



HAL
open science

La charge mentale lors de tâches cognitives

Lina-Estelle Louis

► **To cite this version:**

Lina-Estelle Louis. La charge mentale lors de tâches cognitives. Psychologie. Nantes Université, 2023. Français. NNT : 2023NANU2041 . tel-04637024

HAL Id: tel-04637024

<https://theses.hal.science/tel-04637024>

Submitted on 5 Jul 2024

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

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

THESE DE DOCTORAT

NANTES UNIVERSITÉ

ECOLE DOCTORALE N° 603

Education, Cognition, Langages, Interactions, Santé

Spécialité : *Psychologie cognitive*

Par **Lina-Estelle LOUIS**

La charge mentale lors de tâches cognitives

Thèse présentée et soutenue à Centrale Nantes, le 13 juin 2023

Unité de recherche : LS2N, UMR CNRS 6004

Rapporteurs avant soutenance :

Catherine GABAUDE Directrice de recherche, Université Gustave Eiffel
Raphaëlle ROY Enseignante-Chercheuse - HDR, ISAE SUPAERO

Composition du Jury :

Président :	Jordan NAVARRO	Professeur des Universités HDR, Lyon 2
Rapporteuses :	Catherine GABAUDE	Directrice de recherche, Université Gustave Eiffel
	Raphaëlle ROY	Enseignante-Chercheuse - HDR, ISAE SUPAERO

Directrice de thèse :	Isabelle MILLEVILLE-PENNEL	Chargée de recherche CNRS - HDR, Université de Nantes
Co-directeur de thèse :	Saïd MOUSSAOUI	Professeur HDR, Central Nantes

Invités

Sébastien RAVOUX Leader Innovation, Onepoint
Vincent ROUALDES Neurochirurgien, CHU de Nantes

Titre

La charge mentale lors de tâches cognitives.

Mots-clés

Charge mentale, NASA-TLX, Workload Profile (WP), Tâches cognitives, Performances, Modèles de classification, Design ludique, Ressenti ludique, Multiplicité de fonctions cognitives sollicitées, EEG.

Résumé

La charge mentale est utilisée comme référence pour évaluer le coût mental d'une activité. Un des défis liés à l'expérience utilisateur est de pouvoir prédire en amont d'une tâche, la charge mentale de l'individu et de la surveiller en cours d'activité. Pour ce faire, la thèse avait pour objectif de répondre à des questionnements autant théoriques que pratiques. Sur le plan théorique, une première étape visait à répertorier et à organiser les divers modèles conceptuels de la charge mentale. Pour proposer, en second lieu, un modèle combinant ces différents cadres théoriques et une nouvelle définition de la charge mentale.

Sur le plan pratique, en nous basant sur le modèle conceptuel proposé lors de la phase théorique, nous avons approfondi la compréhension du lien entre la charge mentale subjective et la performance, mais également le lien entre l'engagement (à travers le ressenti ludique), une variable interne à l'individu, et la charge mentale. De plus, nous avons accru les connaissances sur les facteurs intrinsèques (complexité de la tâche et multiplicité de fonctions cognitives sollicitées) et extrinsèques (design ludique) aux exigences de la tâche qui influencent la charge mentale. Enfin, nous avons identifié une nouvelle tâche cognitive pour laquelle nous avons pu établir un modèle de classification supervisée de la charge mentale subjective, proposant ainsi un début de réponse au besoin de prédiction de la charge mentale. Un modèle dont nous avons testé la validité inter-tâches (pour un même contexte de passation) et la validité inter-contextes (pour une même tâche donnée).

Cette thèse propose donc un regard enrichi et novateur sur la charge mentale, alliant travail théorique et expérimental.

Title

Mental workload during cognitive tasks.

Keywords

Mental workload, NASA-TLX, Workload Profile (WP), Cognitive tasks, Performances, Classification models, Fun design, Playful feeling, Multiplicity of cognitive functions solicited, EEG.

Abstract

Mental workload (MWL) is a concept used to assess the mental cost of activities. Nowadays, challenges related to user experience involve predicting an individual's MWL before a task and monitoring it during the activity. To achieve this, the thesis addresses both theoretical and practical questions. On a theoretical level, the first step was to compile and organize various conceptual models of MWL. The second step was to propose a model that combines these different theoretical frameworks and offers a new, more general, and operational definition of MWL.

On a practical level, based on the conceptual model proposed in the theoretical phase, we deepened the understanding of the link between subjective MWL and performance. We also explored the relationship between engagement (through the playful feeling), an internal variable for the individual, and MWL. Moreover, we increased the knowledge of the intrinsic (complexity of the task and multiplicity of cognitive functions solicited) and extrinsic (fun design) factors to the task requirements that influence MWL. Finally, we identified a new cognitive task for which we were able to establish a supervised classification model of subjective MWL. We proposed this as an initial response to the need for MWL prediction. A model for which we tested the inter-task validity (for the same context) and the inter-context validity (for the same task).

Therefore, this thesis offers an enriched and innovative view on MWL that combines theoretical and experimental approaches.

Remerciements

Lorsque je compris ce qu'était la recherche, réaliser une thèse était devenu un rêve.

En premier lieu, mes remerciements s'adressent aux membres de mon jury, mes deux rapporteuses, Catherine Gabaude et Raphaëlle Roy, pour l'intérêt témoigné envers ce travail de doctorat et le temps consacré à l'expertise de ce manuscrit. Je remercie également Jordan Navarro d'en être l'examineur. Par ailleurs, j'adresse mes sincères remerciements à mes encadrants de thèse. Mon co-directeur, Saïd Moussaoui, pour ton humilité. Ma directrice principale, Isabelle Milleville-Pennel, pour ta fidélité et ta bienveillance. Entre les lignes professionnelles, j'ai commencé à connaître une personne pleine d'humour et de respect. J'espère que je t'ai apporté ne serait-ce que le reflet d'un des innombrables aspects positifs que tu as eu dans mon travail de thèse. Ainsi, mes plus grandes espérances sont que tu continues à t'épanouir dans la recherche et que tu puisses transmettre tes qualités à d'autres chanceux.

De plus, je remercie l'ANRT et l'entreprise onepoint qui ont financé ce travail de recherche, plus particulièrement Sébastien Ravoux pour sa confiance. Plus largement, je remercie les membres de la R&D dont Sébastien Maire, Alexandra Delmas et Thomas Le Jan, pour la qualité de leurs rendus. J'adresse également mes remerciements à Vincent Roualdes et Aurélien Van Langenhove, pour leur expertise et remercie les personnels de l'administration et de l'entretien des structures que j'ai fréquentées. Ils effectuent un travail de l'ombre qui par leur humilité nous éclaire sans que l'on s'en aperçoive. Enfin, je salue mes collègues de bureau pour les moments de poésie du quotidien, ainsi que mes camarades à quatre pattes (Cookie, Roxy, Nifty et Emsy) qui étaient d'attendrissants collègues de télétravail.

Pour finir, je remercie ma mère pour son amour inconditionnel, ainsi que mes proches qui m'ont permis d'avoir des instants d'insouciance et de relâchement, indispensables en parallèle d'un travail de thèse. J'aimerai également saluer les situations qui ont tenté de ralentir, voire de mettre en péril ce projet. Je vous remercie car, j'ai redécouvert ma ténacité, ma confiance en moi-même et à l'égard des personnes ressources de mon

entourage. Cette thèse, un travail qui passionne et qui consume, mais surtout éclaire sur les véritables piliers de la vie. Finalement, c'est vrai, que les amis ne sont pas plus du nombre des doigts d'une seule main. Mais ceux qui le sont te remplissent de joie, d'amour et de tranquillité.

En dernier lieu, j'aimerais me remercier d'avoir réalisé ce rêve qui, malgré la confrontation au réel, est resté un rêve.

À ma mère et amie Esther.

Table des matières

Mots-clés	1
Résumé	1
Keywords	2
Abstract	2
Remerciements	3
Table des matières	5
Introduction générale	8
PARTIE THÉORIQUE	13
Chapitre 1 : Modèles et définitions de la charge mentale	14
1.1 Une charge mentale subie par l'individu	15
1.1.1 Approches centrées sur les exigences de la tâche.....	16
1.1.2 Approches centrées sur les performances.....	18
1.1.3 Approches centrées sur les capacités limitées de l'individu	20
1.2 Individu acteur de sa charge mentale	24
1.2.1 Approches centrées sur l'effort mental fourni par l'individu	25
1.2.2 Approches incluant les stratégies de l'individu.....	28
1.2.3 Proposition d'un modèle et d'une définition de la charge mentale	33
Chapitre 2 : Mesures de la charge mentale	42
2.1 Mesures subjectives	44
2.1.1 Échelles unidimensionnelles	44
2.1.2 Échelles multidimensionnelles.....	46
2.1.2.1 NASA-TLX.....	46
2.1.2.2 Workload Profile	47
2.1.2.3 SWAT.....	48
2.1.2.4 Comparaisons entre les mesures subjectives.....	49
2.1.3 Conclusion sur les mesures subjectives	50
2.2 Performances	52
2.2.1 Évaluation directe de la performance.....	52
2.2.2 Évaluation indirecte de la performance.....	54
2.2.3 Conclusion sur les mesures de performances.....	57
2.3 Mesures physiologiques	57
2.3.1 Les mesures périphériques.....	58
2.3.2 Les mesures centrales.....	61
2.3.3 Conclusion sur les mesures physiologiques.....	64
2.4 Conclusion sur l'ensemble des mesures	65
Chapitre 3 : Applications, manques et problématique	67
3.1 Applications	68
3.2 Manques	70
3.3 Problématique	74

