propagate action potentials. Ions of similar charge repel each other, and when many ions are pushed out of many neurons at the same time, they can push their neighbors, who push their neighbors, and so on, in a wave. This process is known as volume conduction. When the wave of ions reaches the electrodes on the scalp, they can push or pull electrons on the metal in the electrodes. Since metal conducts the push and pull of electrons easily, the difference in push or pull voltages between any two electrodes can be measured by a voltmeter. Recording these voltages over time gives us the EEG (Benbadis, Husain, Kaplan, & Tatum, 2007).

The electric potential generated by an individual neuron is far too small to be picked up by EEG (Nunez & Srinivasan, 2006). EEG activity therefore always reflects the summation of the synchronous activity of thousands or millions of neurons that have similar spatial orientation. If the cells do not have similar spatial orientation, their ions do not line up and create waves to be detected. Pyramidal neurons of the cortex are thought to produce the most EEG signal because they are aligned and fire at the same time. Scalp EEG activity shows oscillations at a variety of frequencies. These oscillations represent synchronized activity over a network of neurons (Klein & Thorne, 2006).

Figure 26. ActiChamp system during installation of a happy subject.

Each electrode has LEDs to visualize the impedance relatively to two previously defined thresholds. Green indicates impedance close to lower threshold, orange indicates impedance between the two thresholds and red indicates impedance close the upper threshold. The software allows modifying these thresholds.

4.5.2 ActiCHamp for acquisition and EEGLAB for treatment

We used the EEG system ActiCHamp manufactured by Brain Products (see Figure 26; Brain Products, 2018). The system is a 24-bit battery-supplied, active channel amplifier coming along with actiCAP active electrodes. The systems uses 64 EEG channels with conductive gel and records at 1000 Hz. We used the BrainVision Recorder software provided with the ActiChamp system to record data. We used Matlab (The Mathworks Inc., 1992) with the EEGLAB (Delorme & Makeig, 2004) toolbox because of the extended possibilities. Moreover, thousands of publications already used EEGLAB (Brunner, Delorme, & Makeig, 2013), which attest its multiple applications and have allowed correcting the majority of the flaws and bugs.

4.5.3 Procedure for electroencephalographic data pre-processing

We used Matlab (The Mathworks Inc., 1992) and EEGLAB (Delorme & Makeig, 2004) to import, re-reference, filter, epoch, remove ICA components and build our design. This MATLAB plugin is the only EEG data treatment plugin to offers a graphic user interface, while still allowing much flexibility with third-party contributions (Brunner et al., 2013). Pre-processing was performed differently according to whether the analysis aimed at Event-Related Potentials (ERPs) or specter. We performed the following steps (between brackets are the *functions* and *plugins* used):

- **Re-referencing** to both mastoids M1 and M2 (*pop_chanedit* and *pop_reref*, *EEGLAB*).
- **Signal filtering** with a two-pass pass-band Butterworth filter (*ft_preprocessing*, *Fieldtrip*). This filter demonstrated good filtering capabilities and very low distortion on trial data. Moreover, the two-pass allows correcting for the time shift induced by the first pass. Cut-off frequencies were [0.01Hz; 30Hz] for ERPs epochs (to suppress ASSR and line noise), and [0.01Hz; 100Hz] for spectral epochs (in order to avoid losing the 40Hz ASSR).
- **Epoching** at [-4s; 1s] for ERPs and [-10s; 0s] for spectral analysis (*pop_epoch*, *EEGLAB*).
- **Baseline removing** for ERP epochs at [-200ms; 0ms] (*pop_rmbase*, *EEGLAB*).
- **Epoch rejection** by eye in order to delete all epochs with bad-quality signal (*pop_rejepoch*, *EEGLAB*). Epochs with blink or eye movements were kept (ICA will remove it).

- **Independent Component Analysis** (*pop_runica* with the 'runica' algorithm, *EEGLAB*).
- **Component rejection** to remove components of blinks and eye movements from the signal (*pop_subcomp*, *EEGLAB*).
- **Epoching** again. We created epochs of [-400ms; 800ms] for ERPs as we were interested in N1 and P3. For spectral epochs, we cut it into 1s epochs for better spectral analysis considering our small number of epochs (it is better to have a small number of epochs with lower resolution than having one big epoch of 0.01Hz resolution).

Finally, epochs were discriminated by attentional state (using the corresponding trigger).

4.6 Thesis progress recap (chapter 4)

After investigating the theoretical aspect of MW and OOTL performance problem relation, we detailed in this chapter all materials and methods which will be used for the three experiments to come. We will now go into the first one, aiming at comparing MW dynamic between manual and automated conditions of the obstacle avoidance task.

5 STUDY OF MIND WANDERING DYNAMIC IN OUT-OF-THE-LOOP SITUATIONS

- *We investigated the* ***dynamic of MW*** *and its* ***oculometric marker*** *within the automated LIPS, compared to the same version manually handled.*

- ***Task-unrelated MW increased significantly in the automated condition*** *after some time on task, which could be due to complacency or a loss of agency, or both.*

- *Low perceived mental demand supported the* ***underload hypothesis*** *regarding the vigilance decrement.*

- ***Pupil diameter decreased during task-unrelated MW****, compared to when being focused.*

- *This reduction was* ***not affected by time nor condition****, suggesting a stable marker of task-unrelated MW.*

5.1 Description of the current chapter

In the chapter 3, pilots' reports pointed a link between task-unrelated MW and problems experienced in the cockpit. In this chapter, we report an experiment to compare MW dynamic in manual and automated environments. Subsection 5.2 - Dynamic of mind wandering when facing automation provides a short context and points the important results of Casner and Schooler (2015) on the same subject. We expose tools used for the experiment in the subsection 5.3 - Material and methods (general explanations and adaptations for this experiment, mainly Task Load Index and Icarus working position; see MATERIAL AND METHODS USED FOR EXPERIMENTATIONS 4 for more details). Subsections 5.4 - Results analysis and 5.5 - Discussion on the results highlight data analysis and a discussion on the implications. We provide a summary of the work presented in the thesis so far in the last subsection 5.6 - Thesis progress recap (chapter 5).

5.2 Dynamic of mind wandering when facing automation

5.2.1 Mind wandering in manual and automated environments

As seen in the previous chapters, OOTL performance problem represents an important challenge for automation designers. While implementing higher levels of automation indeed improves the efficiency and capacity of a system, it also creates new challenges for human operators. MW may play an important role in such problems. Indeed, MW is more likely to occur in monotonous environments (Eastwood et al., 2012). Its occurrence favors a decoupling from the ongoing task at perceptual and stimuli processing levels (Kam et al., 2012; Schooler et al., 2011), which can be seen both on behavioral and physiological data. Reading tasks were particularly used to uncover the influence of MW on oculometric markers like blink frequency (Smilek et al., 2010b), fixation duration and saccade frequency (Uzzaman & Joordens, 2011). In simulators, Yanko and

Spalek (2014) studied MW influence on driving performance. They observed a longer reaction time to unexpected events, a shorter headway distance and a higher velocity. Their results were corroborated by other studies in driving environments (Dündar, 2015; He et al., 2011; Lerner, Baldwin, Higgins, Lee, & Schooler, 2015).

Casner and Schooler (2015) studied the impact of automation on MW in an aeronautical context. Their results on 16-minute sessions did not show a significant correlation between automation and the frequency of MW reports. However, their experiment allowed pilots to choose the automation level freely. Pilots could have used this possibility to match automation level with MW propensity to stay focused. Moreover, the propensity to mind wander appeared to increase when everything seemed under control. Supervising ultra-reliable systems could encourage operators to decrease cognitive resources allocated to the monitoring task. In that context, time saved by automation, which should normally be used for other productive tasks and for monitoring, could instead be filled by task-unrelated thoughts. Operators in such a state would not be prepared to regain manual control over the system in response to rare critical events. Such analysis is already considered in the debate regarding the origin of the vigilance decrement (Fraulini, Hancock, Neigel, Claypoole, & Szalma, 2017; Pattyn et al., 2008; Thomson, Besner, & Smilek, 2016), recent evidences showing that both phenomena share many features (see chapter 2).

5.2.2 Leading hypotheses for the experiment

In this experiment, we focus on the difference of MW dynamic between manual and automated environments. We believe automation might influence MW during sessions longer than in Casner and Schooler' (2015) study when operators cannot adapt the automation level. Our hypotheses are: (1) MW frequency increases in automated environments, and (2) oculometric signal allows discriminating attentional states of MW or focusing on the task. Our experiment addresses these hypotheses.

5.3 Material and methods

5.3.1 Participants

Seventeen participants (five female) performed the experiment (ages ranging from 21 to 42; *M* = 27.3, *SD* = 6.0). We performed an *a posteriori* power analysis using η^2 reported by Unsworth and Robison (2016). This study was the only one, to our knowledge, to investigate MW and oculometry using linear models, which was our intention (see the subsection 5.3.3 - Data collection and filtering). They reported $\eta^2 = 0.32$ and $\eta^2 = 0.34$ for the influence of time on MW frequency and pupillometry respectively. Using the *pwr* function in R (Champely, 2017; R Core Team, 2016), we calculated $N_{lim} = 15$ (i.e., we needed at least 15 participants) for standard values of $alpha = .05$, $beta = .80$ for two-tailed tests.

The participants enrolled in this study were volunteers from our company (ONERA, the French Aerospace Lab). All participants were unfamiliar with the concepts used in this study. They had normal or corrected-to-normal visual acuity. All participants signed a written declaration of informed consent. The protocol was approved by the ONERA organization and was conducted in accordance with the World Medical Association's Declaration of Helsinki.

5.3.2 Task description

5.3.2.1 Environment used

We used the Icarus working position for this experiment (see Figure 27; SIMPIT Technologies, 2018). We considered the position both adapted for the context of UAV piloting and comfortable for participants. Our laboratory acquired this working position as an interface prototyping environment, and not specifically for this thesis. Therefore, the hardware and configuration choices of this working position are out of this thesis.

The 16/9 21'' central screen displayed the LIPS interface powered by a regular DELL PC. The 16/9 9'' secondary screen on the right displayed attentional probes using a Raspberry Pi. Both screens allowed touchable interactions. Cameras of the eye-tracker were placed above the central screen, on the left of the left screen and on the right of the right screen.

We used an obstacle avoidance task programmed with the LIPS (Laboratoire d'Interactions Pilote-Système, or Pilot-System Interactions Laboratory; see subsection 4.1 – The obstacle avoidance task).

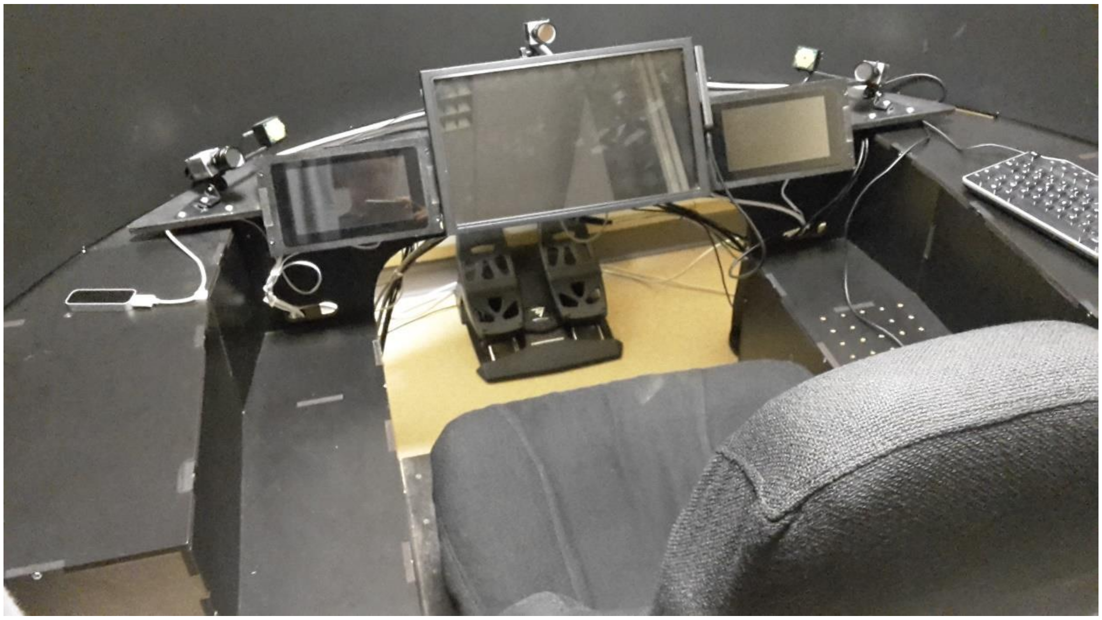


Figure 27. Icarus environment.

Central screen is used to display the LIPS; right screen is used to display questionnaires. Cameras are places above central screen, on the left of the left screen and right of the right screen. Illuminators are joined with left and right cameras (SIMPIT Technologies, 2018).

5.3.2.2 Conditions "Manual" and "Automated"

Both conditions consisted in avoiding incoming objects (each avoidance consisting in a "conflict"), either manually or supervising the autopilot doing it. Clusters of obstacles appeared along the way (every 45s on average). Each cluster could contain between 1 to 5 obstacles, including one on the trajectory.

The first condition was called “Manual” and required participants to avoid manually obstacles. The system-detected obstacles on the trajectory 13s before impact, at which point an orange circle appeared around the UAV and the participant could initiate an avoidance maneuvers. Participants were able to choose the way in which they wished to avoid the obstacle by clicking on “*Evitement Gauche*” (left maneuver) or “*Evitement Droite*” (right maneuver). Once they clicked, the simulator turned the trajectory of the UAV on the chosen side, following a predefined angle. Each obstacle had a safe circle similar to that of the UAV (see Figure 28). A collision warning (i.e., an orange circle around both the UAV and the obstacle with the message “Collision”) was displayed if the UAV safe circle penetrated the obstacle safe circle. A trial with a collision warning triggered was marked as failed. To resume the initial trajectory, the participants were required to click on the “*Retour trajectoire*” (return to original trajectory) button. If no action was taken within 16 seconds after the first change in trajectory, the aircraft automatically resumed the trajectory and the trial was marked as failed.

The second condition was called “Automated”. Participants were required to monitor the system avoiding obstacles. They were instructed to click an “*Acquittement*” (acknowledgement) button to acknowledge automated avoidance decisions as soon as they noticed it (twice per trial, once to acknowledge avoidance of the object and once to acknowledge the return to the normal trajectory after avoiding the object). A feedback message was displayed to the participants. The acknowledgement ensured that participants would have the same motor input under both the manual and automated conditions. If participants detected an automation error, i.e., choosing the wrong obstacle avoidance trajectory, they were instructed to click the button “*Changement d’altitude*” (change altitude) so that the UAV would perform an emergency descent. A feedback message was displayed in that case as well. The altitude change ensured that participants were facing a supervision task.

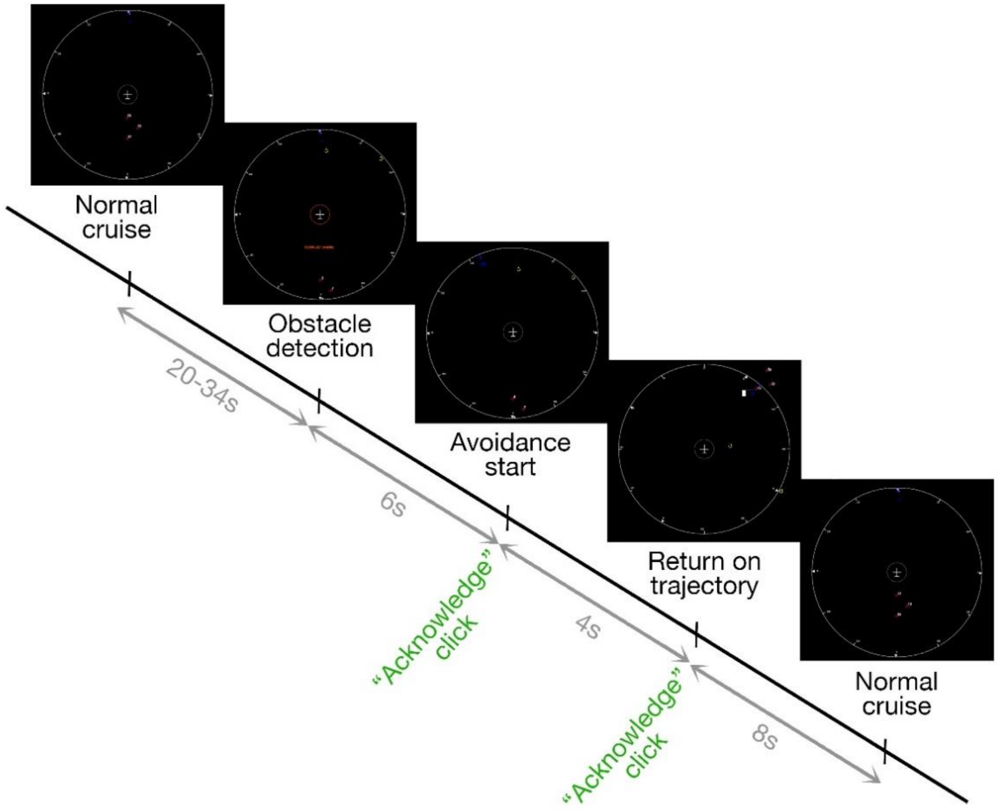


Figure 28. Step by step explanation of a trial in the "Automated" condition.

The UAV moves forward cruising without events for 27s on average. The automated pilot will detect any obstacle along the way and decide which way to go (left or right). Once it decides the direction, participants must click on "Acknowledge". When the automated pilot decides that the obstacle is not on the trajectory anymore, it heads to the next checkpoint and participants must once again click on "Acknowledge". However, when the automated pilot choses the wrong side, participants must click on "Altitude Change" to avoid the collision. At any randomly selected moment, an experience-sampling probe may appear.

5.3.2.3 Attentional probes

On average every 2 minutes, attentional probes appeared on the 10-inch right secondary screen of the Icarus working position displayed attentional probes. These probes enquired for where was participants' attention directed. Participants could answer "On the task", "Something related to the task", "Something unrelated to the task" or "External distraction" (see the subsection 4.3

- Attentional probes for more details). For technical reasons, the obstacle avoidance task was not paused when the probe was displayed. Participants were asked to fill it as soon as it appeared, and any successful or failed trial during this interval would not be taken into account. Participants were informed that the probe was not part of the evaluation to lower the impact of instructions on their natural propensity to mind wander.

5.3.2.4 Task-Load Index questionnaire

We used a validated French version of the NASA Task Load Index (TLX) questionnaire to evaluate task load (see appendix I; Cegarra & Morgado, 2009; Hart & Staveland, 1988). This questionnaire includes questions pertaining to mental load, time pressure, physical strain, effort, frustration, and perceived performance. Participants were asked to answer each question using a horizontal line, ranging from 0 to 20.

We were primarily interested in the “perceived mental demand” dimensions for three reasons. Firstly, trying to measure “workload” or “task load” is hazardous as the terms themselves are open to criticism (Dekker & Hollnagel, 2004). Secondly, we wanted in subsequent experiments to measure perceived mental demand (see chapitre 6). Finally, we wanted to focus on this term as an important one in the underload – overload debate (Pattyn et al., 2008; Warm et al., 2008).

Each participant filled two TLX questionnaires (one after each session). Although a TLX questionnaire completed at each block would allow precise workload monitoring, we believe that MW would have been artificially lower due to the disruption. Therefore, the TLX was only filled at the end of each session.

5.3.2.5 Procedure of the experiment

We made explicit to participants that detection accuracy was more important than speed in button clicks. Each participant performed the two conditions on two separate days in a counterbalanced way. Each day started with an explanation of the task, followed by a 10-minute training period and a 44-

minutes session under the proper condition. Each session contained 60 clusters of obstacles. Each cluster was considered a trial. They were separated by 45 seconds on average. 25 probes were answered under each condition. The distribution of probes was not correlated with events on the obstacle avoidance task in order to avoid performance to influence MW reports (Head & Helton, 2016). The manual and automated condition included respectively seven and eight conflicts with a probe within the 10-seconds interval following the conflict.

Participants encountered one automation error (where they had to click on the "*Changement d'altitude*" button) during training for the "Automated" condition and another during the automated condition at the end of the third block. Under the manual condition, participants encountered at the end of the third block a conflict impossible to avoid. No attentional probe followed both the automation error and this conflict for at least 10 seconds after.

5.3.3 Data collection and filtering

We used R-Studio 1.0.143 and R 3.4.1 (R Core Team, 2016; RStudio Team, 2015) to analyze behavioral and oculometric data.

5.3.3.1 Attentional probes

Considering that the two available screen on the Icarus environment were two Raspberry Pi, we used Python 3.6 to program attentional probes. We used Comma Separated Value (CSV) text files to store all answers. The file contained the exact appearance time along with each answer, in order to synchronize probes data with the pupillometric signal.

5.3.3.2 Oculometry

The 10 seconds preceding each probe were extracted from oculometric data. Extracts before "On the task" and "Something related to the task" were classified as "Focus" to avoid any influence of poor performance on mind wandering related measures (Head & Helton, 2016). Extracts before "Something unrelated to the task" were classified as "MW". Extracts before "External distraction" were discarded as noise.

We performed filtering on pupillometry using R and the procedure detailed in the subsection 4.4.3 - Procedure for oculometric data pre-processing. Firstly, we filtered the signal. Pupil diameter had to be between 1 and 10 mm, had to be less than 80% different from the preceding value and had to be of a quality (computed by the SmartEye software) over 0.01. Extracts were discarded if their resulting pupil diameter series consisted of less than 70% compliant values. The proportion of extracts excluded due to low quality (9.6%) is in line with that excluded in other investigations (Smallwood et al., 2011). Resulting extracts were completed using basic linear interpolation. A second filtering pass was applied with a median filter (moving window of 50 frames). Finally, the data of each participant were normalized by subtracting the mean and dividing by the root mean square of all good-enough quality extracts for this participant.

5.4 Results analysis

All linear mixed-effect analyses used the *lme* (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017) function to create the models. We chose this function since it has been widely used in the literature of mixed-effect models and handles a wide range of data structures (Field, Miles, & Field, 2012). We used the *Anova* (R Core Team, 2016) function to assess the influence of predictors. We chose type 2 sum of squares, or type 3 sum of squares when there were one or more interactions between predictors to consider. Contrary to type 1 sum of squares, type 2 and type 3 are independent of predictor order in the model and thus provide more robust results.

All confidence interval (*95% CI*) reported hereafter were computed using the *boot* package (Canty & Ripley, 2017; Davison & Hinkley, 1997) with 10000 iterations under the first under normal bootstrap approximation.

5.4.1 Mind wandering frequency evolution

We split the 40-minute sessions into four blocks lasting approximately 10 minutes and containing five reports each. MW propensity was calculated as a percentage of all reports in the block (see Figure 29). Participants reported MW episodes for 41% of the probes (*95% CI* = [36; 45] %). This rate is consistent with previous studies on the subject (Kam et al., 2011; Smallwood & Schooler, 2006, 2015). Participants reported on average 4% "Noise" reports. Thus, we discarded these reports and approximated attentional states to being either focus, task-related MW or task-unrelated MW.

We performed a linear mixed-effect model analysis with subject intercepts and by-subject random slopes for the effect of condition (no other random slope due to convergence problems of the model). Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 5, bold values being significant.

Without specific *a priori* predictions on the evolution of MW frequency over time, we conducted Tukey HSD post-hoc tests on the complete model, including the block variable for each condition separately. For the manual condition, all differences were non-significant (*p* > .366). For the automated condition, the third (*M* = 64%, *95% CI* = [51; 76] %) and fourth blocks (*M* = 62%, *95% CI* = [50; 73] %) had significantly higher MW frequency compared to the first block (*M* = 32%, *95% CI* = [22; 42] %), *p* = .001, *d* = 0.54 and *p* = .003, *d* = 0.32, respectively. Similarly, third and fourth blocks had significantly higher MW frequency compared to block 2 (*M* = 36%, *95% CI* = [28; 45] %), *p* = .007, *d* = 0.12 and *p* = .016, *d* = 0.12, respectively.

Table 5. Influence of block and condition on task-unrelated MW frequency.

		Task-unrelated MW	
Effect added	**df**	χ	***p*-value**
Block	3	**28.23**	**< .001**
Condition	4	**21.84**	**< .001**
Block:Condition	6	**13.24**	**.004**

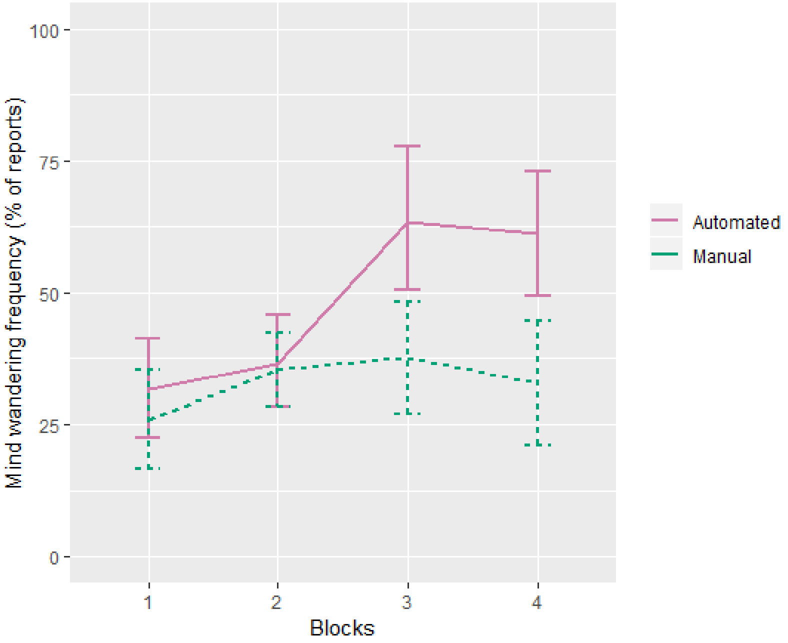


Figure 29. MW frequency evolution for each condition.

Error bars show the 95% confidence intervals based on bootstrapping

5.4.2 Perceived mental demand scores between conditions

Each participant filled in two TLX questionnaires (one after each session). The mean score for perceived mental demand for each subject (see Figure 30) varied substantially (ranging from 1 to 14, *M* = 6.50, *95% CI* = [5.21; 7.80]). Shapiro-Wilk's test indicated that the assumption of normality had been violated for the TLX values, *W* = .918, *p* = .015. Therefore, we used Wilcox's robust version of the *t*-test proposed in the WRS2 package (Mair, Schoenbrodt, & Wilcox, 2017). On average, participants perceived that the automated (*M* = 4.76, *95% CI* = [3.31; 6.22]) condition required less cognitive resources than the manual (*M* = 8.23, *95% CI* = [6.39; 10.07]) condition, *t*(10) = -3.35, *p* = .007, *d* = 0.78. Ratings show that our automated condition succeeded in lowering perceived mental demand. Subsequent analysis revealed no influence of the order of conditions on the ratings (e.g. participants going through the automated condition first did not report significantly different mental demand than did participants going through the automated condition after the manual one).

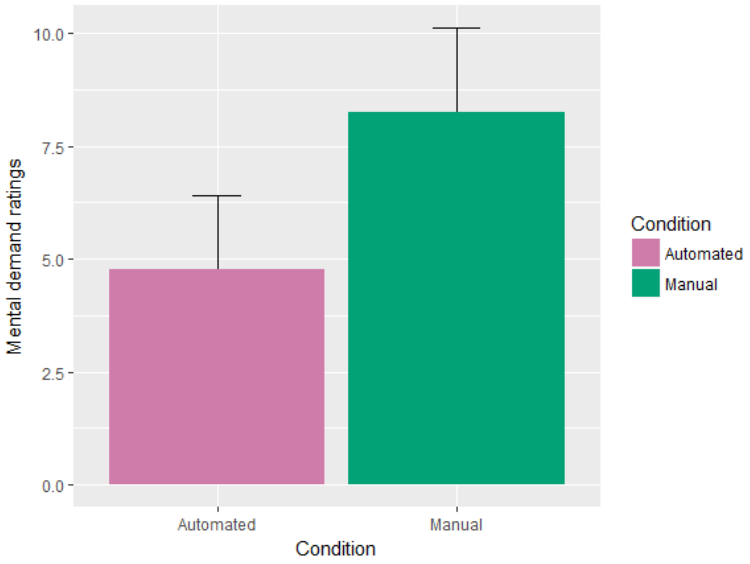


Figure 30. Mental demand ratings for each condition.

Error bars show the 95% confidence intervals based on bootstrapping.

5.4.3 Oculometric measures between attentional states

5.4.3.1 Influence of attentional states on oculometric measures

We discarded the epochs containing participant actions. Following this procedure, eight reports were discarded under the automated condition and nine under the manual condition. We plotted pupil diameter during the 25 seconds preceding reports (see Figure 31). This confirmed that our 10-second interval was appropriate.

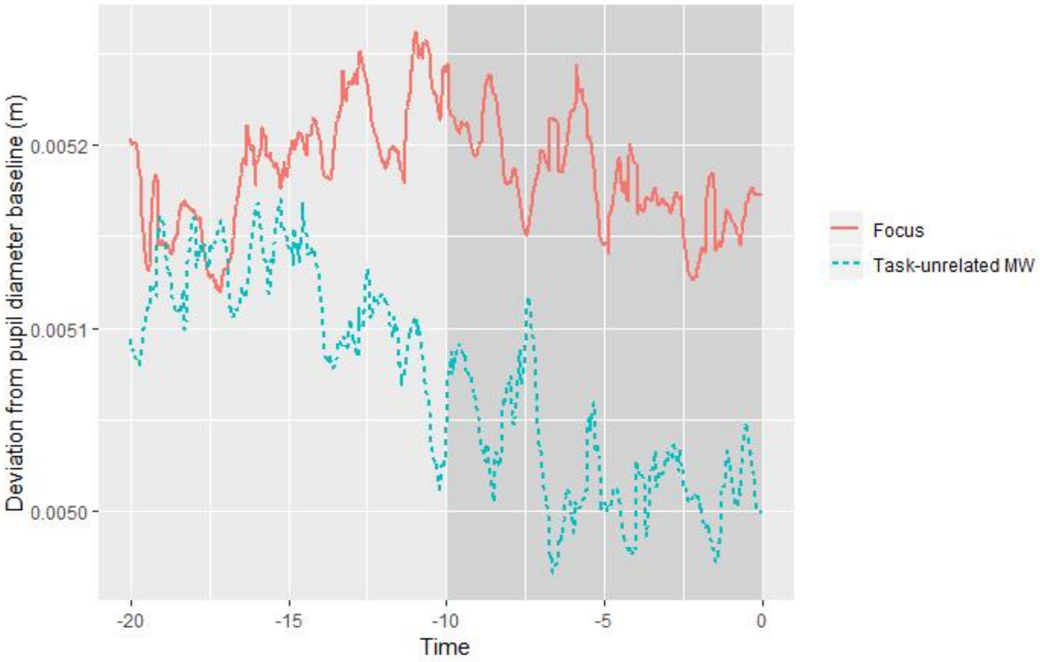


Figure 31. Normalized pupil diameter evolution during the 25-second interval preceding probes display for each attentional state.

The grey part of the signal is used for computation.

We performed a linear mixed-effect analysis to assess the influence of attentional states on oculometric markers. We defined random intercepts for subjects and random slopes for conditions (no other random slope possible due to convergence problems of the model). Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. We reported generalized eta squared (η^2_G) for effect size using *aov_car* (Singmann, Bolker, Westfall, & Aust, 2018). This metric provides comparability across between-subject and within-subject designs (Bakeman, 2005; Olejnik & Algina, 2003). The results are shown in Table 6. On average, participants showed a significantly smaller pupil during MW episodes.

Table 6. Influence of attentional states on oculometric measures

Parameter	MW values		Focus values		Attentional State model		
	M	*95% CI*	*M*	*95% CI*	χ (1)	*p*-value	η^2_G
Pupil size (mm)	5.02	[4.90; 5.20]	5.18	[5.10; 5.30]	**4.47**	**.035**	**.010**
Saccade frequency (sacc/s)	3.95	[3.62; 4.28]	3.70	[3.37; 4.03]	0.13	.717	.002
Mean fixation duration (s)	0.29	[0.19; 0.39]	0.39	[0.27; 0.50]	0.69	.407	< .001
Blink frequency (blink/s)	0.07	[0.06; 0.09]	0.06	[0.04; 0.07]	3.36	.067	.020

Table 7. Influence of time and condition on the oculometric marker difference between attentional states.

Parameter	Time model			Condition model			Time/Condition interaction model		
	χ(3)	*p*-value	η^2_G	χ(1)	*p*-value	η^2_G	χ(3)	*p*-value	η^2_G
Pupil size	1.51	.680	.06	1.72	.189	.04	3.27	.352	//
Saccade frequency (sacc/s)	1.52	.677	.36	<0.01	.974	<.01	0.25	.970	//
Mean fixation duration (s)	6.16	.104	.14	0.07	.780	.05	2.94	.400	//
Blink frequency (blink/s)	5.30	.151	.06	1.66	.197	.03	.829	.843	//

Note. η^2_G could not be computed for the interaction model because of uneven number of reports in each block (rejected data during filtering).

5.5 Discussion on the results

We studied the impact of automated compared to manual environments on MW and its behavioral and oculometric markers. The automated condition revealed significantly lower TLX scores compared to the manual condition, showing a protocol in line with the usual goals regarding automation introduction (Wiener, 1988). Three main results have been shown: (1) MW increases after some time has elapsed in an automated environment, (2) there is a difference in pupil diameter between MW and focus episodes but not for other oculometric markers and (3) pupillometric difference between attentional states remains stable across times and conditions. We discuss these results below.

5.5.1.1 Influence of time and automation on oculometric differences

We looked for any influence of time or automation on oculometric markers using a linear-model analysis. As fixed effects, we entered block, condition and their interaction. As random effects, we had intercepts for subjects but no random slope for convergence problems. Visual inspection of residual plots did not reveal any obvious deviations from normality and homoscedasticity. η^2_G could not be computed for the interaction model due to an uneven number of reports in each block (rejected data during filtering). Results are displayed in Table 7. Oculometric markers remained stable through time and condition.

5.5.2 Complacency or agency may explain mind wandering frequency

5.5.2.1 Comparing mind wandering frequency during both conditions

The first result is a significant increase in the MW frequency under the automated condition between Blocks 2 and 3. No significant time-related evolution of MW was observed under the manual condition. Since both conditions lasted the same amount of time, contained a similar number of actions and pursued the same goal (avoid incoming obstacles), time-related phenomena (drowsiness, habituation, or tiredness) cannot entirely explain the fact that MW increased only under the automated condition. Similarly, the level of automation alone cannot explain the observed data, since the trend did not evolve linearly with time-on-task and showed no difference between conditions for the first two blocks (task complexity remained constant throughout each condition). Finally, since we only took intervals without actions, the MW report difference could not have been biased by a desire to justify low performance (Head & Helton, 2016).

On the one hand, the absence of MW increase under the manual condition is interesting, considering the well-established vigilance decrement observed during sustained attention (Cabon, Coblentz, Mollard, & Fouillot, 1993; D. R. Davies & Parasuraman, 1982; Jeroski, Miller, Langhals, & Tripp, 2014; Mackworth, 1948). Moreover, multiple previous studies exist showing an increase of MW frequency with time (McVay & Kane, 2012; Thomson et al., 2014; Unsworth & Robison, 2016). An explanation may lie in the difference between our protocol and traditional protocols studying mind wandering. Studies reported mostly GO/NOGO and reading tasks. GO/NOGO tasks use frequent stimuli to determine how mechanical behavior induced by MW could affect performance. These tasks have low-complexity. Participants quickly realize how much cognitive resource to put into the task and become rapidly familiar with stimuli. The second common task, reading, forces participants to be active and read the given text. It is a familiar activity for most subjects. Unlike for the GO/NOGO and the reading task, participants in our experiment were completely unfamiliar with the LIPS. Trials were extensive (at least 40 seconds per trial), which also contributed to the impossibility of the task becoming familiar within a few minutes.

On the other hand, the mechanism responsible for the increase in MW frequency under the automated condition should have prevailed over this possible effect of unfamiliarity. There are two constructs, maybe complementary, which may account for this interaction between time and level of automation over MW frequency.

5.5.2.2 Complacency to explain mind wandering frequency

First, complacency might be generated by the high reliability of the system and lower monitoring performance. Complacency is an issue of monitoring automation generated by an uncritical reliance on the system (Parasuraman et al., 1993b). Complacency has been linked to longer reaction time (Bahner, Hüper, & Manzey, 2008; Manzey et al., 2006), loss of situation awareness (Endsley & Kiris, 1995) and failures of detection (Parasuraman et al., 1993b).

In our experiment, participants encountered no error during the first three blocks. Given that the system never committed any miss or error, participants may have thought that it would remain perfectly reliable. In this context, their perception of the required workload might evolve: since the automated system does not seem to require their attention to function properly, participants would redirect their cognitive resources towards more personal matters and mind wander more. The higher perceived workload under the manual condition supports our analysis. Moreover, this could explain why participants, who were novices in supervising the system, exhibited an increase in MW frequency only after some time, while pilots with thousands of hours flying time in the study by Casner and Schooler (2015) experienced MW immediately without temporal evolution. These evidences suggest a mediating influence of system familiarity in MW frequency temporal evolution.

This position would introduce a third possibility within the overload/underload theory debate (Pattyn et al., 2008; Warm et al., 2008). Although the task complexity does not change, the operator’s perception could evolve based on their perception of the system and the overall situation (e.g., trust, familiarity). As pointed out by Seli and colleagues (Seli, Carriere, & Smilek, 2015; Seli, Risko, & Smilek, 2016), there is strong evidence that people can exert some control over their MW. This is in accordance with Casner and Schooler’s (2015) results, who demonstrated that cognitive resources freed by automation in peaceful situations are not allocated to planning the mission ahead, but rather to MW. Moreover, our analysis is in line with studies that observed MW increase in a low-probability-signal environment (Berthié et al., 2015; Galera et al., 2012), with the time elapsed performing the task (McVay & Kane, 2009; Smallwood, Baracaia, Lowe, & Obonsawin, 2003; Smallwood, Riby, Heim, & Davies, 2006) and the view of complacency as a multiple-task strategy (Bahner et al., 2008; Moray & Inagaki, 2000). Operators save cognitive resources allocated to the low-event automated task, in order to perform better on another task (MW), which is considered more interesting or useful, regardless of experiment instructions.

5.5.2.3 Agency to explain mind wandering frequency

The second possible explanation is a disengagement from the task linked to a loss of agency. When dealing with automation, operators give up their direct control over the system for a monitoring role in the supervisory control loop (Moray, 1986b; Sheridan, 1992). Such new role decreases their sense of agency (i.e., the ability to feel in control; (Obhi & Hall, 2011; Wegner, 2002). Multiple studies pointed to a limit in the automation level beyond which users felt less in control (Berberian et al., 2012; Coyle, Moore, Kristensson, Fletcher, & Blackwell, 2012), leading to a form of disengagement from the task at hand (Haggard, 2017).

Interestingly, Szalma (2014) described a similar disengagement when applying the Self-Determination Theory (Ryan & Deci, 2000) to human-system interactions. The inability of a system to support autonomous behavior may lower motivation and create an externalization of task goals (i.e., a process by which operators reject the value of a goal). In our experiment, given that participants did not validate, but rather only acknowledged the system's actions, they initially experienced a loss of agency, causing a decrease in their motivation, leading to a faint sense of responsibility. This process chain could lead participants to reallocate cognitive resources from the task to MW, unconsciously trying to optimize time and mental resources from their perspective. Further studies are needed to distinguish the respective impacts of agency decrease and complacency on MW emergence.

5.5.3 Pupil diameter is lower during mind wandering

Our second result concerns oculometric measures. Given that we only took intervals without actions, pupil diameter signals were only influenced by tonic pupil activity (i.e., the sustained component of the pupillary response, expressed as an absolute pupil diameter, as opposed to phasic activity reflecting changes linked to events). We highlighted a lower pupil diameter during MW, as did several studies on MW (Faber et al., 2017; Grandchamp et al., 2014; Mittner et al., 2014). Moreover, the literature already showed a link between a smaller pupil and periods of lower sensibility to external stimuli (K. McIntire et al., 2014; Nishiyama et al., 2007). Taken together, these results are in line with the view of MW as a phenomenon inducing a decoupling from the environment. However, other research linked large pupils to slow and inaccurate responses (Gilzenrat et al., 2010; Smallwood et al., 2011), or more directly to MW during a word-by-word reading task (Franklin et al., 2013). Addressing this debate, two recent studies by Unsworth and Robison (2016) and Konishi et al. (2017a) observed an inverse U-curve relationship between pupil diameter and performance. In both experiments, smaller pupil diameter was linked to drops in performance and MW episodes associated with internally directed cognition. In contrast, a larger pupil diameter was correlated with external distractions (e.g., conversation, noise, itching) and was accompanied by a decrease in performance. These results corroborate our study and stress the need to investigate these attentional states.

In contrast to pupillometry, other oculometric measures did not exhibit significant sensitivity to MW. This could be due to the relatively low number of participants (N = 17), which yielded statistical power levels of .82, .53 and .49 for saccade frequency, blink frequency and fixation duration models, respectively (all calculated with standard values of = .05 for two-tailed tests). However, one should be careful with an *a posteriori* power analysis using output data, as pointed out by Hoenig & Heisey (2001). Nevertheless, many notable studies have used a number of participants within our range (Smilek et al. (2010b), 12 subjects; Uzzaman and Joordens (2011), 22 subjects; He et al. (2011), 18 participants; Franklin et al. (2013), 13 participants; Braem et al. (2015), 20 participants). They used reading tasks (with the notable exception of Grandchamp et al., 2014), which make extensive use of eye movement. Our task does not require constant saccades between words, since obstacles only appear at a slow rate. Our result could point to important task mediators of MW influence on oculometric markers, such as event rate or cognitive demands. However, we need more studies to make sure that these results are not due to a lack of participants.