PARTIE EXPÉRIMENTALE.....	80
Chapitre 4 : Modélisation de la charge mentale lors de tâches cognitives .	81
4.1 Hypothèses de l'étude et choix des tâches	82
4.1.1 Hypothèses opérationnelles de l'étude 1	82
4.1.2 Choix des tâches cognitives.....	82
4.2 Matériel et méthodes	87
4.2.1 Participants.....	87
4.2.2 Tâches et procédures	87
4.2.2.1 Tâche de N-back (NBT).....	90
4.2.2.2 Test de Corsi	91
4.2.2.3 Test de Go/No-Go (GNG)	93
4.2.2.4 Wisconsin Card Sorting Test (WCST).....	95
4.2.2.5 Double tâche (DT).....	97
4.2.3 Questionnaires	100
4.2.4 Procédure expérimentale.....	102
4.3 Résultats de l'étude 1	103
4.3.1 Correspondance entre les niveaux de charge mentale subjective et les niveaux de complexité.....	106
4.3.2 Clustering de la charge mentale subjective et vérification de la corrélation entre les classes subjectives et le niveau de complexité.....	115
4.3.3 Classification supervisée de la charge mentale subjective basée sur les niveaux de complexité et les performances.....	127
4.3.4 Classification de la charge mentale subjective basée sur les performances uniquement	134
4.3.5 Transférabilité d'un modèle de classification inter-tâches	136
4.3.6 Lien entre le ressenti ludique, la performance et la charge mentale subjective	138
4.4 Discussion de l'étude 1	142
Chapitre 5 : Transférabilité inter-contextes et impact sur la charge mentale de variables intrinsèques et extrinsèques.....	154
5.1 Hypothèses de l'étude et choix des tâches	157
5.1.1 Hypothèses opérationnelles de l'étude 2	157
5.1.2 Choix des tâches cognitives.....	158
5.2 Matériel et méthodes	160
5.2.1 Participants.....	160
5.2.2 Tâches et procédures	160
5.2.2.1 Tâches cognitives classiques	160
5.2.2.2 Tâches cognitives ludiques.....	162
5.2.2.3 Tâches cognitives combinées	167
5.2.3 Mesures subjectives et physiologiques	169
5.2.4 Procédure expérimentale.....	170
5.3 Résultats de l'étude 2	173
5.3.1 Transférabilité d'un modèle de classification inter-contextes.....	174
5.3.2 Impact du design ludique	178
5.3.2.1 Impact d'un design ludique avec des items unicolores	179
5.3.2.2 Impact d'un design ludique avec des items multicolores.....	183
5.3.3 Impact de la multiplicité de fonctions cognitives	186
5.4 Discussion de l'étude 2	202
Discussion générale.....	214
1. Étudier la charge mentale à travers des tâches cognitives	215

2. Discussions sur les modèles théoriques.....	217
3. Discussions sur les mesures de la charge mentale.....	225
4. Principale limite de nos études expérimentales	228
5. Applications envisagées	229
6. Perspectives	230
7. Conclusion générale.....	234
<i>Contributions.....</i>	<i>236</i>
<i>Bibliographie.....</i>	<i>237</i>
<i>Annexes.....</i>	<i>251</i>

Introduction générale

Le cadre général et les différentes parties prenantes

Cette thèse s'inscrit dans le cadre des recherches sur la charge mentale. Elle vise à améliorer les connaissances sur ce concept en se basant sur des considérations théoriques et des études expérimentales.

Ce travail est lié à une Convention Industrielle de Formation par la REcherche (CIFRE) financée par l'entreprise onepoint qui est depuis 2019 dans un Groupement d'Intérêt Scientifique (GIS) nommé ICONIC (Interface Cerveau-Ordinateur Neurochirurgie Ingénierie Clinique). Ce groupement associe trois parties prenantes, l'entreprise onepoint, le CHU de Nantes et l'École Centrale de Nantes (ECN). *Onepoint* est une entreprise de conseil dans le numérique et Sébastien Ravoux supervise la thèse côté entreprise. *Le Centre Hospitalier Universitaire (CHU) de Nantes*, est impliqué spécifiquement à travers l'équipe du GIS ICONIC composée du Dr. Vincent Roualdes et de l'ingénieur Aurélien Van Langhenhoven. De plus, la doctorante est associée au *laboratoire du LS2N* (Laboratoire des Sciences du Numérique de Nantes) composé de plusieurs équipes dont *l'équipe PACCE* (Perception, Action, Cognition pour la Conception et l'Ergonomie) où la doctorante et la Directrice de thèse Isabelle Milleville-Pennel sont membres. Le co-directeur Saïd Moussaoui est, quant à lui, membre du LS2N et intégré à *l'ECN*.

Le point de départ de la thèse a été l'essai clinique *GHOST* porté par l'équipe du GIS ICONIC du CHU de Nantes. Dans ce cadre clinique, la thèse a débuté dans la perspective de mesurer en continu la charge mentale.

L'interaction avec les tâches cognitives peut se faire par Imagerie Mentale Motrice (IMM) qui consiste à s'imaginer en train d'effectuer une action, sans la faire réellement (Cf. **Figure 1**). Cette IMM peut être guidée par une Interface Cerveau-Ordinateur (ICO) qui permet un échange d'informations entre le cerveau d'un individu et une machine extérieure à celui-ci en utilisant comme entrée l'activité cérébrale de la personne (Wolpaw et al, 2002). Une ICO peut être active. Dans ce cas, il est nécessaire que l'utilisateur se mette volontairement dans un état mental particulier, afin d'envoyer une commande à l'application (Lotte, 2012). Une ICO peut également être passive. Dans ce cas, l'interface va suivre et contrôler en continu l'état mental de l'utilisateur, pour par

exemple, mesurer son niveau de charge mentale, afin de changer dynamiquement et implicitement l'application, le tout sans aucune intervention volontaire de l'utilisateur (Lotte, 2012 ; Roy et al., 2013). Dans le cadre de départ de la thèse lié au projet *GHOST*, l'ICO est à la fois active (le patient réalise volontairement de l'IMM, afin d'envoyer une commande à l'application) et passive (mesure en continu la charge mentale du patient à travers l'ÉlectroEncéphaloGraphie de l'ICO).

En effet, une ICO peut être basée sur de l'ÉlectroEncéphaloGraphie (EEG) par exemple, qui permet de mesurer l'activité cérébrale au niveau de la zone motrice pour permettre un retour visuel de l'IMM. Cette possibilité, a été envisagée entre autres par l'équipe du GIS ICONIC du CHU de Nantes. Afin, d'améliorer les rééducations des zones cérébrales motrices. Cependant, ce type de dispositif génère deux sources de charge mentale :

- La charge mentale due à l'IMM
- La charge mentale due à la réalisation de la tâche cognitive (Cf. **Figure 1**).

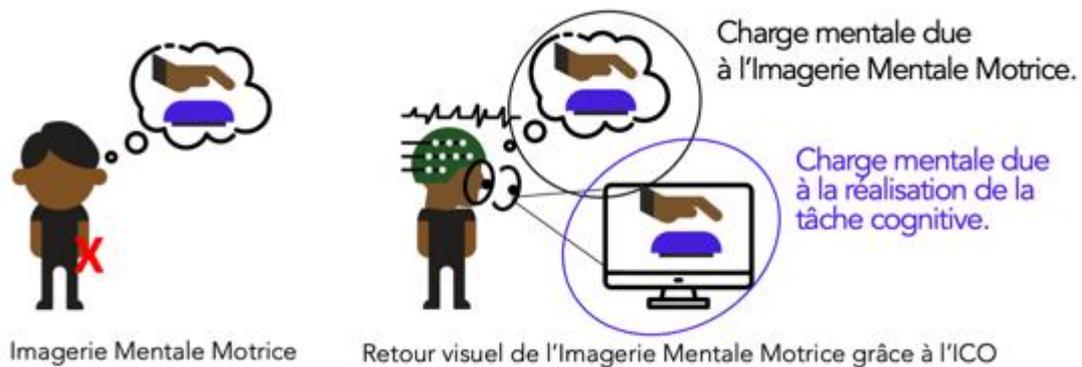


Figure 1 : Schéma explicatif de l'Imagerie Mentale Motrice (IMM- à gauche) et du retour visuel de l'IMM grâce à une Interface Cerveau-Ordinateur (à droite).

Dans le cadre de cette thèse, je me suis intéressée à la seconde source de charge mentale, car nous avons souhaité adapter le niveau de complexité d'une tâche cognitive à la charge mentale due à la réalisation de la tâche cognitive, afin que l'individu ne soit ni en sous-charge mentale ni en surcharge mentale lors de la réalisation de la tâche cognitive, mais dans une zone de confort mentale.

Ainsi, mesurer la charge mentale en continu permettrait d'adapter le niveau de complexité de la rééducation médicale au niveau de charge mentale souhaité, afin

d'améliorer la prise en charge médicale, notamment par la maximisation de l'engagement du patient.

La problématique de recherche

Adapter le niveau de complexité en temps réel grâce à un indicateur physiologique (tel que l'EEG) est nommé boucle biocybernétique (Ewing et al., 2016). Bien que prometteuses, ces techniques ont un certain nombre de limites, telles que le coût (technique, expertise, intrusivité) que la mesure des performances à une tâche n'a pas. Par conséquent, un des objectifs principaux de ce travail de thèse est de proposer une meilleure compréhension du lien entre la charge mentale subjective et celle issue des performances pour pouvoir utiliser ces dernières dans la mesure de la charge mentale en cours d'activité. Ce travail applicatif se situera dans un premier temps dans un cadrage théorique de la charge mentale, sur lequel nous nous appuierons pour sélectionner dans un second temps les mesures subjectives et objectives les plus adéquates pour modéliser la charge mentale.

Plusieurs auteurs (Berka et al., 2007 ; Radüntz, 2017 ; Guan et al., 2021) recommandent d'étudier la charge mentale à travers des tâches cognitives, puisque les situations réelles peuvent être décomposées en plusieurs fonctions cognitives. Ainsi, nous proposons d'étudier la charge mentale à travers plusieurs tâches sollicitant des fonctions cognitives distinctes.

La démarche et la structure du manuscrit de thèse

Tout d'abord, ce travail de thèse s'articulera en deux grandes parties : une réflexion théorique (Chapitres 1 et 2) et une partie pratique (Chapitres 4 et 5 ; Cf. **Figure 2**).

Dans le détail, à la suite d'une revue de la littérature articulée en fonction de la passivité ou non de l'opérateur vis-à-vis de sa charge mentale, notre démarche consistera en premier lieu à re-questionner les différents modèles conceptuels de la charge mentale. Pour proposer, en second lieu, un modèle combinant ces différents cadres théoriques et une nouvelle définition de la charge mentale (**Chapitre 1**). Puis, nous recenserons les mesures de la charge mentale en pointant leurs avantages et leurs limites (**Chapitre 2**). Par la suite, nous identifierons de nouvelles possibilités d'utilisation du concept de la charge mentale lors d'applications émergentes, en nous basant sur les manques identifiés (**Chapitre 3**). Ensuite, nous aborderons la partie pratique où nous proposerons d'approfondir la compréhension du lien entre la charge mentale subjective et la performance lors d'une première étude, mais également le lien entre l'engagement (à travers le ressenti ludique), une variable interne à l'individu, et la charge mentale. Dans cette même expérimentation, nous étudierons l'impact sur la charge mentale d'une variable intrinsèque aux exigences de l'activité, la complexité de la tâche. De plus, nous proposerons des modèles de classification, pour déduire les classes de charge mentale subjective, lors de la réalisation de tâches cognitives et nous étudierons la transférabilité de modèles de classification inter-tâches (**Chapitre 4**). Lors d'une seconde étude, nous proposerons de tester la transférabilité de modèles de classification (pour déduire les classes de charge mentale subjectives) inter-contextes pour une même tâche cognitive. Au cours de cette étude, nous étudierons l'impact sur la charge mentale d'une autre variable intrinsèque aux exigences de la tâche, la multiplicité des fonctions cognitives sollicitées. Enfin, nous étudierons des variables extrinsèques aux exigences de la tâche, des designs ludiques, pouvant exercer une influence sur la charge mentale (**Chapitre 5**).