5.5.4 Pupil diameter decrease remains stable

Finally, the last result is the stability of pupillometric markers with respect to automation and time. Cheyne and colleagues (2009) recently proposed the integration of intensity of environment decoupling as a characteristic of MW episodes. They used a Sustained Attention to Response Task (SART, a form of GO/NOGO task; see (Robertson et al., 1997) to match errors and reaction time evolution with each level of their model. If this model were true, there is little doubt that physiological markers would show some sensibility to intensity of MW. However, no influence on oculometric markers emerged. Several explanations could explain this result. First, our protocol, which differ from previous protocols, may not be able to uncover such a tendency. Second, intensity may not regulate MW impact on pupillometry. Third, there may not be any intensity in MW episodes, each inducing the same environment decoupling. Indeed, the study Cheyne and colleagues (2009) falls under the concerns expressed by Head and Helton (2016), see next paragraph). Further neuronal studies are necessary to answer this question.

5.6 Thesis progress recap (chapter 5)

Our results show that automation increases MW frequency after some time for non-pilot participants. The MW literature in ecological tasks already highlighted how the phenomenon increases the risks in critical environments. Moreover, we demonstrated the possibility to track MW through its pupillometric signature in a complex ecological task.

In compliance with the conjectures previously detailed, we will now investigate the influence of reliability (as one of the most important factors impacting trust and complacency) on MW rate.

6 STUDY OF THE ATTENTIONAL DECOUPLING WITH AUTOMATION VARYING IN RELIABILITY

- *We investigated the **decoupling from the environment** induced by MW and the link between **system reliability** and MW increase over time.*

- *Task-unrelated MW frequency **increased with time** in both conditions, but not task-related MW frequency.*

- *Perceived mental demand, pupil diameter and blink frequency were **lower during task-unrelated MW**, compared to focus, highlighting a new evidence for the decoupling hypothesis.*

- *Perceived mental demand and pupil diameter were **lower during task-related MW** compared to focus, supporting the possibility of intensity as a dimension of MW episodes.*

- ***Trust was not influenced** by attentional states, suggesting that complacency does not play a role in mind wandering emergence.*

6.1 Description of the current chapter

The previous experiment demonstrated an increase in MW frequency during automated conditions compared to manual conditions. Based on the literature (Bahner et al., 2008; Berberian et al., 2012; Pattyn et al., 2008), we suggested two mechanisms as to explain this result: complacency and loss of agency. In this experiment, we want to explore the link of reliability (as the basic requirement for complacency) with MW frequency and the decoupling induced. We detail important concepts on the subsection 6.2 - Influence of mind wandering on operational safety. Subsection 6.3 - Material and methods points to modifications brought to the last experiment according to our hypothesis (mainly reliability of the LIPS and attentional probes). Without surprises, subsections 6.4 - Results analysis and 6.5 - Discussion and conclusion of the experiment provide detailed analysis and discuss how these fit with the current literature. We provide a summary of the work presented in the thesis so far in the last subsection 6.6 - Thesis progress recap (chapter 6).

6.2 Influence of mind wandering on operational safety

6.2.1 Trust in automation may influence mind wandering

An important issue concerns the automation features causing an increase in MW frequency. Automated environments are generally repetitive and monotonous with very few target events, all characteristics known to increase MW. However, they are not the only features of automation that could influence MW.

Among the most important characteristics of automated systems, reliability is considered as one of the causes of the observed vigilance decrement in OOTL situations (Metzger & Parasuraman, 2001). Low reliability may lower the trust operators have concerning the system abilities. Low trust in turn could create skepticism and lead operators to spend too many resources in verifying the validity of automation outputs, or even disengage automation (Andersson, 2008). On the contrary, the regularly pointed paradox of ultra-safe systems is that the absence of any failure for a prolonged period of time will lead operators to make commission errors (i.e. accept automation recommendation despite the fact that it may be wrong, Amalberti, 2001). This phenomenon is called automation-induced complacency (Raja Parasuraman, Molloy, & Singh, 1993a). Complacency is the adoption of a non-optimal information sampling behavior based on overtrust regarding system's capabilities due to a minimization of automation failure probability (Innes-Jones & Scandpower, 2012; Moray & Inagaki, 2000). Even though it can emerge unconsciously, complacency may be a multiple task strategy to optimize the global output when supervising an automated system while also performing a more engaging task. However, this strategy can sometimes lead to dramatic failures in safety critical environments. Multiple meta-studies reported complacency as being one of the main reasons for an important number of crashes (Funk et al., 1999; Raja Parasuraman & Riley, 1997; Wiener, 1981). Complacency may lead operators to disengage from the task and reallocate their cognitive resources to more personal matters, increasing MW frequency, explaining the results of the first experiment.

6.2.2 Mind wandering negatively impact short-term performance

A second issue concerns the impact of MW on safety. One of the most threatening aspect of MW for safety is the decoupling from the environment (Schooler et al., 2011). Operators engaged in an episode of MW will see their encoding of external information degraded. MW disrupts visual information flow by reducing pupil diameter (K. McIntire et al., 2014; Smallwood et al., 2011) and increasing blink frequency (Smilek et al., 2010b). Neuronal studies demonstrated during MW an increase of alpha waves power, linked with sensory suppression (Foxe & Snyder, 2011; O'Connell et al., 2009), and a reduction of Event Related Potentials linked to external information perception and processing (Smallwood, Beach, et al., 2008).

At the behavioral level, this decoupling translates into a decrease in short-term performance. Reaction exhibits higher variability (Bastian & Sackur, 2013), while omissions and anticipations are more common (Cheyne, Carriere, Solman, & Smilek, 2011). Accuracy was shown to decrease in both simple paradigms (Kam et al., 2012) and more ecological ones (Yanko & Spalek, 2014). These evidences demonstrate that MW disrupts online adjustment of behavior (Kam et al., 2012). Eventually, information processing impairment will flow into higher cognition levels and affect operators' model of the situation. Particularly, MW induced decoupling might lead supervisors to disengage from the task and overlook some failures, leading to OOTL performance problem. Such disengagement should be observed both at the behavioral and physiological levels.

6.2.3 Leading hypotheses for the experiment

Even though multiple studies have investigated MW induced perceptual decoupling, no attempt has been made, to our knowledge, to do so when supervising automation. We report in this chapter an experiment on the evolution and consequences of MW within an automated environment of varying reliability. Our hypotheses are that (1) higher reliability increases task-unrelated MW and (2) MW induced decoupling impacts operators' engagement (perceived mental demand and oculometric signal) in operational conditions.

6.3 Material and methods

6.3.1 Participants

16 participants (3 female) performed the experiment (age ranging from 22 to 43 years old; *M* = 29.0, *SD* = 5.8). The participants enrolled in this study were volunteers from our company (ONERA, the French Aerospace Lab). All participants had normal or corrected-to-normal visual acuity. All participants were unfamiliar with the concepts at hand and the LIPS environment. All participants signed a written declaration of informed consent. The procedure was approved by ONERA and conducted in accordance with the World Medical Association Declaration of Helsinki.

6.3.2 Task description

6.3.2.1 Environment used

We used an obstacle avoidance task programmed with the LIPS (Laboratoire d'Interactions Pilote-Système, or Pilot-System Interactions Laboratory). Participants had to avoid incoming obstacles by supervising an UAV in autopilot mode (see Figure 32; for more details, see 4.1 – The obstacle avoidance task). The LIPS was displayed on the left screen.

6.3.2.2 Conditions "Risky" and "Safe"

Both conditions were based on the automated mode introduced in the previous experiment (see subsection 5.3.2.2 - Conditions "Manual" and "Automated").

Participants were required to monitor the autopilot avoiding obstacles. Clusters of obstacles appeared along the way (every 45s on average). Each cluster could contain between 1 to 5 obstacles, including one on the trajectory. When an obstacle was present on the trajectory (this situation is called "conflict"), the autopilot detected it and initiated a deviation automatically. Participants had to click on an "*Acquittement*" (acknowledgement) button to acknowledge automated avoidance decisions as soon as they saw it (twice per trial, once to acknowledge avoidance of the object and once to acknowledge the return to normal trajectory after avoiding the object). A feedback message was displayed to the participants. Finally, if participants detected an incoming collision warning, they were instructed to click on the button "*Changement d'altitude*" (change height) so that the UAV would perform an emergency descent to avoid colliding with the obstacle. Collisions could occur during the avoidance trajectory, if there was another obstacle on the bypass trajectory chosen by the autopilot. Two conditions were proposed.

Figure 32. SmartEye setup for the experiment.

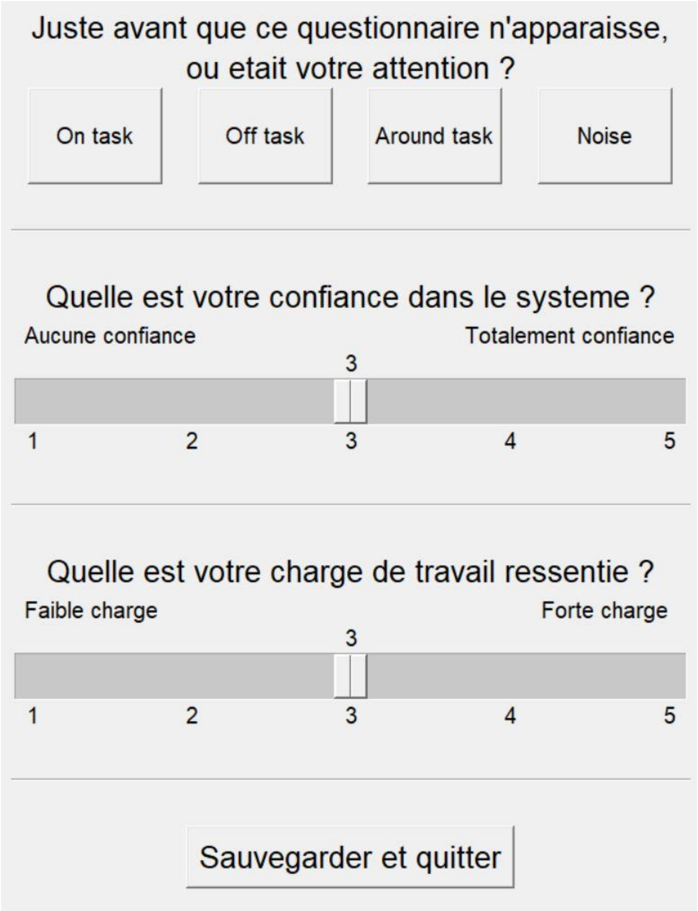
Three cameras and two illuminators are mounted on a rail to keep elements immobile relative to each other. The rail has also Velcro elements to avoid it to move on the table. The three cameras are placed to enhance signal quality relative to the left screen (LIPS screen).

Under the “Risky” condition, the autopilot made an error (choosing the wrong side) leading to a collision in 40% of the trials (27 errors in total) selected randomly. This number was chosen so that there would be a significant number of collisions, while keeping the automated system performance above the chance expectation (50%). Under the other “Safe” condition, the autopilot made 7% errors. All decisions and collisions were predefined and, therefore, they were the same for all subjects.

6.3.2.3 Attentional probes

On average every 2 minutes, an attentional probe appeared on the right screen see Figure 32 and Figure 33). These probes enquired for where was participants’ attention directed. Participants could answer “On the task”, “Something related to the task”, “Something unrelated to the task” or “External distraction” (see the subsection 4.3 - Attentional probes for more details).

The second item was “how much do you trust the system?” We made explicit that this was the trust in the ability of the system to perform its task without errors. Answers ranged from “no trust” to “total trust” on a 5-point Likert scale. Finally, the third item was: “what is your perceived workload?” Even though we wanted to measure mental demand, we used the word “workload” (“*charge de travail*” in French) because the term is generally understood by everyone. Nevertheless, we made explicit to particpants that the question enquired about the amount of mental ressources needed to fulfill the objectives. Perceived mental demand was measured as an important aspect of task engagement (Raja Parasuraman & Riley, 1997). We refer to this item as “perceived mental demand” throughout the rest of the paper. Answers ranged from “low mental demand” to “high mental demand” on a 5-point Likert scale.



Juste avant que ce questionnaire n'apparaisse, ou etait votre attention ?

Quelle est votre confiance dans le systeme ? 3

Quelle est votre charge de travail ressentie ? 3

Figure 33. Screenshot of the experience-sampling probes in French.

6.3.2.4 Procedure of the experiment

Participants were explicitly instructed that detection accuracy was more important than speed of response. Each participant performed under the two conditions on two separate days in a counterbalanced way. Each day started with an explanation of the task, followed by a 10-minute training period and a 50-minute session under the proper condition. Each session contained 67 clusters of obstacles, totaling 201 obstacles. Each cluster contained between one and five obstacles, including one on the trajectory. Clusters were separated by 45 seconds on average. 20 probes were responded to under each condition. The distribution of the experience-sampling probes was not correlated with events on the obstacle-avoidance task, in order to minimize performance influence on experience-sampling reports (Head & Helton, 2016). The “Risky” condition included six conflicts with a probe presented within the 10-second interval following the conflict, while the “Safe” condition included seven conflicts with a probe presented within the 10-second interval following the conflict.

Under the “Risky” condition, the autopilot made an error (choosing the wrong side) leading to a collision in 40% of the trials (27 errors in total), selected randomly. Under the other “Safe” condition, the autopilot made five errors (7% errors; errors on trials 24, 40, 56, 62 and 64). Each condition contained 67 clusters of obstacles. All decisions and collisions were predefined and, therefore, they were the same for all subjects.

6.3.3 Data collection and filtering

We used R-Studio 1.0.143 and R 3.4.1 (R Core Team, 2016; RStudio Team, 2015) to filter and analyze the data.

6.3.3.1 Attentional probes

Python 3.6 was used to program attentional probes with Comma Separated Value (CSV) text files to store all answers from each session with each subject. The exact appearance time was saved along with each answer, in order to synchronize the questionnaire data with the oculometric signal.

6.3.3.2 Oculometry

The 10 seconds preceding each probe were extracted from oculometric data. Extracts before “On the task” and “Something related to the task” were classified as “Focus” to avoid any influence of poor performance on mind wandering related measures (Head & Helton, 2016). Extracts before “Something unrelated to the task” were classified as “MW”. Extracts before “External distraction” were discarded as noise.

We performed filtering on pupillometry using R and the procedure detailed in the subsection 4.4.3 - Procedure for oculometric data pre-processing. Firstly, we filtered the signal. Pupil diameter had to be between 1 and 10 mm, had to be less than 80% different from the preceding value and had to be of a quality (computed by the SmartEye software) over 0.01. Extracts were discarded if their resulting pupil diameter series consisted of less than 70% compliant values. The proportion of extracts excluded due to low quality (5.5%) is in line with that excluded in other investigations (Smallwood et al., 2011). Resulting extracts were completed using basic linear interpolation. A second filtering pass was applied with a median filter (moving window of 50 frames). Finally, the data of each participant were normalized by subtracting the mean and dividing by the root mean square of all extracts of good-enough quality for this participant.

6.4 Results analysis

All linear mixed-effect analyses used the *lme* function to create the models. We used the *Anova* function to assess the influence of predictors. We chose type 2 sum of squares, or type 3 sum of squares when there were interactions to consider between predictors. Post-hoc tests were conducted using the *glht* and *mes* of R to perform Tukey HSD on the complete model.

All confidence interval (*95% CI*) reported hereafter were computed using the *boot* package with 10000 iterations under the first under normal bootstrap approximation.

6.4.1 Mind wandering frequency analysis

We split the 50-minute sessions into five blocks of 10 minutes containing five experience-sampling probes each. Participants reported on average 27.6% task-related MW (*95% CI* = [24; 31] %) and 36% task-unrelated MW (*95% CI* = [31; 41] %). This rate is consistent with previous studies (Kam et al., 2011; Smallwood & Schooler, 2006, 2015). Each participant reported on average 3% external distractions, in line with the previous experiment. Therefore, we approximated attentional state variable as a ternary state (i.e., as being either in focused, task-related MW and task-unrelated MW states).

We investigated the first hypothesis (influence of trust on MW rates) by looking at task-related and task-unrelated MW frequency evolution over time and conditions (see Figure 34 and Figure 35). We performed a linear mixed-effect analysis. We considered Blocks as a 5-level categorical variable. We defined a random intercept for subjects to consider our repeated-measure design. No random slope was possible because of convergence problems. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 8, bold values being significant.

Table 8. Influence of blocks and condition on task-related and unrelated MW frequency.

		Task-related MW		Task-unrelated MW	
Effect added	**df**	χ	***p*-value**	χ	***p*-value**
Block	4	8.71	.069	**14.50**	**.006**
Condition	5	0.69	.406	0.42	.518
Block:Condition	9	**12.28**	**.015**	4.57	.358

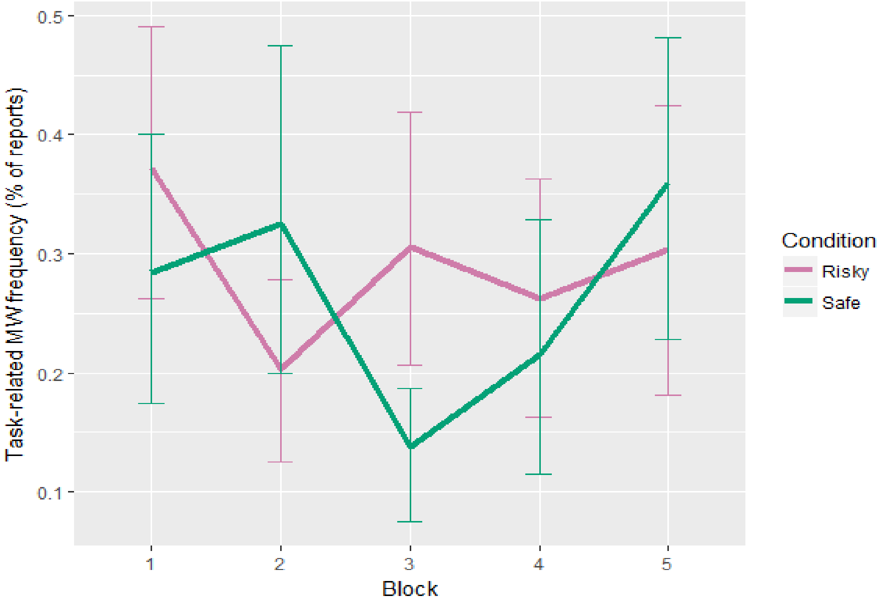


Figure 34. Task-related MW frequency evolution for each condition.

Error bars show the 95% confidence intervals based on bootstrap.

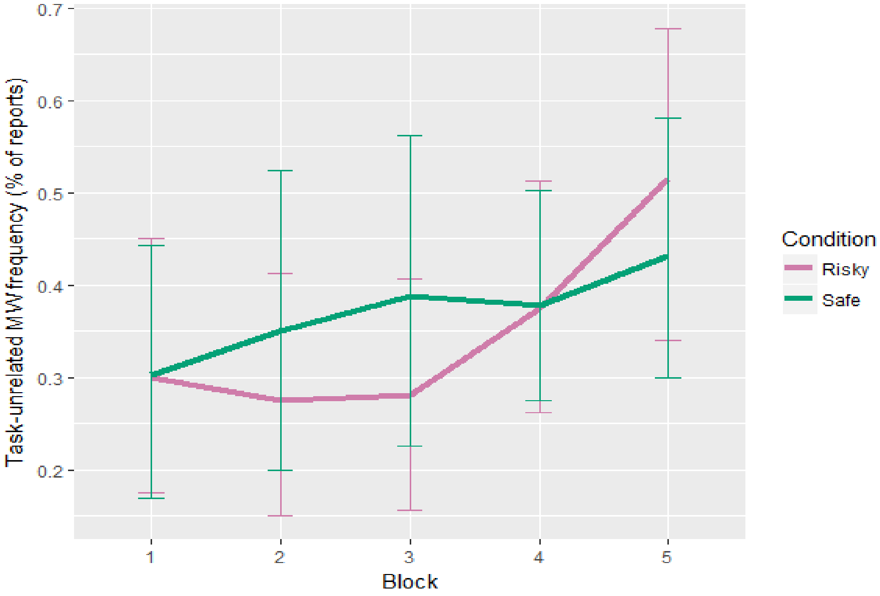


Figure 35. Task-unrelated MW frequency evolution for each condition.

Error bars show the 95% confidence intervals based on bootstrap.

Blocks did not significantly influence task-related MW. There was a significant interaction between blocks and conditions, χ = 12.28, *p* = .015. Without specific *a priori* predictions regarding the block-by-block evolution, we conducted Tukey's post-hoc tests on the complete model. Tests revealed that task-related MW frequencies were significantly higher under the "Risky" condition during block 1 (*M* = 37%, *95% CI* = [27; 48] %) compared to the "Safe" condition during block 3 (*M* = 14%, *95% CI* = [8; 20] %), *p* = .010, *d* = 1.31. However, with only this significant result, no general trend can be observed regarding task-related MW in both condition. We can only say that task-related MW seems to decrease in the middle of the "Safe" condition, whereas no particular trend can be seen for the "Risky" condition.

Task-unrelated MW frequency also changed with time-on-task independently of conditions, χ = 14.50, *p* = .006. Without specific *a priori* predictions regarding the block-by-block evolution, we used Tukey HSD for post-hoc tests. It revealed that task-unrelated MW frequencies were significantly higher in block 5 (*M* = 47%, *95% CI* = [36; 59] %) compared to block 1 (*M* = 30%, *95% CI* = [20; 40] %), *p* = .006, *d* = 0.55, Block 2, *p* = .013, *d* = 0.51, and Block 3, *p* = .047, *d* = 0.43. This demonstrates a significant increase in the task-unrelated MW frequency towards the end of each session, which is consistent with the existing literature (Krimsky et al., 2017). On the contrary, task-unrelated MW did not show any influence by the condition on its levels, nor on its timely evolution. Given that conditions varied with regard to reliability and elicited varying trust (see the following analysis of trust ratings), this result argues against any influence of trust on task-unrelated MW levels.

We continued our analysis by looking at correlations between task-unrelated MW rates, trust and perceived mental demand for each subject. We performed a linear mixed-effect analysis. We defined a random intercept using "Subjects" and a random slope using "Condition". Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 9, bold values being significant.

Table 9. Influence of trust and perceived mental demand on task-unrelated MW frequency

Effect added	**df**	χ	***p*-value**
Trust	1	0.02	.895
Perceived mental demand	2	2.48	.115

Overall, the analysis of the task-unrelated MW frequency showed that there was no significant interaction between trust ratings nor perceived mental demand ratings with task-unrelated MW frequency. However, task-unrelated MW frequency increased significantly at the end of the session for both conditions.

6.4.2 Influence of condition and attentional state on trust

Trust ratings varied substantially between subjects (ranging from 2.12 to 4.58, *M* = 3.38, *95% CI* = [3.30; 3.46]). We continued to investigate our first hypothesis (influence of trust on MW rates) by looking at the trust evolution between conditions and attentional states (see Figure 36). We performed a linear mixed-effect analysis with type 2 sum of squares. We defined a random intercept using “Subjects” and a random slope using “Condition”. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 10, bold values being significant.

Table 10. Influence of predictors on trust ratings

Effect added	**df**	χ	***p*-value**
Condition	1	**14.18**	**< .001**
Attentional State	3	4.47	.512
Condition:Attentional State	5	2.09	.663

The difference in system reliability significantly influenced trust, since trust ratings reported during the “Risky” condition (*M* = 2.93, *95% CI* = [2.82; 3.04]) were significantly lower than during the “Safe” condition (*M* = 3.82, *95% CI* = [3.73; 3.92]), *b* = -0.95, *t* (766) = 4.75, *p* < .001. On the contrary, attentional states did not significantly influence trust, χ = 4.47, *p* = .512. In order to determine whether the absence of difference was due to a lack of power, we computed the Type II error using the *pwr* function (Champely, 2017) and *lmer* function (Bates, Maechler, Bolker, & Walker, 2017, p. 4) given that the *lme* function did not provide the necessary information. Computation yielded a probability for type II error *p* < .001, which indicated a very low risk of accepting the null hypothesis, even though there was a significant effect (however, see the critics of *a posteriori* power analysis using the data by Hoenig and Heisey (2001)). As expected, manipulating system reliability modified trust in the system capabilities. On the contrary, attentional states demonstrated no influence on trust ratings. Subsequent analysis revealed no influence of the order of conditions on the ratings.

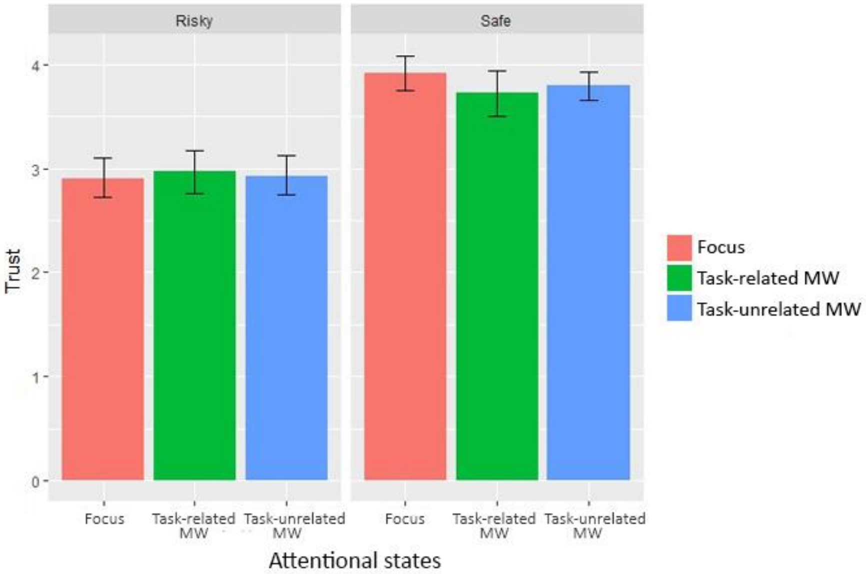


Figure 36. Trust for each condition and attentional state.

Error bars show the 95% confidence intervals based on bootstrap.

6.4.3 Influence of condition and attentional state on perceived mental demand

Perceived mental demand ratings varied between subjects (ranging from 1.02 to 3.39, *M* = 1.78, *95% CI* = [1.71; 1.83]). We investigated our second hypothesis (decoupling hypothesis within automated environments) by looking at perceived mental demand evolution between conditions and attentional states. We performed a linear mixed-effect analysis with type 2 sum of squares. We defined a random intercept for subjects and a random slope for condition. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 11, bold values being significant.

The difference in system reliability produced a significant effect on perceived mental demand (see Figure 37). Perceived mental demand were significantly lower during “Safe” Condition (*M* = 1.66, *95% CI* = [1.58; 1.75]) than during “Risky” condition (*M* = 1.88, *95% CI* = [1.79; 1.98]), *b* = -0.23, *t*(766) = -2.32, *p* = .021. Our protocol validate our hypothesis of decrease perceived mental demand when working with reliable automation. Similarly, we used Tukey’s post-hoc tests to break down the effect of attentional states. Perceived mental demand reports when focused (*M* = 1.97, *95% CI* = [1.86; 2.08]) were significantly higher than those associated with task-related MW (*M* = 1.75, *95% CI* = [1.64; 1.86]), *p* = .029, *d* = 0.25, and task-unrelated MW (*M* = 1.59, *95% CI* = [1.49; 1.69]), *p* < .001, *d* = 0.43. However there was only a non-significant tendency for perceived mental demand reports associated with task-unrelated MW to be lower than those associated with task-related MW, *p* = .073, *d* = -0.19.

Table 11. Influence of predictors on perceived mental demand ratings

Effect added	df	χ	*p*-value
Condition	1	**5.94**	**.015**
Attentional State	3	**23.97**	**< .001**
Condition: Attentional State	5	0.89	.827

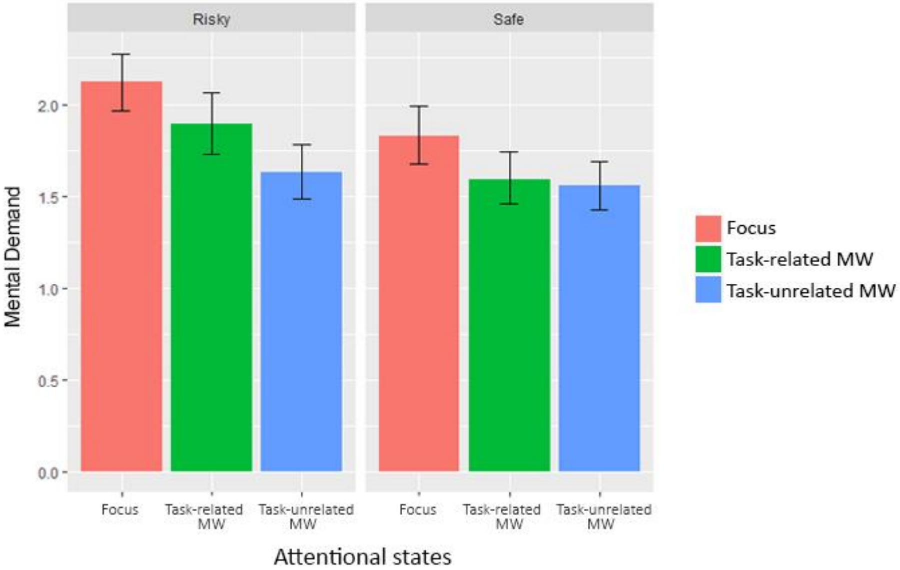


Figure 37. Perceived mental demand for each condition and attentional state.

Error bars show the 95% confidence intervals based on bootstrap.

Even though the interaction between condition and attentional states was not significant, the Figure 37 seemed to show an absence between conditions only for task-unrelated MW. We computed the 95% CI for perceived mental demand during task-unrelated MW, which substantially overlapped between conditions, *95% CI* = [1.48; 1.78] for "Risky" and *95% CI* = [1.42; 1.69] for "Safe". On the contrary, perceived mental demand during focus episodes *95% CI* did not overlap as much, *95% CI* = [1.97; 2.27] for "Risky" and *95% CI* = [1.68; 1.98] for "Safe", nor during task-related MW, *95% CI* = [1.72; 2.07] for "Risky" and *95% CI* = [1.46; 1.73] for "Safe".

Both the lower perceived mental demand during task-unrelated MW compared to other attentional states and the absence of difference between conditions support a decoupling from the task. Subsequent analysis revealed no influence of the order of conditions on the ratings.

6.4.4 Influence of attentional state on oculometry

In order to investigate our second hypothesis (decoupling hypothesis within automated environments) from the physiological aspect, we looked at oculometric data through attentional states. After looking at pupil diameter data, we took the 10 seconds preceding each questionnaire for further analysis. We computed a linear mixed-effect analysis. We defined a random intercept for subjects. No random slope was possible because of the convergence problems due to the quantity of data. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 12, bold values being significant.

Table 12. Influence of time and condition on oculometric markers

	Focus values		Task-related MW values		Task-unrelated MW values		Attentional state (AS)	
Parameter	***M***	***SD***	***M***	***SD***	***M***	***SD***	**χ (2)**	***p*-value**
Pupil size (mm)	3.93	0.69	3.88	0.62	3.75	0.58	**7.48**	**.024**
Saccade frequency (sacc/s)	2.25	1.37	2.10	1.11	2.16	1.38	1.74	.418
Mean fixation duration (s)	0.65	1.28	0.56	0.87	0.56	0.83	0.53	.767
Blink frequency (blink/s)	0.08	0.09	0.09	0.12	0.11	0.16	**7.77**	**.021**

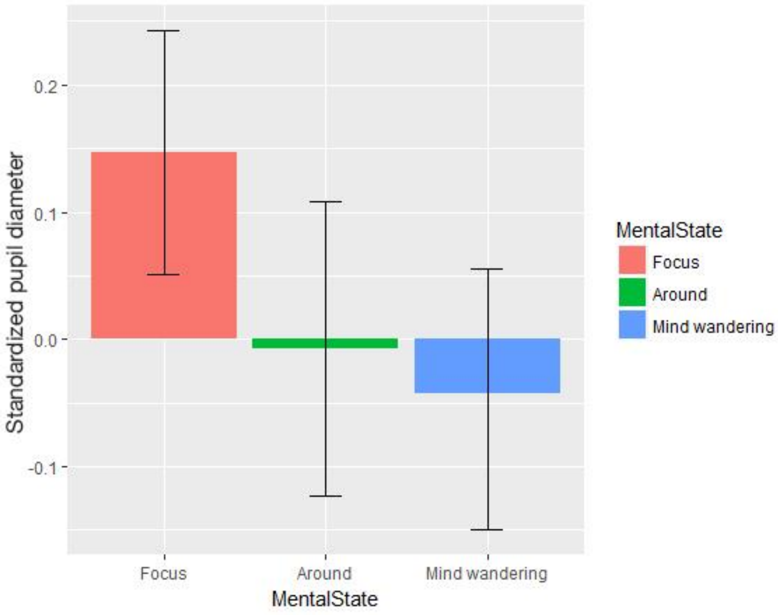


Figure 38. Pupil diameter standardized for each attentional state.

Error bars show the 95% confidence intervals based on bootstrap.

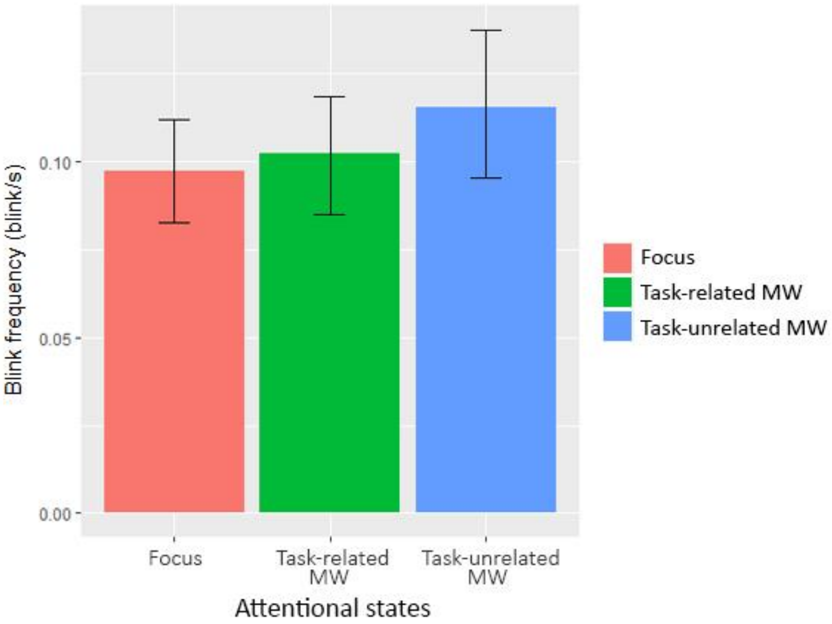


Figure 39. Blink frequency for each attentional state.

Error bars show the 95% confidence intervals based on bootstrap.

Attentional states showed a significant influence on pupil size, χ (4) = 7.97, *p* = .019 (see Figure 38). Without specific *a priori* predictions on the evolution of pupil diameter through attentional states, we conducted Tukey's post-hoc tests on the model. We saw that pupil diameter when focused was significantly higher than during task-related MW, *p* = .036, *d* = 0.08, and task-unrelated MW, *p* = .005, *d* = 0.30. On the contrary, blink frequency was significantly higher during task-unrelated MW than when focused, *p* = .012, *d* = 0.11 (see Figure 39).

In other words, oculometric measures revealed that pupil diameter decreased and blink rate increased when subjects' thoughts were distant from the task. On the contrary, no influence of attentional states was observed for saccade frequency and mean fixation duration.

6.5 Discussion and conclusion of the experiment

We studied the impact of automation reliability on task-unrelated MW frequency and the influence of the MW induced perceptual decoupling on task engagement. We reproduced the increase in task-unrelated MW observed in the last experiment (see 5.4.1 - Mind wandering frequency evolution). Our protocol succeeded in inducing significant differences in trust and perceived workload ratings.

Three main results have been shown: (1) task-unrelated MW induced a decoupling from the task that lowered engagement, (2) the expression of perceptual decoupling extended to task-related MW episodes and (3) task-unrelated MW propensity was not linked with trust in the system reliability. We discuss these results below.

6.5.1 Task-unrelated mind wandering decouples from the task

The first result is the behavioral and physiological evidences supporting an impact of the perceptual decoupling associated with task-unrelated MW on operators' engagement. According to the decoupling hypothesis (Schooler et al., 2011), our mind decouples attention from sensory information to sustain prolonged MW. With minimum impact of external information, it becomes dramatically more difficult for operators to perceive and encode external information during task-unrelated MW episodes. We highlighted the effects of this perception decoupling on perceived mental demand, pupil diameter and blink frequency.

Firstly, perceived mental demand decreased when participants reported task-unrelated MW. Participants may have experienced a reduced sensitivity to the characteristics of the task and not updated their perceived mental demand. Another possibility is that they answered the probes with limited attention, again relying on information gathered while they were focused. Either way, participants did not spend more cognitive resources on updating their mental model of the situation. This could explain why task-unrelated MW has been shown to disrupt online adjustment of behavior (Kam et al., 2012), since participants might have been operating with an out-of-date model of the situation. Thoughts not directly linked to current task decisions also decreased pupil diameter. This is in line with studies investigating the trade-off between exploration-exploitation (Jepma & Nieuwenhuis, 2011). Indeed, MW is a characterized state of exploitation of information already acquired (e.g., memories) and does not use sensory information except for its ignition point (Seli, Risko, & Smilek, 2016).

Moreover, the literature already linked a lower pupil baseline to periods of lower sensibility to external stimuli (K. McIntire et al., 2014; Nishiyama et al., 2007). It should be noted that some studies highlighted a higher pupil baseline during task-unrelated MW (Franklin et al., 2013; Smallwood et al., 2011). Nevertheless, two recent studies by Konishi et al. (2017a) and Unsworth and Robison (2016) observed an inverse U-curve relationship between the pupil diameter and performance. A smaller pupil diameter was linked with a decrease in performances and MW episodes as internally directed cognition. In contrast, a larger pupil diameter was correlated with external distractions (e.g., conversation, noise, or itching) and was accompanied with a decrease in performance. These studies corroborate our results, while explaining apparent contradictory results.

Finally, blink frequency increased during MW episodes. Blinks are known to disrupt visual information processing on two levels: they occlude the retina and they trigger cortical deactivation of the areas responsible for visual information processing (Bristow, Frith, & Rees, 2005). Overall, these three measures support the decoupling induced by task-unrelated MW, for both the behavioral and physiological aspects. Far from being anecdotal, the perception of task demands by operators disengaged from the task was found to not be aligned with reality, and these might be unable to perform efficiently. This could lead to automation issues, as described by Parasuraman and Riley (1997). If they had the possibility of performing some tasks or letting the automation handle it, their inaccurate evaluation of the situation may lead them to either choose to handle something manually even though they do not have the cognitive resources for it (disuse), or let the automation do it despite some previous errors (misuse).

6.5.2 Task-related mind wandering may also create a decoupling

Our second result is the extension of the decoupling evidence to the task-related MW. Both perceived mental demand and pupil diameter were significantly lower when participants reported task-related MW compared to being focused. All measures influenced by attentional states (perceived mental demand, pupil diameter and blink frequency) showed the same linear pattern, placing measures linked to task-related MW between those associated with being focused and with task-unrelated MW. Such results are supported by the three-state engagement model of MW (Cheyne et al., 2009). This model proposes three states of MW corresponding to three intensities of decoupling from the task. The model revealed consistent temporal associations between performance and MW levels. The model also revealed bidirectional effects between MW and performance, suggesting that MW can lower performance via the decoupling effect, but also that poor performance can create task-related MW. However, one must remain cautious about the extension of the decoupling hypothesis to MW that includes thoughts related to the task. Blink frequency, which was significantly different between the “Focus” and MW states, was not significantly different between the “Focus” and “Around” states. Further studies are needed to assess the range of thoughts inducing perceptual decoupling, and whether MW episodes indeed possess a depth-modulating perceptual decoupling.

Overall, our results contradict Head and Helton (2016). Even though we took into account their results in different aspects of our protocol, we found proof of the perceptual decoupling in both behavioral and physiological levels. Three aspects of the paradigm freed us from the rationalization hypothesis consequences. First, the distribution of experience-sampling probes was not correlated with events during the obstacle avoidance task, in order to minimize performance influence on experience-sampling reports. Second, we only kept in our analysis epochs without actions (intervals where participants did not click on any button). Third, we introduced among the attentional probe answers the possibility of reporting "task-related MW", which we treated separately. Nevertheless, our results remain in line with the literature, supporting the decoupling hypothesis for both task-related and task-unrelated MW.

6.5.3 Task-related mind wandering may not be linked to trust

Finally, our third result concerns the converging evidence that task-unrelated MW frequency may not be linked to trust. Correlation tests did not show any association between task-unrelated MW frequency and trust ratings. Multilevel regression showed no influence of attentional states on trust ratings with significantly low type II error. Even though we cannot assert that trust is not linked to attentional states, our result supports this hypothesis. This result may seem in contradiction to the results obtained from the questionnaire, which seemed to display a negative correlation between problems in the cockpit and perceived usability of the autopilot (see chapter 3). However, the questions measured usability in general, and not trust. When isolating the question "I trust the autopilot", there was no obvious tendency in one direction or another.