Ce mémoire de thèse est donc organisé en cinq chapitres et chacun d'entre eux est associé à un code couleur (Cf. **Figure 2**) qui nous guidera tout au long du manuscrit.



Figure 2 : Plan général du manuscrit articulé en deux grandes parties

Pour finir, au niveau de la **Discussion générale**, nous réaliserons un bilan de nos recherches en mettant en lumière les contributions apportées au domaine de la charge mentale tout en pointant les limites de notre travail. Enfin, nous proposerons un ensemble de recommandations et de perspectives pour les futures recherches.

Mais avant toute chose, revenons à la base et intéressons-nous aux modèles explicatifs de la charge mentale, qui sont présentés ci-après, dans le premier chapitre de la partie théorique.

PARTIE THÉORIQUE

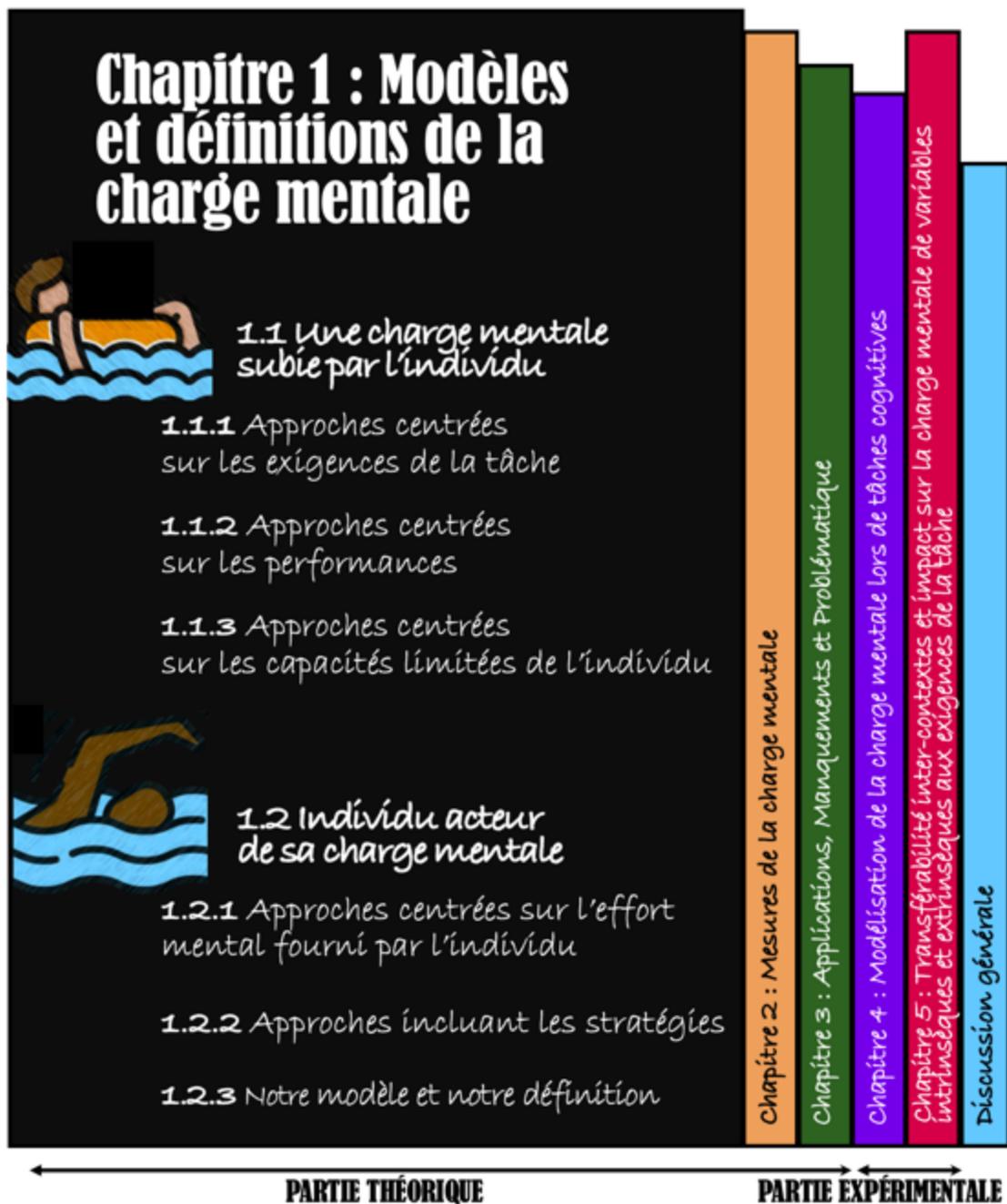


Figure 1 : Plan général du manuscrit et détail des sous-parties du chapitre 1.

Le premier chapitre du manuscrit introduit les modèles à la base de la conceptualisation de la charge mentale (Cf. **Figure 1**). Il en existe une multitude. Ce qui explique qu'il n'y a, à l'heure actuelle, aucune définition consensuelle de ce concept. Par conséquent, à la suite d'une revue de la littérature, nous proposerons un modèle conceptuel

combinant différents modèles théoriques, en plus d'une définition de la charge mentale qui nous guidera tout au long du manuscrit.

Avant toute chose, la charge mentale est un concept introduit dans les années 40 (Radüntz, 2017 ; Hancock et al., 2021), notamment par Borneman (1942), afin d'optimiser les Interactions Humain-Machines (IHM). Dès lors, ce concept a été repris et enrichi par de nombreux auteurs qui ont vu l'utilité de sa mesure dans diverses situations. En effet, la charge mentale permet de quantifier le coût mental lors de la réalisation d'une activité, afin d'anticiper les performances de l'individu et du système avec lequel il interagit (Cain, 2007). Cependant, en l'absence d'un cadre conceptuel consensuel, la charge mentale s'est vu offrir une multitude de modèles et de définitions. Dans ce chapitre, nous les présentons en les distinguant par la passivité ou la non-passivité de l'opérateur à l'égard de sa charge mentale. En effet, au regard de la variété des approches, cette distinction est apparue comme la plus pertinente à la suite d'une revue de la littérature sur la charge mentale (Cain, 2007 ; Martin, 2013 ; Young et al., 2015 ; Kostenko, 2017 ; Van Acker, 2018 ; Mallat, 2019 ; Hancock et al., 2021 ; Le Gonidec, 2022 ; Longo et al., 2022).

1.1 Une charge mentale subie par l'individu

Dans un premier temps, en se référant à l'analogie de la charge physique, les modèles de la charge mentale se sont focalisés sur la relation entre les exigences de la tâche et la



charge mentale. En parallèle, les modèles se sont également intéressés au lien entre ce concept et l'impact sur les performances de l'opérateur. Par la suite, la relation entre les capacités limitées du traitement de l'information et la charge mentale a été mise en exergue. Pour ces trois cas, l'individu est continuellement passif à l'égard de sa charge mentale. Autrement dit, il expérimente la charge mentale sans aucune marge de manœuvre pour y faire face ou pour la réguler. Dans cette section, nous présentons ces modèles à travers trois approches : l'approche centrée sur les exigences de la tâche, l'approche mettant au centre les performances de l'opérateur et l'approche centrée sur les capacités limitées de l'individu.

1.1.1 Approches centrées sur les exigences de la tâche

Premièrement, la notion des exigences de la tâche est centrale dans la conceptualisation et la définition de la charge mentale. En effet, parmi plusieurs définitions, Le Gonidec (2022) a constaté la présence de la notion d'exigences de la tâche dans plus de 60 % des définitions de ce concept. Parmi celles-ci, Hancock et Caird (1993) définissent la charge mentale comme un concept multidimensionnel qui est largement dicté par les caractéristiques des exigences de la tâche. Plus tard, Colombi et al. (2012) la définissent comme l'expression des exigences imposées à un opérateur.

En fonction des modèles conceptuels de la charge mentale, les exigences de la tâche sont également nommées *stress/contraintes* (Sperandio, 1988 ; Schlegel, 1993), *charge imposée* (Hart et Staveland, 1988), *charge prescrite* (Falzon et Sauvagnac, 2004), *ressources exigées* (Wickens et al., 2015), ou encore *inputs/charge d'entrée* (Longo et al., 2022). Elles correspondent à différents aspects de la tâche, tels que les objectifs, la complexité de la tâche, les exigences de performances, ou encore les exigences temporelles.

Modèle temps/charge mentale

Concernant ce dernier aspect, les exigences temporelles, le modèle temps/charge mentale (Hancock et Chignell, 1988 ; Hancock et Caird, 1993) postule que la charge mentale augmente lorsque la pression temporelle augmente et/ou lorsque l'individu progresse vers la fin de la tâche.

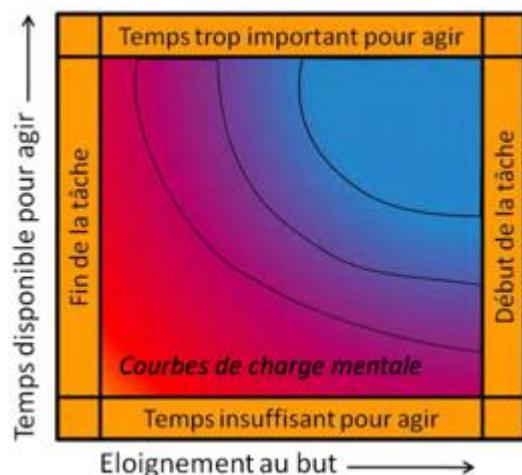


Figure 2 : Modèle illustrant les zones de charge mentale selon les exigences temporelles, proposé par Hancock et Chignell (1988), puis modifié par Hancock et Caird (1993) ; traduit en français par Mallat (2019).

Ainsi, lorsqu'un individu est au début de la tâche et qu'il dispose d'un temps considérable pour agir, la charge mentale est faible (zone bleue ; **Figure 2**). À l'inverse, lorsque l'individu est en fin de tâche et que le temps disponible est insuffisant, la charge mentale est élevée (zone rouge ; **Figure 2**). Enfin, pour l'opérateur, il existerait une zone

de confort de la charge mentale (zone en violet ; **Figure 2**) où il disposerait de possibilités d'action optimales.

Modèle contrainte/astreinte

Toujours au cours du XX^{ème} siècle, le modèle contrainte/astreinte (Karasek, 1979 ; Sperandio, 1988 ; Schlegel, 1993 ; Raufaste et al., 2004) émerge et met au centre les exigences mentales dans leur globalité. Il n'est donc pas limité qu'aux exigences temporelles comme le modèle temps/charge mentale (Hancock et Chignell, 1988 ; Hancock et Caird, 1993).