It is possible that our paradigm failed to highlight the influence of reliability on MW. Complacency may have a dynamic necessitating more time to take place. Operators generally are subjected to thousands of working hours when supervising their system, whereas in this case we tested novices. Investigating experts in similar settings could reveal different results. On the other hand, the behavioral correlates of complacency are not well defined, as it is a broad concept (see 2.2.2 - Complacency as a possible explanatory mechanism). The possibility exists that other concepts could be linked to complacency and mediate its link with MW. For example, we measured "trust", but what about "perceived reliability"?

Nevertheless, we still observed an increase in MW frequency, corroborating the results of the first experiment (see chapitre 5). If not trust, another possible candidate mecanism would be the impact of a loss of agency (see subsection 2.2.3.1 - Agency and the decoupling of human observer from the task). Agency is the feeling of control produced by the idea that our actions are producing the observed effect. Several studies showed that one's feeling of agency decreased as automation level increased, compared to the same task done manually (Berberian et al., 2012; Obhi & Hall, 2011). Knowing that a decrease in the feeling of agency leads to the operator's disengagement from the task (Haggard, 2017), human operators might disengage from the task and allocate a lower amount of cognitive resources to the task. Resources could then be used for task-unrelated MW maintenance. This hypothesis is tightly linked with motivation and the Self-Determination Theory (Ryan & Deci, 2000; Szalma, 2014). Even though participants were volunteers, the task proposed was purposely boring and did not produce much motivation. The inability of the task to support autonomous behavior and internalization of the goal may lower motivation and create an externalization of task goals (i.e., a process by which operators reject the intrinsic value of a goal). Ultimately, participants could voluntarily redirect their attention and cognitive resources whenever possible toward more personally interesting and useful matters, increasing MW frequency. Further studies building on agency and task-unrelated MW literatures should investigate this hypothetical link.

6.6 Thesis progress recap (chapter 6)

Contrary to our hypothesis, reliability did not influence MW rates. However, we observed an important disengagement of the operator from the task from both behavioral and oculometric aspects. Coupled with the significant increase of task-unrelated MW in automated environment and its influence on reported detection problems, our results underline the importance of decoupling characterization needed to prevent OOTL performance problems. Therefore, we will now investigate the possibility to use neuronal markers of MW in complex environments.

7

STUDY OF NEURONAL MARKERS OF MIND WANDERING IN OUT-OF-THE-LOOP SITUATIONS

- *We investigated the* ***viability of MW neuronal markers*** *in complex ecological bimodal environments with the LIPS and random beeps.*

- *There was* ***no increase*** *of both task-defined MW, although more* ***task-unrelated MW*** *emerged during the second block.*

- ***N1 ERP component*** *had lower amplitude during task-unrelated MW, while* ***P3 component*** *had higher amplitude during task-related MW, compared to other attentional states.*

- ***Alpha waves*** *activity was higher in parieto-occipital regions during task-unrelated MW, while* ***ASSR amplitude*** *was not influenced by attentional states.*

- *Results underline the* ***complex influence*** *of the MW perceptual decoupling on operators behavior in ecological environments.*

7.1 Description of the current chapter

The last two experiments (detailed in chapters 5 and 6) examined MW dynamic in automated environments and its influence on operators' engagement in the task. In this chapter, we delve into neuronal measures to explore the possibility to track MW (and therefore OOTL performance problem risk) without disrupting the operator. We firstly review the studies investigating MW in multimodal environments in the subsection 7.2.1 - Mind wandering influence on multimodal sensory integration. We present in the subsection 7.2 - Neuronal why EEG signal may help to go further in such environments, closer to operational ones. We detail the use of electroencephalogram setup chosen in the subsection 7.3 - Material and methods. Finally, results are detailed and analyzed in both subsections 7.4 – Results and 7.5 - Discussion and conclusion of the experiment. We provide a summary of the work presented in the thesis so far in the last subsection 7.6 - Thesis progress recap (chapter 7).

7.2 Neuronal markers in automated environments

7.2.1 Mind wandering influence on multimodal sensory integration

OOTL performance problem is difficult to study in laboratories. The phenomenon can have multiple causes, necessitate time to emerge and is strongly influenced by the environment. Studies try more and more to switch from simple tasks to simulators in order to limit the gap between their findings and real conditions. Nevertheless, with complexity comes difficulty to analyze measures. Previous attempts to study MW influence on multimodal tasks have only measured behavioral markers, sometimes also oculometric signal (Casner & Schooler, 2014, 2015; Durantin et al., 2015; Yanko & Spalek, 2014). Unfortunately, behavioral measures like accuracy measurement and subjective reports are difficult to integrate, and constrain paradigms while disrupting participants. Oculometric signal is sensitive to a wide range of internal states, making it difficult to attribute a modulation to one specific change in operators' internal state. Neuronal measures could overcome these issues and provide a more accurate picture of MW influence on operators. Kam et al. (2011) reported the first attempt to measure MW influence on EEG signal during a task with multimodal stimuli. However, the target stimuli were only visual, while beeps irrelevant to the task were here to assess MW emergence. When considering the influence of MW perceptual decoupling, the fact that a modality supports target stimuli may highly influence its inhibition. To our knowledge, no previous attempt has been made to study MW in environments with multimodal target stimuli.

7.2.2 Neuronal aspects of mind wandering decoupling

Neuronal measures can provide much information directly coming from the very structures involved in internal states changes. Researchers could use the markers of MW already identified. Kam and colleagues pioneered the study of MW through Event Related Potentials (ERPs), i.e. neuronal responses to discrete events. In a series of experiments using a visual SART (Kam et al., 2011), they showed that the amplitude of P1, N1 and P3 components (respectively associated with visual perception, auditory perception and external stimuli processing) were all lower during MW. This effect held true whether stimuli were the SART stimuli or irrelevant to the task. Kam and colleagues similarly highlighted lower P3 component during a time-estimation task (Kam et al., 2012), and demonstrated that the lower processing of external stimuli induced by MW also impacted the emotional aspects (Kam, Xu, & Handy, 2014).

On top of ERPs, MW literature also identified the influence of MW on brain waves. Particularly, MW is linked with an increase in alpha waves power (10-12Hz) in the occipital sites during a visual task (O'Connell et al., 2009), but with lower alpha during a breath counting with eyes closed (Braboszcz & Delorme, 2011). Such results can be explained by the inhibitory function of alpha waves (Benedek et al., 2014; Bonnefond & Jensen, 2012; Villena-González, López, & Rodríguez, 2016). As MW induces a decoupling from the task, it could effectively deactivate occipital sites (involved in visual perception) for a visual task, but stop the alpha wave inhibition of these areas when the eyes are closed. Several aspects of brain waves relation to MW remain to be explored. However, other results could come from Steady State Responses.

7.2.3 Steady-state auditory-evoked potential

A steady-state response (SSR, also called steady-state evoked potential) is an evoked potential emerging from periodical external stimulus and whose phase and amplitude remains constant (Picton, John, Dimitrijevic, & Purcell, 2003). In practice, researchers set a periodical signal with a fixed frequency, which will amplify the EEG spectral power around the stimulus frequency. SSR response can be distinguished from natural brain waves in that they are externally triggered and more stable. Auditory Steady-State Responses (ASSR) have higher evoked power at frequencies between 40 and 45 Hz (Geisler, 1960). The technique is nowadays used to objectively assess hearing ability for a wide range of populations (Korczak, Smart, Delgado, Strobel, & Bradford, 2012).

ASSR revealed interesting properties related to attention. Picton et al. (2003) highlighted lower ASSR amplitude during drowsiness, a result later reproduced by Griskova et al. (2007). Saupe et al. (2009) investigated the influence of sensory modality and attention on ASSR amplitude. In their experiment, participants listened to bursts of 500 Hz tones modulated at 40 Hz while looking at a cross in the middle of a screen. In the “attend” condition, participants had to listen to the auditory stimuli and push a button whenever the modulation frequency changed to 30 Hz. In the “non-attend” condition, participants had to ignore auditory stimuli and push a button whenever the cross changed in size. Saupe et al. (2009) highlighted a significant increase in ASSR amplitude when participants’ attention was focused on the auditory modality (“attend” condition). They later reproduced the SSR amplification with shifting attention between VSSR and ASSR (Keitel, Schröger, Saupe, & Müller, 2011).

More directly related to MW, O’Connell et al. (2009) investigated the possibility to use VSSR to track MW in real-time. Participants watched a 25 Hz VSSR in the form of patterned stimuli where they detect a stimulus lasting 40% longer than the others. They assumed that detection failures in such a simple task would result from attention lapses, i.e. a form of MW. Results showed no difference in VSSR amplitude between correct detection and misses. It could be that MW only inhibits transient stimuli and not periodical ones, or that the paradigm used did not allow pointing such inhibition.

7.2.4 Leading hypotheses for the experiment

Our purpose in this experiment is to demonstrate the viability of MW neuronal markers in complex ecological automated environments. Our hypothesis are (1) attentional decoupling caused by MW influences ERP components, (2) MW attentional decoupling can also be observed in alpha waves increase and (3) ASSR can be seen on EEG signal in complex environments.

7.3 Material and methods

7.3.1 Participants

18 participants (12 female, all right-handed) performed the experiment (age ranging from 21 to 45 years old; *M* = 25, *95% CI* = [22; 29]). The participants enrolled in this study were volunteers from our company (ONERA, the French Aerospace Lab) or the Marseille University. They received 20€ vouchers for the experiment. All participants had normal or corrected-to-normal visual acuity. We verified prior to the experiment that all participants could hear the beeps (same fixed volume for all participants). All participants signed a written declaration of informed consent. The procedure was approved by ONERA and conducted in accordance with the World Medical Association Declaration of Helsinki.

After pre-processing the data, three subjects were discarded, either because they did not fully understand the task, or because their EEG data was too noisy. Data analysis is conducted with 15 subjects.

7.3.2 Task description

7.3.2.1 Environment used

The visual task was an obstacle avoidance task programmed with the LIPS (Laboratoire d'Interactions Pilote-Système, or Pilot-System Interactions Laboratory). Participants had to avoid incoming obstacles by supervising an UAV in autopilot mode (see Figure 40; for more details, see 4.1 – The obstacle avoidance task). The LIPS was displayed on the right screen. On the left screen, attentional probes appeared semi randomly. Speakers on the left and right side of the LIPS screen sent a continuous modulated brown noise with discrete beeps at semi-random intervals.

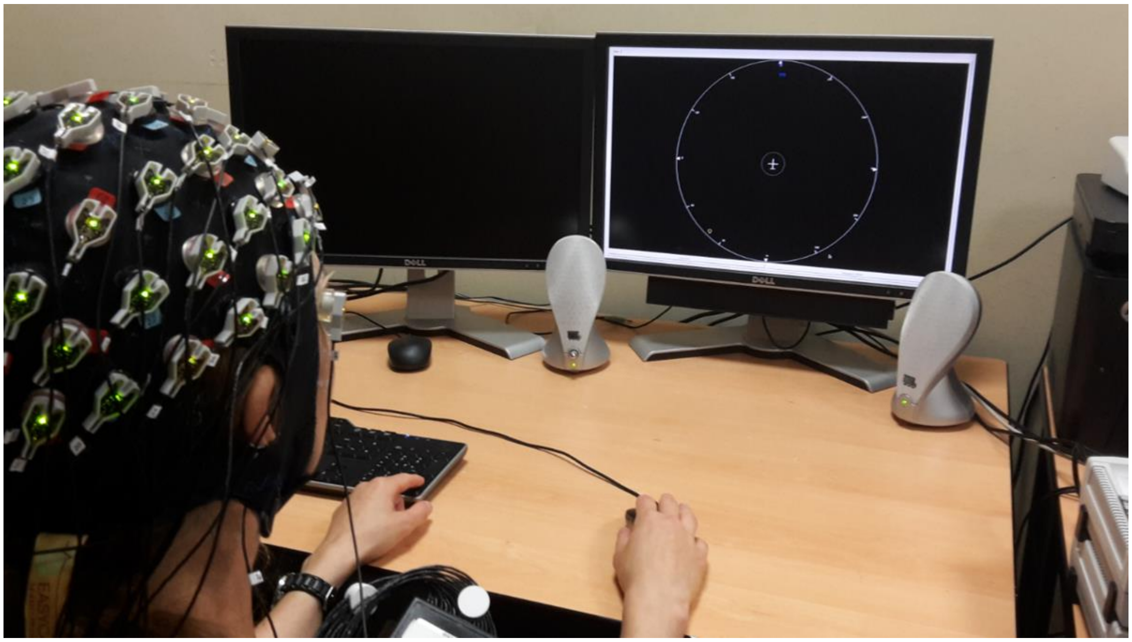


Figure 40. Experimental setup.

The participant is equipped with the EEG system and sits in front of the right screen (LIPS screen). Speakers are on both sides of the right screen. The left screen is used to display attentional probes

7.3.2.2 Attentional probes

Attentional probes were identical to these of the first experiment (see subsection 5.3.2.3 - Attentional probes). On average every 2 minutes, an attentional probe appeared on the left screen (see Figure 40). These probes enquired for where was participants' attention directed. Participants could answer "On the task", "Something related to the task", "Something unrelated to the task" or “External distraction” (see the subsection 4.3 - Attentional probes for more details).

7.3.2.3 Visual part of the task

We used the “Automated” mode introduced in the previous experiment described in chapter 5. Participants were required to monitor the autopilot avoiding obstacles. Clusters of obstacles appeared along the way (every 45s on average). Each cluster could contain between 1 to 5 obstacles, including one on the trajectory. When an obstacle was present on the trajectory (this situation is called “conflict”), the autopilot detected it and initiated a deviation automatically. Participants had to click on an “*Acquittement*” (acknowledgement) button to acknowledge automated avoidance decisions as soon as they saw it (twice per trial, once to acknowledge avoidance of the object and once to acknowledge the return to normal trajectory after avoiding the object). A feedback message was displayed to the participants. Finally, if participants detected an incoming collision warning, they were instructed to click on the button “*Changement d’altitude*” (change height) so that the UAV would perform an emergency descent to avoid colliding with the obstacle. Collisions could occur during the avoidance trajectory, if there was another obstacle on the bypass trajectory chosen by the autopilot. The autopilot made 3% errors. All decisions and collisions were predefined and, therefore, they were the same for all subjects.

7.3.2.4 Auditory part of the task

We used Python programming language to generate the auditory stimuli with the packages *acoustics* (felipeacsi & Rietdijk, 2018), *wave*, *math* and *random* (Python Software Foundation, 2018). Python was chosen as it allowed us much flexibility to test different options when creating the soundtracks. Acoustics generated brown noise using the *acoustics.generator.brown* function. This signal was then modulated in amplitude with a sinusoidal wave with a peak of 0.5 (50% amplitude modulation). Three different 5-seconds soundtracks were randomly created. The first soundtrack was fused with a 100-milliseconds 1000 Hz beep using the Audacity software (Audacity, 2018). The beep played at the beginning of the 5s soundtrack.

7.3.2.5 Procedure of the experiment

Participants were explicitly instructed that detection accuracy was more important than speed of response. The sessions started with an explanation of the task, followed by a 10-minute training period and a 55-minute session. The session contained 70 clusters of obstacles, totaling 210 obstacles. Each cluster contained between one and five obstacles, including one on the trajectory. Clusters were separated by 45 seconds on average. 32 probes were displayed during the whole session. The distribution of the experience-sampling probes was not correlated with events on the obstacle-avoidance task, in order to minimize performance influence on experience-sampling reports. The autopilot made two errors initially placed randomly (3% errors; errors on trials 31 and 52 for all subjects).

Parallel to the visual task, subjects had to listen to beeps and push the "Enter" button every time they heard it as fast as possible. They were explicitely told that beeps were to be treated as fast as possible, whatever was happening on the visual task. Beeps were presented between 20 and 40 seconds. One every three beep on average was followed by an attentional probe.

7.3.3 Data collection and filtering

We used R-Studio 1.1.456, R 3.5.1 (R Core Team, 2016; RStudio Team, 2015), Matlab 2016a and 2018a (The Mathworks Inc., 1992) and EEGLAB (Delorme & Makeig, 2004) to filter and analyze the data.

7.3.3.1 Attentional probes

We used E-Prime 2.0 (Psychology Software Tools, 2018) to program attentional probes. We chose E-Prime because it allows millisecond accurate stimulus presentation and offers to log various useful behavioral information (e.g.

reaction time, accuracy, stimulus time presentation, stimulus duration). Each time the probe appeared, a signal was sent to the ActiCHamp software to record a trigger on the EEG signal. Similarly, another trigger was sent when participants answered the probe. This last trigger depended on which attentional state participants reported. We used R-Studio 1.1.456 and R 3.5.1 (R Core Team, 2016; RStudio Team, 2015) to analyze the data.

7.3.3.2 Auditory stimuli

We used E-Prime 2.0 (Psychology Software Tools, 2018) to present auditory stimuli. A trigger was recorded by E-Prime and sent to the EEG system each time a soundtrack or a beep started. Soundtracks played in a predefined random order as to produce a continuous non-predictable background noise. We observed no problems of "clicks" between soundtracks. When subjects were told about concatenated soundtracks, none reported having already understood the nature of the seemingly continuous sound (Agus & Pressnitzer, 2013).

7.3.3.3 Electroencephalogram

We used the ActiCHamp system and Brain Vision software (Brain Products, 2018) to record scalp potentials. 64 Ag-Cl electrodes were mounted on a standard elastic cap (see subsection 4.5 - Electroencephalography) at the standard sites of the 10-10 International system (see Figure 41 and Oostenveld & Praamstra, 2001).

Fpz was used as the ground electrode. We used electrooculography sites in order to capture eye movements. We chose the left mastoid M1 as a reference for recording. We then re-referenced using left and right mastoids by using FT9 and FT10 electrodes. We chose this reference according to the literature (Griskova et al., 2007; Kam et al., 2012, 2011). Both mastoids are inactive zones regarding neuronal activity, while still being close to other electrodes. This ensures that the signal recorded by mastoids electrodes is free from activity that we want to study, while being able to record any electromagnetic perturbation that would affect all electrodes. Finally, averaging M1 and M2 correct for reference lateralization effects (Yao et al., 2005). Impedance was kept below 5kΩ for all electrodes.

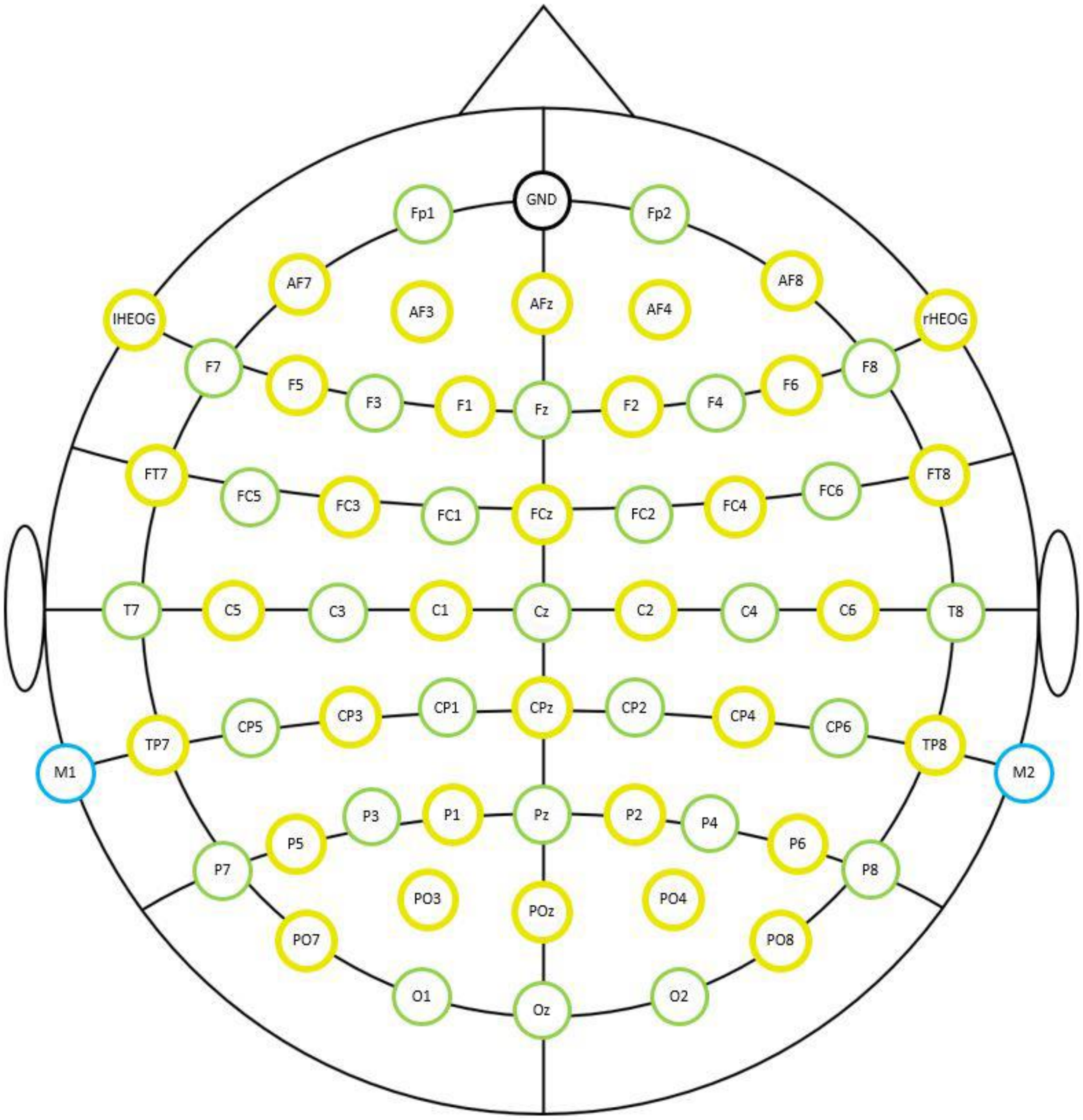


Figure 41. Electrodes sites used for the experiment.

We used Matlab and EEGLAB to import, re-reference, filter, epoch, remove ICA components and build our design. We then exported data to R in order to perform statistical analysis.

7.3.3.4 Performance

In order to assess performance in the auditory condition, we recorded reaction time to beeps through the E-Prime software (difference between start of the beep and button press).

7.4 Results analysis

All linear mixed-effect analyses used the *lme* function to create the models. We used the *Anova* function to assess the influence of predictors. We chose type 2 sum of squares, or type 3 sum of squares when there were interactions to consider between predictors. Post-hoc tests were conducted using the *glht* and *mes* functions of R on the complete model.

All confidence interval (95% CI) reported hereafter were computed using the boot package with 10000 iterations under the first under normal bootstrap approximation.

7.4.1 Mind wandering frequency analysis

We split the 55-minute sessions into four blocks of approximately 14 minutes containing eight experience-sampling probes each. Participants reported on average 31.3% task-related MW (*SD* = 4.4%) and 36.6% task-unrelated MW (*SD* = 5.0%, see Figure 42). This rate is consistent with our previous studies. Each participant reported on average 1.5% “Noise” reports (*SD* = 1.21), lower than both previous experiments. Therefore, we discarded “Noise” reports and adopted the ternary approximation of attentional states (i.e., either focused, task-related MW or task-unrelated MW).

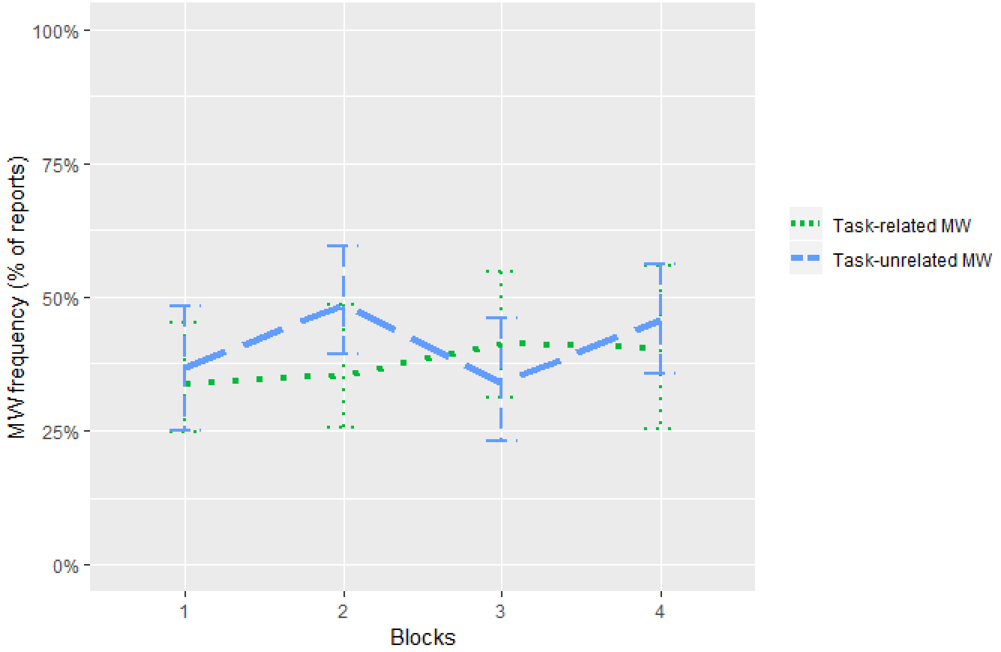


Figure 42. Task-related and task-unrelated MW evolution through blocks.

Error bars show the 95% confidence intervals based on bootstrap.

We focused on task-related and task-unrelated MW frequency evolution over time and conditions. We performed a linear mixed-effect analysis. We considered Blocks as a 4-level categorical variable. We defined a random intercept for subjects to consider our repeated-measure design. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 13, bold values being significant.

Blocks did not significantly influence task-related MW. On the contrary, blocks significantly influenced task-unrelated MW rates, χ = 12.28, p = .015. Without specific *a priori* predictions regarding the block-by-block evolution, we conducted Tukey's post-hoc tests on the complete model. Tests revealed that task-unrelated MW rate were significantly higher under the second block compared to the first and third blocks, p = .021, d = 0.55, p = .010, d = 0.62, respectively. The placement of system errors could explain the drop in task-unrelated MW rates during block 3 (the first one was at the end of block 2, the other one in the middle of block 3).

Table 13. Influence of blocks on task-related and unrelated MW frequency.

		Task-related MW		Task-unrelated MW	
Effect added	**df**	χ	*p*-value	χ	*p*-value
Block	3	0.30	.828	**12.13**	**.007**

Another possibility could be that the task we designed required too much attentional resources for participants to engage in MW. We compared the attentional state rates with previous experiments (see Figure 43). Subjects reported equivalent percentage for each attentional state, ruling out the possibility that the lack of task-unrelated MW may be due to a more engaging paradigm.

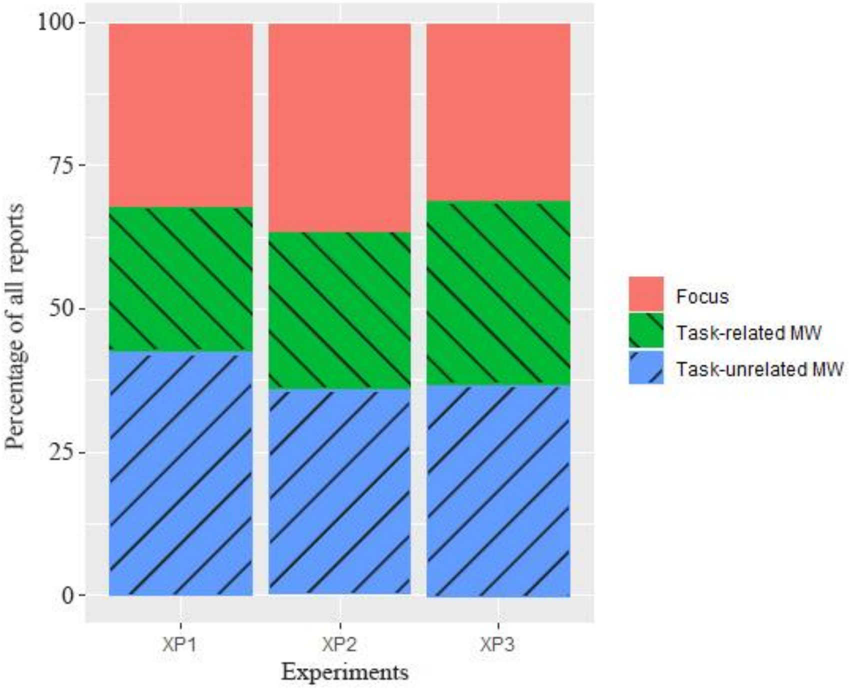


Figure 43. Attentional state percentage for each experiment.

Overall, the analysis of MW rates showed a variable task-unrelated MW propensity that could be linked to visual task features. This variability mirrored variability of focus reports frequency, as task-related MW remained stable throughout the session.

7.4.2 Reaction time to beeps

We measured the reaction time to beeps (see Figure 44). The influence of attentional states and blocks were analyzed using a linear mixed-effect analysis. We took all subjects except the one reporting too much noise (may have passed the experiment with high distractibility). Results are gathered in Table 14.

Attentional states did not influence reaction time to beeps. On the contrary, the analysis highlighted a significant influence of blocks, $\chi(3) = 25.52$, $p < .001$. Without *a priori* information about the evolution between blocks, we used Tukey HSD to break the effect. Without specific *a priori* predictions regarding the block-by-block evolution, we conducted Tukey's post-hoc tests on the complete model. Tests revealed that participants were significantly slower during the fourth block, compared to the first and third blocks, $p = .007$, $d = 0.48$, $p = .016$, $d = 0.28$, respectively. Overall, beep reaction time analysis shows a tendency for participants to slow over time, independently of attentional states. The decrease non-significant tendency to accelerate between blocks 2 and 3 may be due to automation errors at the end of block 2.

Table 14. Influence of attentional states and blocks on beep reaction time.

Effect added	df	χ	*p*-value
Attentional states	2	2.89	0.24
Block	3	**25.52**	**< .001**
Attentional states:Blocks	6	10.09	.121

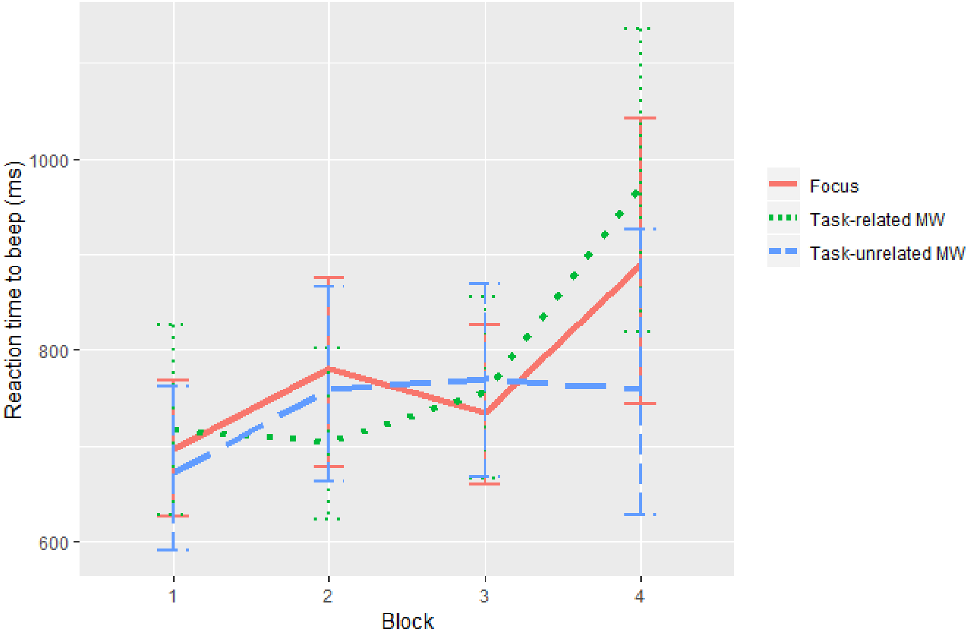


Figure 44. Influence of blocks and attentional states on beep reaction time.

Error bars show the 95% confidence intervals based on bootstrap.

7.4.3 Influence of attentional states on ERPs

As a marker of perception, we investigated the influence of attentional states on the amplitude of the N1 ERP (see Figure 45). The peak of the N1 component emerged approximately at 190ms after the beeps. Therefore, we took the mean of the interval [180-200] ms for computation. Similarly, as a marker of stimuli processing, we looked at the P3 component amplitude. The P3 component peaked around 400ms. Therefore, we chose the mean of the interval [380; 420] ms for computation. We used the electrodes Fz, Pz and Cz. Both intervals and electrodes chosen are in line with the literature (Kam et al., 2011, 2014; Kam & Handy, 2013).

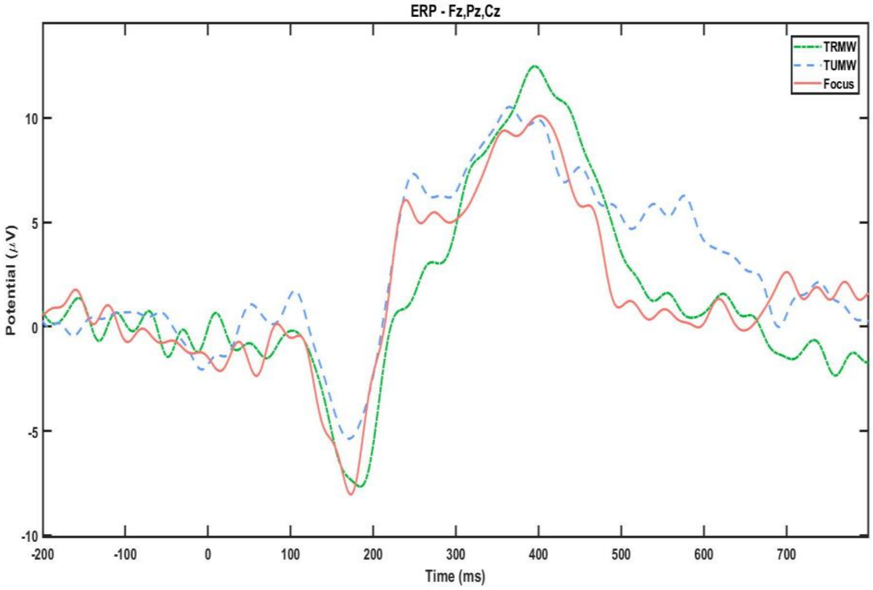


Figure 45. Beep ERP signal for each attentional state.

We performed a linear mixed-effect analysis. We defined a random intercept for subjects to consider our repeated-measure design. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 15, bold values being significant.

Table 15. Influence of attentional states on the amplitude of the ERP components N1 and P3.

Effect added	df	N1 component		P3 component	
		χ	*p*-value	χ	*p*-value
Attentional states	2	**9.41**	**.009**	8.83	.012

Without *a priori* concerning the influence of attentional states on both ERPs amplitude, we conducted post-hoc tests using Tukey HSD. For the N1 component, reports of task-unrelated MW were accompanied with a lower amplitude (*M* = -6.06 μV, *95% CI* = [-8.01; -4.12] μV) compared to periods of focus

(*M* = -9.39 μV, *95% CI* = [-12.21; -6.60] μV), *p* = .024, *d* = 0.36. For the P3 component, the statistics showed a significantly higher amplitude for task-related MW (*M* = 12.69 μV, *95% CI* = [9.28; 16.13] μV) compared to focus periods (*M* = 8.20 μV, *95% CI* = [5.54; 10.85] μV), *p* = .009, *d* = 0.16. Both results point to an impact of MW. The lower perception of auditory stimuli has been observed in a visual task where subjects had to ignore the beeps (Kam et al., 2011). The absence of difference in P3 amplitude between focus and task-unrelated MW periods may be because participants during focus were also inhibiting processing of other stimuli than the visual ones. On the contrary, task-related MW may be decoupled from the visual task but not the global environment to perceive and treat auditory stimuli better than in both other attentional states.

7.4.4 Influence of attentional states on brain wave amplitude

After looking at ERPs, we investigated the amplitude of alpha waves and ASSR (see Figure 46 and Figure 47). We computed the mean of the frequencies between 10 and 12 Hz for alpha waves. Since previous studies repeatedly revealed consistent results for the lower and upper alpha band (e.g. Benedek, Bergner, Könen, Fink, & Neubauer, 2011; Jauk, Benedek, & Neubauer, 2012), we did not perform additional analyses for alpha sub-bands in this study. We chose the electrodes Pz, P1/2, P3/4, P5/6, Poz, PO3/4, Oz, O1/2 for alpha in order to cover parieto-occipital region. Previous studies observed higher alpha amplitude linked with visual sensory inhibition in this region, in line with the MW perceptual decoupling (Benedek et al., 2014; Foxe et al., 1998; O'Connell et al., 2009). We used the sites FCz, FC1/2 for ASSR. Those sites had already been used by Saupe and colleagues in experiments investigating ASSR and attention (Keitel et al., 2011; Saupe et al., 2009).

We used a linear mixed-effect analysis to look at the influence of attentional states on alpha and ASSR amplitude. We defined a random intercept for subjects to consider our repeated-measure design. Visual inspection of residual plots did not reveal any obvious deviations from normality or homoscedasticity. All results are gathered in Table 16, bold values being significant.

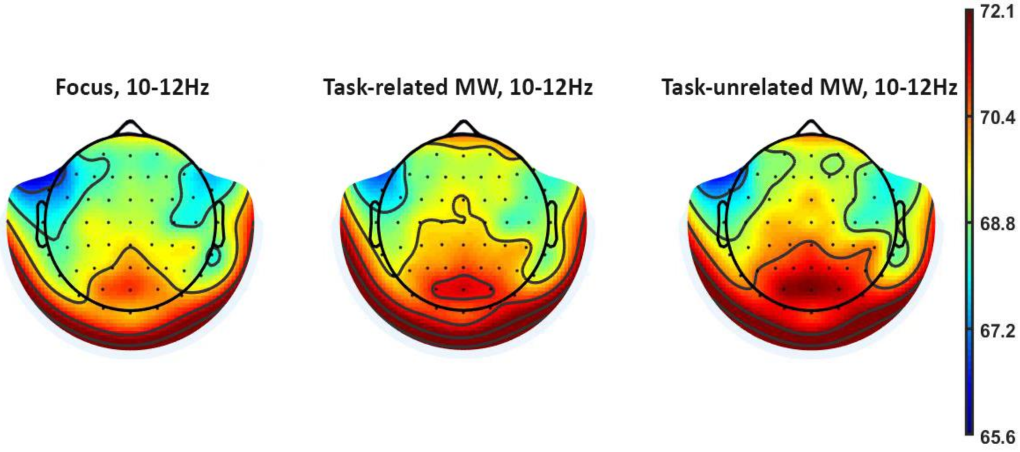


Figure 46. Topography of alpha frequency for each attentional state.

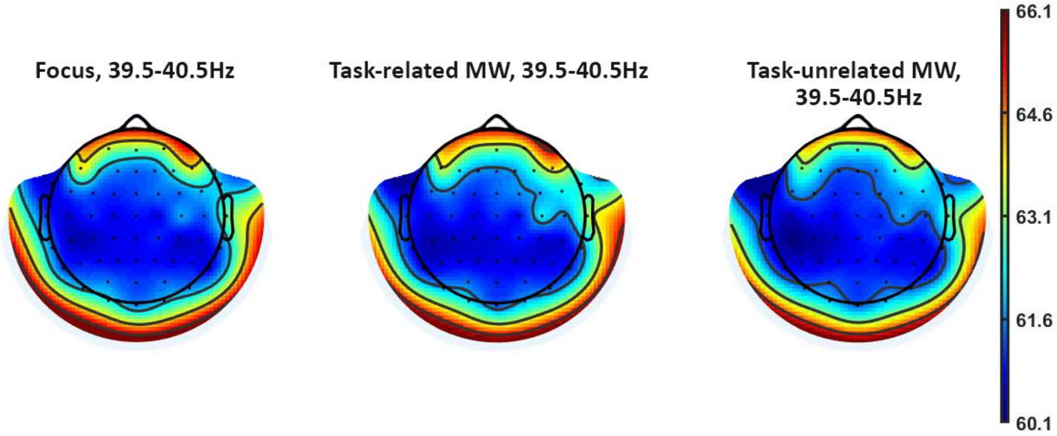


Figure 47. Topography of ASSR frequency for each attentional state.

Table 16. Influence of attentional states on alpha and ASSR amplitude.

Effect added	df	Alpha		ASSR	
		χ	*p*-value	χ	*p*-value
Attentional states	2	**8.35**	**.015**	2.55	.279

Results showed a significant influence of attentional states on alpha amplitude, $\chi(2) = 8.35$, $p = .015$. Without *a priori* concerning the influence of attentional states on alpha amplitude, we conducted post-hoc tests using Tukey HSD. Alpha wave had significantly higher amplitude during task-unrelated MW

(*M* = 53.83 µV /Hz, *95% CI* = [52.35; 55.31] µV /Hz) compared to focus episodes (*M* = 53.03 µV /Hz, *95% CI* = [51.90; 54.16] µV /Hz), *p* = .014, *d* = 0.27. All other comparisons were not significant.

On the contrary, no influence of attentional states on ASSR amplitude was uncovered. However, specter plot still revealed a peak at 40 Hz, showing that the ASSR was visible on participants' specter even during this complex task (see Figure 48). It is worth noting that after the experiment, we asked participants "do you have any comment about the background noise". 12 participants out of 18 reported that they felt the noise had aeronautical connotation, similar to a propeller airplane.

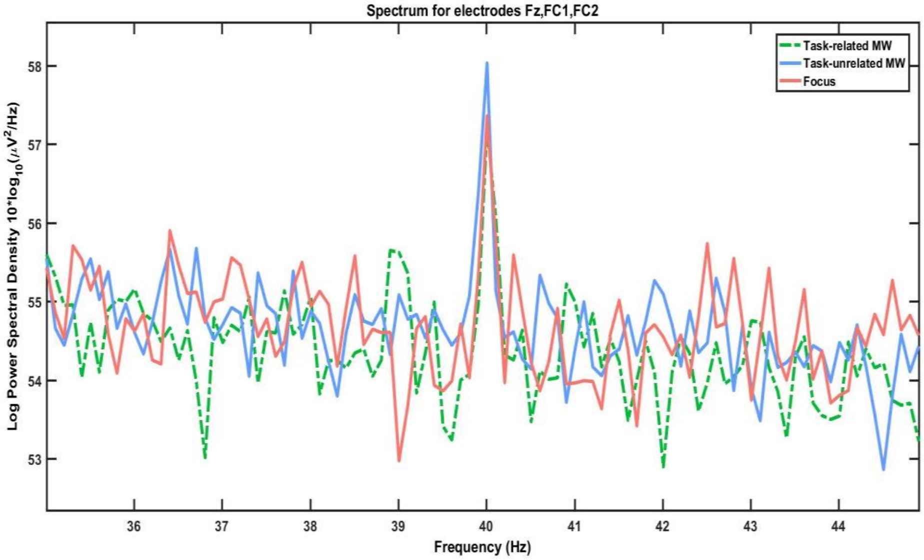


Figure 48. Specter of 35-45 Hz interval for each attentional state.

7.5 Discussion and conclusion of the experiment

7.5.1 Multimodal tasks may thwart task-unrelated mind wandering

Contrary to the experiment 1 and 2, task-unrelated MW did not increase with time on task in this experiment (see subsections 5.4.1 - Mind wandering frequency evolution and 6.4.1 - Mind wandering frequency analysis). Only the second block exhibited higher task-unrelated MW rates compared to the first and third ones. This evolution may be cause by the automation errors, placed at the ends of the second and third blocks. It may also be that the beeps allowed for a reengagement of participants throughout the task, acting like “reminders” to spend attentional resources in the task and not in MW. A final explanation may be that the use of the EEG setup impacted the natural tendency to mind wander in repetitive tasks. Nevertheless, delving into the cause of such absence of MW increase may reveal important factors influencing MW propensity.

7.5.2 Task-unrelated mind wandering decouples from visual task

Task-unrelated MW induced lower perception and sensory integration, highlighted by lower N1 and P3 components. Moreover, our results also showed a significant increase of alpha activity in parieto-occipital regions during task-unrelated MW, compared to focus episodes. On the contrary, task-related episodes were not significantly different from other attentional states. Both results provide new evidences supporting the decoupling hypothesis, and extend the literature in more complex environments.

Nevertheless, the exact impact of task-related MW was not clear. Based on ERP components, task-related MW seemed to stop participants from focusing too much on the visual task and participated in a more balanced distribution of attentional resources. However, this was not followed by higher performances on the auditory task, i.e. lower reaction time to beeps. This result highlights the complex impact of MW perceptual decoupling on operators. Different types of MW may affect the attentional resources distribution differently. Sensory modality conveying targets may also mediate the link.

Notwithstanding, our results revealed the possibility to conduct both spectral and ERP analysis in complex environments. This prospect is the first step to conduct other ecological studies in order to understand the action of MW on operators in automated environments.

7.5.3 Steady-State Responses may not help to track mind wandering

ASSR was visible on participants' specter. However, the direction of attention (external or internal) did not modulate ASSR amplitude. This outcome completes the results of O'Connell et al. (2009) concerning the impossibility to use steady-state responses to study MW perceptual decoupling, whether the SSR supports the task targets or not. Saupe and colleagues demonstrated that the SSR amplitude is amplified if the participant is focusing on the modality supporting the SSR (when the SSR itself conveys the target signal, see Keitel et al., 2011; Saupe et al., 2009). It may be possible that attentional focus has two orthogonal dimensions: external/internal and sensory modality. Whereas differences in sensory modality seems to impact SSR amplitude, our results may underline that external/internal changes may not have the same impact. More work in needed, which could highlight the basic mechanisms of attention modulation and its impact on sensory integration. A straightforward possibility could be to use the same paradigm (concurrent attention between VSSR and ASSR) as reported by Keitel et al. (2011), but with MW probes.

7.6 Thesis progress recap (chapter 7)

This experiment completed the experiments 1 and 2 and offered more accurate tools to assess MW influence on operators. We highlighted the possibility to use traditional EEG markers of MW (ERP components, alpha waves) into complex multimodal tasks. This chapter concludes the experimental part of this thesis, linking the state of the art evidences and the practical questionnaire results. We will continue with the discussion about the results of this thesis work and the future research tracks.

8 DISCUSSION AND CONCLUSION

- **Task-unrelated MW increases in OOTL situations** *for non-expert operators independently of automation reliability. Agency still remains as a possible mechanism able to explain task-unrelated MW increase.*

- ***Task-related MW showed varying impact*** *on behavior and physiological signal, underlining the importance of task proximity in the link between MW and performances.*

- *Contracted pupil, higher blink rates, attenuated N1 and P3 components and higher alpha activity all highlighted a* ***perceptual decoupling*** *from the task during task-unrelated MW.*

- *Systems able to* ***facilitate transitions between focus periods and MW episodes*** *may be the best option to use MW benefits while tempering the consequences of the perceptual decoupling.*

- *Aside from MW, the development of* ***human-machine joint-agency*** *and the understanding of* ***system performance monitoring*** *by operators are other important aspects for mitigating OOTL performance problem.*

8.1 Description of the current chapter

The purpose of this thesis was to investigate the dynamic of MW in automated environments and highlight its possible influence on the OOTL performance problem. This last chapter will start by a summary of our contribution to the literature in the subsection 8.2 - Contribution to the literature. Then, we present in the section 8.3 - Perspectives and future research tracks some promising avenues for future research in order to mitigate OOTL performance problem risk on safety-critical environments. We finish with our traditional 8.5 - Thesis progress recap (chapter 8).

8.2 Contribution to the literature

8.2.1 Task-unrelated mind wandering rate increases in out-of-the-loop situations

In the two experiments reported in chapters 5 and 6, we observed an increase of task-unrelated MW frequency with time on task. On the contrary, task-related MW did not exhibit any linear increase in all three experiment (stable in chapters 5 and 7, chaotic in chapter 6). Time saved by automation should theoretically be used for planning the mission ahead (task-related MW in our experiments). Instead, participants progressively shift their attentional resources from focusing on the task to unrelated matters (voluntarily or not). Moreover, the tendency observed on pilots' scores questionnaire correlated with experimental results (see chapter 3).

Alongside those three experiments, we designed and validated a paradigm to study MW in aeronautical context with complex intructions. The environment demonstrated compatibility with a variety of measures (behavioral, oculometric or EEG), stable results and important flexibility. Further experiments using the LIPS (or more generally any obstacle avoidance task using aeronautical context) could unveil new features of MW in ecological context.

Nevertheless, the exact mechanism responsible for this increase remains unknown. We tested the influence of reliability, however without finding any significant result. Without accepting the null hypothesis from only one paradigm, our results do not support trust (and therefore complacency) as a mechanism explaining high task-unrelated MW frequency when supervising a system. This conclusion may seem in apparent contradiction with the possible relationship between perceived autopilot usability and human-automation problems frequency observed in subsection 3.5.3 - Part 3 and 5 influence human-automation problems. However, only one question inquired about trust in the part 3 “Autopilot usability”. When considering only this question, no obvious positive or negative slope appeared from the graph.

We only tested one system feature (reliability), while many others may mediate MW emergence. Nevertheless, the second hypothetical mechanism that we proposed remains untested (see 5.5.2.3 - Agency to explain mind wandering frequency). Indeed, the very nature of automation environments may create higher rates of MW because of the loss of the feeling of agency. Our lack of influence of trust on MW rates while those rates increased two times in automated conditions back this hypothesis up. Moreover, recent literature on agency and automation demonstrated how automation lowers the feeling of agency and the sensitivity to external stimuli, thus setting all conditions for MW emergence (Berberian et al., 2012; Caspar et al., 2016; Obhi & Hall, 2011; Sahaï et al., 2017b).

8.2.2 MW decouples operators from their supervision

Multiple measures demonstrated in the three experiments that task-unrelated MW decouples operators from the task. During task-unrelated episodes, participants reported similar mental demand for systems of different reliability, showing that participants did not update their model of the situation (chapter 5). We observed how lower pupil size (chapters 5 and 6) and higher blink rate (chapter 6) during task-unrelated MW episodes compared to focus periods betrayed lower sensibility to external stimuli. Finally, EEG signal showed smaller N1 components and higher alpha activity, already identified in the literature as characteristic of the perceptual decoupling (chapter 7). Even though critics against MW research point to important flaws in some paradigms and data analysis (Dang et al., 2018; Head & Helton, 2016), we overall confirmed the results of the literature and demonstrated that the decoupling hypothesis is also at hand in more complex environments. Moreover, we extended the traditional view of the decoupling on a perceptual level to other executive functions needed for subjective reports (chapter 6).

Although the decoupling seems clear for task-unrelated MW, task-related MW evolution and effect on both behavioral and physiological measures was not as visible. Firstly, the absence of clear evolution in both chapter 5 and chapter 6 remains intriguing. In both chapters, automation errors seemed to have little to no effect on task-related MW rates. It is possible that task-related MW frequency depends less on transient features (learning, errors) and more on permanent features of the human-system cooperation (system features, operators' personality). Concerning the effects of task-related MW, chapter 7 revealed that it may help operators to stop focusing too much on one aspect of the task or to loose themselves in internal cognition. This aspect of task-related MW may highlight a different aspect of the perceptual decoupling, closer to balanced attentional resources distribution.

Our results demonstrate the impact of the relation of thoughts to the task when considering the influence of MW on operators. This supports the vision of a graded MW with progressive decoupling from the environment (Cheyne et al., 2009). Moreover, we experimentally confirmed literature evidences that the link between MW and OOTL situations exists. Finally, throughout three experiments, we demonstrated the possibility to use oculometric and electroencephalographic MW markers in multimodal environments more complex than reading or GO/NOGO tasks. Even though our tasks remain performed in controlled laboratory conditions, our work extends the range of such markers along with the experimental possibilities to characterize the MW decoupling in ecological environments. This work is a first step to understand and characterize the influence of MW on operators. However, many aspects of this phenomenon remain unknown.

8.3 Perspectives and future research tracks

There exist many different research axes to explore the link between OOTL performance problem and MW phenomenon in order to reduce the risks of accidents. We identified three different tracks that we believe could be fruitful for the field.

8.3.1 Train operators to control their MW

Casner and Schooler (2014) were the first to propose the hypothesis that some populations may be able to exert more control than others on their MW emergence. If assessed, such advantage would be of high interest for OOTL situations. However, could it be taught? Should recruiters start testing MW propensity when recruiting operators?

Meditation may be the perfect starting point to investigate technics to better control one’s flow of thoughts. Mindfulness meditation training demonstrated an important impact on task-unrelated MW rates. Mrazek et al. (2013) observed a significant decrease of distracting thoughts after a two-weeks mindfulness training. They measured thoughts during an OSPAN task (presentation of stimuli to be remembered alternated with an unrelated processing task). In parallel with the decrease of task-unrelated MW rate, participants exhibited higher performances for the task. Improvements were partly attributed to low MW frequency, but also to higher working memory capacity. Xu et al. (2017) obtained similar results during a SART after mindfulness training on highly anxious undergraduates. Other forms of meditation also demonstrated interesting results. Brandmeyer and Delorme (2016) witnessed fewer reports of MW for expert Yoga meditators during a meditation task, compared to novice Yoga meditators.

Interestingly, a substantial number of studies investigated the impact of meditation on soldiers. This population is particularly affected by various degraded internal states when operating, which in turn can dramatically lower safety on the battlefield. Significant various results were demonstrated, mainly for stress reduction (Cheema & Grewal, 2013; Pokorski & Jayatunge, 2018; Rees, 2011), but also for enhanced working memory capacity (Jha, Stanley, Kiyonaga, Wong, & Gelfand, 2010; Jha, Witkin, Morrison, Rostrup, & Stanley, 2017) and increase melatonin levels (Harinath et al., 2004) with various meditation technics. However, the effects on MW propensity on the same population have been rarely assessed. Jha et al. (2015) reported an experiment using two versions of the mindfulness training on soldiers on predeployment phase. Versions of the training differed in their distribution of didactic (e.g. speaking about wellness) versus practice (e.g. practicing mindfulness) exercises, and were compared to a control group without training. At the end of the training, the practical training group performed significantly better on the SART while reporting more awareness of their own thoughts and fewer task-unrelated MW episodes. Such results demonstrate that OOTL research could use meditation in operators' training as a tool to increase safety.

However, the long term effects were not assessed. Should people continue regular meditation exercises in order to maintain these lower MW rates? More importantly, at what cost does MW reduction come? To answer this question, we need to better assess the link between task-unrelated MW and long-term performances.

8.3.2 Characterization of mind wandering in out-of-the-loop situations

We showed that the perceptual decoupling induced by MW also impact operators in automated conditions. Coupled with results on the disruption of short-term performances by MW, there is little doubt that MW could lower operators' short-term performances. However, little is known about influence of MW on long-term performances. Driven by evolutionary psychology (Buss, 1995), we can safely say that if MW was detrimental to our survival, evolution would have removed it (Schooler et al., 2014).

A first possible benefit of MW may be that it could help to cope with boredom, thus reducing cognitive fatigue. Boredom has been recently defined by Cummings (2015) as “an affective state of low arousal and dissatisfaction caused by a lack of interest in an inadequately stimulating environment”. Autonomous environments are known to produce high levels of boredom and stress on operators after long periods of supervision (Cummings et al., 2015; Warm et al., 2008). The outcome of such environments is generally operators experiencing cognitive fatigue that lowers their performances. In operational conditions, people who have to maintain high levels of externally directed attention in monotonous environments (truck and train drivers, pilots, sailors) reported thinking regularly about personal matters or fantasize to cope with boredom (see subsection 3.4.3 - Distribution of time when in the cockpit and Cummings et al., 2015; 1991). Moreover, several studies by Cheyne, Carriere and Smilek (Carriere et al., 2008; Cheyne et al., 2006) found a significant increase in everyday attentional failures for individuals more prone to boredom. However, it could also be that MW increases boredom. The boredom emerging from the discrepancy between what operators want and their immediate task could be amplified by the difference between engaging task-unrelated thoughts and the monotonous operators have actions to perform.

Notwithstanding, MW may also facilitate operators’ work by acting as a problem incubator and a future planning mechanism. Future-oriented and self-related MW episodes form the “autobiographical planning” aspect of MW (Mooneyham & Schooler, 2013). Research show that autobiographical planning represent an important proportion of MW episodes, and benefit people in the long term (Baird et al., 2011; Stawarczyk, Majerus, Maj, Van der Linden, & D’Argembeau, 2011). Long-term planning and mental simulation of future events may counter-balance the short-term performance cost associated with MW. Similarly, several papers have highlighted the benefits of MW for creative problem solving (Zedelius & Schooler, 2015, 2016) and emphasized the incubator effect of MW on current problems. By allowing us to think freely, MW could allow us to make links between concepts and generate “out of the box” ideas for our current problems. Baird et al. (2012) investigated MW incubation effect by asking participants to perform the Unusual Uses Task (Guilford, 1967). This task requires participants to generate as many unusual uses of a common object in a limited period of time. Participants performed the Unusual Uses Task during 2 blocks of 2 minutes each, separated by an incubation time. They were exposed to 4 different conditions, depending on how was the incubation time: (a) performing a demanding task between the 2 blocks, (b) performing a less-demanding task, (c) resting and (d) no break. Consistent with their hypothesis, the less-demanding task increased performances regarding the generation of novel uses than did both the demanding task condition and the no-break condition. Surprisingly, the less demanding also elicited more novel use generation than did the resting condition. However, this effect only manifested when the problem was the same before and after the incubation period. Sio and Ormerod (2009) confirmed with a meta-analysis the divergent thinking generation properties of MW and its benefits for humans.

Boredom reduction, problem incubation and future planning may be three useful aspects of MW. Those aspects could even counterbalance MW negative impact on short-term performances and explain why we mind wander so much. Although still barely known, they may be of high importance for operators. Pilots in the cockpit reported copious task-related MW, which may help them envision different possible issues of the next landing while helping them to cope with boredom during cruise. Provided, however, that short term errors do not suppress definitely any notion of “long term” for the operators. Although research generally points to reducing MW during critical tasks, it may be more beneficial to design systems optimizing operators’ MW and reducing the cost of focusing back on the task.

8.3.3 Conception of systems adapted to mind wandering

Although experiments performed in laboratory conditions (e.g. reading and simulators experiments) have produced useful results, they were all performed in controlled environment. In order to avoid future accidents due to MW, we need to explore ways to apply experimental research to real conditions.

The first step towards ecological research on OOTL performance problem and MW is the technical feasibility. Indeed, adaptive systems should provide the same comfort as manual ones, and adapt to operational environments.

Addressing ease of implementation, dry electrodes measure EEG signal without need for skin preparation (Taheri et al., 1994). Although the signal-to-noise signal is lower and requires further improvement, it could be implemented in operational environments with little disruption for the user, especially if they already wear a helmet, such as jet pilots. Recent researches and advances in high-tech industry displayed applications opportunities for real-time neuroimaging monitoring (Mullen et al., 2015; OpenBCI, 2016; This Place, 2016). Several studies completed these results with evidences that EEG can be used for arousal monitoring in operational environments (Dussault, Jouanin, Philippe, & Guezennec, 2005; Jeroski et al., 2014). Nevertheless, EEG is not the only promising physiological signal measure. Khan and Hong (2015) used functional Neuro InfraRed Spectroscopy recorded with a BCI to detect drowsiness with a success rate of 84%. Oculometry has also been substantially improved over the past decade, producing efficient, small, and cheap devices. Scanella et al. (2015) showed that flight phases could be differentiated using an eye tracker while demonstrating a remarkable independence regarding inter- and intra-subject variability. Dehais et al. (2008, 2010) found that an embedded eye tracker allowed detection of gaze features during flight in both nominal and degraded conditions. However, in most of the previously cited research, acceptability of the new system by participants (i.e. the capacity of the system to fulfill user's needs and be accepted for a regular use) was not evaluated. Still, these results demonstrate the possibility of building better human-machine interfaces, which could potentially prevent many related accidents.

As physiological measurement tools grow in their range of applications and ease of use, automation research could try to mitigate OOTL performance problems by engaging the operators more. As high automation handles many aspects of the task, operators tend to experience drifts of attention initiated by MW or drowsiness. The idea was then to design automation that could dynamically change according to operators' measured internal state. This type of automation is called "adaptive automation" and relies on negative feedback, i.e. the less the operator is engaged, the less automated aids are provided, and conversely. As Air Traffic Control officers (ATCo) remain seated in calm environments (low noise, no changes in light), they are particularly suited to develop such systems. Abbass et al. (2014) investigated the possibility to use both EEG and task-complexity to trigger maneuver's proposals and help ATCo. They found that triggers based on EEG brain waves enhanced ATCo's performance, compared to no help at all. More importantly, EEG triggered help also increased performance compared to task complexity triggered help. In their study, Prinzel et al. (2003) investigated the use of Event Related Potentials (ERPs), brain waves and heart rate to vary the level of automation. All three measures demonstrated the possibility to vary automation level and improve performances, compare to a manually operated system or to a static automation system. Such systems demonstrate promising results to mitigate OOTL performance problem. Markers of MW could be used to enhance such systems (see subsection 2.3 - Mind wandering markers to study out-of-the-loop).

The MINIMA project (Detecting and Mitigating the Negative Impact of Automation) included the first insights of this thesis work and gave interesting results (Berberian et al., 2017; Kraemer, De Crescenzio, Berberian, Ohneiser, & Di Flumeri, 2018). The MINIMA project developed a prototype of adaptive automation working position for Air Traffic Control officers (ATCo). The system used an eye-tracker and an EEG helmet with 16 electrodes to detect the level of arousal of ATCo and change the display of automated aids accordingly. 15 professional ATCo performed two conditions each, BASELINE with a fixed high automation level, and SOLUTION with adaptive automation. After both conditions, ATCo filled the NASA Task Load Index questionnaire (questionnaire we used in chapter 5 experiment) and the Thinking Content component of the Dundee Stress State Questionnaire (DSSQ, Matthews et al., 1999) to measure task-unrelated MW. Results replicated the vigilance decrement when operating the BASELINE system. On the contrary, the dynamic deactivation of automated aids during low-arousal periods succeeded in countering the overall vigilance decrement. Moreover, the adaptive automation reduced detection time to incoming aircraft (showed by eye-tracker data) and task-unrelated MW propensity. Finally, ATCo reported the SOLUTION condition as being more mentally demanding than the BASELINE solution (as they were more involved in it) while being also less frustrated. Taken together, these results are a first step towards the possibility to mitigate OOTL problem through adaptive working positions.

We developed in the previous subsection 8.3.2 - Characterization of mind wandering in out-of-the-loop situations the many aspects of MW influence on performances. It is far from obvious that a lower rate of task-related MW and higher engagement at all times may produce less errors and higher performances after hours of supervision. Another possibility to mitigate the influence of MW on OOTL situations may be, not to stop MW emergence, but to facilitate transitions between MW and focus periods. On one hand, such system could lower problems in case of automation disconnection when a problem occurs. On the other hand, it would allow operators to MW during idle phases of their missions, which may allow them to reduce their own cognitive fatigue. The critical feature of such system would be a careful examination of the information needed to come back in the loop and their display. It is not surprising that those requirements come back at the very basic of system supervision, i.e. operator-system communication. As stated in the subsection 2.2.3 - Out-of-the-loop problems and mind wandering: some similarities, OOTL performance problem is fundamentally a problem of cooperation between operators and machines due in most cases to a lack of proper communication.

8.4 Other factors may contribute to out-of-the-loop problem

As exposed in the subsection 1.2.2 - Automation moves operators out of the control loop, OOTL performance problem is a complex construct defined by impaired supervision psychological mechanisms due to either (or both) system features or operators' internal states. Even though we highlighted that MW is an internal state that could be (in short-term) increasing OOTL performance problem, it is not the only validated concept that could be used to study OOTL situations. We hereafter review recent findings and promises of both joint-agency and performance-monitoring.

8.4.1 Joint-agency and cooperative understandable systems

In the subsection 8.2.1 - Task-unrelated mind wandering rate increases in out-of-the-loop situations, we detailed how MW could be related to agency and explain higher MW rates in OOTL situations. However the loss of agency alone has already been pointed as having a debilitating effect on automation supervisors, thwarting efforts for human-machine cooperation and coordination (Norman, 1990; Norman, 2010). Operators even declared that they prefer to deactivate higher automation during critical phases rather than not knowing what it does s (Andersson, 2008). This is a clear example of operator disengagement (and OOTL risk increase) due to a loss of agency.

Interestingly, it has been proposed that the cognitive mechanisms that are involved in the sense of agency during individual actions are of the same kind as those that underlie the sense of agency during joint-actions of other humans (Pacherie, 2012). Called we-agency or joint-agency (Crivelli & Balconi, 2010), this new agentive identity leads individuals to experience agency as soon as one of the two had performed a goal directed action. Unfortunately, people do not experience the sense of we-agency when interacting with machine as much as when interacting with other humans (Glasauer, Huber, Basili, Knoll, & Brandt, 2010; Obhi & Hall, 2011; Sahaï, Pacherie, Grynszpan, & Berberian, 2017; Sahaï et al., 2017). This lack of joint-agency in human-system interactions may be due to the impossibility for operators' neuronal mirror system to simulate machines' behavior. Motor simulation was shown to support our understanding of the low-level motor intentions of others, i.e., the type of action he or she is doing (Rizzolatti, Fogassi, & Gallese, 2001), and also our understanding of others' higher-level prior intentions, i.e., why he or she is doing this action (Iacoboni et al., 2005). Without such mechanism, building a team understanding between operators and automation may prove excessively difficult.

Even though knowledge from social robotic research keep trying to optimize the interactions between social robots and humans, investigations about the sense of agency during these joint actions are still missing. Maximizing human-automation joint-agency may help to build more understandable and enhance systems acceptability. For example, human-like features like voice, motion, or reasoning are already explored. Jaguar (Six, 2018) recently equipped autonomous vehicle prototypes with eyes to keep watching pedestrians when they cross. People reported having more trust that the vehicle actually saw them before crossing, increasing the trust in autonomous cars. Following this philosophy, Le Goff et al. (2018) demonstrated an increase in user acceptance, feeling of agency and performances when operators' were provided with information about what the system is about to do next. Optimizing human-automation interaction is a crucial issue for the design of future technological systems given that humans will be increasingly involved in tasks where they need to interact with highly automated environment. The science of agency can help understand how to develop better cooperating systems (see the full body of work by Sahaï and colleagues for more details).

8.4.2 Monitoring of automation performance

As mentioned in the 1.2.2 - Automation moves operators out of the control loop, OOTL performance problem is directly caused by impaired supervision mechanisms. Among them, the performance monitoring is vital for efficient automation supervision (Somon et al., 2017).

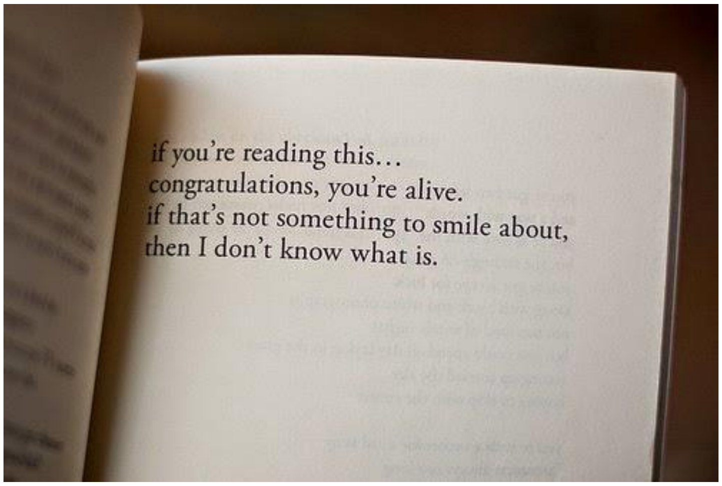
At the neurocognitive level, performance monitoring is defined as "a set of cognitive and affective functions determining whether adaptive control is needed and, if so, which type and magnitude is required" (Ullsperger, Danielmeier, & Jocham, 2014). This cognitive function is used both for monitoring of our own action, but also for the actions of another agent or a system. Nowadays, self-performance monitoring seems to be quite well understood in theory and is applied in several contexts (aviation, see Shappell & Douglas, 2007; medicine, see Taylor, Stern, & Gehring, 2007).

Some studies suggest that observation of our own errors (error execution monitoring) and ones of others (error observation monitoring) would involve the performance monitoring system in the same way (Carp, Halenar, Quandt, Sklar, & Compton, 2009; Jääskeläinen et al., 2016; Koban, Pourtois, Vocat, & Vuilleumier, 2010; van Schie, Mars, Coles, & Bekkering, 2004; Yu & Zhou, 2006). However, this view is still under debate. Moreover, several authors addressed the question of a similarity in the cognitive processes involved in the supervision of human and artificial agents. For example, Padrão and colleagues (2016) used avatar embodiment through virtual reality to study system failure perception. Participants had to perform a task (i.e. goal-directed behavior) while their virtual reality avatar performed the same movements. On several occasions, the avatar would perform an erroneous action creating a system failure. Electrophysiological recording showed performance monitoring activity at fronto-central locations, similar to self-performance monitoring, but delayed.

As the OOTL performed problem is created by a decreased ability to monitor the automated system and detect its errors, deciphering the neural correlates of other's performance monitoring and how they can be degraded could permit to understand and maybe counter OOTL problem risks. Such results could help to characterize one aspect of the OOTL phenomenon at a physiological level (see the full body of work by Somon and colleagues for more details).

8.5 Thesis progress recap (chapter 8)

Previous chapters presented our whole work, starting with theoretical insights inspired by MW and OOTL literatures, delving into pilots' subjective reports concerning autopilot interaction problems, to finish with three experiments in automated environments. We hope you appreciated the journey, and congratulate you for reaching at the end of this thesis (see Figure 49).



if you're reading this…
congratulations, you're alive.
if that's not something to smile about,
then I don't know what is.

Figure 49. Congratulations, you're alive.

From (Saying Images, 2018).

9 REFERENCES

Abbass, H. A., Tang, J., Amin, R., Ellejmi, M., & Kirby, S. (2014). Augmented Cognition using Real-time EEG-based Adaptive Strategies for Air Traffic Control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 230–234. https://doi.org/10.1177/1541931214581048

Ackerman, E. (2017, April 17). Cadillac Adds Level 2 Highway Autonomy With Super Cruise. Retrieved July 19, 2017, from http://spectrum.ieee.org/cars-that-think/transportation/self-driving/cadillac-adds-level-2-highway-autonomy-with-super-cruise

Agus, T. R., & Pressnitzer, D. (2013). The detection of repetitions in noise before and after perceptual learning. *The Journal of the Acoustical Society of America*, *134*(1), 464–473. https://doi.org/10.1121/1.4807641

Amalberti, R. (1999). Automation in aviation: A human factors perspective. In D. J. Garland, J. Wise, & D. Hopkin (Eds.), *Aviation Human Factors* (pp. 173–192). Hillsdale, New Jersey: Lawrence Erlbaum Associates. Retrieved from http://www.irit.fr/SIGCHI/old/docs/debat/Automation.in.Avia.LEA98%27.doc

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, *37*(2), 109–126.

Andersson, J. (2008). *Levels of Automation and User Control - Evaluation of a Turbine Automation Interface* (p. 30). Chalmers: Nordic nuclear safety research.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, *28*, 403–450.

Audacity. (2018). Audacity ® | Free, open source, cross-platform audio software for multi-track recording and editing. (Version 2.2.2). Retrieved from https://www.audacityteam.org/

Bagheri, N., & Jamieson, G. A. (2004). Considering subjective trust and monitoring behavior in assessing automation-induced “complacency.” In A. Vicenzi, M. Mouloua, & O. A. Hancock (Eds.), *Human performance, situation awareness, and automation: Current research and trends* (pp. 54–59). Mahwah, New Jersey: Lawrence Erlbaum Associates. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a426023.pdf#page=64

Bahner, J. E., Hüper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, *66*(9), 688–699. https://doi.org/10.1016/j.ijhcs.2008.06.001

Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, *8*(4), 321–348. https://doi.org/10.1080/14639220500535301

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779.

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by Distraction: Mind Wandering Facilitates Creative Incubation. *Psychological Science*, *23*(10), 1117–1122. https://doi.org/10.1177/0956797612446024

Baird, B., Smallwood, J., & Schooler, J. W. (2011). Back to the future: Autobiographical planning and the functionality of mind-wandering. *Consciousness and Cognition*, *20*(4), 1604–1611. https://doi.org/10.1016/j.concog.2011.08.007

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384. https://doi.org/10.3758/BF03192707

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual Word Recognition: The Journey from Features to Meaning (A Travel Update). In *Handbook of psycholinguistics* (2nd ed). Boston, MA: Elsevier.

Bastian, M., Lerique, S., Adam, V., Franklin, M. S., Schooler, J. W., & Sackur, J. (2017). Language facilitates introspection: Verbal mind-wandering has privileged access to consciousness. *Consciousness and Cognition*, *49*, 86–97. https://doi.org/10.1016/j.concog.2017.01.002

Bastian, M., & Sackur, J. (2013). Mind wandering at the fingertips: automatic parsing of subjective states based on response time variability. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00573

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2017). lme4: Linear Mixed-Effects Models using “Eigen” and S4 (Version 1.1-15).

Baxter, G., Besnard, D., & Riley, D. (2007). Cognitive mismatches in the cockpit: Will they ever be a thing of the past? *Applied Ergonomics*, *38*(4), 417–423.

Baxter, G., Rooksby, J., Wang, Y., & Khajeh-Hosseini, A. (2012). The ironies of automation … still going strong at 30? In *Proceedings of ECCE 2012 Conference*. Edinburgh, North Britain.

Benbadis, S. R., Husain, A. M., Kaplan, P. W., & Tatum, W. O. (2007). *Handbook of EEG Interpretation*. Baltimore, MD: Demos Medical Publishing.

Benedek, M., Bergner, S., Könen, T., Fink, A., & Neubauer, A. C. (2011). EEG alpha synchronization is related to top-down processing in convergent and divergent thinking. *Neuropsychologia*, *49*(12), 3505–3511. https://doi.org/10.1016/j.neuropsychologia.2011.09.004

Benedek, M., Jauk, E., Beaty, R. E., Fink, A., Koschutnig, K., & Neubauer, A. C. (2016). Brain mechanisms associated with internally directed attention and self-generated thought. *Scientific Reports*, *6*, 22959. https://doi.org/10.1038/srep22959

Benedek, M., Schickel, R. J., Jauk, E., Fink, A., & Neubauer, A. C. (2014). Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia*, *56*, 393–400. https://doi.org/10.1016/j.neuropsychologia.2014.02.010

Berberian, B., Ohneiser, O., De Crescenzio, F., Babiloni, F., Di Flumeri, G., & Hasselberg, A. (2017). MINIMA Project: Detecting and Mitigating the Negative Impact of Automation. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 87–105). Springer.

Berberian, B., Sarrazin, J.-C., Le Blaye, P., & Haggard, P. (2012). Automation Technology and Sense of Control: A Window on Human Agency. *PLoS ONE*, *7*(3), e34075. https://doi.org/10.1371/journal.pone.0034075

Berthié, G., Lemercier, C., Paubel, P.-V., Cour, M., Fort, A., Galera, C., … Maury, B. (2015). The restless mind while driving: drivers' thoughts behind the wheel. *Accident Analysis & Prevention*, *76*, 159–165. https://doi.org/10.1016/j.aap.2015.01.005

Billings, C. E. (1991). *Human-centered aircraft automation: A concept and guidelines*. Retrieved from http://ntrs.nasa.gov/search.jsp?R=19910022821

Bixler, R., & D'Mello, S. (2014). Toward Fully Automated Person-Independent Detection of Mind Wandering. In *User Modeling, Adaptation, and Personalization* (pp. 37–48). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-08786-3_4

Bixler, R., & D'Mello, S. (2015). Automatic gaze-based user-independent detection of mind wandering during computerized reading. In *Proceedings of the 23rd User Modeling and User-Adapted Interactions conference* (pp. 31– 43). Dublin, Ireland. https://doi.org/10.1007/s11257-015-9167-1

Blanchard, N., Bixler, R., Joyce, T., & D'Mello, S. (2014). Automated physiological-based detection of mind wandering during learning. In *Intelligent Tutoring Systems* (pp. 55–60). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-07221-0_7

Boeing. (2011, April). 757 Model Summary. Retrieved September 8, 2018, from http://active.boeing.com/commercial/orders/displaystandardreport.cfm?cboCurrentModel=757&optReportType=AllModels&cboAllModel=757&ViewReportF=View+Report

Bonnefond, M., & Jensen, O. (2012). Alpha Oscillations Serve to Protect Working Memory Maintenance against Anticipated Distracters. *Current Biology*, *22*(20), 1969–1974. https://doi.org/10.1016/j.cub.2012.08.029

Boucsein, W., Haarmann, A., & Schaefer, F. (2007). Combining skin conductance and heart rate variability for adaptive automation during simulated IFR flight. In *Engineering Psychology and Cognitive Ergonomics* (pp. 639–647). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-73331-7_70

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: Neural markers of low alertness during mind wandering. *NeuroImage*, *54*(4), 3040–3047. https://doi.org/10.1016/j.neuroimage.2010.10.008

Braem, S., Coenen, E., Bombeke, K., van Bochove, M. E., & Notebaert, W. (2015). Open your eyes for prediction errors. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 374–380. https://doi.org/10.3758/s13415-014-0333-4

Brain Products, 2018. (2018). Brain Products GmbH / Products & Applications / actiCHamp. Retrieved September 7, 2018, from https://www.brainproducts.com/productdetails.php?id=42&tab=3

Brandmeyer, T., & Delorme, A. (2016). Reduced mind wandering in experienced meditators and associated EEG correlates. *Experimental Brain Research*, *236*(9), 2519–2528. https://doi.org/10.1007/s00221-016-4811-5

Bristow, D., Frith, C., & Rees, G. (2005). Two distinct neural effects of blinking on human visual processing. *NeuroImage*, *27*(1), 136–145. https://doi.org/10.1016/j.neuroimage.2005.03.037

Broadbent, D. E., Cooper, P. F., FitzGerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*(1), 1–16.

Brunner, C., Delorme, A., & Makeig, S. (2013). Eeglab – an Open Source Matlab Toolbox for Electrophysiological Research. *Biomedical Engineering / Biomedizinische Technik*. https://doi.org/10.1515/bmt-2013-4182

Bureau d’Enquête et d’Analyse. (1992). *Rapport final sur l’accident survenu le 20 janvier 1992 à l’Airbus A320 exploité par Air Inter immatriculé F-GGED sur l’approche de l’aéroport Strasbourg-Entzheim*. Bureau d’Enquête et d’Analyse.

Bureau d’Enquête et d’Analyse. (2002). *Rapport final de l’accident survenu le 2 octobre 1996 au Boeing 757 exploité par Aeroperu vol Lima - Santiago*. Paris: Bureau d’Enquête et d’Analyse.

Bureau d’Enquête et d’Analyse. (2012a). *Rapport final de l’accident survenu le 1er juin 2009 à l’Airbus A330-203 immatriculé F-GZCP exploité par Air France vol AF 447 Rio de Janeiro - Paris*. Paris: Bureau d’Enquête et d’Analyse.

Bureau d’Enquête et d’Analyse. (2012b). *Rapport final sur l’incident grave survenu le 27 février 2012 à l’Airbus A330 exploité par Air France immatriculé F-GZCG sur le vol Antananarivo - Paris*. Bureau d’Enquête et d’Analyse.

Bureau d’Enquête et d’Analyse. (2016). *Rapport final sur l’incident grave survenu le 24 mai 2011 sur le Dassault Falcon 7X, immatriculé HB-JFN à Subang*. Paris: Bureau d’Enquête et d’Analyse.

Buss, D. M. (1995). Evolutionary Psychology A New Paradigm for Psychological Science. *Psychological Inquiry*, *6*(1), 1–30.

Cabon, P., Coblentz, A., Mollard, R., & Fouillot, J. P. (1993). Human vigilance in railway and long-haul flight operation. *Ergonomics*, *36*(9), 1019–1033. https://doi.org/10.1080/00140139308967974

Callard, F., Smallwood, J., Golchert, J., & Margulies, D. S. (2013). The era of the wandering mind? Twenty-first century research on self-generated mental activity. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00891

Canty, A., & Ripley, B. (2017). boot: Bootstrap R (S-Plus) Functions (Version 1.3-20).

Carmody, M. A., & Gluckman, J. P. (1993). Task Specific Effects of Automation and Automation Failure on Performance, Workload, and Situational Awareness. In R. S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology*. Columbus, Ohio: Department of Aviation, The Ohio State University.

Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., & Compton, R. J. (2009). Perceived similarity and neural mirroring: Evidence from vicarious error processing. *Social Neuroscience*, *4*(1), 85–96. https://doi.org/10.1080/17470910802083167

Carriere, J. S. A., Cheyne, J. A., & Smilek, D. (2008). Everyday attention lapses and memory failures: The affective consequences of mindlessness. *Consciousness and Cognition*, *17*(3), 835–847. https://doi.org/10.1016/j.concog.2007.04.008

Casner, S. M., & Schooler, J. W. (2014). Thoughts in Flight Automation Use and Pilots' Task-Related and Task-Unrelated Thought. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(3), 433– 442. https://doi.org/10.1177/0018720813501550

Casner, S. M., & Schooler, J. W. (2015). Vigilance impossible: Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. *Consciousness and Cognition*, *35*, 33–41. https://doi.org/10.1016/j.concog.2015.04.019

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion Changes the Sense of Agency in the Human Brain. *Current Biology*, *26*(5), 585–592. https://doi.org/10.1016/j.cub.2015.12.067

Cegarra, J., & Morgado, N. (2009). Étude des propriétés de la version francophone du NASATLX. In B. Cahour, F. Anceaux, & A. Giboin (Eds.), *Acte du 5ème colloque de psychologie ergonomique (EPIQUE'2009)* (pp. 233–239). Nice, France.

Champely, S. (2017). pwr: Basic Functions for Power Analysis (Version 1.2-1).

Cheema, S. S., & Grewal, D. S. (2013). Meditation for stress reduction in Indian Army- An Experimental Study. *IOSR Journal of Business and Management*, *10*(2), 27–37. https://doi.org/10.9790/487X-1022737

Cheyne, J. A., Carriere, J. S. A., & Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and Cognition*, *15*(3), 578–592. https://doi.org/10.1016/j.concog.2005.11.009

Cheyne, J. A., Carriere, J. S. A., Solman, G. J. F., & Smilek, D. (2011). Challenge and error: Critical events and attention-related errors. *Cognition*, *121*(3), 437–446. https://doi.org/10.1016/j.cognition.2011.07.010

Cheyne, J. A., Solman, G. J. F., Carriere, J. S. A., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*(1), 98–113. https://doi.org/10.1016/j.cognition.2008.12.009

Chialastri, A. (2012). Automation in aviation. In F. Kongoli (Ed.), *Automation*. InTech. https://doi.org/10.5772/49949

Christoff, K. (in press). Mind-wandering as a scientific concept: cutting through the definitional haze. *Trends in Cognitive Sciences*.

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, *106*(21), 8719–8724.

Christoff, K., Irving, Z. C., Fox, K. C. R., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: a dynamic framework.

Nature Reviews Neuroscience, *17*(11), 718–731. https://doi.org/10.1038/nrn.2016.113

Christoff, K., Ream, J. M., & Gabrieli, J. D. E. (2004). Neural basis of spontaneous thought processes. *Cortex*, *40*(4), 623–630.