En effet, Schlegel (1993) considère que la charge mentale comprend deux composantes nommées en anglais : le *stress* (exigences de la tâche) et le *strain* (conséquences du *stress* sur l'individu). Une approche partagée par Sperandio (1988), qui quelques années plus tôt, avait décrit cette relation dichotomique en ces termes français : la contrainte (pour le *stress*) et l'astreinte (pour le *strain*). Selon Falzon et Sauvagnac (2004), la contrainte peut varier à tout moment, du fait de la fluctuation des exigences instantanées au cours d'une activité donnée. La théorie de l'activité (Vygotsky, 1980) approfondit cette réflexion en précisant qu'il existe des facteurs dynamiques qui ne peuvent pas être anticipés avant l'exécution de la tâche et qu'ils sont stochastiques, autrement dit, aléatoires. Parmi ces facteurs, on peut citer les règles, la communauté dans un environnement de l'opérateur ou encore les incidents du système.

Cette approche centrée sur les exigences de la tâche, qui a vu le jour il y a plus de trente ans, est toujours d'actualité. La preuve en est avec la norme internationale sur la charge mentale (ISO 1075-1 : 2017) qui s'appuie sur le modèle de contrainte/astreinte.

Cependant, cette approche a une limite majeure, celle d'envisager la mesure de la contrainte d'une tâche, comme suffisante pour en déduire l'astreinte. Or, le postulat selon lequel il existerait une relation linéaire entre la contrainte et l'astreinte est questionnable. Ainsi, d'autres modèles de la charge mentale, où l'individu est toujours passif à l'égard de sa charge mentale, ont tenté de pallier cette limite, en prenant en compte les performances de l'individu.

1.1.2 Approches centrées sur les performances

Les performances permettent d'évaluer l'atteinte des objectifs associés aux exigences de la tâche (Delignières, 2004). Elles peuvent donc s'évaluer à travers des indicateurs de réussite de la tâche réalisée (Bayssié et Chaudron, 2002). Par ailleurs, les critères de performance peuvent être prescrits par la tâche et/ou être fixés par l'opérateur lui-même (Hancock et Caird, 1993).

Modèle performance/charge mentale en U inversé

Un des modèles de la charge mentale centré sur les performances est issu des travaux de Yerkes et Dodson (1908). Ces derniers formulent une loi lors de paradigmes de conditionnement où ils observent une diminution des performances lorsque les stimuli sont faibles ou à l'inverse intenses. D'après Klein (1982), la fonction de Hebb (1955) reprend la loi de Yerkes et Dodson (1908) à travers une courbe curvilinéaire (en U inversé) qui relie la performance et l'éveil physiologique. Selon Boyer (2016), l'éveil physiologique étant un concept large, la fonction de Hebb (1955) peut être appliquée à des constructions théoriques proches, comme celle de la charge mentale. Par conséquent, dans ce modèle, la charge mentale est liée à la performance à travers une courbe en U inversé (Cf. **Figure 3**). Il existe une zone de *confort* (en violet ; Cf. **Figure 3**) pour laquelle les niveaux de charge mentale sont acceptables et les performances sont optimales. À l'inverse, que ce soit lors de la surcharge mentale (en rouge) ou lors de la sous-charge mentale (en bleu), les performances se dégradent.

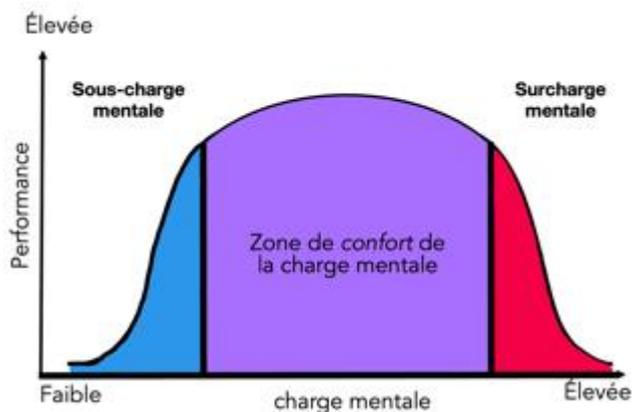


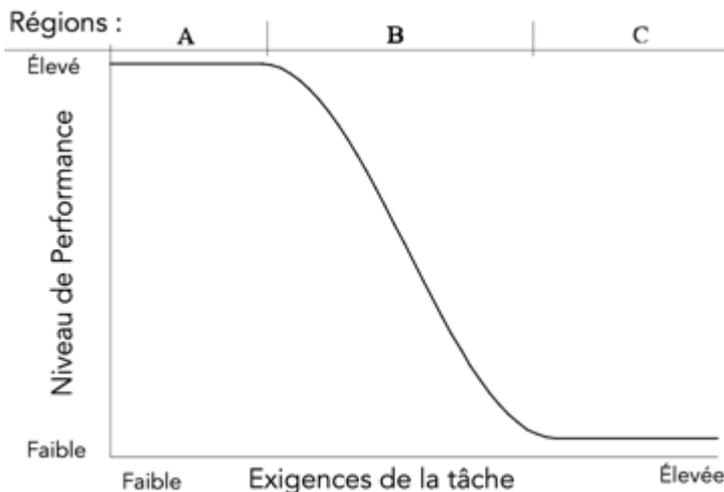
Figure 3 : Relation, entre la performance et le niveau de charge mentale, adaptée des travaux de Yerkes et Dodson (1908) et de la représentation en U inversé de la fonction de Hebb (1955) appliquée au concept de la charge mentale (Boyer, 2016).

C'est ainsi que Young et Stanton (2002) précisent que la sous-charge mentale est tout aussi délétère que la surcharge mentale, car les deux peuvent conduire à une dégradation des performances. Ce dernier point est une des contributions majeures de la fonction de

Hebb (1955) appliquée à la charge mentale qui est toujours valable dans les conceptions actuelles de la charge mentale.

Modèle de régions de Meister

Toujours au XX^{ème} siècle, un autre modèle de la charge mentale centré sur les performances a vu le jour. Il s'agit du modèle de *régions* de Meister (1976) qui met en lien deux dimensions dans la conceptualisation de la charge mentale : les performances et les exigences de la tâche. Trois régions A, B et C sont définies (Cf. **Figure 4**). La région A correspond à une charge mentale faible avec des performances élevées et au niveau de laquelle une augmentation des exigences de la tâche n'entraîne pas une diminution des performances. Pour la région B, le niveau de performance diminue avec l'augmentation des exigences de la tâche et l'augmentation de la charge mentale. Enfin, la région C correspond à des niveaux extrêmes de charge mentale qui diminuent les



performances jusqu'à un niveau minimum (un niveau plancher). Par ailleurs, la transition entre la région A et B est considérée comme la ligne rouge de la charge mentale (basculement vers la surcharge mentale) selon Rueb et al. (1992).

Figure 4 : Illustration du modèle de régions de Meister (1976) qui met en relation les exigences de la tâche, la performance et la charge mentale.

Le modèle de région de Meister (1976) a néanmoins une limite. Si les niveaux de surcharge mentale peuvent être situés dans la région C de la **Figure 4**, il n'est pas évident de concevoir le domaine de la sous-charge mentale. De plus, bien que ces deux modèles, la fonction de Hebb (1955) appliquée à la charge mentale et le modèle de Meister (1976), aient été repris par plusieurs auteurs du XXI^{ème} siècle (Andre, 2001 ; Cuvelier, 2012 ;

Mandrick, 2013), ces modèles se heurtent aux différences inter-individuelles, car l'effet de la contrainte (exigences de la tâche) n'est pas le même d'un individu à un autre. Ainsi, la charge mentale ne peut être décorrélée de facteurs individuels, notamment de la notion de capacités de traitement de l'information limitées et qui est sujette à des différences inter-individuelles. Ce constat constitue la base de l'origine de l'approche centrée sur les capacités limitées de l'opérateur.

1.1.3 Approches centrées sur les capacités limitées de l'individu

Les approches centrées sur les capacités limitées considèrent la charge mentale comme le rapport entre les exigences de la tâche et la capacité disponible de traitement de l'information pour réaliser la tâche. Par conséquent, ces approches considèrent toujours l'individu comme étant passif à l'égard de sa charge mentale.

Ces approches pour modéliser la charge mentale ont eu un impact sur la manière de la définir, car plus de la moitié des définitions de ce concept se base sur la notion de capacités limitées (Le Gonidec, 2022). La charge mentale est alors considérée comme un produit des ressources disponibles ou de capacités cognitives nécessaires pour répondre aux demandes de la tâche (Welford, 1978 ; Young et al., 2008 ; Brouwer et al., 2012 ; Di Stasi et al., 2013). Une distinction peut être réalisée entre les notions de *capacités* et/ou de *ressources*. Ces dernières sont en quantité limitée et sont déployées pour répondre à une tâche. Elles sont intimement liées au concept d'attention, étant donné son rôle dans le traitement de l'information (Hancock et al., 2021). En effet, l'attention est la capacité à sélectionner une ou plusieurs informations pertinentes au milieu d'un flux important d'informations (Boyer, 2016).

Les *capacités*, quant à elles, représentent le seuil maximal de *ressources* que l'individu peut engager, pour réaliser un traitement de l'information à un instant donné (De Waard, 1996 ; Chanquoy et al., 2007 ; Sperandio, 2019). La capacité de traitement de l'information est rattachée aux capacités de la Mémoire de Travail (MdT). Effectivement, la MdT est la partie de la mémoire, où est stockée pour une brève période (de quelques secondes à quelques minutes) un nombre limité d'informations (Baddeley et Hitch, 1974 ; Baddeley, 1986). Modélisée par Baddeley et Hitch (1974) ; Baddeley (1986), la MdT pallie une limite de la Mémoire à Court Terme (MCT) du modèle

d'Atkinson et Shiffrin (1968), qui ne permettait pas d'expliquer que l'on puisse manipuler des informations en plus de les stocker (Baddeley et Hitch, 1974 ; Baddeley, 1986). La MdT a donc des capacités limitées de l'ordre de sept plus ou moins deux éléments (nommés *chunks*) selon Miller (1956), voire de quatre *chunks* selon Cowan (2001).

La notion de *capacités* englobe donc celle liée aux *ressources*. Le concept de capacité de traitement limitée, repris par la suite par des modèles de la charge mentale, a été développé par de nombreuses théories, à commencer par les théories dites *des ressources*. Elles suggèrent qu'il existe un ou des réservoirs de ressources mentales dans lequel/lesquels il est possible de puiser pour aider l'organisme à répondre aux exigences de la tâche (Staal, 2004). Elles ajoutent que plus les exigences de la tâche sont élevées, plus les ressources limitées sont sollicitées (Hancock et al., 2021).

Modèles de la ressource unique

Un des premiers modèles à voir le jour est le modèle de la ressource unique de Moray (1967). Il postule que chaque individu a une capacité de traitement limitée dans la mesure où les activités mentales partagent la même ressource centrale qui est limitée. Dans le cadre de la conceptualisation de la charge mentale, le modèle de Wickens et al. (2015) reprend l'idée d'un système cognitif limité en ressources en le mettant en relation avec les exigences de la tâche et les performances. Ainsi, les ressources atteignent un seuil lorsque les exigences dépassent le stock de ressources (Cf. **Figure 5**). À gauche de la ligne rouge, les individus peuvent avoir une charge mentale dite « *capacité de réserve* » qui augmente sans impact sur les performances. Dès que la demande sera égale au stock de ressources, une nouvelle augmentation de la demande entraînera de nouvelles baisses de performance. Dans ce cas, l'individu est dans un état de surcharge mentale.

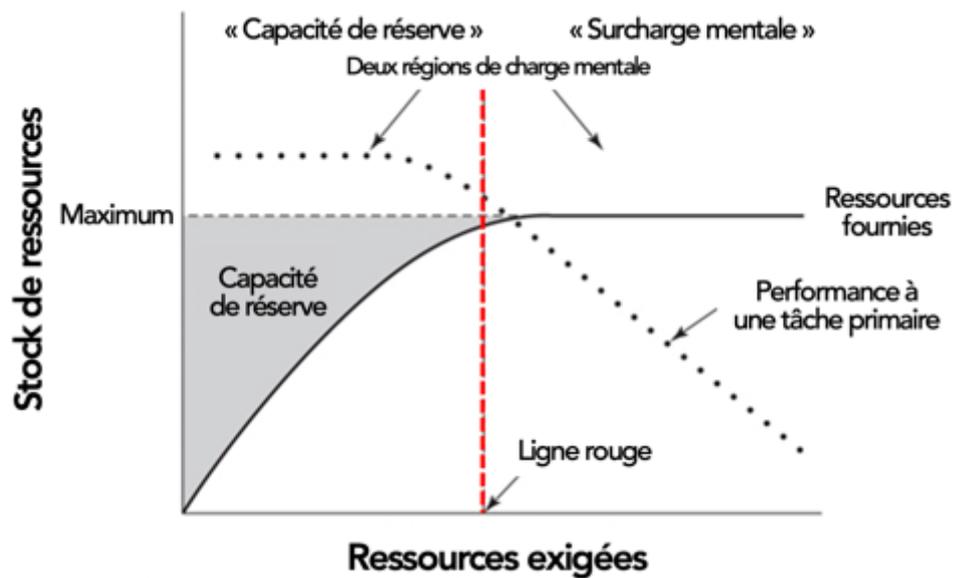


Figure 5 : Relation entre les ressources exigées, les ressources fournies, la charge mentale et les performances à la tâche primaire (traduction française dans le cadre de cette thèse du schéma de Wickens et al., 2015).

Modèles de la ressource unique malléable

Cependant, les modèles de Moray (1967) et de Wickens et al. (2015) ont une limite, celle de supposer que quel que soit le niveau d'exigences de la tâche (faibles ou élevées), la taille de la réserve des ressources est fixe (Young et Stanton, 2002). La théorie des ressources attentionnelles malléables (MART - *Malleable Attentional Resources Theory*) proposée par Young et Stanton (2002) pallie cet écueil. En effet, la MART postule que la capacité des ressources attentionnelles peut varier en réponse à des changements dans les exigences de la tâche (Cf. **Figure 6**). Ainsi, les ressources ne sont pas fixes, mais s'adaptent pour répondre aux exigences de la tâche (et pas uniquement lors des demandes élevées de la tâche), d'où la notion de *malléabilité*. Lorsque les exigences de la tâche sont faibles, comme dans le cas d'une situation monotone, les capacités des ressources attentionnelles s'accommodent en diminuant. Cette diminution est à l'origine des effets de la sous-charge mentale sur la performance (Nguma, 2020). Dans le cas contraire, lorsque les exigences de la tâche sont élevées, les capacités de ressources attentionnelles s'accommodent en augmentant.

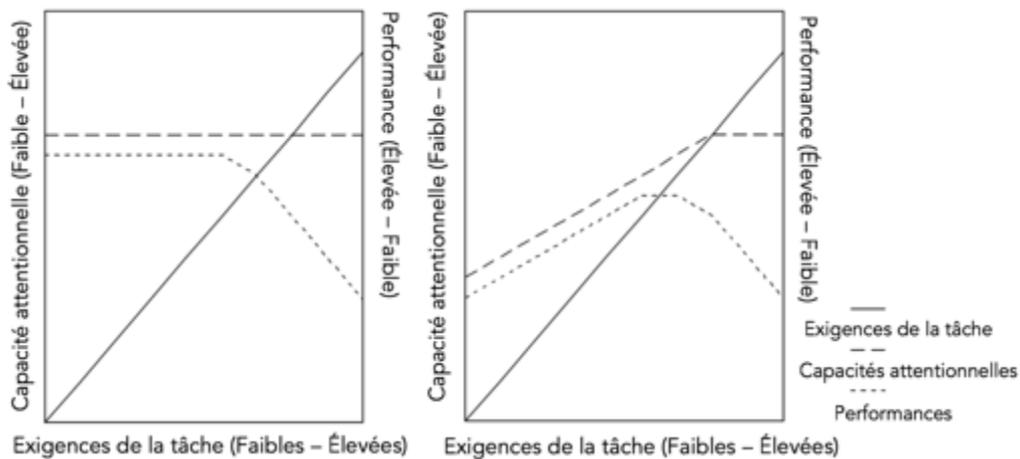


Figure 6 : Relation entre les exigences de la tâche, les capacités attentionnelles et la performance dans le modèle des ressources attentionnelles fixes (gauche) et des ressources attentionnelles malléables (droite) proposé par Nguma (2020).

Modèles des ressources multiples

Les modèles de ressources à réservoir unique (Moray, 1967 ; Kahneman, 1973 ; Young et Stanton, 2002 ; Wickens et al., 2015) ont été enrichis par les modèles de ressources à réservoirs multiples. En effet, Sanders (1983) postule que lors du traitement de l'information, il existe plusieurs étapes (allant du pré-traitement du signal à la préparation de la réponse motrice) ayant chacune leur propre réservoir de ressources. Dans la continuité des travaux de Sanders (1983), Wickens (1984, 1987, 2002, 2008) défend également l'idée qu'il existe plusieurs réservoirs de ressources et propose un modèle avec quatre modules du traitement de l'information. Les modules (ayant chacun leur propre réservoir de ressources) sont : la modalité de la tâche, le code de traitement, les étapes de traitement et enfin la modalité de la réponse (Cf. **Figure 7**).

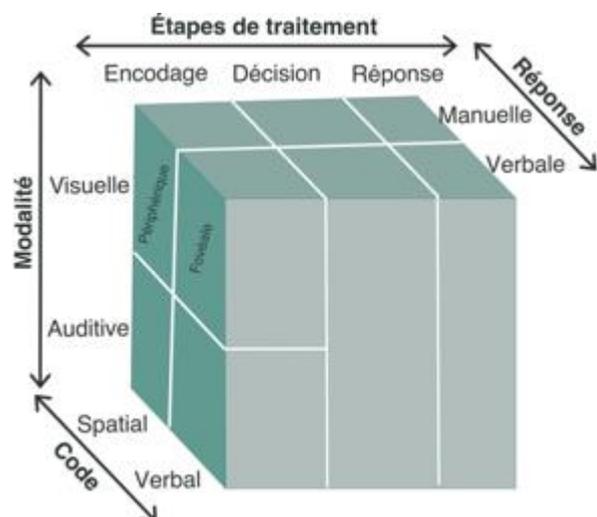


Figure 7 : Représentation du modèle de Wickens (2008) traduit par Le Gonidec (2022).

La modélisation de ces différents réservoirs de ressources met en lumière l'existence de charges de natures différentes. Un opérateur peut être en surcharge mentale sur certaines ressources, et être en sous-charge mentale sur d'autres (Kostenko, 2017). Ainsi, lorsque la tâche fait appel à des ressources similaires, les traitements mis en œuvre rentrent en compétition et la charge mentale augmente (Paxion, 2014). Par ailleurs, deux tâches complexes peuvent être réalisées parallèlement sans interférence, si elles n'utilisent pas les mêmes réservoirs de ressources (Le Gonidec, 2022).

En conclusion de la section 1.1 *Une charge mentale subie par l'individu*, il existe de nombreux modèles de la charge mentale où l'individu subit sa charge mentale. Par conséquent, cette dernière est directement liée à des variables externes comme les exigences de la tâche. Or, le rapport binaire "exigences de la tâche/charge mentale" est restreint et l'ajout d'une variable liée à l'individu, telle que les conséquences sur les performances a pu pallier cette limite. Enfin, les modèles centrés sur les capacités limitées ont permis de prendre en considération les différences inter-individuelles.

Par ailleurs, au fil de notre lecture, nous nous sommes aperçus que la plupart des modèles où l'individu est passif à l'égard de sa charge mentale, sont des modèles non systémiques. Ce type de modèles correspond à la mise en relation entre deux, voire trois dimensions de la charge mentale et s'opposent aux modèles systémiques ayant la volonté d'avoir une vue d'ensemble du système. Néanmoins, les approches non-système ont leur importance puisqu'elles permettent d'établir ou d'approfondir une relation entre un nombre limité de dimensions. Ainsi, les approches systémiques s'appuient sur et synthétisent plusieurs modèles non-système. Dans la prochaine section, nous allons exposer plusieurs approches systémiques de la charge mentale où, cette fois-ci, l'individu est actif à l'égard de ce concept.

1.2 Individu acteur de sa charge mentale

Désormais, nous savons que la charge mentale est impactée par les exigences de la tâche. Néanmoins, cette relation est à nuancer, car elle peut être régulée par l'effort mental et des changements de stratégies de la part de l'opérateur. Par conséquent, vis-à-vis



de la charge mentale, l'opérateur détient une marge de manœuvre pour y faire face ou pour la réguler. L'individu peut donc être acteur de sa propre charge mentale. Une des définitions qui se réfère à ce constat précise que la charge mentale est une structure multidimensionnelle reflétant l'interaction d'éléments, tels que les exigences de la tâche et du système, les capacités de traitement et d'effort de l'opérateur, les critères de performance subjectifs, le processus de traitement de l'information et les stratégies de l'opérateur, ainsi que sa formation et son expérience antérieure (Meshkati, 1988, p. 305). Plusieurs modèles se basent sur la nature dynamique et adaptative de l'individu à l'égard de sa charge mentale. Certains modèles se focalisent sur la prise en compte de l'effort mental, et d'autres sur les stratégies de régulation de l'opérateur.