Cicchetti, D. V. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychological Assessment*, *6*(4), 284–290.

Cleeremans, A. (2006). Implicit Learning Models. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*. Chichester: John Wiley & Sons, Ltd. https://doi.org/10.1002/0470018860.s00097

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Cowley, J. A. (2013). Off Task Thinking Types and Performance Decrements During Simulated Automobile Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1214–1218. https://doi.org/10.1177/1541931213571270

Cox, D. R., & Isham, V. (1980). *Point processes* (Vol. 12). New York, NY: Chapman and Hall.

Coyle, D., Moore, J., Kristensson, P. O., Fletcher, P., & Blackwell, A. (2012). I did that! Measuring users' experience of agency in their own actions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2025–2034). Austin, TX: ACM. https://doi.org/10.1145/1518701.1518766

Crivelli, D., & Balconi, M. (2010). Agency and Inter-agency, Action and Joint Action: Theoretical and Neuropsychological Evidence. In M. Balconi (Ed.), *Neuropsychology of the Sense of Agency*. Milano, Italy: Springer.

Cummings, M. L. (2004). Automation bias in intelligent time critical decision support systems. *American Institute of Aeronautics and Astronautics*, *2*, 557–562. https://doi.org/10.2514/6.2004-6313

Cummings, M. L., Gao, F., & Thornburg, K. M. (2015). Boredom in the Workplace: a new look at an old problem. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*, 279–300. https://doi.org/10.1177/0018720815609503

Dang, J. S., Figueroa, I. J., & Helton, W. S. (2018). You are measuring the decision to be fast, not inattention: the Sustained Attention to Response Task does not measure sustained attention. *Experimental Brain Research*, *236*(8), 2255–2262. https://doi.org/10.1007/s00221-018-5291-6

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. London, England: Academic Press.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge: Cambridge University Press.

de Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 196–217. https://doi.org/10.1016/j.trf.2014.06.016

Degani, A., Chappell, S. L., & Hayes, M. S. (1991). Who or what saved the day? A comparison of traditional and glass cockpits. In *Proceedings of the Sixth International Symposium on Aviation Psychology*. Columbus, Ohio. Retrieved from http://asrs.arc.nasa.gov/docs/rs/41_Who_or_What_Saved_The_Day.pdf

Degani, A., & Heymann, M. (2000a). Pilot-autopilot interaction: A formal perspective. *Abbott et Al.[1]*, 157–168.

Degani, A., & Heymann, M. (2000b). Pilot-autopilot interaction: A formal perspective. *Abbott et Al.*, 157–168. https://doi.org/10.1207/s15327108ijap0902_3

Dehais, F., Causse, M., & Pastor, J. (2008). Embedded eye tracker in a real aircraft: new perspectives on pilot/aircraft interaction monitoring. In *Proceedings from The 3rd International Conference on Research in Air Transportation. Fairfax, USA: Federal Aviation Administration*. Retrieved from http://mickael.causse.free.fr/medias/publis/Embedded%20eye%20tracker%20in%20a%20real%20aircraft%20new%20perspectives%20on%20pilot%20aircraft%20interaction%20monitoring.pdf

Dehais, F., Causse, M., & Pastor, J. (2010). Toward the definition of a pilot's physiological state vector through oculometry: a preliminary study in real flight conditions. In *Proceedings of HCI Aero*. Retrieved from http://mickael.causse.free.fr/medias/publis/HCI_dehais_causse_pastor.pdf

Dekker, S. W. A. (2015). The danger of losing situation awareness. *Cognition, Technology & Work*, *17*(2), 159–161. https://doi.org/10.1007/s10111-015-0320-8

Dekker, S. W. A., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work*, *6*(2), 79–86. https://doi.org/10.1007/s10111-003-0136-9

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Desmond, P., Hancock, P., & Monette, J. (1998). Fatigue and Automation-Induced Impairments in Simulated Driving Performance. *Transportation Research Record: Journal of the Transportation Research Board*, *1628*, 8–14. https://doi.org/10.3141/1628-02

Dijksterhuis, A. (2004). Think Different: The Merits of Unconscious Thought in Preference Development and Decision Making. *Journal of Personality and Social Psychology*, *87*(5), 586–598. https://doi.org/10.1037/0022-3514.87.5.586

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On Making the Right Choice: The Deliberation-Without-Attention Effect. *Science*, *311*(5763), 1005–1007. https://doi.org/10.1126/science.1121629

Dündar, C. (2015). *The effects of mind wandering on simulated driving performance*. Retrieved from http://etd.lib.metu.edu.tr/upload/12618921/index.pdf

Durantin, G. (2015). *Contrôle de l'état attentionnel en conditions aéronautiques écologiques par imagerie cérébrale* (PhD Thesis). Toulouse, ISAE.

Durantin, G., Dehais, F., & Delorme, A. (2015). Characterization of mind wandering using fNIRS. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00045

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., & Nikolic, D. (1999). *Situation Awareness As a Predictor of Performance in En Route Air Traffic Controllers.* DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA360807

Dussault, C., Jouanin, J.-C., Philippe, M., & Guezennec, C.-Y. (2005). EEG and ECG changes during simulator operation reflect mental workload and vigilance. *Aviation, Space, and Environmental Medicine*, *76*(4), 344–351.

Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The Unengaged Mind: Defining Boredom in Terms of Attention. *Perspectives on Psychological Science*, *7*(5), 482–495. https://doi.org/10.1177/1745691612456044

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). In *Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National* (pp. 789–795). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=195097

Endsley, M. R. (1996). Automation and Situation Awareness. In *Automation and human performance: Theory and applications* (pp. 163–181).

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394.

Endsley, M. R., & Rodgers, M. D. (1998). Distribution of Attention, Situation Awareness & Workload in a Passive Air Traffic Control Task. *Air Traffic Control Quarterly*, *6*(1), 21–44.

Ernest, N., & Carroll, D. (2016). Genetic Fuzzy based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions. *Journal of Defense Management*, *06*(01). https://doi.org/10.4172/2167-0374.1000144

Esterman, M., Noonan, S. K., Rosenberg, M., & DeGutis, J. (2013). In the Zone or Zoning Out? Tracking Behavioral and Neural Fluctuations During Sustained Attention. *Cerebral Cortex*, *23*(11), 2712–2723. https://doi.org/10.1093/cercor/bhs261

Faber, M., Bixler, R., & D’Mello, S. K. (2017). An automated behavioral measure of mind wandering during computerized reading. *Behavior Research Methods*. https://doi.org/10.3758/s13428-017-0857-y

Farley, J., Risko, E. F., & Kingstone, A. (2013). Everyday attention and lecture retention: the effects of time, fidgeting, and mind wandering. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2013.00619

Federal Aviation Administration. (2009). Chapter 04: Automated Flight Control. In *Advanced Avionics Handbook* (Vol. 1). U.S. Department of Transportation. Retrieved from https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/advanced_avionics_handbook/media/aah_ch04.pdf

Federal Aviation Authority. (1972, December 29). Eastern Airlines Flight 401, L-1011, accident near Miami - Accident Overview. Retrieved September 20, 2016, from http://lessonslearned.faa.gov/ll_main.cfm?TabID=1&LLID=8

Federal Aviation Authority. (1995, December 20). American Airlines Flight 965, B-757, accident near Cali - Accident Overview [Lessons Learned FAA]. Retrieved August 18, 2016, from http://lessonslearned.faa.gov/ll_main.cfm?TabID=3&LLID=43

felipeacsi, & Rietdijk, F. (2018). *Python acoustics*. Retrieved from https://github.com/python-acoustics

Feng, S., D'Mello, S., & Graesser, A. C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review*, *20*(3), 586–592. https://doi.org/10.3758/s13423-012-0367-y

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R* (SAGE Publications Ltd, Vol. 1). London, England.

Fisher, C. D. (1991). Boredom at work: a neglected concept. *Human Relations*, *46*(3), 395–417.

Forster, S., & Lavie, N. (2014). Distracted by your mind? Individual differences in distractibility predict mind wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1), 251–260. https://doi.org/10.1037/a0034108

Foxe, J. J., Simpson, G. V., & Ahlfors, S. P. (1998). Parieto-occipital ~ 10 Hz activity reflects anticipatory state of visual attention mechanisms. *NeuroReport*, *9*, 3929–3933.

Foxe, J. J., & Snyder, A. C. (2011). The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00154

Franklin, M. S., Broadway, J. M., Mrazek, M. D., Smallwood, J., & Schooler, J. W. (2013). Window to the wandering mind: Pupillometry of spontaneous thought while reading. *The Quarterly Journal of Experimental Psychology*, *66*(12), 2289–2294. https://doi.org/10.1080/17470218.2013.858170

Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: Using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin & Review*, *18*(5), 992–997. https://doi.org/10.3758/s13423-011-0109-6

Fraulini, N. W., Hancock, G. M., Neigel, A. R., Claypoole, V. L., & Szalma, J. L. (2017). A critical examination of the research and theoretical underpinnings discussed in Thomson, Besner, and Smilek (2016). *Psychological Review*, *124*(4), 525–531. https://doi.org/10.1037/rev0000066

Fried, R., Schettlinger, K., & Borowski, M. (2014). robfilter: Robust Time Series Filters (Version 4.1). Retrieved from https://CRAN.R-project.org/package=robfilter

Frith, C. D. (2014). Action, agency and responsibility. *Neuropsychologia*, *55*, 137–142. https://doi.org/10.1016/j.neuropsychologia.2013.09.007

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & Owen, G. (1999). Flight Deck Automation Issues. *The International Journal of Aviation Psychology*, *9*(2), 109–123.

Galera, C., Orriols, L., M'Bailara, K., Laborey, M., Contrand, B., Ribereau-Gayon, R., … Lagarde, E. (2012). Mind wandering and driving: responsibility case-control study. *British Medical Journal*, *345*(dec13 8), e8105–e8105. https://doi.org/10.1136/bmj.e8105

Geisler, C. D. (1960). *Average responses to clicks in man recorded by scalp electrodes* (p. 166). Cambridge, MA: Massachussets Institute of Technology.

Gerbert, K., & Kemmler, R. (1986). The causes of causes: determinants and background variables of human factor incidents and accidents. *Ergonomics*, *29*(11), 1439–1453. https://doi.org/10.1080/00140138608967257

Giambra, L. M. (1993). The influence of aging on spontaneous shifts of attention from external stimuli to the contents of consciousness. *Experimental Gerontology*, *28*(4), 485–492.

Gilbert, S. J., Dumontheil, I., Simons, J. S., Frith, C. D., & Burgess, P. W. (2007). Wandering Minds: The Default Network and Stimulus-Independent Thought. *Science*, *317*(5834), 393–395. https://doi.org/10.1126/science.1140801

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*, *10*(2), 252–269. https://doi.org/10.3758/CABN.10.2.252

Glasauer, S., Huber, M., Basili, P., Knoll, A., & Brandt, T. (2010). Interacting in time and space: Investigating human-human and human-robot joint action. In *19th International Symposium in Robot and Human Interactive Communication* (pp. 252–257). Viareggio, Italy: IEEE. https://doi.org/10.1109/ROMAN.2010.5598638

Glick, W. H. (1985). Conceptualizing and Measuring Organizational and Psychological Climate: Pitfalls in Multilevel Research. *The Academy of Management Review*, *10*(3), 601. https://doi.org/10.2307/258140

Golchert, J., Smallwood, J., Jefferies, E., Seli, P., Huntenburg, J. M., Liem, F., … Margulies, D. S. (2016). Individual variation in intentionality in the mind-wandering state is reflected in the integration of the default-mode, fronto-parietal, and limbic networks. *NeuroImage*, *146*, 226–235. https://doi.org/10.1016/j.neuroimage.2016.11.025

Gramann, K., Fairclough, S. H., Zander, T. O., & Ayaz, H. (2017). Editorial: Trends in Neuroergonomics. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00165

Grandchamp, R., Braboszcz, C., & Delorme, A. (2014). Oculometric variations during mind wandering. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00031

Grégoire, S., & Lachance, L. (2013, April). *La méditation au travail : pourquoi et comment ?* Presented at the 1er Symposium du Groupe de recherche et d'intervention sur l'insertion et le maintien durable en emploi, Université Laval, Québec. Retrieved from https://www.crievat.fse.ulaval.ca/fichiers/site_crievat/documents/Equipe_FQRSC/Symposium/GREGOIRE_LACHANCE_Symposium_12-19_avril_2013.pdf

Griskova, I., Morup, M., Parnas, J., Ruksenas, O., & Arnfred, S. M. (2007). The amplitude and phase precision of 40 Hz auditory steady-state response depend on the level of arousal. *Experimental Brain Research*, *183*(1), 133–138. https://doi.org/10.1007/s00221-007-1111-0

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*(4), 196–207. https://doi.org/10.1038/nrn.2017.14

Harinath, K., Malhotra, A. S., Pal, K., Prasad, R., Kumar, R., Kain, T. C., … Sawhney, R. C. (2004). Effects of Hatha Yoga and Omkar Meditation on Cardiorespiratory Performance, Psychologic Profile, and Melatonin Secretion. *The Journal of Alternative and Complementary Medicine*, *10*(2), 261–268. https://doi.org/10.1089/107555304323062257

Harris, W. C., Goernert, P. N., Hancock, P. A., & Arthur, E. (1994). The comparative effectiveness of adaptive automation and operator initiated automation during anticipated and unanticipated taskload increases. In *M. Mouloua and R. Parasuraman (Eds.). Human performance in automated systems: Current research and trends.* (pp. 40–44).

Hart, S. G., & Staveland, L. E. (1988). Development of NASA- TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–203.

He, J., Becic, E., Lee, Y.-C., & McCarley, J. S. (2011). Mind Wandering Behind the Wheel: Performance and Oculomotor Correlates. *Human Factors:*

The Journal of the Human Factors and Ergonomics Society, *53*(1), 17–21. https://doi.org/10.1177/0018720810391530

Head, J., & Helton, W. S. (2016). The troubling science of neurophenomenology. *Experimental Brain Research*. https://doi.org/10.1007/s00221-016-4623-7

Helton, W. S., & Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta Psychologica*, *129*(1), 18–25. https://doi.org/10.1016/j.actpsy.2008.04.002

Hoenig, J. M., & Heisey, D. M. (2001). The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis. *The American Statistician*, *55*(1), 19–24. https://doi.org/10.1198/000313001300339897

Hu, N., He, S., & Xu, B. (2012). Different efficiencies of attentional orienting in different wandering minds. *Consciousness and Cognition*, *21*(1), 139–148. https://doi.org/10.1016/j.concog.2011.12.007

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the Intentions of Others with One's Own Mirror Neuron System. *PLoS Biology*, *3*(3), 7.

Innes-Jones, G., & Scandpower, L. (2012). Complacency as a causal factor in accidents - fact or fallacy? *IChemE*, (158). Retrieved from https://www.icheme.org/~/media/Documents/Subject%20Groups/Safety_Loss_Prevention/Hazards%20Archive/XXIII/XXIII-Paper-18.pdf

International Civil Aviation Organization. (2017). *Safety Report* (p. 25). Retrieved from https://www.icao.int/safety/Documents/ICAO_SR_2017_18072017.pdf

International Society of Women Airline Pilots. (2018, April). Airline numbers and women pilots - beginning of 2018. Retrieved from https://s3.amazonaws.com/ClubExpressClubFiles/658242/documents/Current_List_of_Airline_numbers_and_women_pilots_beginning_of_2018_2120066331.pdf?AWSAccessKeyId=AKIAIB6I23VLJX7E4J7Q&Expires=1532007444&respo

nse-content-disposition=inline%3B%20filename%3DCurrent_List_of_Airline_numbers_and_women_pilots_beginning_of_2018.pdf&Signature=6v1phR4ZneRP8%2BpysAxC5dYlEJI%3D

Jääskeläinen, I. P., Halme, H.-L., Agam, Y., Glerean, E., Lahnakoski, J. M., Sams, M., … Manoach, D. S. (2016). Neural mechanisms supporting evaluation of others' errors in real-life like conditions. *Scientific Reports*, *6*, 18714. https://doi.org/10.1038/srep18714

Jardina, J. R., & Chaparro, B. S. (2013). Usability, Engagement, and Satisfaction of Two e-Textbook Applications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 482–486. https://doi.org/10.1177/1541931213571104

Jauk, E., Benedek, M., & Neubauer, A. C. (2012). Tackling creativity at its roots: Evidence for different patterns of EEG alpha activity related to convergent and divergent modes of task processing. *International Journal of Psychophysiology*, *84*(2), 219–225. https://doi.org/10.1016/j.ijpsycho.2012.02.012

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration–exploitation trade-off: evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*(7), 1587–1596.

Jeroski, J., Miller, M. E., Langhals, B., & Tripp, L. (2014). Impact of Vigilance Decrement upon Physiology Measures. In *IIE Annual Conference. Proceedings* (p. 1184). Institute of Industrial Engineers-Publisher. Retrieved from http://www.xcdsystem.com/iie2014/abstract/finalpapers/I471.pdf

Jha, A. P., Morrison, A. B., Dainer-Best, J., Parker, S., Rostrup, N., & Stanley, E. A. (2015). Minds "At Attention": Mindfulness Training Curbs Attentional Lapses in Military Cohorts. *PLOS ONE*, *10*(2), e0116889. https://doi.org/10.1371/journal.pone.0116889

Jha, A. P., Stanley, E. A., Kiyonaga, A., Wong, L., & Gelfand, L. (2010). Examining the protective effects of mindfulness training on working memory capacity and affective experience. *Emotion*, *10*(1), 54–64. https://doi.org/10.1037/a0018438

Jha, A. P., Witkin, J. E., Morrison, A. B., Rostrup, N., & Stanley, E. (2017). Short-Form Mindfulness Training Protects Against Working Memory Degradation over High-Demand Intervals. *Journal of Cognitive Enhancement*, *1*(2), 154–171. https://doi.org/10.1007/s41465-017-0035-2

Johnson, A., & Proctor, R. W. (2013). *Neuroergonomics: a cognitive neuroscience approach to human factors and ergonomics*. Springer International Publishing.

K. McIntire, L., P. McIntire, J., Mckinley, R., & Goodyear, C. (2014). Detection of vigilance performance with pupillometry. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 167–174). Safety Harbor, FL, USA: ACM. https://doi.org/10.1145/2578153.2578177

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113–153. https://doi.org/10.1080/1463922021000054335

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *10*(4), 409–430. https://doi.org/10.1002/1520-6564(200023)10:4<409::AID-HFM4>3.0.CO;2-V

Kam, J. W. Y., Dao, E., Blinn, P., Krigolson, O. E., Boyd, L. A., & Handy, T. C. (2012). Mind wandering and motor control: off-task thinking disrupts the online adjustment of behavior. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00329

Kam, J. W. Y., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow fluctuations in attentional control of sensory cortex. *Journal of Cognitive Neuroscience*, *23*(2), 460–470.

Kam, J. W. Y., & Handy, T. C. (2013). The neurocognitive consequences of the wandering mind: a mechanistic account of sensory-motor decoupling. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00725

Kam, J. W. Y., Xu, J., & Handy, T. C. (2014). I don't feel your pain (as much): The desensitizing effect of mind wandering on the perception of others' discomfort. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 286–296. https://doi.org/10.3758/s13415-013-0197-z

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169–183. https://doi.org/10.1037//0096-3445.130.2.169

Kane, M. J., Brown, L. H., McVay, J. C., Silvia, P. J., Myin-Germeys, I., & Kwapil, T. R. (2007). For whom the mind wanders, and when an experience-sampling study of working memory and executive control in daily life. *Psychological Science*, *18*(7), 614–621.

Keitel, C., Schröger, E., Saupe, K., & Müller, M. M. (2011). Sustained selective intermodal attention modulates processing of language-like stimuli. *Experimental Brain Research*, *213*(2–3), 321–327. https://doi.org/10.1007/s00221-011-2667-2

Khan, M. J., & Hong, K.-S. (2015). Passive BCI based on drowsiness detection: an fNIRS study. *Biomedical Optics Express*, *6*(10), 4063. https://doi.org/10.1364/BOE.6.004063

Killingsworth, M. A., & Gilbert, D. T. (2010). A Wandering Mind Is an Unhappy Mind. *Science*, *330*(6006), 932–932. https://doi.org/10.1126/science.1192439

Klein, S. B., & Thorne, M. B. (2006). *Biological Psychology* (1st ed.). New York, NY: hand.

Koban, L., Pourtois, G., Vocat, R., & Vuilleumier, P. (2010). When your errors make me lose or win: Event-related potentials to observed errors of cooperators and competitors. *Social Neuroscience*, *5*(4), 360–374. https://doi.org/10.1080/17470911003651547

Konishi, M., Brown, K., Battaglini, L., & Smallwood, J. (2017a). When attention wanders: Pupillometric signatures of fluctuations in external attention. *Cognition*, *168*, 16–26. https://doi.org/10.1016/j.cognition.2017.06.006

Konishi, M., Brown, K., Battaglini, L., & Smallwood, J. (2017b). When attention wanders: Pupillometric signatures of fluctuations in external attention. *Cognition*, *168*, 16–26. https://doi.org/10.1016/j.cognition.2017.06.006

Konishi, M., McLaren, D. G., Engen, H., & Smallwood, J. (2015). Shaped by the Past: The Default Mode Network Supports Cognition that Is Independent of Immediate Perceptual Input. *PLOS ONE*, *10*(6), e0132209. https://doi.org/10.1371/journal.pone.0132209

Koopman, P., & Wagner, M. (2016). Challenges in Autonomous Vehicle Testing and Validation. *SAE International Journal of Transportation Safety*, *4*(1), 15–24. https://doi.org/10.4271/2016-01-0128

Korczak, P., Smart, J., Delgado, R., Strobel, T. M., & Bradford, C. (2012). Auditory Steady-State Responses. *Journal of the American Academy of Audiology*, *23*(3), 146–170. https://doi.org/10.3766/jaaa.23.3.3

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 857–875. https://doi.org/10.1098/rstb.2007.2093

Kraemer, J., De Crescenzio, F., Berberian, B., Ohneiser, O., & Di Flumeri, G. (2018). *MINIMA Project: Final Project Results Report*. SESAR Joint Undertaking. Retrieved from http://www.minima-project.eu/wp/wp-content/uploads/2018/06/MINIMA_D5.2-Final-Project-Results-Report-V2.pdf

Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. P. (2017). The influence of time on task on mind wandering and visual working memory.

Cognition, *169*(Supplement C), 84–90. https://doi.org/10.1016/j.cognition.2017.08.006

Lau, N., Skraaning Jr, G., & Jamieson, G. A. (2009). Metacognition in Nuclear Process Control. In *Proceedings of the 17th Triennial World Congress on Ergonomics* (p. 8).

Le Goff, K., Rey, A., Haggard, P., Oullier, O., & Berberian, B. (2018). Agency modulates interactions with automation technologies. *Ergonomics*, *61*(9), 1282–1297. https://doi.org/10.1080/00140139.2018.1468493

Lebreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Lemercier, A. (2014). *Développement de la pupillométrie pour la mesure objective des émotions dans le contexte de la consommation alimentaire* (Psychologie). Université Paris 8.

Lerner, N., Baldwin, C., Higgins, J. S., Lee, J., & Schooler, J. (2015). Mind Wandering While Driving: What Does it Mean and What do we do about it? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1686–1690. https://doi.org/10.1177/1541931215591364

Leveson, N. G., & Turner, C. S. (1993). An investigation of the Therac-25 accidents. *Computer*, *26*(7), 18–41. https://doi.org/10.1109/MC.1993.274940

Liu, Y., Fuld, R., & Wickens, C. D. (1993). Monitoring behavior in manual and automated scheduling systems. *International Journal of Man-Machine Studies*, *39*, 1015–1029.

Lorenz, B., Di Nocera, F., Roettger, S., & Parasuraman, R. (2002). Automated fault management in a simulated space flight micro-world. *Aviation, Space, & Environmental Medicine*, *73*, 886–897.

Lorenz, B., Di Nocera, F., Röttger, S., & Parasuraman, R. (2001). The effects of level of automation on the out-of-the-loop unfamiliarity in a complex dynamic fault-management task during simulated spaceflight operations. In

Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 45, pp. 44–48). SAGE Publications. Retrieved from http://pro.sagepub.com/content/45/2/44.short

Louw, T., Kountouriotis, G., Carsten, O., & Merat, N. (2015). Driver Inattention During Vehicle Automation: How Does Driver Engagement Affect Resumption Of Control? In *Proceedings of the 4th International Conference on Driver Distraction and Inattention (DDI2015)*. Sydney, Australia: ARRB Group. Retrieved from http://eprints.whiterose.ac.uk/91858/

Louw, T., & Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transportation Research Part C: Emerging Technologies*, *76*, 35–50.

Lowenstein, O., & Loewenfeld, I. E. (1962). The pupil. In *The eye* (Vol. 3, pp. 231–267).

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge: MIT Press.

Lund, A. M. (2001). Measuring usability with the USE questionnaire. *Usability Interface*, *8*(2), 3–6.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quaterly Journal of Experimental Psychology*, *1*, 6–14. https://doi.org/10.1080/17470214808416738

Mair, P., Schoenbrodt, F., & Wilcox, R. (2017). *WRS2: Wilcox robust estimation and testing*.

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind: further investigations of sustained attention to response. *Neuropsychologia*, *37*(6), 661–670.

Manzey, D., Bahner, J. E., & Hüper, A.-D. (2006). Misuse of automated aids in process control: Complacency, automation bias and possible training interventions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, pp. 220–224). Los Angeles, CA: Sage Publications.

Retrieved from http://journals.sagepub.com/doi/abs/10.1177/154193120605000303

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science*, *315*(5810), 393–395.

Mathewson, K. E. (2011). *Pulsed out of awareness: EEG alpha oscillations represent a pulsed inhibition of ongoing cortical processing*. University of Illinois at Urbana-Champaign. Retrieved from https://www.ideals.illinois.edu/handle/2142/26295

Matthews, G., Joyner, L., Gilliland, K., Campbell, S., Falconer, S., & Huggins, J. (1999). Validation of a Comprehensive Stress State Questionnaire: Towards a State "Big Three." In I. Mervielde, I. J. Deary, F. De Fruyt, & F. Ostenderf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 335–350). Tilburg, The Netherlands: Tilburg: Tilburg University Press.

Matthews, M. D., & Beal, S. A. (2002). *Assessing situation awareness in field training exercises*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA408560

McGuinness, B. (2004). *Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self-ratings*. BAE SYSTEMS BRISTOL (UNITED KINGDOM) ADVANCED TECHNOLOGY CENTRE. Retrieved from http://www.dtic.mil/docs/citations/ADA465817

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 196.

McVay, J. C., & Kane, M. J. (2010). Does mind wandering reflect executive function or executive failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin*, *136*, 188–207. https://doi.org/10.1037/a0018298

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*(2), 302–320. https://doi.org/10.1037/a0025250

Melinscak, F., Montesano, L., & Minguez, J. (2014). Discriminating between attention and mind wandering during movement using EEG. In *Proceedings of the 6th International Brain–Computer Interface Conference*. Graz: Graz University of Technology Publishing House. https://doi.org/10.3217/978-3-85125-378-8-60

Merat, N., & Jamson, A. H. (2008). The Effect of Stimulus Modality on Signal Detection: Implications for Assessing the Safety of In-Vehicle Technology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(1), 145–158. https://doi.org/10.1518/001872008X250656

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., … Keinath, A. (2018). The “Out-of-the-Loop” concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*. https://doi.org/10.1007/s10111-018-0525-8

Methot, L. L., & Huitema, B. E. (1998). Effects of signal probability on individual differences in vigilance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(1), 102–110.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 519–528.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Mittner, M., Boekel, W., Tucker, A. M., Turner, B. M., Heathcote, A., & Forstmann, B. U. (2014). When the Brain Takes a Break: A Model-Based Analysis of Mind Wandering. *Journal of Neuroscience*, *34*(49), 16286–16295. https://doi.org/10.1523/JNEUROSCI.2062-14.2014

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(1), 11–18. https://doi.org/10.1037/a0031569

Moray, N. (1986a). Monitoring behavior and supervisory control. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. II, pp. 40/1-40/51). New York, NY: Wiley & Sons.

Moray, N. (1986b). Monitoring behavior and supervisory control. In K. Boff (Ed.), *Handbook of perception and human performance* (Vol. 2, pp. 1–51). New York, NY: Wiley.

Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science*, *1*(4), 354–365. https://doi.org/10.1080/14639220052399159

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*(6684), 467–470. https://doi.org/10.1038/30976

Morrison, J. G., Cohen, D., & Gluckman, J. P. (1993). Prospective Principles and Guidelines for the Design of Adaptively Automated Crewstations. In R. S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 172–177). Columbus, Ohio: Department of Aviation, The Ohio State University.

Mosier, K. L., Skitka, L. J., & Korte, K. J. (1994). Cognitive and social psychological issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: current research and trends* (pp. 191–197). Cambridge: University Press.

Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness Training Improves Working Memory Capacity and GRE

Performance While Reducing Mind Wandering. *Psychological Science*, *24*(5), 776–781. https://doi.org/10.1177/0956797612459659

Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M., & Schooler, J. W. (2013). Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00560

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration Without Awareness: Expanding the Limits of Unconscious Processing. *Psychological Science*, *22*(6), 764–770. https://doi.org/10.1177/0956797611408736

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., … Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2553–2567. https://doi.org/10.1109/TBME.2015.2481482

Murphy, T. (2013, April 1). The First Level of Super Mario Bros. is Easy with Lexicographic Orderings and Time Travel. . . after that it gets a little tricky. Retrieved from https://www.cs.cmu.edu/~tom7/mario/mario.pdf

National Transportation Safety Board. (1975). *Aircraft accident report, Northwest Airlines Incorporated, Boeing 727-25, N26 4US near Thiells, New York, December 1st 1974*. National Transport Safety Board.

National Transportation Safety Board. (2010). *Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River US Airways Flight 1549 Airbus A320 - 214, N106US Weehawken, New Jersey January 15, 2009*. Washington, D.C.

NDTV. (2016, August 5). Artificial Intelligence Used To Detect Rare Leukemia Type In Japan. Retrieved August 27, 2018, from https://www.ndtv.com/health/artificial-intelligence-used-to-detect-rare-leukemia-type-in-japan-1440789

Niedermeyer, E., & da Silva, F. L. (Eds.). (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Baltimore, MD: Lippincott Williams & Wilkins.

Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Communications of the ACM*, *37*(4), 66–75. https://doi.org/10.1145/175276.175282

Nishiyama, J., Tanida, K., Kusumi, M., & Hirata, Y. (2007). The pupil as a possible premonitor of drowsiness. In *29th Annual International Conference of the IEEE* (pp. 1586–1589). Lyon, France: IEEE. https://doi.org/10.1016/j.neuron.2015.09.012

Norman, D. A. (1990). The “Problem” with Automation: Inappropriate Feedback and Interaction, not “Over-Automation.” *Philosophical Transactions of the Royal Society B: Biological Sciences*, *327*(1241), 585–593. https://doi.org/10.1098/rstb.1990.0101

Norman, D. A. (2010). *Living with complexity*. MIT Press.

Normand, J.-M. (2017, April 4). Tesla en pilotage (presque) automatique. *Le Monde.fr*. Retrieved from http://www.lemonde.fr/m-voiture/article/2017/04/04/tesla-en-pilotage-presque-automatique_5105772_4497789.html

Nunez, P. L., & Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. USA: Oxford University Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd edition)* (Vol. 226). New York, NY: McGraw-Hill.

Obhi, S. S., & Hall, P. (2011). Sense of agency in joint action: influence of human and computer co-actors. *Experimental Brain Research*, *211*(3–4), 663–670. https://doi.org/10.1007/s00221-011-2662-7

O’Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of* *the American Society for Information Science and Technology*, *59*(6), 938–955. https://doi.org/10.1002/asi.20801

O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., & Kelly, S. P. (2009). Uncovering the Neural Signature of Lapsing Attention: Electrophysiological Signals Predict Errors up to 20 s before They Occur. *Journal of Neuroscience*, *29*(26), 8604–8611. https://doi.org/10.1523/JNEUROSCI.5967-08.2009

O'Hanlon, J. F. (1981). Boredom: practical consequences and a theory.pdf. *Acta Psychologica*, *49*, 52–82.

Olejnik, S., & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, *8*(4), 434–447. https://doi.org/10.1037/1082-989X.8.4.434

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*, 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

OpenBCI. (2016, May 23). OpenBCI - Open Source Biosensing Tools (EEG, EMG, EKG, and more). Retrieved May 23, 2016, from http://openbci.com/

Overgaard, M., & Fazekas, P. (2016). Can no-report paradigms extract true consciousness.pdf. *Trends in Cognitive Sciences*, *20*(4), 241–242.

Padrao, G., Gonzalez-Franco, M., Sanchez-Vives, M. V., Slater, M., & Rodriguez-Fornells, A. (2016). Violating body movement semantics: Neural signatures of self-generated and external-generated errors. *NeuroImage*, *124*, 147–156. https://doi.org/10.1016/j.neuroimage.2015.08.022

Parasuraman, R. (2011). Neuroergonomics: Brain, Cognition, and Performance at Work. *Current Directions in Psychological Science*, *20*(3), 181–186. https://doi.org/10.1177/0963721411409176

Parasuraman, Raja. (1979). Memory load and event rate control sensitivity decrements in sustained attention. *Science*, *205*(4409), 924–927.

Parasuraman, Raja. (1987). Human-computer monitoring. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *29*(6), 695–706.

Parasuraman, Raja, Molloy, R., & Singh, I. L. (1993a). Automation Induced Complacency. *The International Journal of Aviation Psychology*.

Parasuraman, Raja, Molloy, R., & Singh, I. L. (1993b). Performance Consequences of Automation-Induced "Complacency." *The International Journal of Aviation Psychology*, *3*(1), 1–23.

Parasuraman, Raja, & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse and Abuse. *Human Factors*, *39*(2), 230–253.

Parasuraman, Raja, Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions On*, *30*(3), 286–297.

Parasuraman, Raja, & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 511–520.

Pattyn, N., Neyt, X., Henderickx, D., & Soetens, E. (2008). Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue? *Physiology & Behavior*, *93*(1–2), 369–378. https://doi.org/10.1016/j.physbeh.2007.09.016

Peysakhovich, V. (2016). *Study of pupil diameter and eye movements to enhance flight safety*. Université Toulouse 3 Paul Sabatier.

Peysakhovich, V., Lefrancois, O., Dehais, F., & Causse, M. (2016). Des apports de l'oculométrie dans la sécurité aérienne. Retrieved from http://oatao.univ-toulouse.fr/17473/

Pham, P., & Wang, J. (2015). AttentiveLearner: Improving Mobile MOOC Learning via Implicit Heart Rate Tracking. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education* (Vol. 9112, pp. 367–

376). Cham: Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-19773-9_37

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses. *International Journal of Audiology*, *42*, 177–219.

Picton, T. W., John, M. S., Purcell, D. W., & Plourde, G. (2003). Human Auditory Steady-State Responses: The Effects of Recording Technique and State of Arousal. *Anesthesia & Analgesia*, 1396–1402. https://doi.org/10.1213/01.ANE.0000082994.22466.DD

Piekniewski, F. (2018, May 29). AI winter is well on its way. Retrieved June 15, 2018, from https://blog.piekniewski.info/2018/05/28/ai-winter-is-well-on-its-way/

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). nlme: Linear and Nonlinear Mixed Effects Models (Version 3.1-131). Retrieved from https://CRAN.R-project.org/package=nlme

Pokorski, M., & Jayatunge, R. M. (2018). Post-traumatic Stress Disorder: A Review of Therapeutic Role of Meditation Interventions. In *Advances in Experimental Medicine and Biology* (pp. 1–7). Retrieved from https://link.springer.com/chapter/10.1007/5584_2018_167

Popper, K. R. (1972). *The logic of scientific discovery*. London, England: Hutchinson.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

Prinzel, L. J., Parasuraman, R., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). *Three Experiments Examining the Use of Electroencephalogram, Event-Related Potentials, and Heart-Rate Variability for Real-Time Human-Centered Adaptive Automation Design*. National Aeronautics and Space Administration.

Psychology Software Tools. (2018). E-Prime ® (Version 2.0). Retrieved from https://pstnet.com/products/e-prime/

Python Software Foundation. (2018). *Python*. Retrieved from https://www.python.org/

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rees, B. (2011). Overview of Outcome Data of Potential Meditation Training for Soldier Resilience. *Military Medicine*, *176*(11), 1232–1242. https://doi.org/10.7205/MILMED-D-11-00067

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye Movements During Mindless Reading. *Psychological Science*, *21*(9), 1300–1310. https://doi.org/10.1177/0956797610378686

Revelle, W. (2017). psych: Procedures for Personality and Psychological Research (Version 1.7.5). Evanston, Illinois, USA: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Rifkin, J. (1995). *The End of Work: The Decline of the Global Labor Force and the Dawn of the Post-Market Era* (Putnam Publishing Group). New York, NY.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670. https://doi.org/10.1038/35090060

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758.

RStudio Team. (2015). RStudio: Integrated Development for R (Version 1.0.143). Boston, MA, USA: RStudio Inc. Retrieved from http://www.rstudio.com/

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68.

Sabatini, N. (2006). *Downward Pressure on the Accident Rate*. Federal Aviation Administration, Herndon, VA.

Sahaï, A., Pacherie, E., Grynszpan, O., & Berberian, B. (2017a). Co-representation of human-generated actions vs. machine-generated actions: Impact on our sense of we-agency? In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 341–345). Lisbon: IEEE. https://doi.org/10.1109/ROMAN.2017.8172324

Sahaï, A., Pacherie, E., Grynszpan, O., & Berberian, B. (2017b). Predictive Mechanisms Are Not Involved the Same Way during Human-Human vs. Human-Machine Interactions: A Review. *Frontiers in Neurorobotics*, *11*. https://doi.org/10.3389/fnbot.2017.00052

Sarter, N. B., & Woods, D. D. (1995a). "From Tool to Agent": The Evolution of (Cockpit) Automation and Its Impact on Human-Machine Coordination. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 39, pp. 79–83). SAGE Publications. Retrieved from http://pro.sagepub.com/content/39/1/79.short

Sarter, N. B., & Woods, D. D. (1995b). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 5–19.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of Human Factors and Ergonomics*, *2*, 1926–1943.

Sauer, J., Nickel, P., & Wastell, D. (2013). Designing automation for complex work environments under different levels of stress. *Applied Ergonomics*, *44*(1), 119–127. https://doi.org/10.1016/j.apergo.2012.05.008

Saupe, K., Widmann, A., Bendixen, A., Müller, M. M., & Schröger, E. (2009). Effects of intermodal attention on the auditory steady-state response and the event-related potential. *Psychophysiology*, *46*(2), 321–327. https://doi.org/10.1111/j.1469-8986.2008.00765.x

Saying Images. (2018, 04). Congratulations Quotes, Cards, Messages, Pictures | SayingImages.com. Retrieved September 15, 2018, from https://sayingimages.com/congratulations-quotes-cards-messages-pictures/

Scanella, S., Peysakhovich, V., Ehrig, F., & Dehais, F. (2015, August). *Can flight phase be inferred using eye movements? Evidence from real flight conditions*. Presented at the 18th European Conference on Eye Movements, Vienna, Austria.

Schad, D. J., Nuthmann, A., & Engbert, R. (2012). Your mind wanders weakly, your mind wanders deeply: Objective measures reveal mindless reading at different levels. *Cognition*, *125*(2), 179–194. https://doi.org/10.1016/j.cognition.2012.07.004

Schooler, J. W., Mrazek, M. D., Franklin, M. S., Baird, B., Mooneyham, B. W., Zedelius, C., & Broadway, J. M. (2014). The Middle Way. In *Psychology of Learning and Motivation* (Vol. 60, pp. 1–33). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/B9780128000908000019

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *5*, 319–326. https://doi.org/10.1016/j.tics.2011.05.006

See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. *Psychological Bulletin*, *117*(2), 230–249. https://doi.org/10.1037/0033-2909.117.2.230

Seli, P., Beaty, R. E., Cheyne, J. A., Smilek, D., Oakman, J., & Schacter, D. L. (2018). How pervasive is mind wandering, really? *Consciousness and Cognition*, *66*, 74–78. https://doi.org/10.1016/j.concog.2018.10.002

Seli, P., Carriere, J. S. A., & Smilek, D. (2015). Not all mind wandering is created equal: dissociating deliberate from spontaneous mind wandering. *Psychological Research*, *79*(5), 750–758. https://doi.org/10.1007/s00426-014-0617-x

Seli, P., Kane, M. J., Metzinger, T., Smallwood, J., Schacter, D. L., Maillet, D., … Smilek, D. (2018). The Family-Resemblances Framework for Mind-Wandering Remains Well Clad. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2018.07.007

Seli, P., Kane, M., Smallwood, J., Schacter, D. L., Maillet, D., Schooler, J., & Smilek, D. (2018). Mind-wandering as a natural kind: A family-resemblances view. *Trends in Cognitive Sciences*, *22*(6). https://doi.org/10.1016/j.tics.2018.03.010

Seli, P., Maillet, D., Schacter, D. L., Kane, M. J., Smallwood, J., Schooler, J. W., & Smilek, D. (2017). What does (and should) “mind wandering” mean? Retrieved from https://osf.io/preprints/psyarxiv/q7s8u/

Seli, P., Risko, E. F., & Smilek, D. (2016). On the necessity of distinguishing between unintentional and intentional mind wandering. *Psychological Science*, *27*(5), 685–691. https://doi.org/10.1177/0956797616634068

Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-Wandering With and Without Intention. *Trends in Cognitive Sciences*, *20*(8), 605–617. https://doi.org/10.1016/j.tics.2016.05.010

Shappell, S., & Douglas, W. (2007). Developing a Methodology for Assessing Safety Programs Targeting Human Error in Aviation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 90–92). Los Angeles, CA: SAGE Publications.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge: MIT Press.