1.2.1 Approches centrées sur l'effort mental fourni par l'individu

Les approches centrées sur l'effort mental fourni par l'individu considèrent la charge mentale comme la quantité de ressources mentales mobilisées volontairement pour répondre aux exigences d'une tâche (Leplat, 1977 ; Le Gonidec, 2022). Cette modulation volontaire de la quantité de ressources mentales investies lors de la réalisation d'une tâche correspond à la notion d'effort mental (Kahneman, 1973 ; Delignières et Deschamps, 2000). L'effort mental est donc spécifique à chaque individu et représente la façon dont l'opérateur perçoit le niveau d'exigences de la tâche. Selon Paas et Van Marriënboer (1994), seuls les processus contrôlés (traitement conscient, lent et intentionnel de l'information par l'individu) et non les processus automatiques (traitement inconscient et rapide de l'information) induisent un effort mental. De plus, l'effort mental mobilisé dépend de variables externes à l'individu, telles que les caractéristiques de la tâche ou internes à l'individu comme l'expertise ou l'état physiologique de l'opérateur, deux variables internes que nous allons détailler.

Selon Raufaste et al. (2004), l'effort mental est influencé par l'expertise. Cependant, cette relation n'est pas linéaire, car l'effort mental est modulé par la motivation. Pour mieux comprendre, nous nous référons à la théorie de l'intensité de la motivation proposée par Brehm et Self (1989). Ils postulent qu'il existe une relation linéaire entre l'effort mental et la difficulté perçue lors de la réalisation d'une tâche (Cf. **Figure 8**).

La motivation potentielle symbolise la limite supérieure de l'effort mental volontaire dans l'atteinte des objectifs de la tâche. Si la motivation potentielle est basse, alors l'opérateur mobilise peu d'effort mental quel que soit le niveau de difficulté perçue.

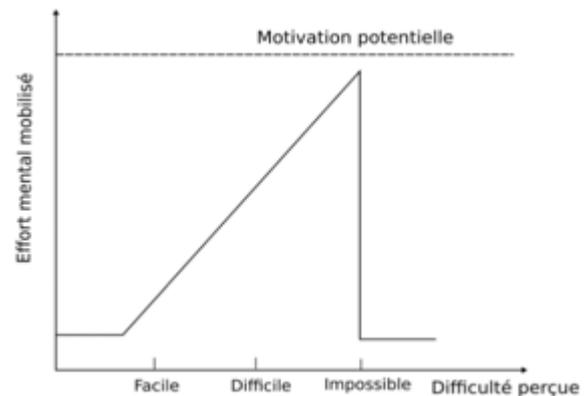


Figure 8 : Relation entre l'effort mental, la motivation et la difficulté perçue proposée par Brehm et Self (1989) traduit par Mallat (2019).

Si la motivation potentielle est élevée, alors l'individu mobilise de l'effort mental selon la difficulté perçue. Enfin, lorsque l'opérateur perçoit la difficulté de la tâche comme impossible à surmonter, il se désengage de l'activité et ne mobilise plus d'effort mental (Mallat, 2019).

Concernant l'état physiologique de l'opérateur, la théorie du contrôle d'état (Hockey, 1986) l'a conceptualisée en lien avec l'effort mental. Lorsqu'il y a un décalage entre l'état physiologique actuel et l'état requis, l'effort mental peut être mobilisé dans la manipulation active de l'état actuel vers l'état requis.

Modèle de De Waard

Concernant spécifiquement les modèles de la charge mentale centrés sur l'effort mental fourni par l'individu, un des plus connus fait suite aux travaux de Meister détaillés à la section précédente (Cf. section 1.1.2 *Approches centrées sur les performances*). En effet, De Waard (1996) reprend le modèle de *régions* (A-B-C) de Meister (1976) en l'enrichissant avec la notion d'effort mental. En parallèle, De Waard (1996) se base également sur les travaux de Mulder (1986). Ce dernier postule l'existence de deux formes d'effort mental : l'effort lié à la tâche qui nécessite un traitement contrôlé de l'information (région A3 sur la **Figure 9**), et l'effort dans le cas où un individu doit modifier l'état actuel vers un état requis, autrement dit un effort compensatoire lié à l'état de l'individu (région A1 sur la **Figure 9**).

Par conséquent, le modèle de De Waard (1996) segmente la région A en trois parties en fonction de l'effort investi par l'opérateur (Cf. **Figure 9**). Lorsque les exigences de la tâche augmentent, à partir de l'état optimal de l'opérateur dans la région A2, une transition de la région A3 vers la région B peut avoir lieu. Au niveau de la région B, la performance est affectée et lorsqu'elle s'est détériorée à un niveau minimum, la région C apparaît. Par conséquent, selon De Waard (1996), il est plus utile de placer une ligne rouge de charge mentale à la transition de la région A2 à A3, plutôt qu'à la transition de la région A (A3) à B, comme l'ont proposé Rueb et al. (1992). En symétrie, De Waard (1996) propose une deuxième ligne rouge de la charge mentale (basculement vers la surcharge mentale) à la transition de la région A2 à la région A1, qui peut survenir, selon l'auteur, lors de tâches monotones.

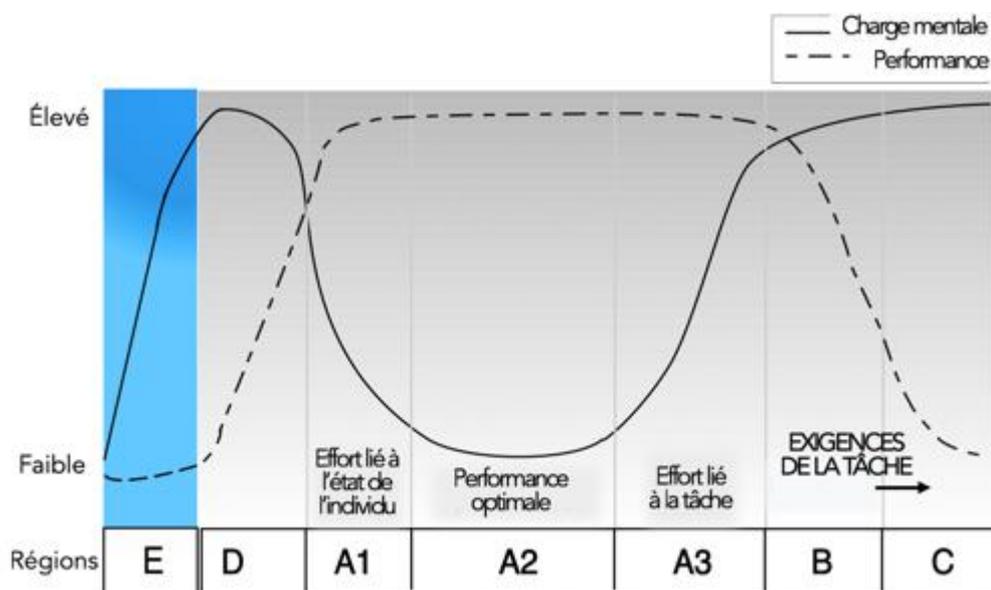


Figure 9 : Relation entre l'effort mental, les exigences de la tâche, la charge mentale et la performance issue du modèle de De Waard (1996) avec notre ajout de la région E correspondant à la sous-charge mentale.

De plus, lorsque l'investissement dans l'effort mental n'est plus efficace, la performance est affectée au niveau de la région D. Ainsi, le deuxième ajout principal de De Waard (1996) à la pensée de Meister (1976) est de proposer la région D (pour « Désactivation ») à l'extrémité gauche de la **Figure 9**. La région D correspond aux tâches monotones, aux faibles exigences, mais entraînant une augmentation de la charge mentale dû à

l'effort mental lié à l'état de l'individu. En effet, l'opérateur doit modifier un état actuel vers un état requis (De Waard, 1996).

Cependant, le modèle de De Waard (1996) a une principale limite, celle du manque de visibilité dans le domaine de la sous-charge mentale, où les exigences de la tâche sont faibles et où les performances se dégradent. La zone E (en bleu sur la **Figure 9**) est un ajout de notre part pour représenter dans la continuité du modèle de De Waard (1996) le domaine de la sous-charge mentale.

Enfin, bien que De Waard (1996) mette en exergue les stratégies mises en place dans la gestion de la charge mentale dans sa revue de la littérature, il n'en fait pas référence dans son modèle. Or, même si le niveau d'exigences de la tâche est élevé, la charge mentale peut être acceptable, donc dans la zone dite *de confort*. Une explication est basée sur l'adoption, de la part de l'individu, de stratégies de régulation de la charge mentale. La section suivante discute de cette approche où l'individu est de nouveau acteur de sa charge mentale, mais cette fois-ci, par la mobilisation de stratégies.

1.2.2 Approches incluant les stratégies de l'individu

Jusqu'à présent, l'ensemble des modèles présentés dans ce chapitre sont en boucles ouvertes. Ils n'ont donc pas de boucle de rétroaction d'une dimension du modèle vers une autre. Selon Loft et al. (2007), ces modèles mettent en relation les exigences de la tâche, la charge mentale et les facteurs médiateurs entre ces deux concepts. Les facteurs médiateurs (ou modérateurs) ont été catégorisés en deux groupes par Van Acker et al. (2018), ceux relatifs à l'individu (tels que l'expertise ou la personnalité) et ceux relatifs à l'environnement (comme le support social ou technologique). Les processus modérateurs et médiateurs portent tous deux sur l'exploration du rôle joué par une troisième variable dans la relation X-Y (X représentant les exigences de l'activité et Y représentant la perception et la conduite de l'opérateur. Des auteurs (Van Acker et al., 2018) en font la distinction et d'autres (Loft et al., 2007) ne la font pas. Dans le cadre de cette thèse, nous reprenons la terminologie de Loft et al. (2007) qui a servi de prémisses aux modèles suivants. Le lien entre les variables internes de l'individu et les médiateurs se fera à partir de la **Figure 12** de ce chapitre.

Par ailleurs, en fonction des exigences de la tâche et de la charge mentale ressentie ou de la charge mentale anticipée, un opérateur peut adopter une stratégie d'adaptation (Paxion, 2014 ; Kostenko, 2017). Cela implique la mise en place de processus de régulation (dits *adaptatifs* ou *compensatoires*), afin de maintenir un niveau de charge mentale acceptable (Laouar-Zouyed, 2021). Selon Sperandio (1971), la stratégie opérationnelle choisie par les individus dépend de trois variables, les caractéristiques de l'opérateur (telles que la motivation, l'âge), les caractéristiques de la tâche (correspondant aux exigences de la tâche), et le niveau de la charge de travail, qui résulte lui-même de la stratégie opérationnelle.

Modèles basés sur les boucles de rétroaction

Sur ce postulat, Sperandio (1971) propose un modèle repris plus tard par Loft et al. (2007) qui décrivent les mécanismes de régulation de l'activité, afin de maintenir un niveau de charge mentale acceptable (dans la *zone de confort*). Pour ce faire, deux boucles de régulations (en boucles fermées ; Cf. **Figure 10**) sont identifiées. Premièrement, la variation de la charge mentale résultant des méthodes de travail a, par rétroaction, un effet régulateur sur le choix des méthodes de travail (boucle de rétroaction 1). Deuxièmement, la méthode de travail utilisée en réponse aux exigences perçues de la tâche régule le futur niveau d'exigence perçue de la tâche (boucle de rétroaction 2). Par la suite, Leplat (2006) propose une troisième boucle de régulation correspondant à la différence entre le niveau de performance atteint et le niveau attendu, ce qui exerce une influence sur la sélection des futures stratégies. Dans le cadre de la thèse, nous avons illustré cette boucle de régulation **en vert** sur la **Figure 10**. En parallèle, Loft et al. (2007) proposent également une autre boucle de régulation correspondant à la différence entre la charge mentale actuelle et la charge mentale désirée, ce qui exerce également une influence sur la sélection des futures stratégies.

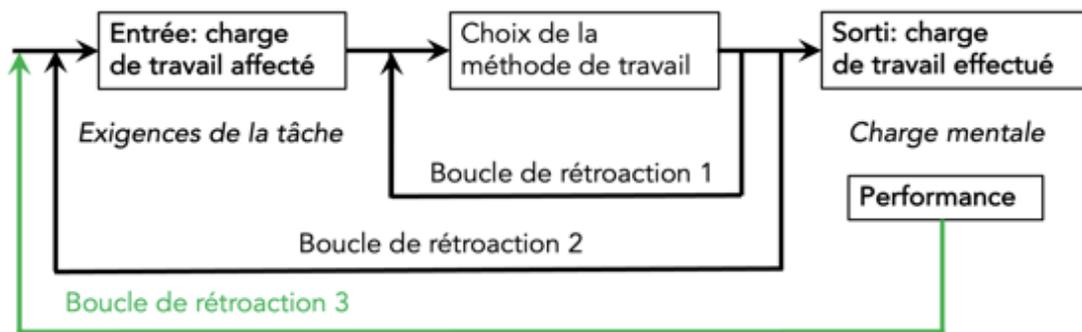


Figure 10 : Les boucles de régulation pour maintenir une charge mentale acceptable issues des travaux de Sperandio, 1971 (boucle 1), adapté par Loft et al., 2007 (boucle 2), puis adapté dans le cadre de la thèse (boucle 3).

Modèle d'Hart et Staveland

Le modèle d'Hart et Staveland (1988) prend également appui sur le caractère actif de l'opérateur en détaillant davantage le lien entre la perception (évaluation de la situation), la conduite de l'opérateur (sélection et implémentation des stratégies) et les résultats (performances). En effet, la conduite de l'individu reflète la perception de ce que l'on attend de lui, ainsi que les stratégies, les efforts et les ressources du système déployés pour atteindre les objectifs de la tâche. Par ailleurs, la perception de l'opérateur affecte son comportement ultérieur (dont ses stratégies), et donc ses performances. Si l'opérateur considère que la charge de travail imposée (comprenant les exigences de la tâche) est excessive, il peut se comporter comme s'il était surchargé, même si la charge de travail imposée est objectivement faible. Par conséquent, il y a la perception (évaluation de la situation), puis la conduite (sélection et implémentation des stratégies) de l'opérateur qui produit des résultats (performances). De plus, il y a une boucle de rétroaction entre les performances et la charge de travail imposée. En effet, les performances impactent la charge de travail imposée qui peut par la suite être plus ou moins élevée. Le modèle d'Hart et Staveland (1988) est donc inséré dans une vision dynamique du système, où la charge de travail imposée n'est pas statique, mais est fluctuante. Enfin, les effets observables de l'activité sur la charge mentale peuvent être mesurés de manière subjective (expérience de la charge mentale), ou de manière objective (conséquences physiologiques) représentées par des flèches noires sur le cadre conceptuel d'Hart et Staveland (1988 ; **Figure 11**).

Selon Kostenko (2017), il existe deux processus dans le modèle d'Hart et Staveland (1988). Le premier (avec les **flèches en rouge** sur la **Figure 11**) est le processus primaire où l'opérateur (perception et conduite) répond à la charge de travail imposée en produisant des résultats. Ces derniers, peuvent à leur tour impacter la charge de travail imposée, dans une vision dynamique du système.

Le second processus (avec les **flèches bleues** sur la **Figure 11**) correspond à un processus de supervision du processus primaire. Effectivement, l'opérateur réactualise en permanence son évaluation de la situation (perception) en tenant compte des changements de la charge imposée et du niveau de performance atteint, d'où la flèche bleue qui prend racine au niveau des performances et se dirige vers les perceptions de l'opérateur. Cette évaluation continue permet à l'opérateur de réguler son effort mental et d'adapter ses stratégies. Par conséquent, cette flèche bleue (des performances aux perceptions de l'individu) correspond à la boucle de régulation des stratégies proposée par Leplat (2006 ; Cf. boucle de couleur verte à la **Figure 10**). Elle correspond donc, en partie, à la différence entre le niveau de performance atteint et le niveau de performance attendu qui aura une incidence sur la sélection des futures stratégies.

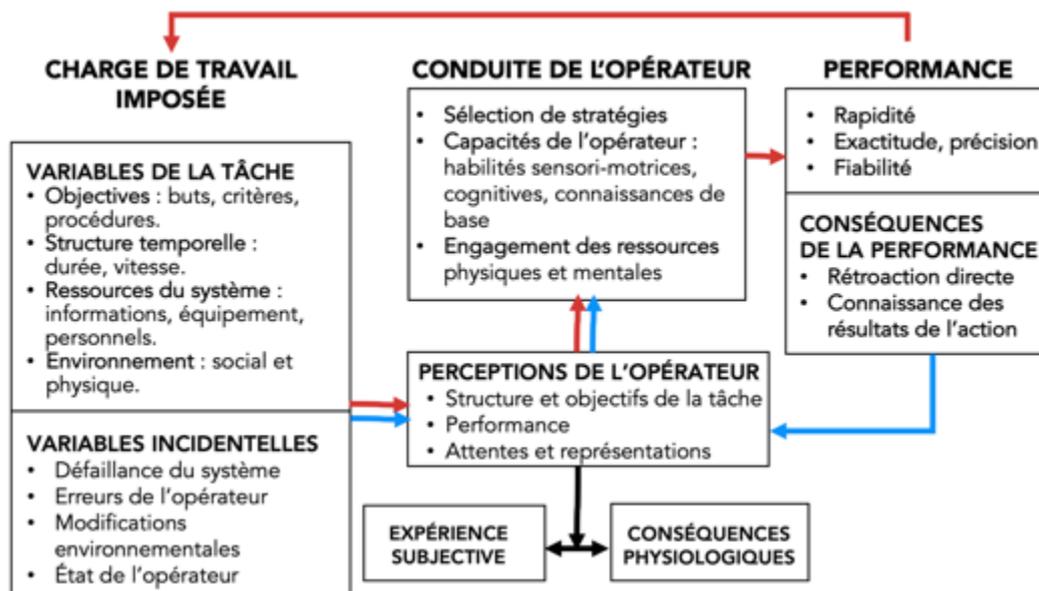


Figure 11 : Modèle d'Hart et Staveland (1988) avec l'ajout de couleurs (rouge pour le processus primaire ; bleu pour le processus de supervision ; Kostenko, 2017).

Enfin, qu'importe le type de stratégies utilisées, certaines sont plus économiques que d'autres. Selon Sperandio (1971), l'opérateur n'utilise les méthodes plus économiques que lorsque les ressources mentales atteignent les capacités limitées. Lorsque la tâche

ne requiert qu'une partie de ses capacités, l'opérateur peut utiliser des modes opératoires moins économiques en termes de charge de travail, mais plus satisfaisants selon d'autres critères, tels que le fait de favoriser des méthodes qui permettront la consolidation de certaines informations en Mémoire à Long Terme (MLT), donc qui permettront un meilleur apprentissage.

Théorie de la charge cognitive

Le modèle de la charge mentale qui établit un lien entre la Mémoire de Travail (MdT) et la MLT, est la théorie de la charge cognitive (Cognitive Load Theory – CLT) de Sweller (1988). Ce concept est lié à celui de charge mentale. Selon Orru et Longo (2019), le premier a évolué au sein de la psychologie de l'éducation, tandis que le second est issu de l'ergonomie et des facteurs humains. Malgré leur évolution indépendante au sein de différentes disciplines, les deux concepts sont fondés sur les mêmes cadres théoriques, tels que le modèle de la mémoire d'Atkinson et Shiffrin (1968). Certains auteurs (Hancock et al., 2021) vont plus loin, en considérant ces deux concepts comme des synonymes.

Selon Young et Sewell (2015), la CLT s'appuie sur le modèle de mémoire développé par Atkinson et Shiffrin (1968). Ces derniers précisent que la mémoire se compose de trois sous-systèmes principaux : la mémoire sensorielle, la mémoire à court terme reconsidérée comme la Mémoire de Travail (MdT ; Baddeley et Hitch, 1974) et la mémoire à long terme (MLT). La mémoire sensorielle perçoit et conserve brièvement les informations sensorielles de l'environnement. Les informations sensorielles portées à la connaissance consciente entrent dans le domaine de la MdT. Cette dernière récupère les connaissances pertinentes que possède l'apprenant et qui sont stockées de manière illimitée dans la MLT sous forme de schémas. La MdT organise et intègre ensuite les nouvelles informations à celles déjà existantes pour faciliter un stockage efficace sous la forme de nouveaux schémas (ou de schémas modifiés).

À partir de ce modèle, la CLT se concentre sur la manière dont les charges *intrinsèque* (associée à la tâche d'apprentissage elle-même), *extrinsèque* (charge non essentielle à la tâche essentiellement reliée aux consignes de la tâche) et *essentielle* (nommée également *utile* - imposée par l'utilisation délibérée de stratégies cognitives par l'apprenant pour

l'apprentissage) ont un impact sur la MdT d'un apprenant (Sweller, 1988 ; Van Merriënboer et Sweller, 2010 ; Leppink et al, 2014 ; Young et Sewell, 2015 ; Young et al., 2021). Plus tard, Sweller (2010) re-conceptualise la CLT en introduisant le concept d'interactivité des éléments. Les *éléments* font référence au contenu d'apprentissage qui peut être traité soit indépendamment dans la MdT (comme une liste de mots à mémoriser), soit simultanément (comme les éléments d'une équation mathématique). Cela détermine le niveau d'interactivité entre les éléments. En outre, la charge *essentielle* représente l'effort supplémentaire requis pour l'apprentissage (construction de schémas). Toutefois, pour que la charge *essentielle* soit efficace, la MdT doit être en capacité suffisante. Si ce n'est pas le cas, la charge *extrinsèque* peut être réduite (par exemple en regroupant les mots par leur sens). En revanche, la charge *intrinsèque* reste constante pour un niveau d'expertise spécifique. Par conséquent, la charge *essentielle* est liée à la charge *intrinsèque*, qui dépend du degré d'interactivité de l'élément de la tâche, et à la charge *extrinsèque*, qui ne doit pas être élevée pour faire de la place à la charge *essentielle*. Cette dernière fait donc référence à l'effort et au traitement cognitif de l'individu, dans la formation de schémas sous la forme de connaissances dans la MLT (Longo et al., 2022). Ce type de charge résulte donc de la mise en relation d'informations pertinentes issues de la MLT ou du contexte avec les nouveaux éléments d'information issus de l'activité à réaliser. Par conséquent, le modèle de la CLT est intéressant dans la conceptualisation de la charge mentale pour mettre en exergue l'importance de la MLT dans ce concept.