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of Human Factors and Ergonomics*, *1*(1), 89–129.

SIMPIT Technologies. (2018). Avenger 2 -180 FOV - Icarus. Retrieved July 31, 2018, from https://www.simpit.co.nz/index.php/icarus-180-avenger

Singer, J. L., & Antrobus, J. S. (1970). *Manual for the Imaginal Processes Inventory*. Princeton, NJ: Educational Testing Services.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). afex: Analysis of Factorial Experiments (Version 0.21-2).

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin*, *135*(1), 94–120. https://doi.org/10.1037/a0014212

Six, N. (2018, September 7). Pourquoi Jaguar et Aurrigo ont doté un véhicule autonome d'un regard mignon [News]. Retrieved September 7, 2018, from https://www.lemonde.fr/pixels/article/2018/09/07/pourquoi-jaguar-et-aurrigo-ont-dote-un-vehicule-autonome-d-un-regard-mignon_5351843_4408996.html

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, *51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Smallwood, J. (2011). Mind-wandering While Reading: Attentional Decoupling, Mindless Reading and the Cascade Model of Inattention. *Language and Linguistics Compass*, *5*(2), 63–77. https://doi.org/10.1111/j.1749-818X.2010.00263.x

Smallwood, J., Baracaia, S. F., Lowe, M., & Obonsawin, M. (2003). Task unrelated thought whilst encoding information. *Consciousness and Cognition*, *12*(3), 452–484. https://doi.org/10.1016/S1053-8100(03)00018-7

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of Cognitive Neuroscience*, *20*(3), 458–469.

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., … Schooler, J. W. (2011). Pupillometric Evidence for the Decoupling of Attention from Perceptual Input during Offline Thought. *PLoS ONE*, *6*(3), e18298. https://doi.org/10.1371/journal.pone.0018298

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*(4), 657–690. https://doi.org/10.1016/j.concog.2004.06.003

Smallwood, J., Fishman, D. J., & Schooler, J. W. (2007). Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic Bulletin & Review*, *14*(2), 230–236.

Smallwood, J., McSpadden, M., & Schooler, J. W. (2008). When attention matters: The curious incident of the wandering mind. *Memory & Cognition*, *36*(6), 1144–1150. https://doi.org/10.3758/MC.36.6.1144

Smallwood, J., Riby, L., Heim, D., & Davies, J. B. (2006). Encoding during the attentional lapse: Accuracy of encoding during the semantic sustained attention to response task. *Consciousness and Cognition*, *15*(1), 218–231. https://doi.org/10.1016/j.concog.2005.03.003

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010a). Failures of sustained attention in life, lab, and brain: Ecological validity of the SART. *Neuropsychologia*, *48*(9), 2564–2570. https://doi.org/10.1016/j.neuropsychologia.2010.05.002

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010b). Out of Mind, Out of Sight: Eye Blinking as Indicator and Embodiment of Mind Wandering. *Psychological Science*, *21*(6), 786–789. https://doi.org/10.1177/0956797610368063

Smith, R. P. (1981). Boredom: A review. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *23*(3), 329–340. https://doi.org/10.1177/001872088102300308

Somon, B., Campagne, A., Delorme, A., & Berberian, B. (2017). Performance Monitoring Applied to System Supervision. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00360

Sonderegger, A., & Sauer, J. (2010). The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics*, *41*(3), 403–410. https://doi.org/10.1016/j.apergo.2009.09.002

Song, X., & Wang, X. (2012). Mind wandering in Chinese daily lives–an experience sampling study. *PLoS One*, *7*(9), e44423.

Stawarczyk, D., Cassol, H., & D’Argembeau, A. (2013). Phenomenology of future-oriented mind-wandering episodes. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00425

Stawarczyk, D., & D’Argembeau, A. (2016). Conjoint influence of mind-wandering and sleepiness on task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1587–1600. https://doi.org/10.1037/xhp0000254

Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D’Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, *136*(3), 370–381. https://doi.org/10.1016/j.actpsy.2011.01.002

Stawarczyk, D., Majerus, S., Maquet, P., & D’Argembeau, A. (2011). Neural Correlates of Ongoing Conscious Experience: Both Task-Unrelatedness and Stimulus-Independence Are Related to Default Network Activity. *PLoS ONE*, *6*(2), e16997. https://doi.org/10.1371/journal.pone.0016997

Strauch, B. (2002). *Investigating human error: Incidents, accidents, and complex systems*. Burlington, VT : Ashgate.

Swartout, W., Paris, C., & Moore, J. (1991). Explanations in knowledge systems: Design for explainable expert systems. *IEEE Expert*, *6*(3), 58–64.

Szalma, J. L. (2014). On the Application of Motivation Theory to Human Factors/Ergonomics: Motivational Design Principles for Human–Technology Interaction. *Human Factors*, *56*(8), 1453 –1471. https://doi.org/10.1177/0018720814553471

Taylor, S. F., Stern, E. R., & Gehring, W. J. (2007). Neural Systems for Error Monitoring: Recent Findings and Theoretical Perspectives. *The Neuroscientist*, *13*(2), 160–172. https://doi.org/10.1177/1073858406298184

Teasdale, J. D., Dritschel, B. H., Taylor, M. J., Proctor, L., Lloyd, C. A., Nimmo-Smith, I., & Baddeley, A. D. (1995). Stimulus-independent thought depends on central executive resources. *Memory & Cognition*, *23*(5), 551–559.

Teichner, W. H. (1974). The detection of a simple visual signal as a function of time of watch. *Human Factors*, *16*(4), 339–352.

Thackray, R. I., Bailey, J. P., & Touchstone, R. M. (1975). *Physiological, subjective and performances correlates of reported boredom and monotony while performing a simulated radar control task*. US Department of Transportation - Federal Aviation Administration.

Thackray, R. I., & Touchstone, R. M. (1989). *A comparison of detection efficiency on an air traffic control monitoring task with and without computer aiding*. US Department of Transportation - Federal Aviation Administration. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA206422

The Economist. (2016, March 12). Showdown. *The Economist*. Retrieved from https://www.economist.com/science-and-technology/2016/03/12/showdown

The Mathworks Inc. (1992). *MATLAB User's Guide*. Natick, MA.

The Tesla Team. (2016, June 30). A Tragic Loss. Retrieved from https://www.tesla.com/blog/tragic-loss

This Place. (2016, May 23). MindRDR. Retrieved May 23, 2016, from http://mindrdr.thisplace.com/static/index.html

Thomson, D. R., Besner, D., & Smilek, D. (2016). A critical examination of the evidence for sensitivity loss in modern vigilance tasks. *Psychological Review*, *123*(1), 70–83. https://doi.org/10.1037/rev0000021

Thomson, D. R., Seli, P., Besner, D., & Smilek, D. (2014). On the link between mind wandering and task performance over time. *Consciousness and Cognition*, *27*, 14–26. https://doi.org/10.1016/j.concog.2014.04.001

Tsuchiya, N., Frässle, S., Wilke, M., & Lamme, V. (2016). No-report and report-based paradigms jointly unravel the NCC: response to Overgaard and Fazekas. *Trends in Cognitive Sciences*, *20*(4), 241–242. https://doi.org/10.1016/j.tics.2016.01.004

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences*, *19*(12), 757–770. https://doi.org/10.1016/j.tics.2015.10.002

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of Performance Monitoring and Adaptive Behavior. *Physiological Reviews*, *94*(1), 35–79. https://doi.org/10.1152/physrev.00041.2012

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(4), 601–615. https://doi.org/10.3758/s13415-016-0417-4

Uzzaman, S., & Joordens, S. (2011). The eyes know what you are thinking: Eye movements as an objective measure of mind wandering. *Consciousness and Cognition*, *20*(4), 1882–1886. https://doi.org/10.1016/j.concog.2011.09.010

van Charante, E. M., Cook, R. I., Woods, D. D., Yue, L., & Howie, M. B. (1993). Human-computer Interaction in Context: Physician Interaction with Automated Intravenous Controllers on the Heart. In *Analysis, Design and Evaluation of Man–Machine Systems 1992* (pp. 263–274). Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/B978008041900850044X

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*(8), 519–534. https://doi.org/10.1016/j.euroneuro.2010.03.008

van Gaal, S., & Lamme, V. A. F. (2012). Unconscious High-Level Information Processing: Implication for Neurobiological Theories of Consciousness. *The Neuroscientist*, *18*(3), 287–301. https://doi.org/10.1177/1073858411404079

van Schie, H. T., Mars, R. B., Coles, M. G. H., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, *7*(5), 549–554. https://doi.org/10.1038/nn1239

Villena-González, M., López, V., & Rodríguez, E. (2016). Orienting attention to visual or verbal/auditory imagery differentially impairs the processing of visual stimuli. *NeuroImage*, *132*, 71–78. https://doi.org/10.1016/j.neuroimage.2016.02.013

Wang, J. X. (2017, January 2). Therac-25 and industrial design engineering of socio-technical systems. Retrieved from https://www.crcpress.com/authors/news/i3158-therac-25-and-industrial-design-engineering-of-socio-technical-systems

Warm, J. S. (1984). *Sustained attention in human performance*. (J. S. Warm, Ed.). Chichester, UK: Wiley.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In M. Mouloua & R. Parasuraman (Eds.), *Automation and human performance: Theory and applications* (pp. 183–200). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 433–441. https://doi.org/10.1518/001872008X312152

Wegner, D. M. (2002). *The Illusion of Conscious Will* (MA: MIT Press). Cambridge.

What is MBCT? Do You Want To Know More About Mindfulness? (2015). Retrieved September 7, 2018, from https://sites.google.com/site/mbct4yppcdevon/do-you-want-to-know-about-mindfulness

Wickens, C. D. (2008). Situation Awareness: Review of Mica Endsley's 1995 Articles on Situation Awareness Theory and Measurement. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 397–403. https://doi.org/10.1518/001872008X288420

Wickens, C. D., & Kessel, C. (1977). *The effects of participatory mode and task workload on the detection of dynamic system failures*. Urbana-Champaign, Illinois: Aviation Research Laboratory.

Wickens, C., Dixon, S., Goh, J., & Hammer, B. (2005). *Pilot dependence on imperfect diagnostic automation in simulated UAV flights: An attentional visual scanning analysis*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA446167

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, *21*(12).

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*.

Wiener, E. L. (1981). Complacency: Is the term useful for air safety. In *Proceedings of the 26th Corporate Aviation Safety Seminar* (Vol. 117, pp. 116–125). Denver, CO: Flight Safety Foundation.

Wiener, E. L. (1987). Application of vigilance research: rare, medium, or well done? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *29*(6), 725–736.

Wiener, E. L. (1988). *Cockpit automation*. National Aeronautics and Space Administration. Retrieved from http://ntrs.nasa.gov/search.jsp?R=19890047073

Wiener, E. L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft* (p. 222). NASA Ames Research Center.

Wikipedia. (2018a). AlphaGo versus Lee Sedol. In *Wikipedia* (Consulted the 2018-08-27). Retrieved from https://en.wikipedia.org/w/index.php?title=AlphaGo_versus_Lee_Sedol&oldid=850185871

Wikipedia. (2018b). Automation. In *Wikipedia* (Consulted the 2018-08-14). Retrieved from https://en.wikipedia.org/w/index.php?title=Automation&oldid=854905390

Willems, B., & Truitt, T. R. (1999). *Implications of reduced involvement in en route air traffic control*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA369584

Wittgenstein, L. (1968). *Philosophical investigations* (3rd Edition). Basil Blackweel Ltd.

Woods, D. D., & Tinapple, D. (1999, September). *W3: Watching human factors watch people at work*. Presented at the 43rd Annual Meeting of the Human Factors and Ergonomics Society, Houston, TX.

Xu, M., Purdon, C., Seli, P., & Smilek, D. (2017). Mindfulness and mind wandering: The protective effects of brief meditation in anxious individuals. *Consciousness and Cognition*, *51*, 157–165. https://doi.org/10.1016/j.concog.2017.03.009

Yanko, M. R., & Spalek, T. M. (2014). Driving With the Wandering Mind: The Effect That Mind-Wandering Has on Driving Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(2), 260–269. https://doi.org/10.1177/0018720813495280

Yao, D., Wang, L., Oostenveld, R., Nielsen, K. D., Arendt-Nielsen, L., & Chen, A. C. N. (2005). A comparative study of different references for EEG spectral mapping: the issue of the neutral reference and the use of the infinity reference. *Physiological Measurement*, *26*(3), 173–184. https://doi.org/10.1088/0967-3334/26/3/003

Yoss, R. E., Moyer, N. J., & Hollenhorst, R. W. (1970). Pupil size and spontaneous pupillary waves associated with alertness, drowsiness, and sleep. *Neurology*, *20*(6).

Yu, R., & Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *Neuroreport*, *17*(16), 1747–1751.

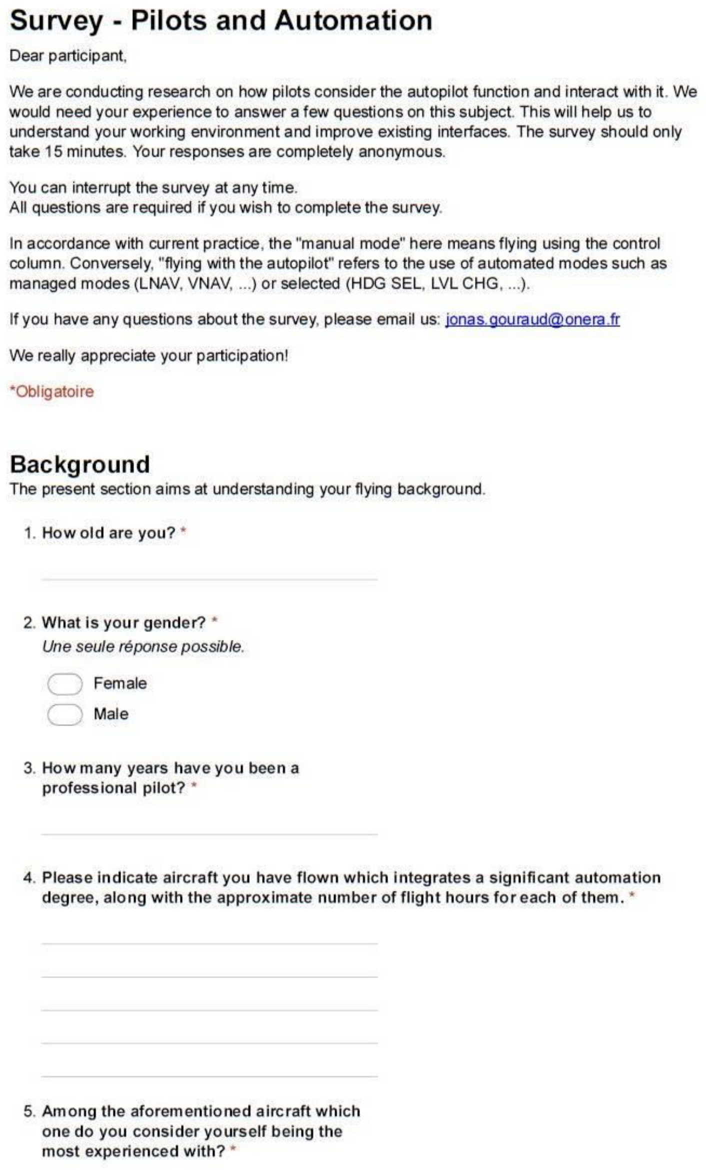
Zedelius, C. M., & Schooler, J. W. (2015). Mind wandering "Ahas" versus mindful reasoning: alternative routes to creative solutions. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00834

Zedelius, C. M., & Schooler, J. W. (2016). The Richness of Inner Experience: Relating Styles of Daydreaming to Creative Processes. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.02063

Zgur_. (2013, November 7). Vous êtes ici ... Retrieved September 14, 2018, from http://zgur.20minutes-blogs.fr/archive/2013/11/07/vous-etes-ici.html

10 APPENDIXES

A. Online questionnaire for pilots in English



Survey - Pilots and Automation

Dear participant,

We are conducting research on how pilots consider the autopilot function and interact with it. We would need your experience to answer a few questions on this subject. This will help us to understand your working environment and improve existing interfaces. The survey should only take 15 minutes. Your responses are completely anonymous.

You can interrupt the survey at any time.
All questions are required if you wish to complete the survey.

In accordance with current practice, the "manual mode" here means flying using the control column. Conversely, "flying with the autopilot" refers to the use of automated modes such as managed modes (LNAV, VNAV, ...) or selected (HDG SEL, LVL CHG, ...).

If you have any questions about the survey, please email us: jonas.gouraud@onera.fr

We really appreciate your participation!

*Obligatoire

Background

The present section aims at understanding your flying background.

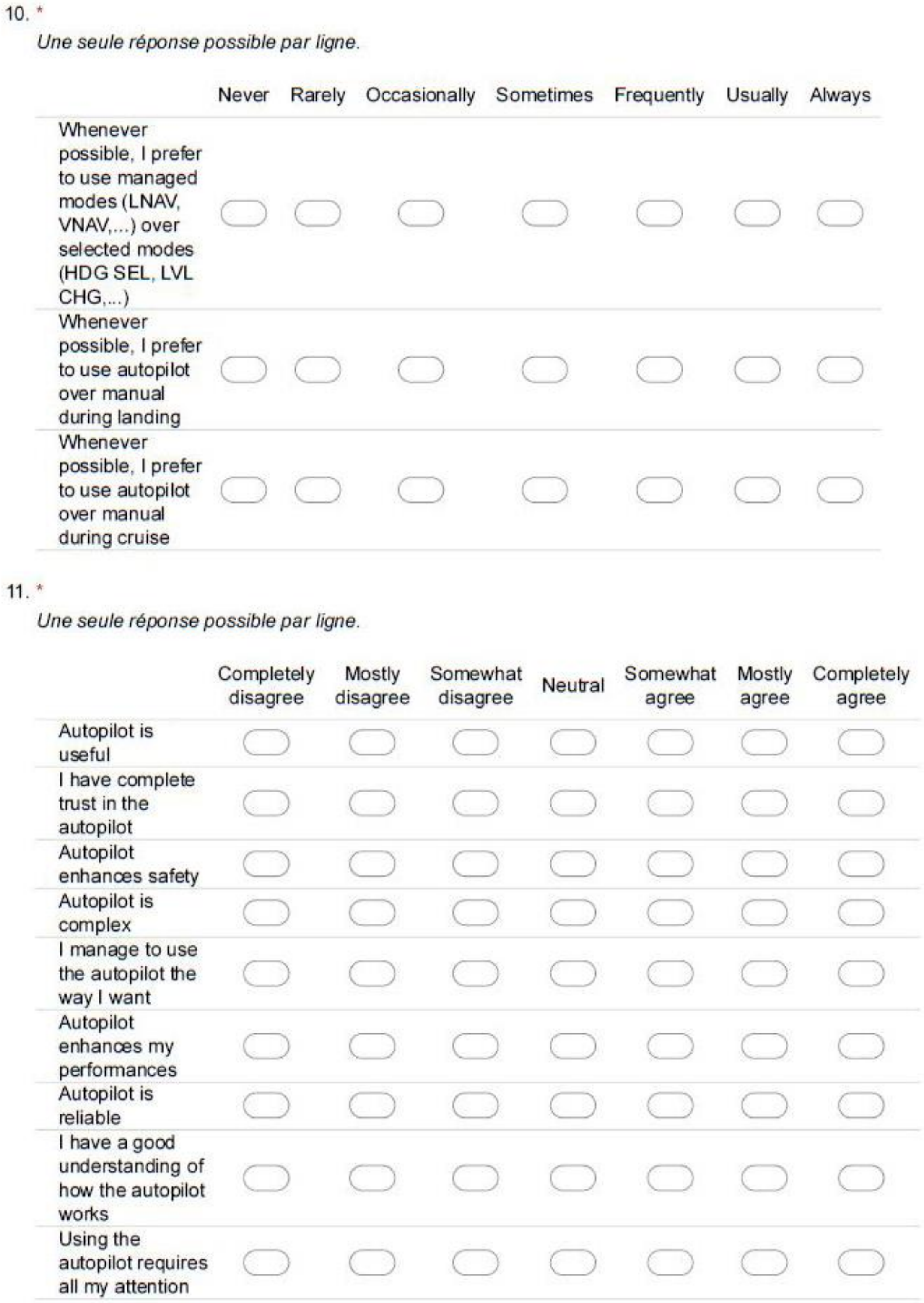
1. **How old are you?** *

2. **What is your gender?** *
   Une seule réponse possible.
   ○ Female
   ○ Male

3. **How many years have you been a professional pilot?** *

4. **Please indicate aircraft you have flown which integrates a significant automation degree, along with the approximate number of flight hours for each of them.** *

5. **Among the aforementioned aircraft which one do you consider yourself being the most experienced with?** *



10. *

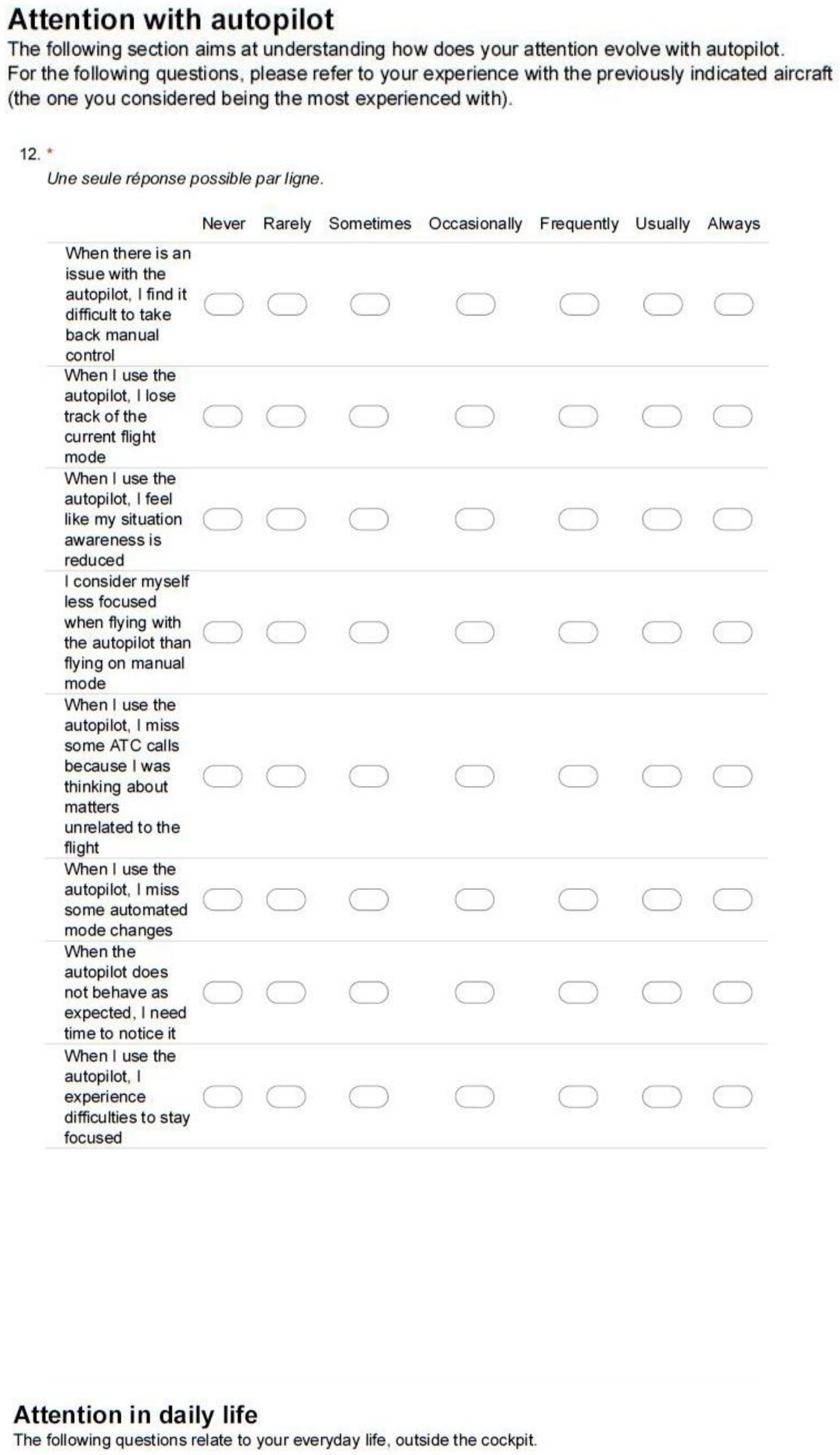
Une seule réponse possible par ligne.

	Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
Whenever possible, I prefer to use managed modes (LNAV, VNAV,...) over selected modes (HDG SEL, LVL CHG,...)	○	○	○	○	○	○	○
Whenever possible, I prefer to use autopilot over manual during landing	○	○	○	○	○	○	○
Whenever possible, I prefer to use autopilot over manual during cruise	○	○	○	○	○	○	○

11. *

Une seule réponse possible par ligne.

	Completely disagree	Mostly disagree	Somewhat disagree	Neutral	Somewhat agree	Mostly agree	Completely agree
Autopilot is useful	○	○	○	○	○	○	○
I have complete trust in the autopilot	○	○	○	○	○	○	○
Autopilot enhances safety	○	○	○	○	○	○	○
Autopilot is complex	○	○	○	○	○	○	○
I manage to use the autopilot the way I want	○	○	○	○	○	○	○
Autopilot enhances my performances	○	○	○	○	○	○	○
Autopilot is reliable	○	○	○	○	○	○	○
I have a good understanding of how the autopilot works	○	○	○	○	○	○	○
Using the autopilot requires all my attention	○	○	○	○	○	○	○



Attention with autopilot

The following section aims at understanding how does your attention evolve with autopilot.
For the following questions, please refer to your experience with the previously indicated aircraft (the one you considered being the most experienced with).

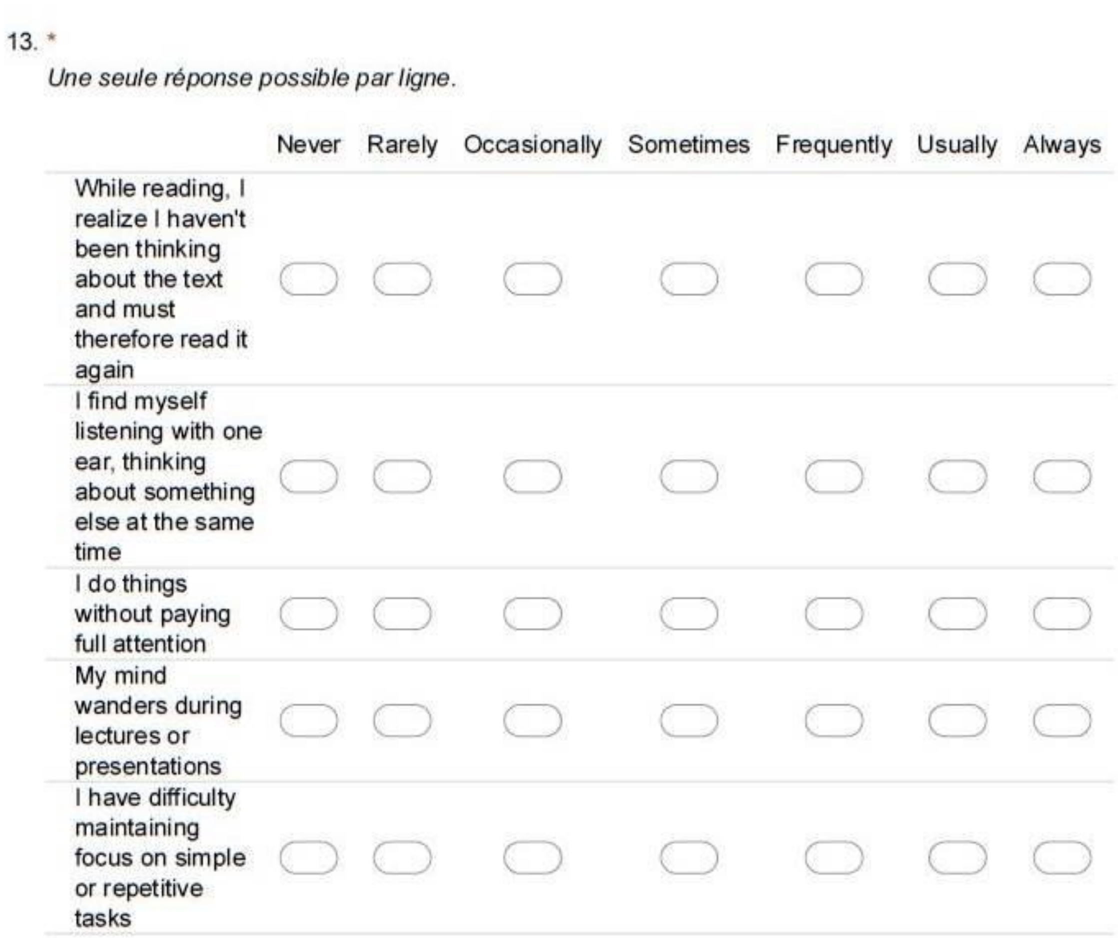
12. *

Une seule réponse possible par ligne.

	Never	Rarely	Sometimes	Occasionally	Frequently	Usually	Always
When there is an issue with the autopilot, I find it difficult to take back manual control	○	○	○	○	○	○	○
When I use the autopilot, I lose track of the current flight mode	○	○	○	○	○	○	○
When I use the autopilot, I feel like my situation awareness is reduced	○	○	○	○	○	○	○
I consider myself less focused when flying with the autopilot than flying on manual mode	○	○	○	○	○	○	○
When I use the autopilot, I miss some ATC calls because I was thinking about matters unrelated to the flight	○	○	○	○	○	○	○
When I use the autopilot, I miss some automated mode changes	○	○	○	○	○	○	○
When the autopilot does not behave as expected, I need time to notice it	○	○	○	○	○	○	○
When I use the autopilot, I experience difficulties to stay focused	○	○	○	○	○	○	○

Attention in daily life

The following questions relate to your everyday life, outside the cockpit.

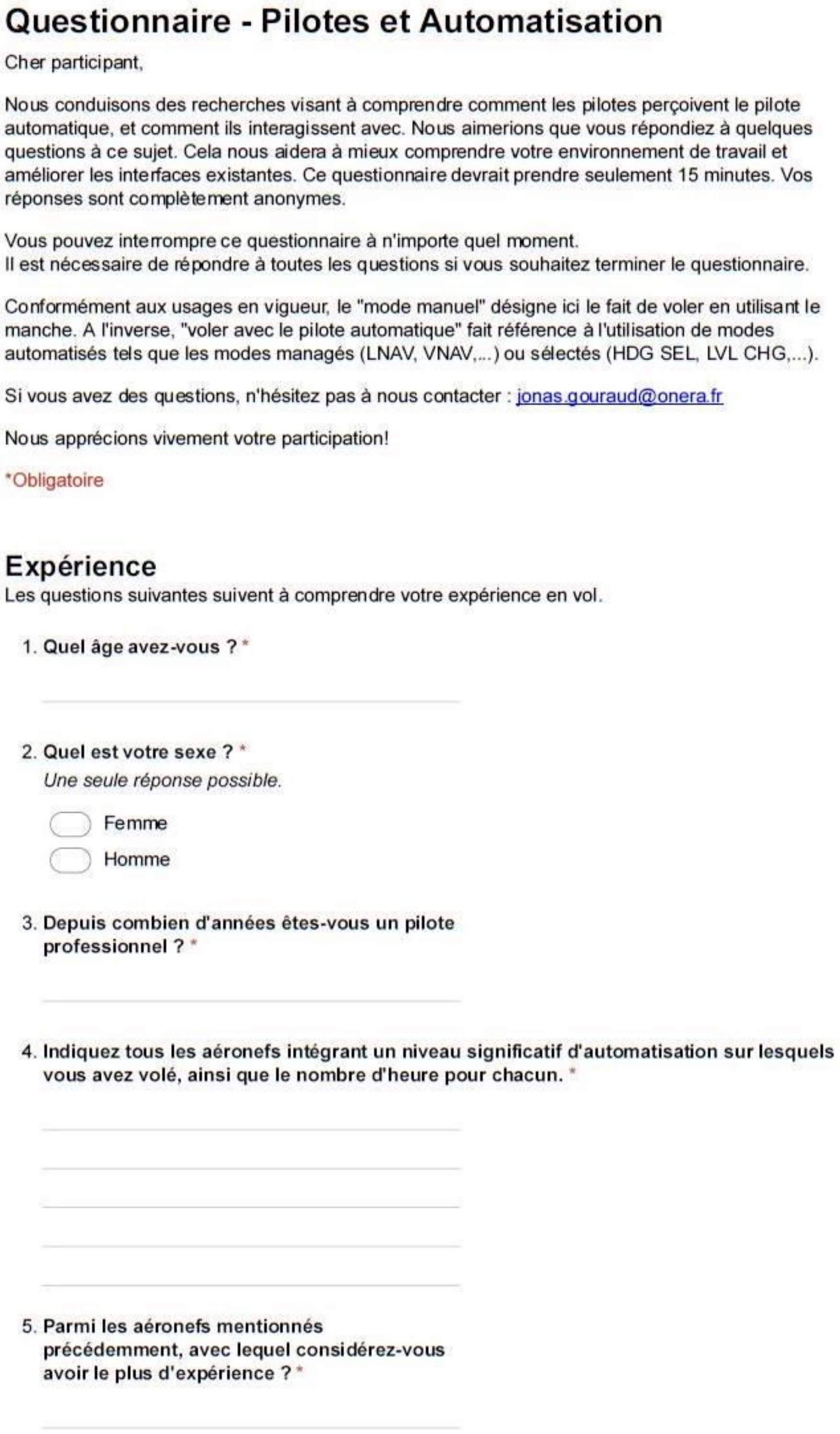


13. *

Une seule réponse possible par ligne.

	Never	Rarely	Occasionally	Sometimes	Frequently	Usually	Always
While reading, I realize I haven't been thinking about the text and must therefore read it again	○	○	○	○	○	○	○
I find myself listening with one ear, thinking about something else at the same time	○	○	○	○	○	○	○
I do things without paying full attention	○	○	○	○	○	○	○
My mind wanders during lectures or presentations	○	○	○	○	○	○	○
I have difficulty maintaining focus on simple or repetitive tasks	○	○	○	○	○	○	○

B. Online questionnaire for pilots in French



Questionnaire - Pilotes et Automatisation

Cher participant,

Nous conduisons des recherches visant à comprendre comment les pilotes perçoivent le pilote automatique, et comment ils interagissent avec. Nous aimerions que vous répondiez à quelques questions à ce sujet. Cela nous aidera à mieux comprendre votre environnement de travail et améliorer les interfaces existantes. Ce questionnaire devrait prendre seulement 15 minutes. Vos réponses sont complètement anonymes.

Vous pouvez interrompre ce questionnaire à n'importe quel moment.
Il est nécessaire de répondre à toutes les questions si vous souhaitez terminer le questionnaire.

Conformément aux usages en vigueur, le "mode manuel" désigne ici le fait de voler en utilisant le manche. A l'inverse, "voler avec le pilote automatique" fait référence à l'utilisation de modes automatisés tels que les modes managés (LNAV, VNAV,...) ou sélectés (HDG SEL, LVL CHG,...).

Si vous avez des questions, n'hésitez pas à nous contacter : jonas.gouraud@onera.fr

Nous apprécions vivement votre participation!

*Obligatoire

Expérience

Les questions suivantes suivent à comprendre votre expérience en vol.

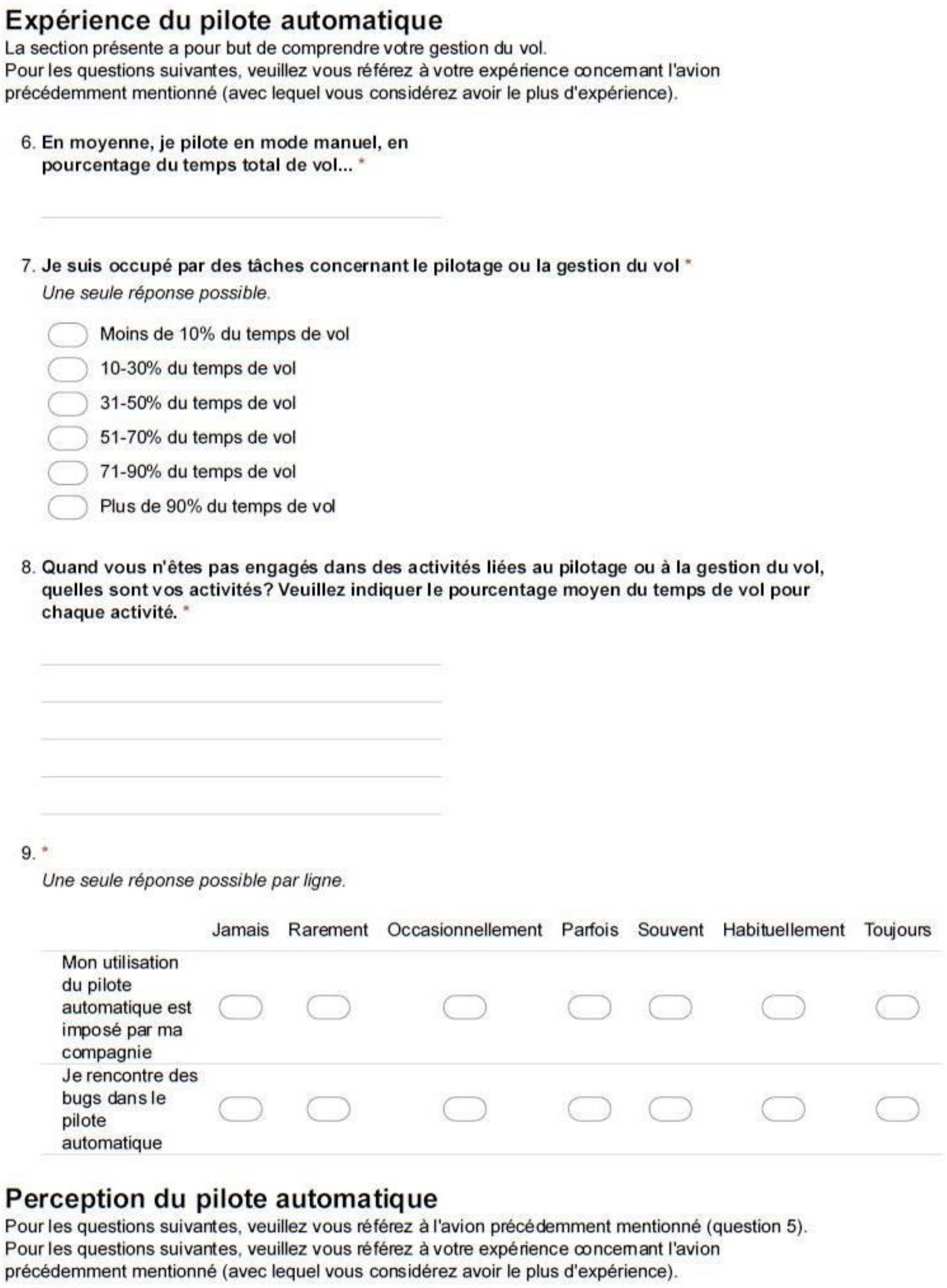
1. **Quel âge avez-vous ?** *

2. **Quel est votre sexe ?** *
   Une seule réponse possible.
   - ○ Femme
   - ○ Homme

3. **Depuis combien d'années êtes-vous un pilote professionnel ?** *

4. **Indiquez tous les aéronefs intégrant un niveau significatif d'automatisation sur lesquels vous avez volé, ainsi que le nombre d'heure pour chacun.** *

5. **Parmi les aéronefs mentionnés précédemment, avec lequel considérez-vous avoir le plus d'expérience ?** *



Expérience du pilote automatique

La section présente a pour but de comprendre votre gestion du vol.
Pour les questions suivantes, veuillez vous référez à votre expérience concernant l'avion précédemment mentionné (avec lequel vous considérez avoir le plus d'expérience).

6. **En moyenne, je pilote en mode manuel, en pourcentage du temps total de vol...** *

7. **Je suis occupé par des tâches concernant le pilotage ou la gestion du vol** *

Une seule réponse possible.

- ○ Moins de 10% du temps de vol
- ○ 10-30% du temps de vol
- ○ 31-50% du temps de vol
- ○ 51-70% du temps de vol
- ○ 71-90% du temps de vol
- ○ Plus de 90% du temps de vol

8. **Quand vous n'êtes pas engagés dans des activités liées au pilotage ou à la gestion du vol, quelles sont vos activités? Veuillez indiquer le pourcentage moyen du temps de vol pour chaque activité.** *

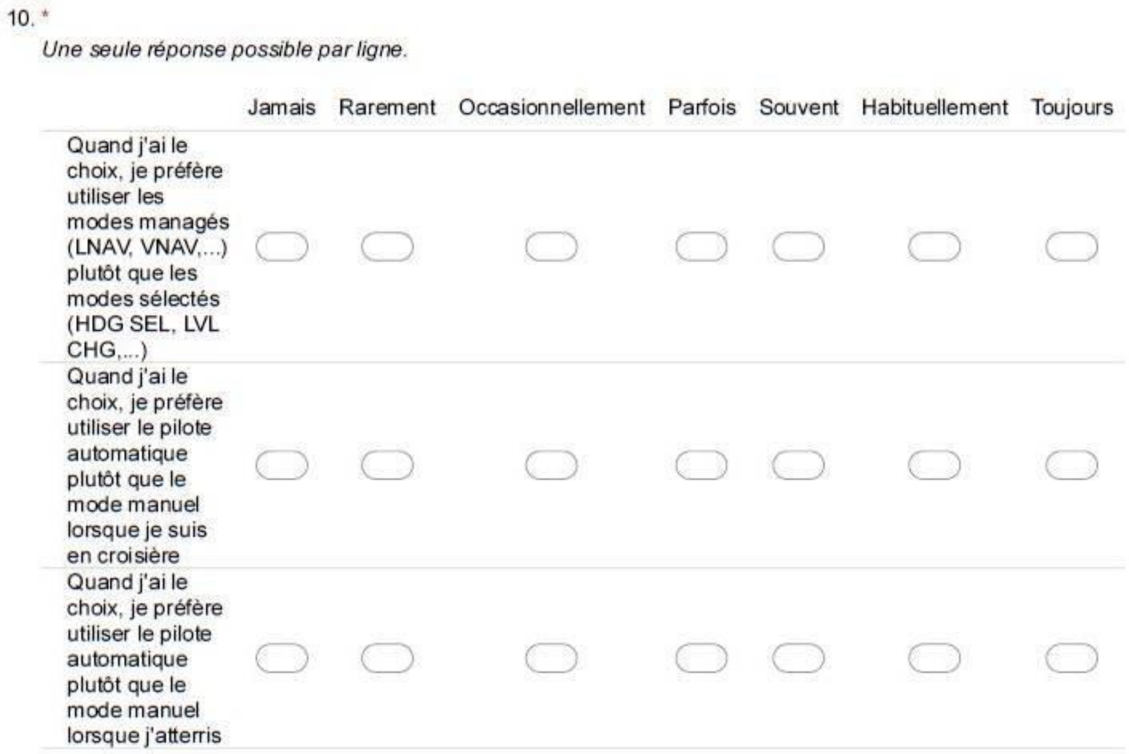
9. *

Une seule réponse possible par ligne.

	Jamais	Rarement	Occasionnellement	Parfois	Souvent	Habituellement	Toujours
Mon utilisation du pilote automatique est imposé par ma compagnie	○	○	○	○	○	○	○
Je rencontre des bugs dans le pilote automatique	○	○	○	○	○	○	○

Perception du pilote automatique

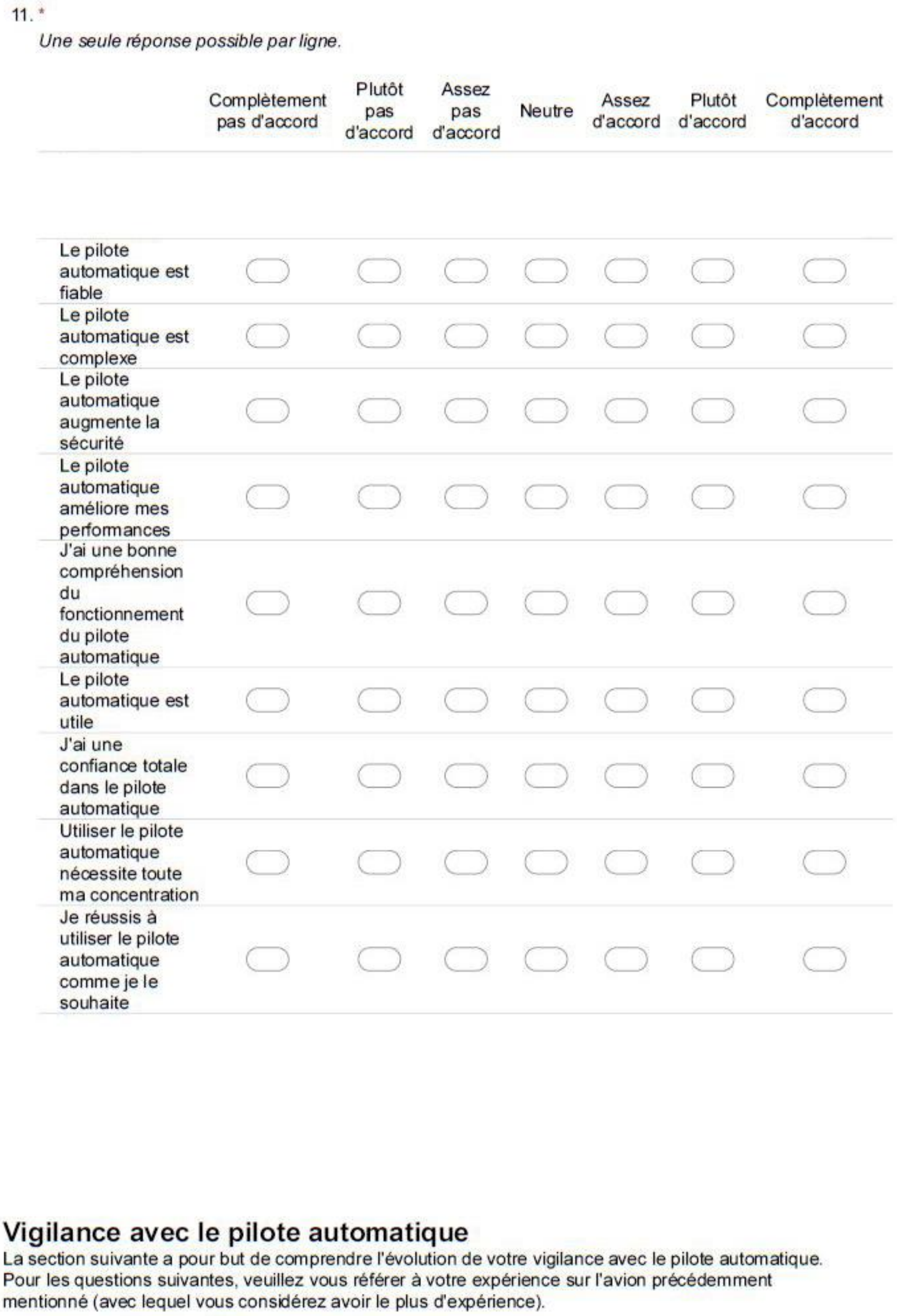
Pour les questions suivantes, veuillez vous référez à l'avion précédemment mentionné (question 5).
Pour les questions suivantes, veuillez vous référez à votre expérience concernant l'avion précédemment mentionné (avec lequel vous considérez avoir le plus d'expérience).



10. *

Une seule réponse possible par ligne.

	Jamais	Rarement	Occasionnellement	Parfois	Souvent	Habituellement	Toujours
Quand j'ai le choix, je préfère utiliser les modes managés (LNAV, VNAV,...) plutôt que les modes sélectés (HDG SEL, LVL CHG,...)	○	○	○	○	○	○	○
Quand j'ai le choix, je préfère utiliser le pilote automatique plutôt que le mode manuel lorsque je suis en croisière	○	○	○	○	○	○	○
Quand j'ai le choix, je préfère utiliser le pilote automatique plutôt que le mode manuel lorsque j'atterris	○	○	○	○	○	○	○



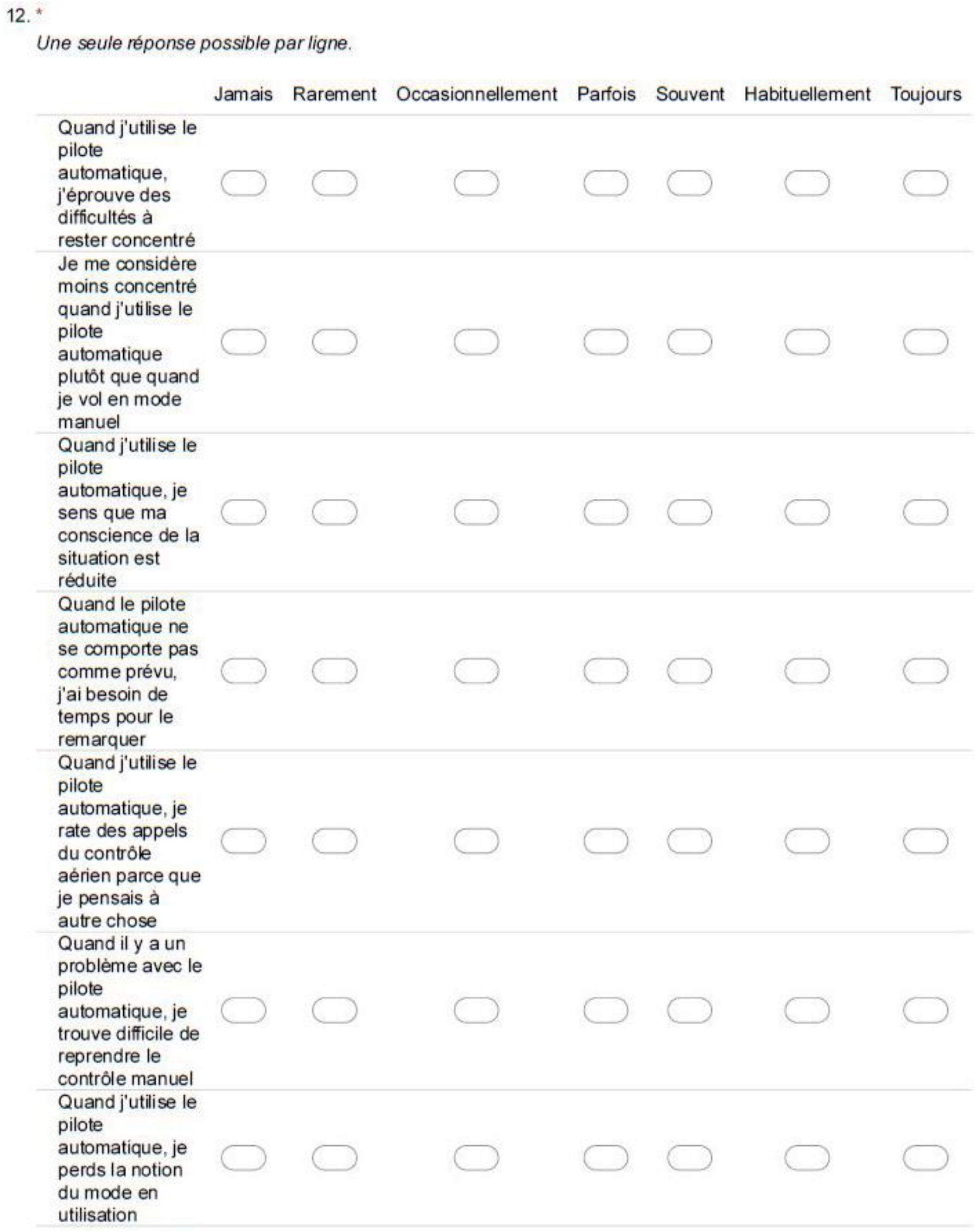
11. *

Une seule réponse possible par ligne.

	Complètement pas d'accord	Plutôt pas d'accord	Assez pas d'accord	Neutre	Assez d'accord	Plutôt d'accord	Complètement d'accord
Le pilote automatique est fiable	○	○	○	○	○	○	○
Le pilote automatique est complexe	○	○	○	○	○	○	○
Le pilote automatique augmente la sécurité	○	○	○	○	○	○	○
Le pilote automatique améliore mes performances	○	○	○	○	○	○	○
J'ai une bonne compréhension du fonctionnement du pilote automatique	○	○	○	○	○	○	○
Le pilote automatique est utile	○	○	○	○	○	○	○
J'ai une confiance totale dans le pilote automatique	○	○	○	○	○	○	○
Utiliser le pilote automatique nécessite toute ma concentration	○	○	○	○	○	○	○
Je réussis à utiliser le pilote automatique comme je le souhaite	○	○	○	○	○	○	○

Vigilance avec le pilote automatique

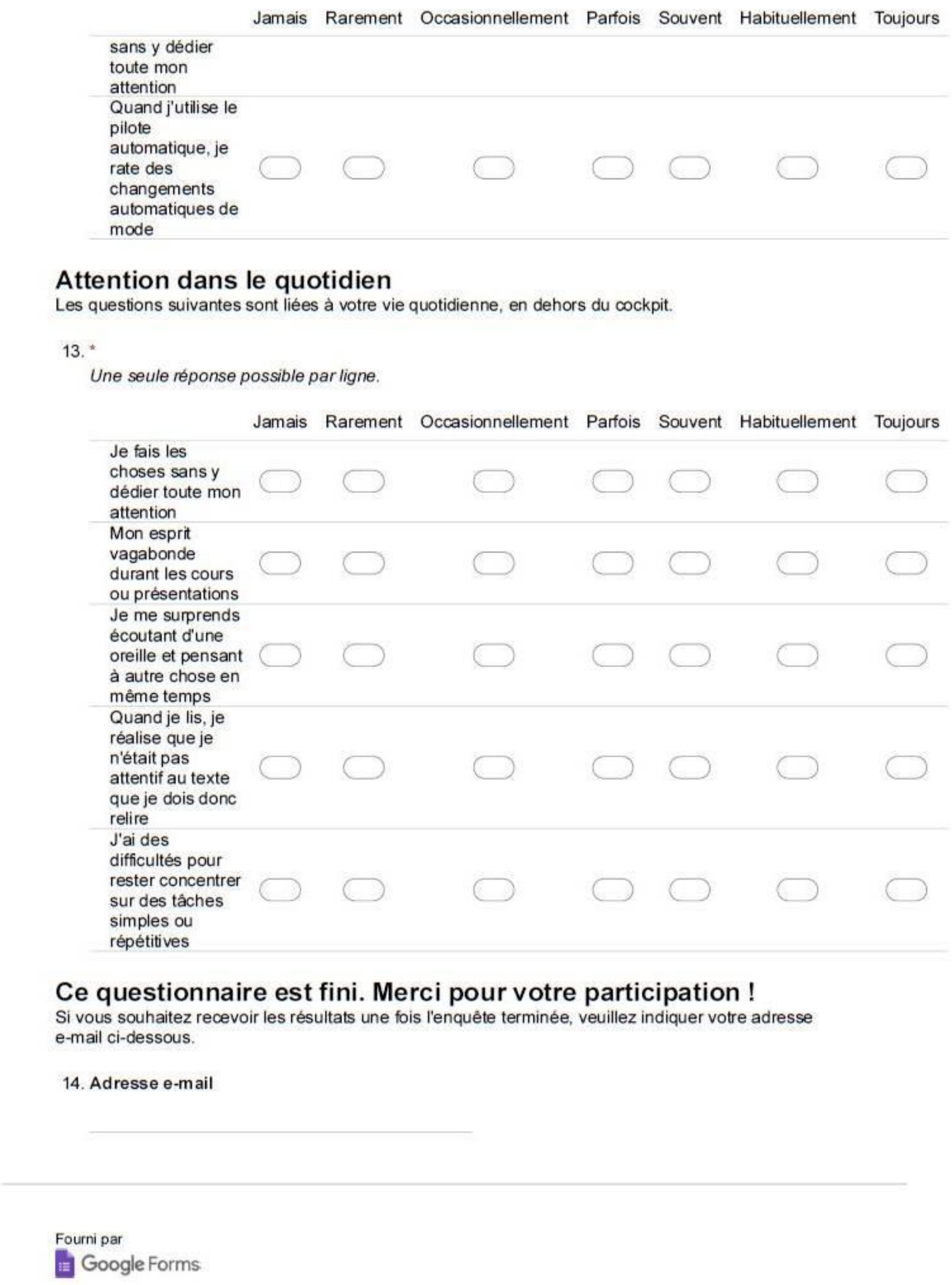
La section suivante a pour but de comprendre l'évolution de votre vigilance avec le pilote automatique. Pour les questions suivantes, veuillez vous référer à votre expérience sur l'avion précédemment mentionné (avec lequel vous considérez avoir le plus d'expérience).



12. *

Une seule réponse possible par ligne.

	Jamais	Rarement	Occasionnellement	Parfois	Souvent	Habituellement	Toujours
Quand j'utilise le pilote automatique, j'éprouve des difficultés à rester concentré	○	○	○	○	○	○	○
Je me considère moins concentré quand j'utilise le pilote automatique plutôt que quand je vol en mode manuel	○	○	○	○	○	○	○
Quand j'utilise le pilote automatique, je sens que ma conscience de la situation est réduite	○	○	○	○	○	○	○
Quand le pilote automatique ne se comporte pas comme prévu, j'ai besoin de temps pour le remarquer	○	○	○	○	○	○	○
Quand j'utilise le pilote automatique, je rate des appels du contrôle aérien parce que je pensais à autre chose	○	○	○	○	○	○	○
Quand il y a un problème avec le pilote automatique, je trouve difficile de reprendre le contrôle manuel	○	○	○	○	○	○	○
Quand j'utilise le pilote automatique, je perds la notion du mode en utilisation	○	○	○	○	○	○	○



	Jamais	Rarement	Occasionnellement	Parfois	Souvent	Habituellement	Toujours
sans y dédier toute mon attention							
Quand j'utilise le pilote automatique, je rate des changements automatiques de mode	○	○	○	○	○	○	○

Attention dans le quotidien

Les questions suivantes sont liées à votre vie quotidienne, en dehors du cockpit.

13. *

Une seule réponse possible par ligne.

	Jamais	Rarement	Occasionnellement	Parfois	Souvent	Habituellement	Toujours
Je fais les choses sans y dédier toute mon attention	○	○	○	○	○	○	○
Mon esprit vagabonde durant les cours ou présentations	○	○	○	○	○	○	○
Je me surprends écoutant d'une oreille et pensant à autre chose en même temps	○	○	○	○	○	○	○
Quand je lis, je réalise que je n'était pas attentif au texte que je dois donc relire	○	○	○	○	○	○	○
J'ai des difficultés pour rester concentrer sur des tâches simples ou répétitives	○	○	○	○	○	○	○

Ce questionnaire est fini. Merci pour votre participation !

Si vous souhaitez recevoir les résultats une fois l'enquête terminée, veuillez indiquer votre adresse e-mail ci-dessous.

14. **Adresse e-mail**

C. Online questionnaire for non-pilots in French

Attention au quotidien - MWQ

Cher participant,

Nous conduisons des recherches visant à comprendre l'attention au quotidien. Ce questionnaire devrait prendre seulement 2 minutes.

Vos réponses sont complètement anonymes. Vous pouvez interrompre ce questionnaire à n'importe quel moment. Il est nécessaire de répondre à toutes les questions si vous souhaitez terminer le questionnaire.

Si vous avez des questions, n'hésitez pas à nous contacter : jonas.gouraud@onera.fr

Nous apprécions vivement votre participation!

*Obligatoire

Profession

Cette section vise à recueillir des informations démographiques et comprendre votre profession.

1. **Quel âge avez-vous ?** *

2. **Quel est votre sexe ?** *

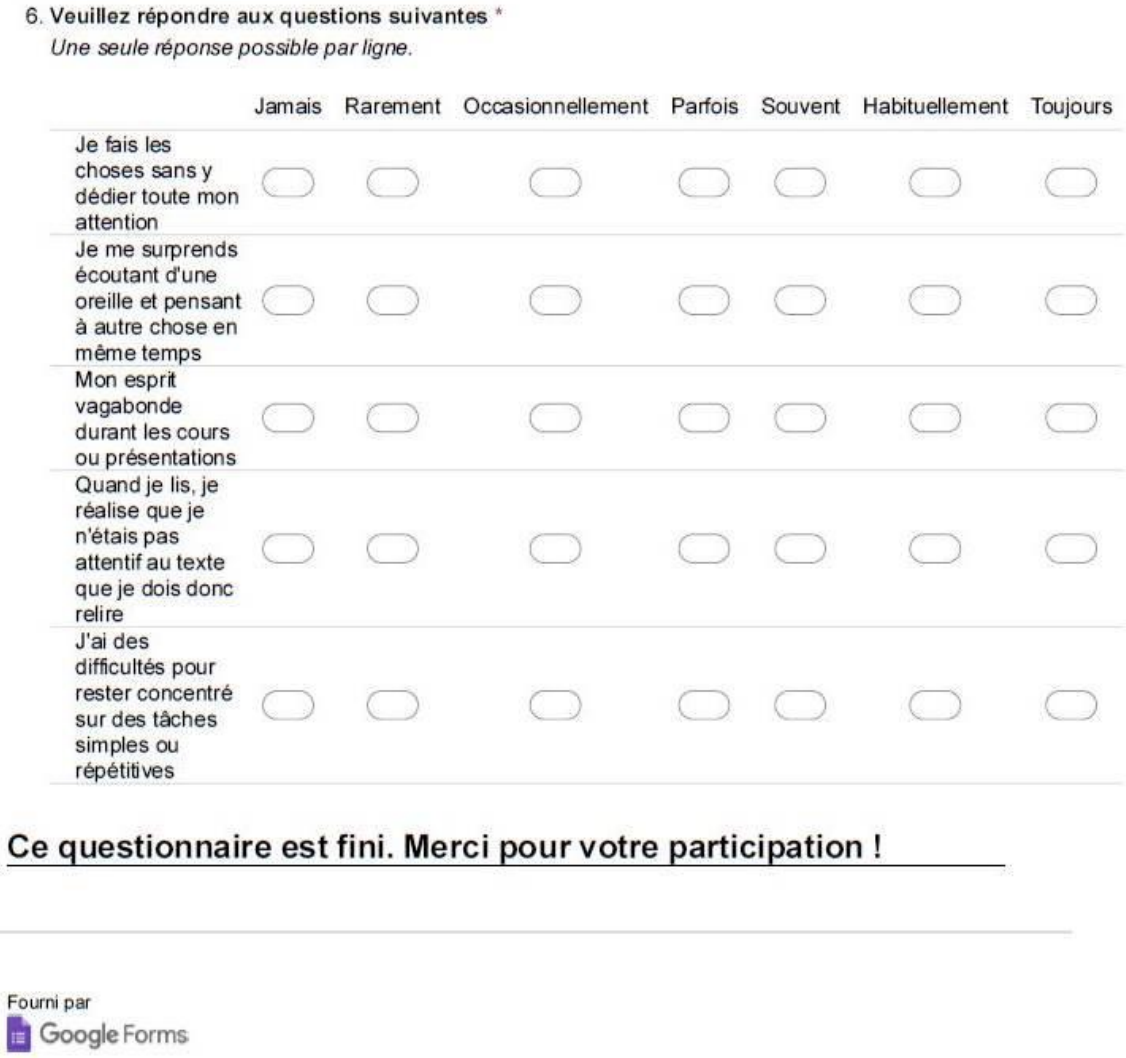
   Une seule réponse possible.

   ○ Femme
   ○ Homme

3. **Quelle est votre profession ?** *

Attention au quotidien

Cette section s'intéresse à votre attention au quotidien.



6. **Veuillez répondre aux questions suivantes** *

Une seule réponse possible par ligne.

	Jamais	Rarement	Occasionnellement	Parfois	Souvent	Habituellement	Toujours
Je fais les choses sans y dédier toute mon attention	○	○	○	○	○	○	○
Je me surprends écoutant d'une oreille et pensant à autre chose en même temps	○	○	○	○	○	○	○
Mon esprit vagabonde durant les cours ou présentations	○	○	○	○	○	○	○
Quand je lis, je réalise que je n'étais pas attentif au texte que je dois donc relire	○	○	○	○	○	○	○
J'ai des difficultés pour rester concentré sur des tâches simples ou répétitives	○	○	○	○	○	○	○

Ce questionnaire est fini. Merci pour votre participation !

Fourni par Google Forms

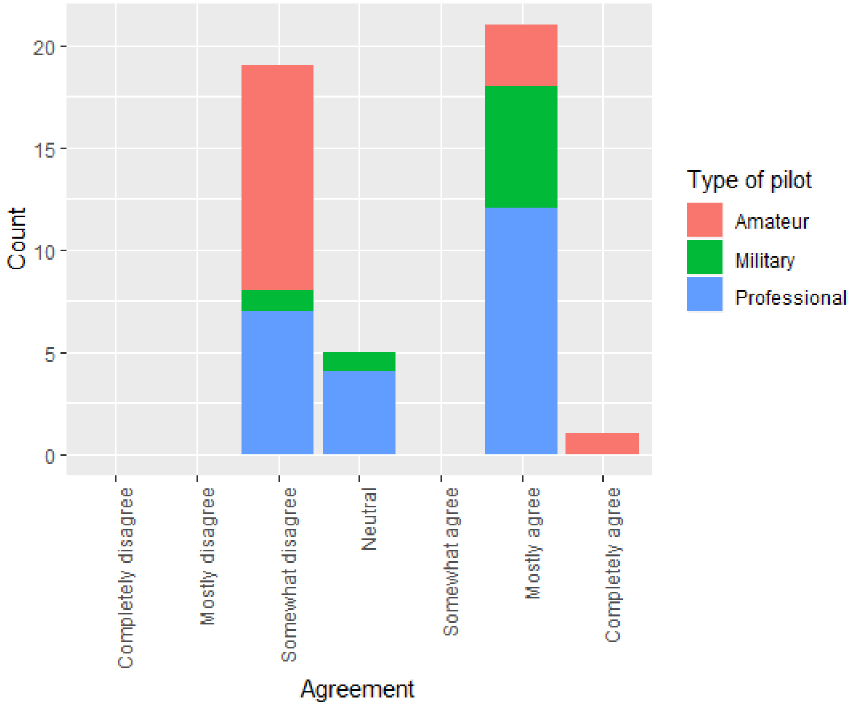
D. Aircraft reported as “reference aircraft”

Aircrafts	Number of reports	Type of pilot
AIRBUS 320	6	Professional
SOCCATA TB20	6	Amateur
DIAMOND DA42	5	Amateur
BOEING 737	3	Professional
TRANSALL C160	2	Military
SOCATA TBM 700	2	Professional
AIRBUS 340	2	Professional
DYNAMIC WT9	1	Amateur
BEECHCRAFT BARON58	1	Amateur
AIRBUS 400M	1	Military
AIRBUS 330 MRTT	1	Military
EMBRAER 110	1	Military
LOCKHEED C130	1	Military
TEJAS LCA	1	Military
EMBRAER 195	1	Professional
CIRRUS SR22	1	Amateur
TWIN OTTER	1	Professional
BOEING 777	1	Professional
CESSNA 172	1	Amateur
CASACN 235	1	Professional
BOEING 767	1	Professional
EMBRAER 120	1	Professional
AIRBUS 330	1	Professional
MD11	1	Professional
BOEING 747	1	Professional

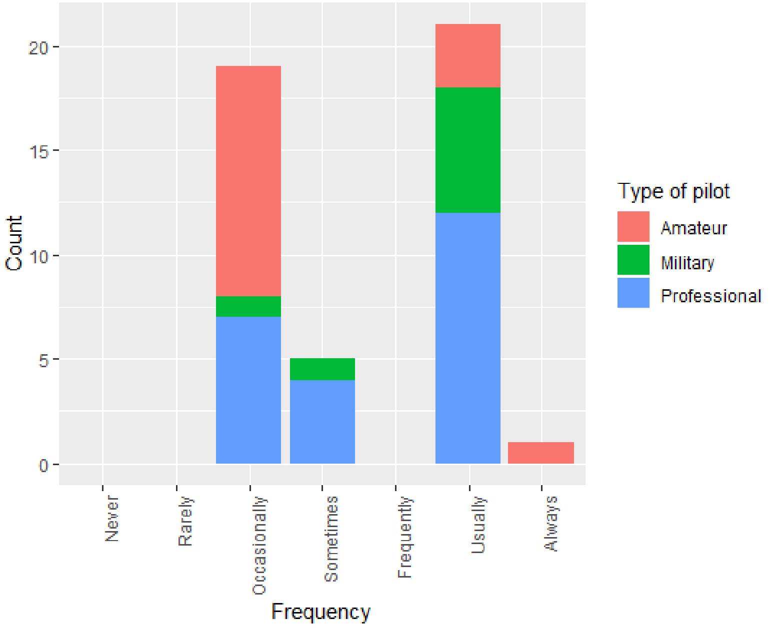
E. Pilots’ answers for part 3 “autopilot usability”

All confidence intervals were computed using bootstrap with 10000 iterations.

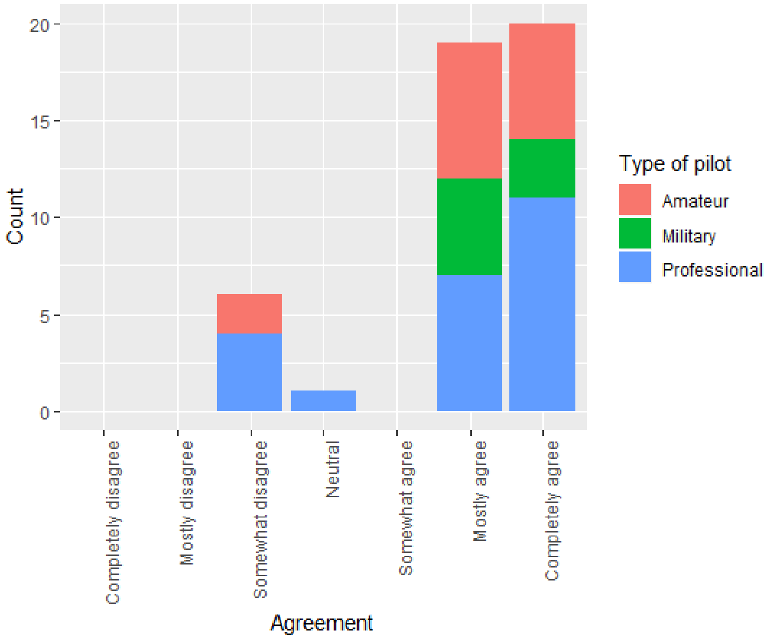
Items	Mean and 95% confidence interval			
	Overall	Amateur	Military	Professional
All part 3 items	5.6 [5.4; 5.8]	5.4 [5.1; 5.7]	5.9 [5.6; 6.2]	5.7 [5.4; 6.0]
Autopilot is useful	6.8 [6.7; 6.9]	6.7 [6.4; 6.9]	3.9 [6.6; 7.1]	6.9 [6.7; 7.0]
I have complete trust in the autopilot	4.6 [4.1; 5.0]	3.9 [3.1; 4.6]	5.4 [4.6; 6.1]	4.7 [4.2; 5.3]
Autopilot enhances safety	6.0 [5.6; 6.4]	6.0 [5.4; 6.6]	6.4 [6.0; 6.7]	5.9 [5.3; 6.5]
Autopilot is complex	4.1 [3.6; 4.6]	3.7 [2.9; 4.5]	4.3 [3.1; 5.4]	4.2 [3.5; 4.9]
I manage to use the autopilot the way I want	6.0 [5.7; 6.3]	5.9 [5.5; 6.4]	6.0 [5.1; 6.8]	6.0 [5.5; 6.6]
Autopilot enhances my performances	6.0 [5.6; 6.3]	5.9 [5.4; 6.5]	6.2 [6.0; 6.6]	5.9 [5.3; 6.4]
Autopilot is reliable	5.9 [5.5; 6.3]	5.1 [4.4; 5.8]	6.1 [5.9; 6.4]	6.3 [5.9; 6.8]
I have a good understanding of how the autopilot works	5.8 [5.4; 6.2]	5.5 [4.8; 6.2]	6.6 [6.3; 7.0]	5.7 [5.1; 6.3]



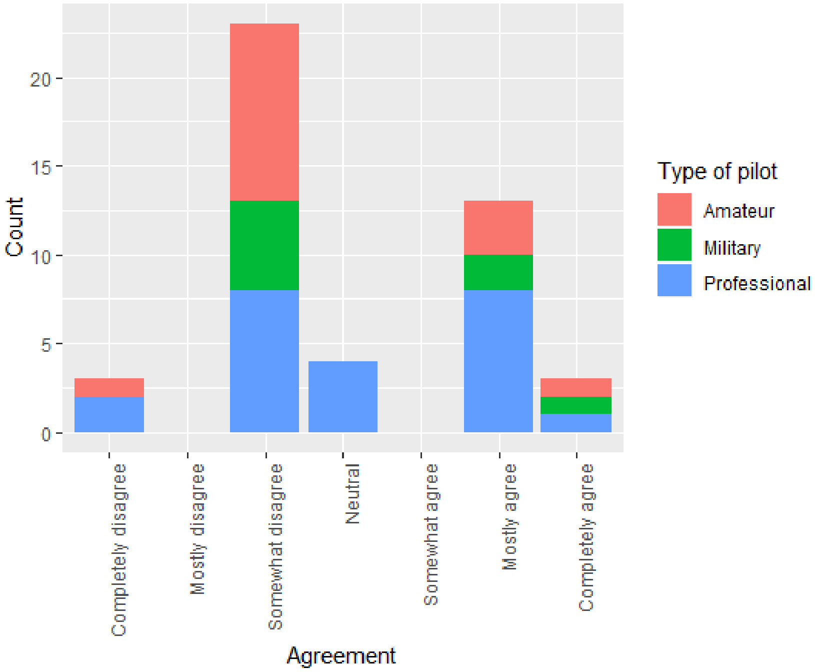
Pilots' answers to "Autopilot is useful"



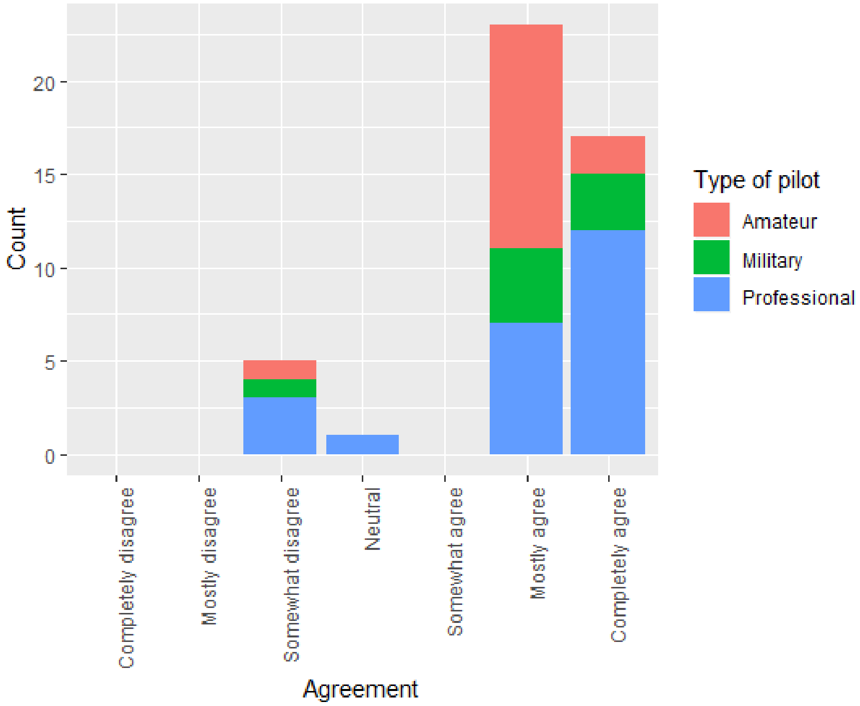
Pilots' answers to "I have complete trust in the autopilot"



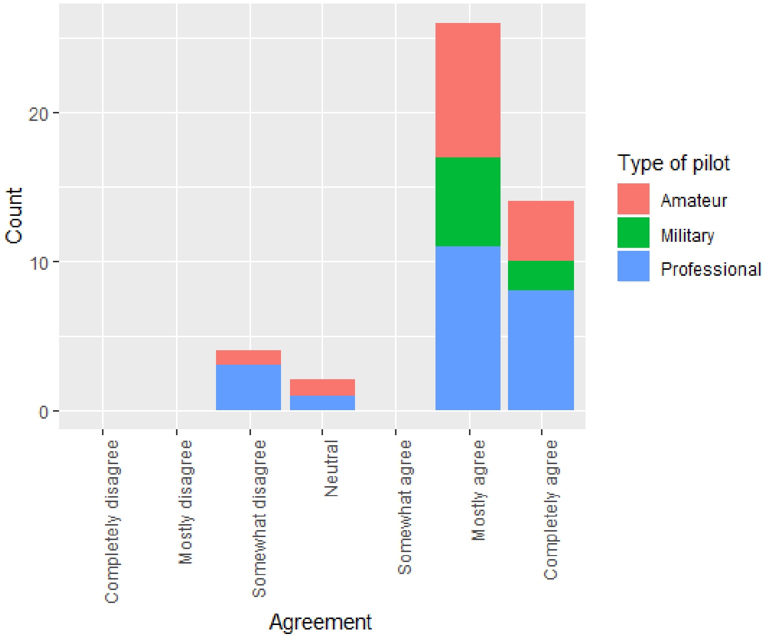
Pilots’ answers to “Autopilot enhances safety”



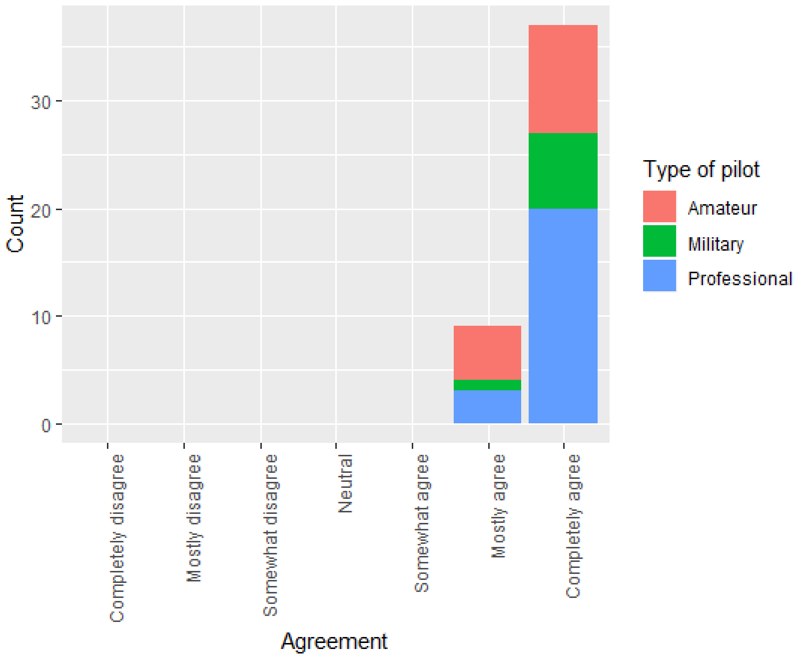
Pilots’ answers to “Autopilot is complex”



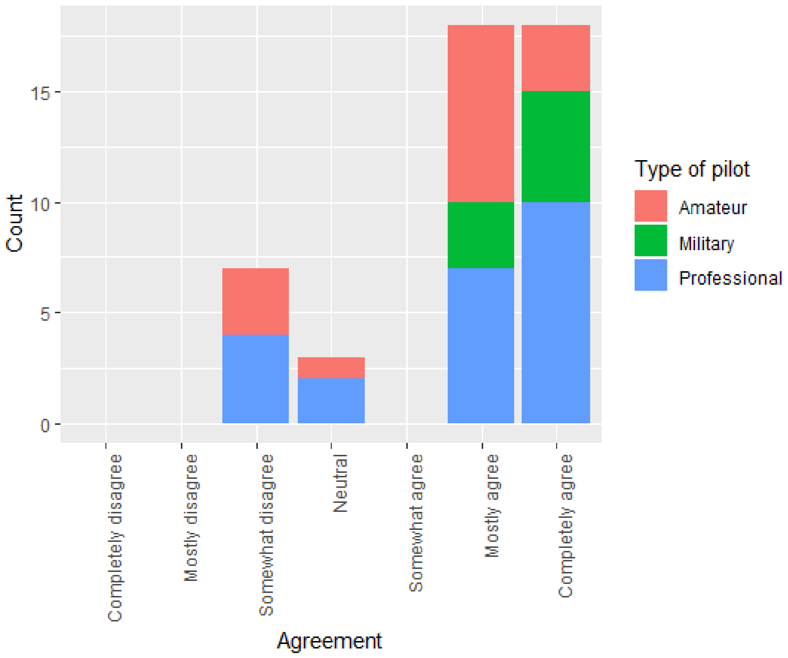
Pilots' answer to "I manage to use the autopilot the way I want"



Pilots' answers to "Autopilot enhances my performances"



Pilots' answers to "Autopilot is reliable"



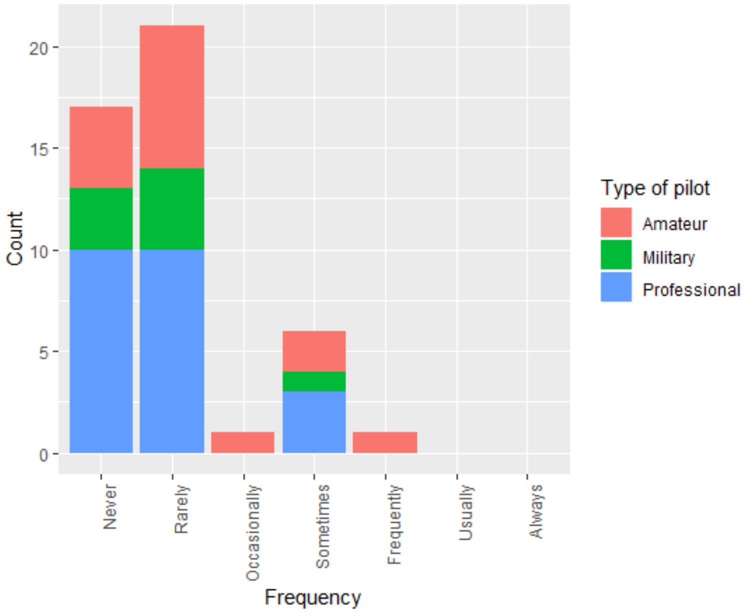
Pilots' answers to "I have a good understanding of how the autopilot works"

F. Pilots' answers for part 4 "human-autopilot problems"

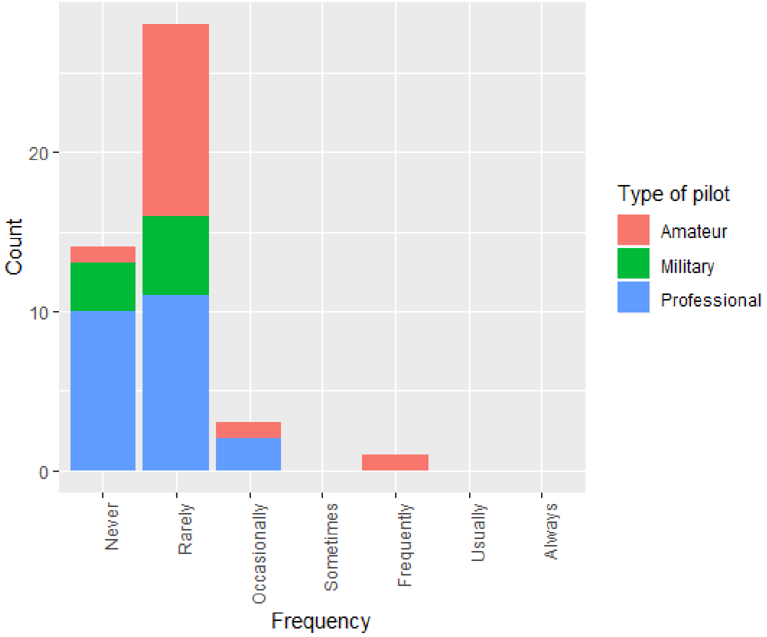
All confidence intervals were computed using bootstrap with 10000 iterations.

Items	Mean and 95% confidence interval			
	Overall	Amateur	Military	Professional
All part 4 items	2.3 [2.1; 2.5]	2.6 [2.2; 3.0]	2.0 [1.7; 2.4]	2.2 [1.9; 2.4]
When there is an issue with the autopilot, I find it difficult to take back manual control	2.0 [1.7; 2.3]	2.3 [1.7; 2.9]	1.9 [1.2; 2.5]	1.8 [1.4; 2.2]
When I use the autopilot, I lose track of the current flight mode	1.8 [1.6; 2.0]	2.2 [1.8; 2.6]	1.6 [1.3; 2.0]	1.6 [1.4; 1.9]
When I use the autopilot, I feel like my situation awareness is reduced	2.2 [1.9; 2.5]	2.5 [1.9; 3.2]	1.9 [1.3; 2.4]	2.1 [1.7; 2.4]
I consider myself less focused when flying with the autopilot than flying on manual mode	3.0 [2.6; 3.4]	3.7 [2.8; 4.5]	2.5 [1.8; 3.2]	2.7 [2.2; 3.2]
When I use the autopilot, I miss some ATC calls because I was thinking about matters unrelated to the flight	2.0 [1.7; 2.2]	2.1 [1.6; 2.5]	1.8 [1.4; 2.1]	2.0 [1.6; 2.3]

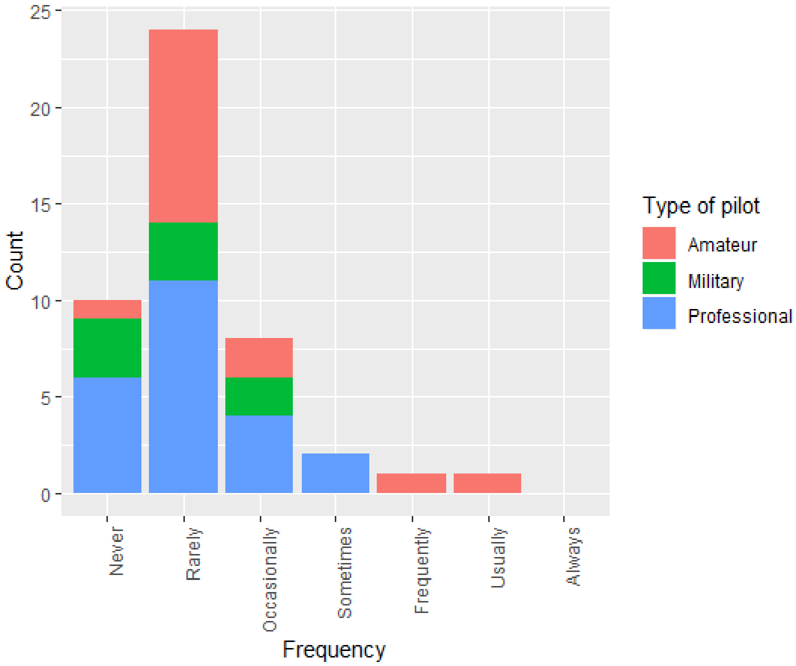
When I use the autopilot, I miss some automated mode changes	2.1 [1.9; 2.3]	2.3 [2.0; 2.5]	1.8 [1.3; 2.2]	2.2 [1.8; 2.5]
When the autopilot does not behave as expected, I need time to notice it	2.7 [2.4; 3.1]	3.1 [2.3; 3.8]	2.6 [1.8; 3.4]	2.6 [2.2; 2.9]
When I use the autopilot, I experience difficulties to stay focused	2.3 [2.0; 2.6]	2.5 [1.9; 3.0]	2.3 [1.7; 2.8]	2.2 [1.7; 2.7]



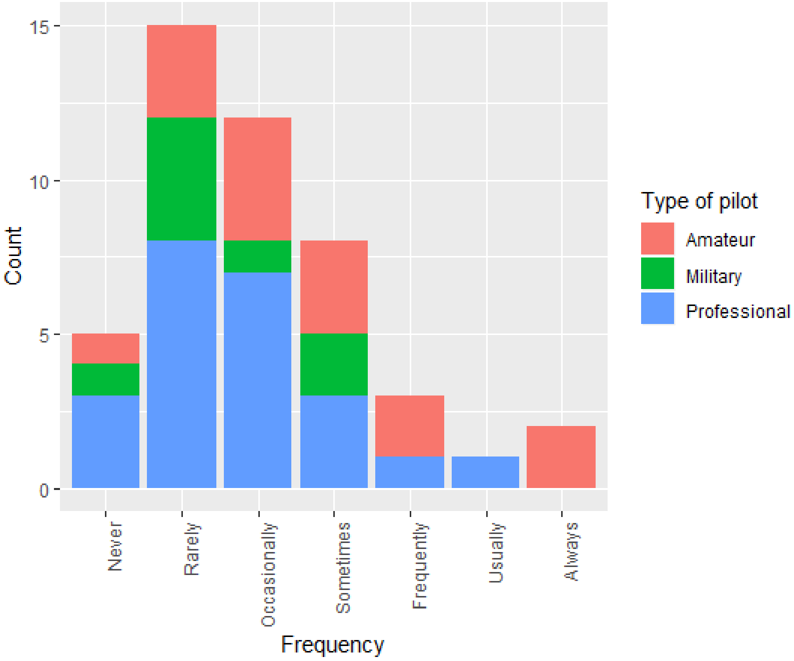
Pilots' answers to "When there is an issue with the autopilot, I find it difficult to take back manual control"



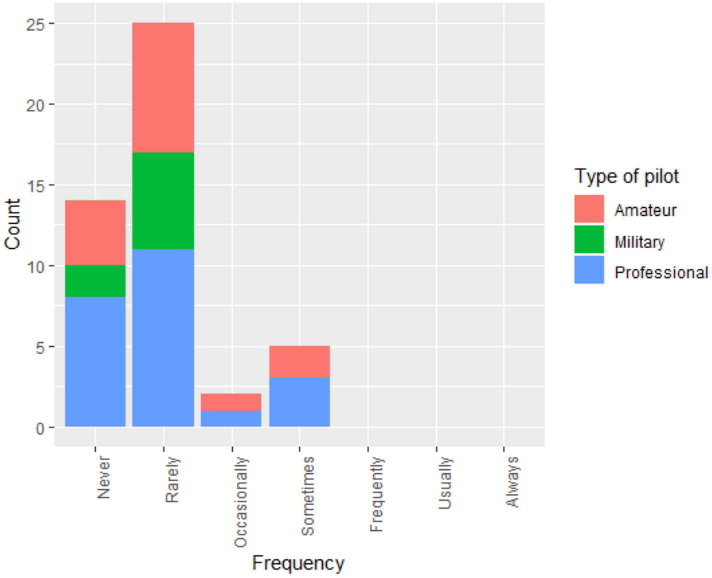
Pilots' answers to "When I use the autopilot, I lose track of the current flight mode"



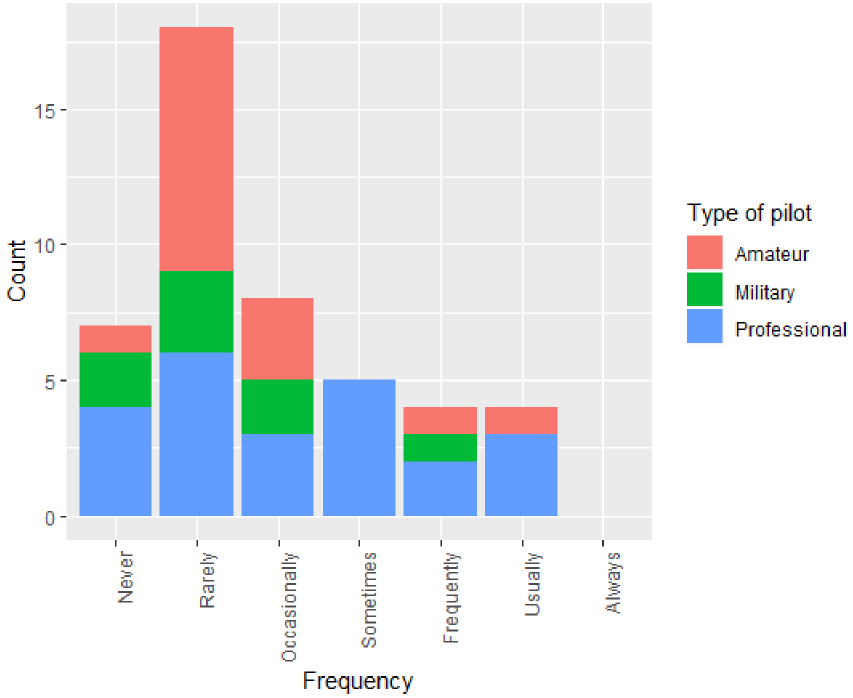
Pilots' answers to "When I use the autopilot, I feel like my situation awareness is reduced"



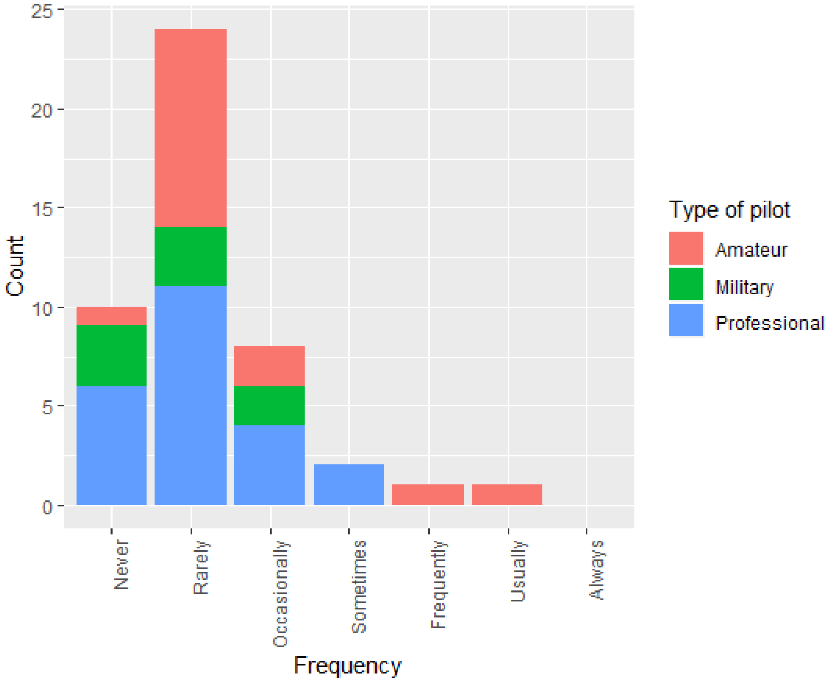
Pilots' answers to "I consider myself less focused when flying with the autopilot than flying on manual mode"



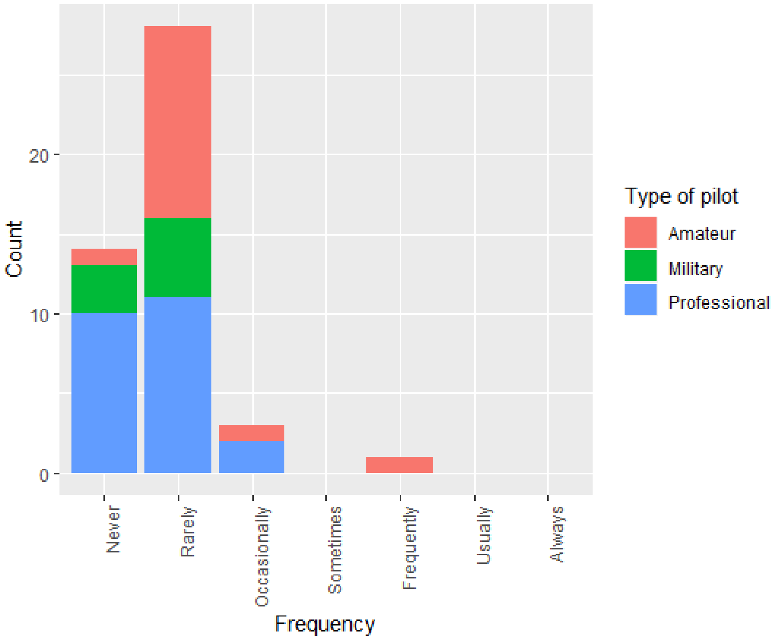
Pilots' answers to "When I use the autopilot, I miss some ATC calls because I was thinking about matters unrelated to the flight"



Pilots' answers to "When I use the autopilot, I miss some automated mode changes"



Pilots' answers to "When the autopilot does not behave as expected, I need time to notice it"

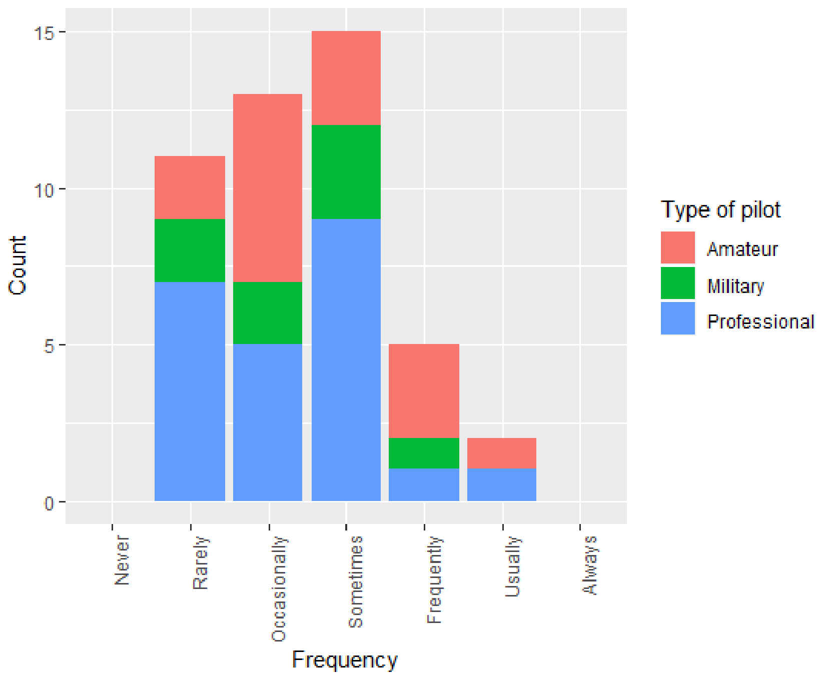


Pilots' answers to "When I use the autopilot, I experience difficulties to stay focused"

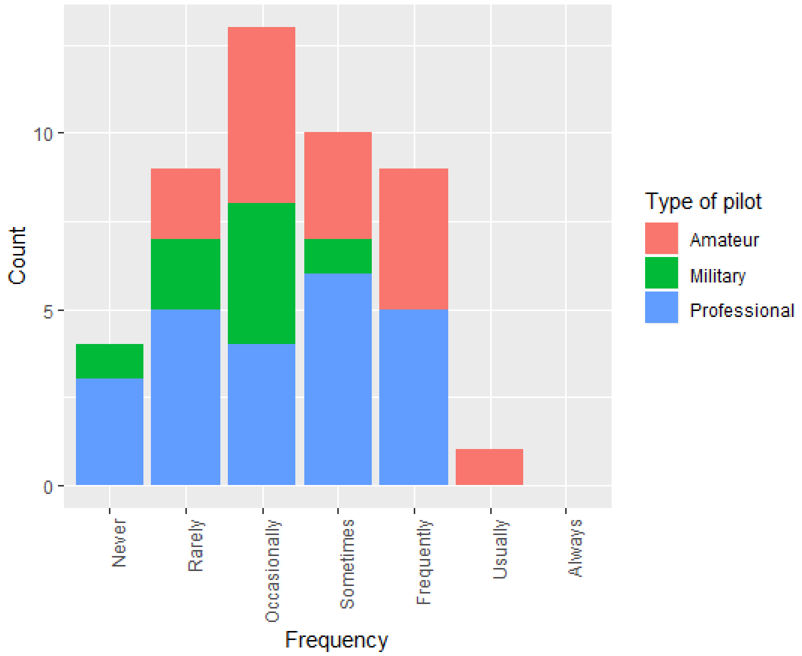
G.Pilots' answers for part 5 "mind wandering questionnaire"

All confidence intervals were computed using bootstrap with 10000 iterations.

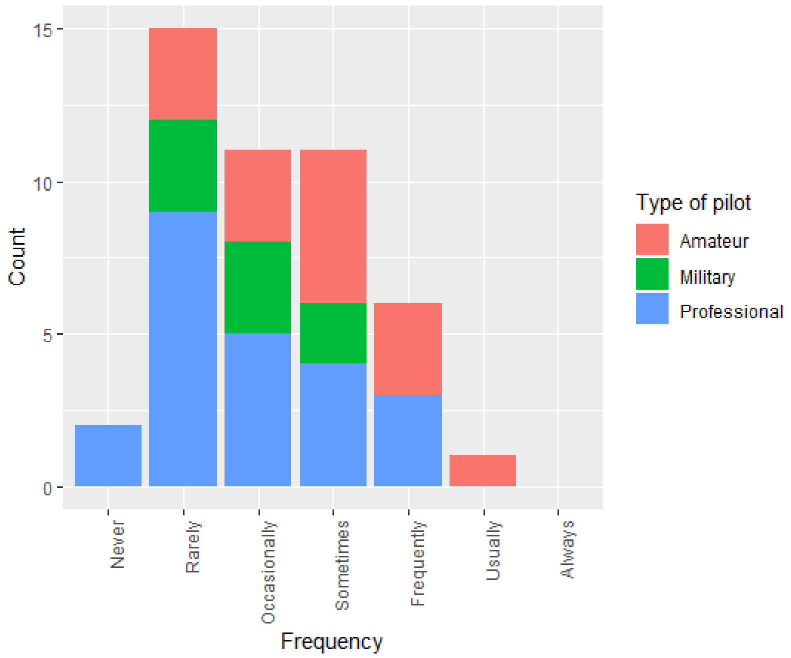
Items	Mean and 95% confidence interval			
	Overall	Amateur	Military	Professional
All part 5 items	3.2 [3.0; 3.5]	3.7 [3.3; 4.1]	2.9 [2.5; 3.3]	3.0 [2.7; 3.4]
While reading, I realize I haven't been thinking about the text and must therefore read it again	3.4 [3.1; 3.8]	3.7 [3.1; 4.2]	3.4 [2.7; 4.1]	3.3 [2.9; 3.7]
I find myself listening with one ear, thinking about something else at the same time	3.3 [2.9; 3.7]	3.8 [3.2; 4.4]	2.6 [2.0; 3.2]	3.2 [2.7; 3.8]
I do things without paying full attention	3.2 [2.8; 3.5]	3.7 [3.1; 4.3]	2.9 [2.3; 3.4]	2.9 [2.4; 3.4]
My mind wanders during lectures or presentations	3.6 [3.3; 4.0]	4.0 [3.4; 4.6]	3.4 [2.7; 4.1]	3.4 [2.9; 4.0]
I have difficulty maintaining focus on simple or repetitive tasks	2.7 [2.4; 3.0]	3.3 [2.7; 3.9]	2.4 [1.8; 3.0]	2.4 [2.0; 2.8]



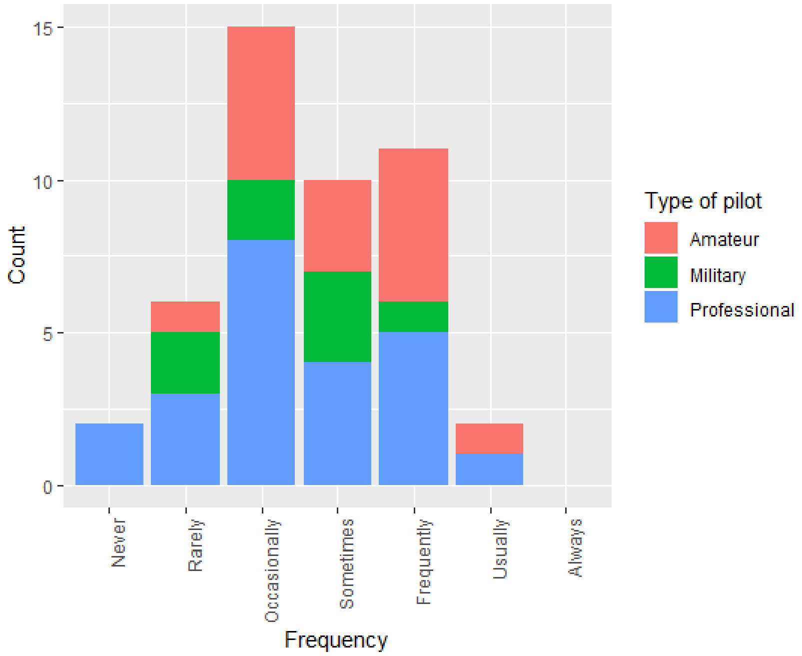
Pilots' answers to "While reading, I realize I haven't been thinking about the text and must read it again"



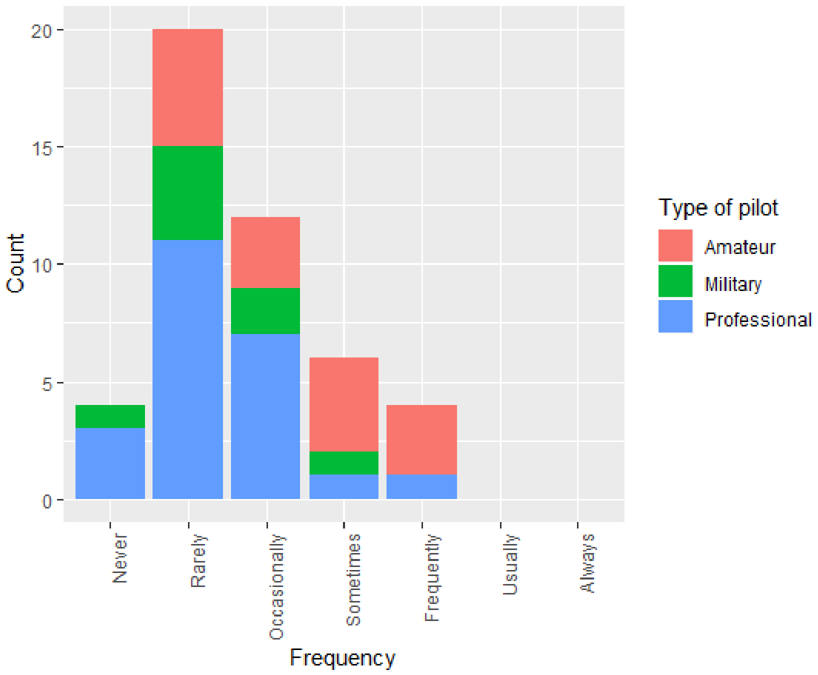
Pilots' answer to "I find myself listening with one ear, thinking about something else at the same time"



Pilots' answers to "I do things without paying full attention"



Pilots' answers to "My mind wanders during lectures or presentations"

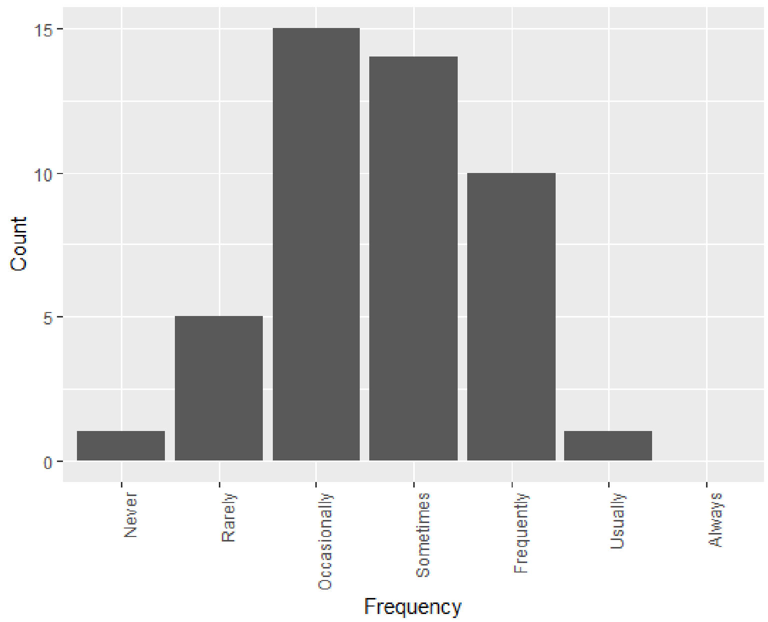


Pilots' answers to "I have difficulty maintaining focus on simple or repetitive tasks"

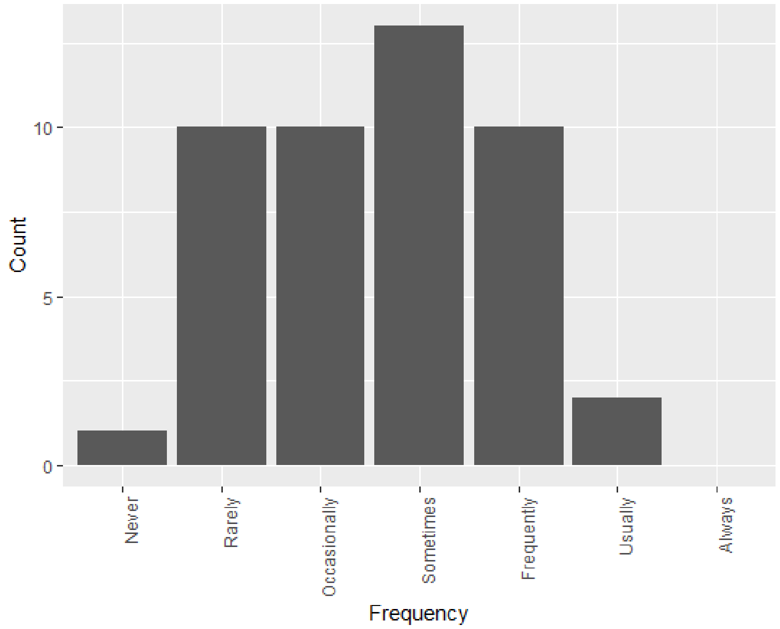
H. Non-pilots' answers for the "mind wandering questionnaire"

All confidence intervals were computed using bootstrap with 10000 iterations.

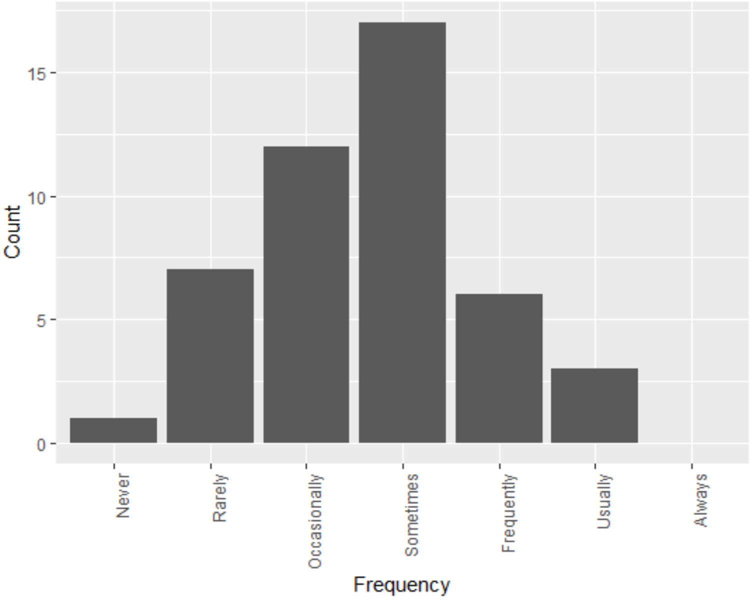
Items	Mean and 95% confidence interval
All part 5 items	3.6 [3.3; 3.8]
While reading, I realize I haven't been thinking about the text and must therefore read it again	3.7 [3.3; 4.0]
I find myself listening with one ear, thinking about something else at the same time	3.6 [3.3; 4.0]
I do things without paying full attention	3.6 [3.3; 4.0]
My mind wanders during lectures or presentations	3.8 [3.4; 4.2]
I have difficulty maintaining focus on simple or repetitive tasks	3.1 [2.8; 3.5]



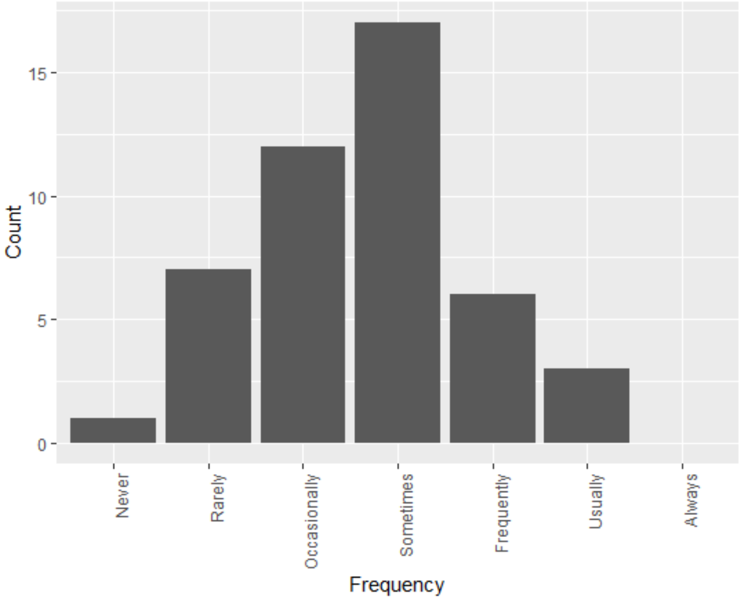
Non-pilots' answers to "While reading, I realize I haven't been thinking about the text and must read it again"



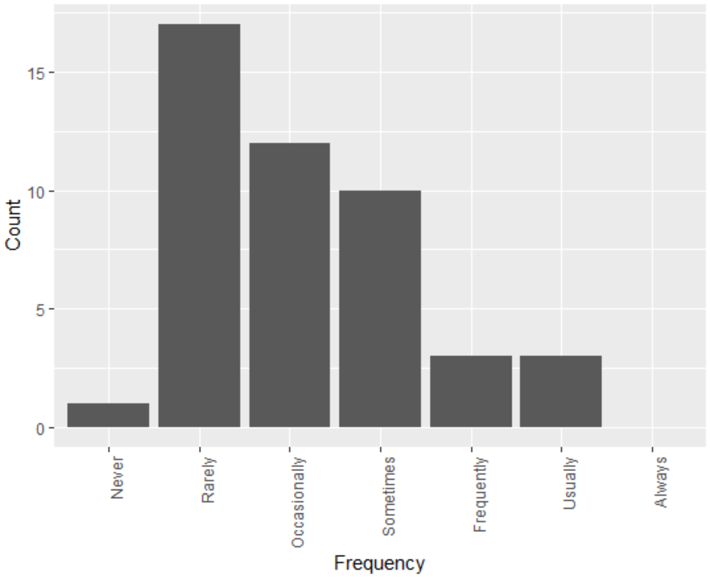
Non-pilots' answers to "I find myself listening with one ear, thinking about something else at the same time"



Non-pilots' answers to "I do things without paying full attention"

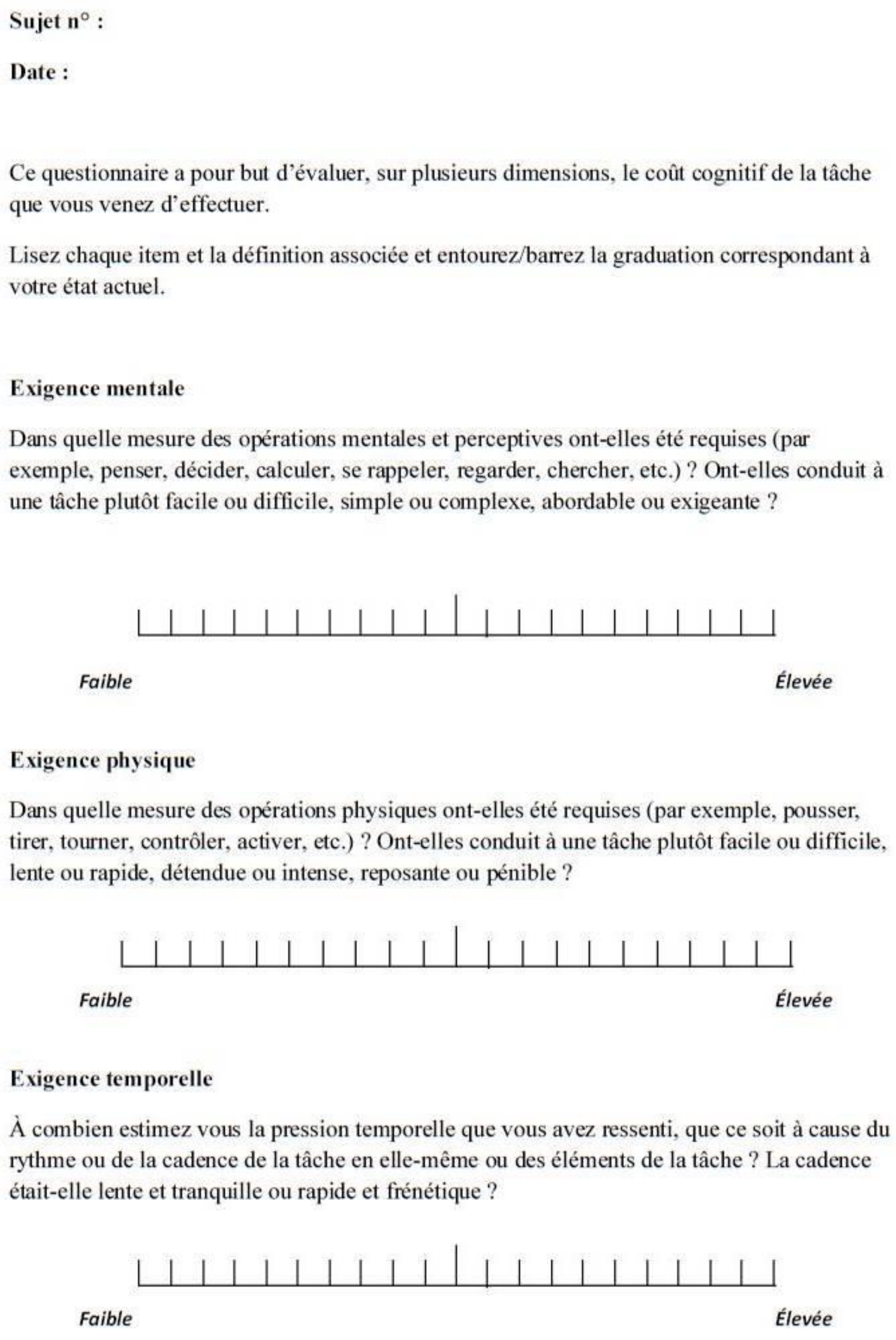


Non-pilots' answers to "My mind wanders during lectures or presentations"



Non-pilots' answers to "I have difficulty maintaining focus on simple or repetitive tasks"

I. NASA Task-Load Index in French



Sujet n° :

Date :

Ce questionnaire a pour but d'évaluer, sur plusieurs dimensions, le coût cognitif de la tâche que vous venez d'effectuer.

Lisez chaque item et la définition associée et entourez/barrez la graduation correspondant à votre état actuel.

Exigence mentale

Dans quelle mesure des opérations mentales et perceptives ont-elles été requises (par exemple, penser, décider, calculer, se rappeler, regarder, chercher, etc.) ? Ont-elles conduit à une tâche plutôt facile ou difficile, simple ou complexe, abordable ou exigeante ?

Faible *Élevée*

Exigence physique

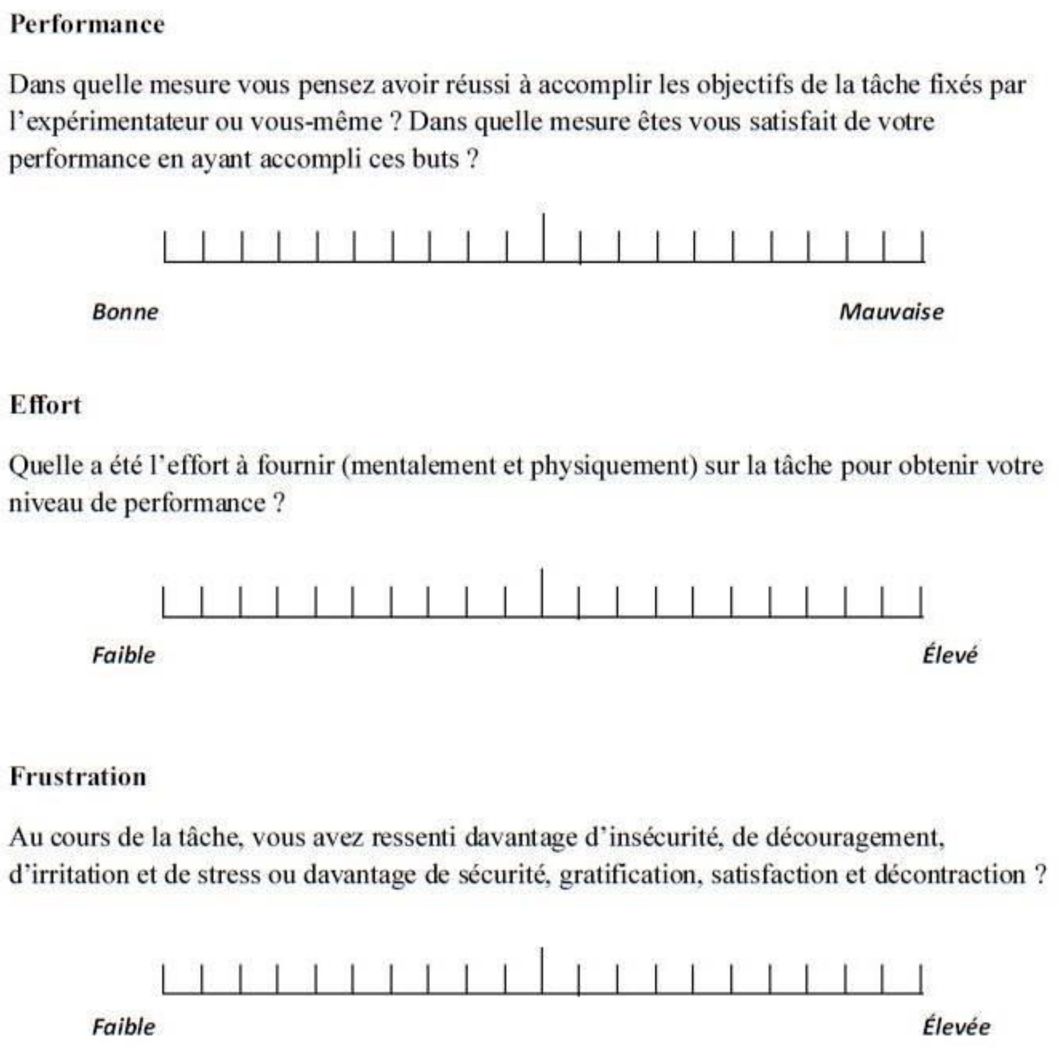
Dans quelle mesure des opérations physiques ont-elles été requises (par exemple, pousser, tirer, tourner, contrôler, activer, etc.) ? Ont-elles conduit à une tâche plutôt facile ou difficile, lente ou rapide, détendue ou intense, reposante ou pénible ?

Faible *Élevée*

Exigence temporelle

À combien estimez vous la pression temporelle que vous avez ressenti, que ce soit à cause du rythme ou de la cadence de la tâche en elle-même ou des éléments de la tâche ? La cadence était-elle lente et tranquille ou rapide et frénétique ?

Faible *Élevée*



Performance

Dans quelle mesure vous pensez avoir réussi à accomplir les objectifs de la tâche fixés par l'expérimentateur ou vous-même ? Dans quelle mesure êtes vous satisfait de votre performance en ayant accompli ces buts ?

Bonne *Mauvaise*

Effort

Quelle a été l'effort à fournir (mentalement et physiquement) sur la tâche pour obtenir votre niveau de performance ?

Faible *Élevé*

Frustration

Au cours de la tâche, vous avez ressenti davantage d'insécurité, de découragement, d'irritation et de stress ou davantage de sécurité, gratification, satisfaction et décontraction ?

Faible *Élevée*