En conclusion, la plupart des études se focalisent sur les exigences de la tâche et les performances pour la modélisation de la charge mentale (Loft et al., 2007). Or, l'individu est actif à l'égard de la charge mentale, que ce soit par l'effort mental déployé ou par les stratégies mises en place pour conserver un niveau acceptable de charge mentale. Par conséquent, lors de la prochaine section, nous proposons un modèle systémique et une définition qui s'ancrent dans une approche multidimensionnelle prenant en compte le caractère actif de l'opérateur à l'égard de sa propre charge mentale.

1.2.3 Proposition d'un modèle et d'une définition de la charge mentale

À la suite de cette revue de la littérature, nous proposons un modèle multidimensionnel (Cf. **Figure 13**) dans lequel l'opérateur est actif à l'égard de sa charge mentale. Même

si nous souhaitons tenir compte de l'ensemble des approches décrites dans les sections précédentes, nous nous sommes principalement basés sur les modèles de Sperandio (1971), d'Hart et Staveland (1988) et de De Waard (1996).

Une des principales différences avec le modèle de Sperandio (1971), concernant le contrôle régulateur par les stratégies, est que nous prenons en considération l'ensemble des conséquences (expérience subjective, performance et physiologique) et pas uniquement les performances, pour l'évaluation continue de la situation faite par l'opérateur pour réguler son effort mental et adapter ses stratégies.

La principale différence avec le modèle de De Waard (1996), est que nous ne détaillons pas les différentes régions (telles que les régions A1 ; A2 ; A3 ; Cf. **Figure 9**) précisant les niveaux de charge mentale. En effet, notre objectif avec le cadre conceptuel présenté en **Figure 13** est d'avoir une vue sur les différentes relations qu'entretiennent les dimensions de la charge mentale et non l'exposé des niveaux de charge mentale qui a déjà été proposé maintes fois dans la littérature (Meister, 1976 ; Hancock et Williams, 1993 ; De Waard, 1996 ; Wickens et al., 2015 ; Boyer, 2016).

Enfin, une des différences principales avec l'approche d'Hart et Staveland (1988), est que dans notre modèle, les conséquences physiologiques découlent de la perception et de la conduite de l'opérateur (comprenant la sélection des stratégies et l'effort mental), et non, uniquement de la perception de l'opérateur. De plus, le modèle d'Hart et Staveland (1988 ; Cf. **Figure 11**) mêle, dans la catégorie *charge de travail imposée*, les variables incidentelles ("défaillances du système"), le contexte ("environnement social et physique"), les exigences de la tâche ("objectifs, structure temporelle") ou encore les variables internes ("état de l'opérateur") et externes ("équipement") à l'individu. Dans le modèle proposé en **Figure 13**, nous en faisons des catégories distinctes. Nous distinguons le contexte de réalisation de la tâche (où nous retrouvons l'environnement physique et social ou les défaillances du système) de la tâche à réaliser.

En outre, la tâche à réaliser possède un certain nombre d'exigences avec des critères objectifs de performance. Cette même tâche possède également un certain nombre de variables qui ne dépendent pas des exigences de la tâche. Nous considérons donc ces variables comme extrinsèques aux exigences de la tâche. Cette distinction se retrouve sur la **Figure 12**. Par ailleurs, si ces variables, intrinsèques ou extrinsèques aux

exigences de la tâche exercent une influence sur la charge mentale, leurs statuts changent. Elles deviennent des facteurs qui impactent la charge mentale, d'où les flèches directionnelles sur la **Figure 12**. De même, l'individu, qui répond aux exigences de la tâche, est soumis à un certain nombre de variables internes ou externes à lui-même. Si ces variables exercent une influence sur la charge mentale, alors elles deviennent des médiateurs de ce concept, d'où les flèches directionnelles sur la **Figure 12**. Comme nous l'avons vu à la section 1.2.2, la notion de médiateurs de la charge mentale a été proposée par Loft et al. (2007) et dans le modèle de Van Acker et al. (2018).

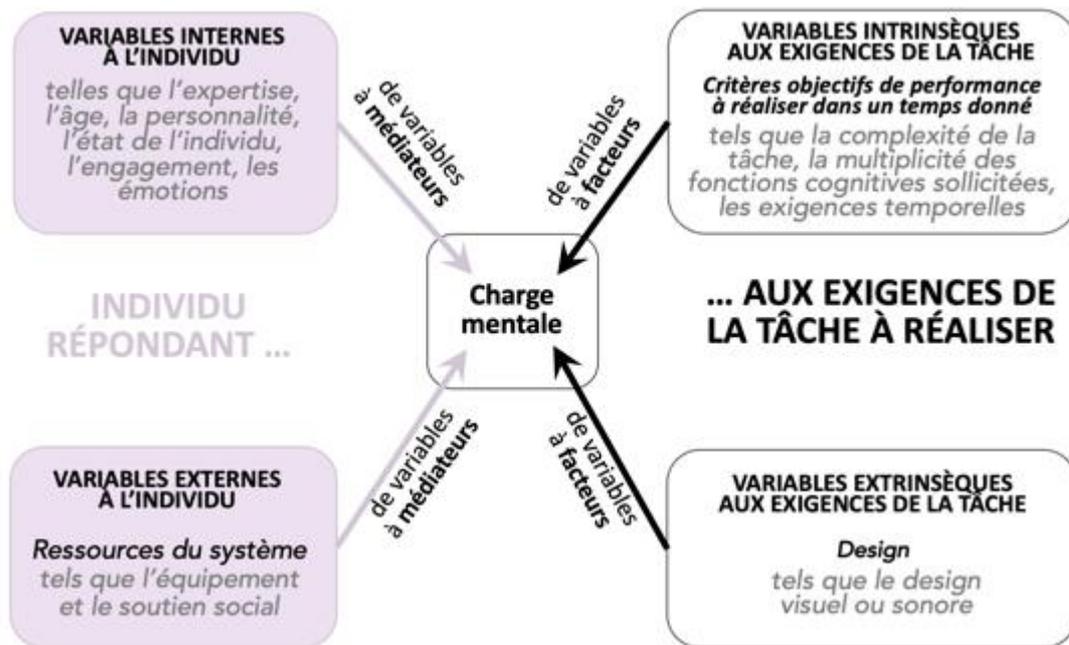


Figure 12 : Les différentes variables internes et externes à l'individu pouvant être des médiateurs de la charge mentale. Ainsi que les différentes variables intrinsèques et extrinsèques aux exigences de la tâche pouvant être des facteurs de la charge mentale.

Le cadre conceptuel de la charge mentale, que nous proposons à la **Figure 13**, illustre donc un individu qui répond aux exigences d'une activité donnée. Étant donné que l'objectif avec ce cadre conceptuel est d'avoir une vue sur les différentes relations qu'entretiennent les dimensions de la charge mentale, nous considérons directement les variables internes et externes à l'individu comme des médiateurs. De même, nous considérons les variables intrinsèques et extrinsèques à la tâche comme des facteurs.

De plus, un code couleur, qui reprend en partie celui de la **Figure 12**, est établi. Les exigences de la tâche sont représentées sous fond blanc. Les processus liés à l'individu

sont sous un fond gris clair. Ils concernent l'attention, les Mémoires Sensorielles, la perception, la conduite de l'opérateur et la Mémoire à Long Terme. Les conséquences sur l'individu sont sous un fond gris foncé. Enfin, les médiateurs (relatifs à l'individu et à l'environnement) sont sous un fond rose.

Par ailleurs, comme pour la **Figure 11**, les flèches de couleurs issues du modèle d'Hart et Staveland (1988) enrichi par Kostenko (2017) correspondent :

- Au processus primaire où l'opérateur répond aux exigences de la tâche en produisant des résultats. Ces derniers, peuvent à leur tour impacter les exigences de la tâche dans une vision dynamique du système (**flèches rouges**).
- Au processus de supervision du processus primaire où l'opérateur réactualise en permanence son évaluation de la situation (**flèches bleues**). Cette réactualisation prend compte des changements des exigences de la tâche et du niveau de performance atteint, d'où la flèche bleue qui prend racine au niveau des performances et se dirige vers les perceptions de l'opérateur. Cette évaluation continue permet à l'opérateur de réguler son effort mental et d'adapter ses stratégies. Par conséquent, cette flèche bleue (des performances aux perceptions de l'individu) correspond à la boucle de régulation des stratégies proposée par Leplat (2006 ; Cf. boucle de couleur verte à la **Figure 10**).

De plus, il y a une double flèche bleue entre la perception et la conduite de l'individu. En effet, la perception de l'individu va impacter sa conduite lors de la réalisation de l'activité. Elle correspond à l'évaluation de l'effort investi et des ressources engagées et disponibles pour la conduite du processus. En retour sa conduite ira impacter ses perceptions.

Par ailleurs, le nouveau cadre conceptuel présenté à la **Figure 13** propose plusieurs flèches inédites, telles que la flèche bidirectionnelle allant de la conduite de l'opérateur à la Mémoire à Long Terme (MLT). En effet, selon la théorie de la charge cognitive (Sweller, 1988 ; Cf. section 1.2.2 à la page 29), il y a une mise en relation d'informations pertinentes issues de la MLT avec les nouveaux éléments d'informations à traiter (phase de récupération). De plus, ces derniers, par leur pertinence, peuvent être encodés, puis stockés en MLT (phase d'encodage et de stockage).

En outre, nous soupçonnons un lien bidirectionnel entre la perception de l'individu et les médiateurs internes à l'individu. En effet, les médiateurs internes, tels que les

émotions ou l'engagement, impactent la perception qu'à l'individu sur les caractéristiques intrinsèques de la situation. En retour, la perception qu'à l'individu sur les caractéristiques intrinsèques de la situation pourrait impacter ses émotions ou son engagement. À notre connaissance, dans un contexte d'étude lié à la charge mentale, ce lien n'est pas encore vérifié expérimentalement. Nous le présentons donc en pointillé au niveau de la **Figure 13**.

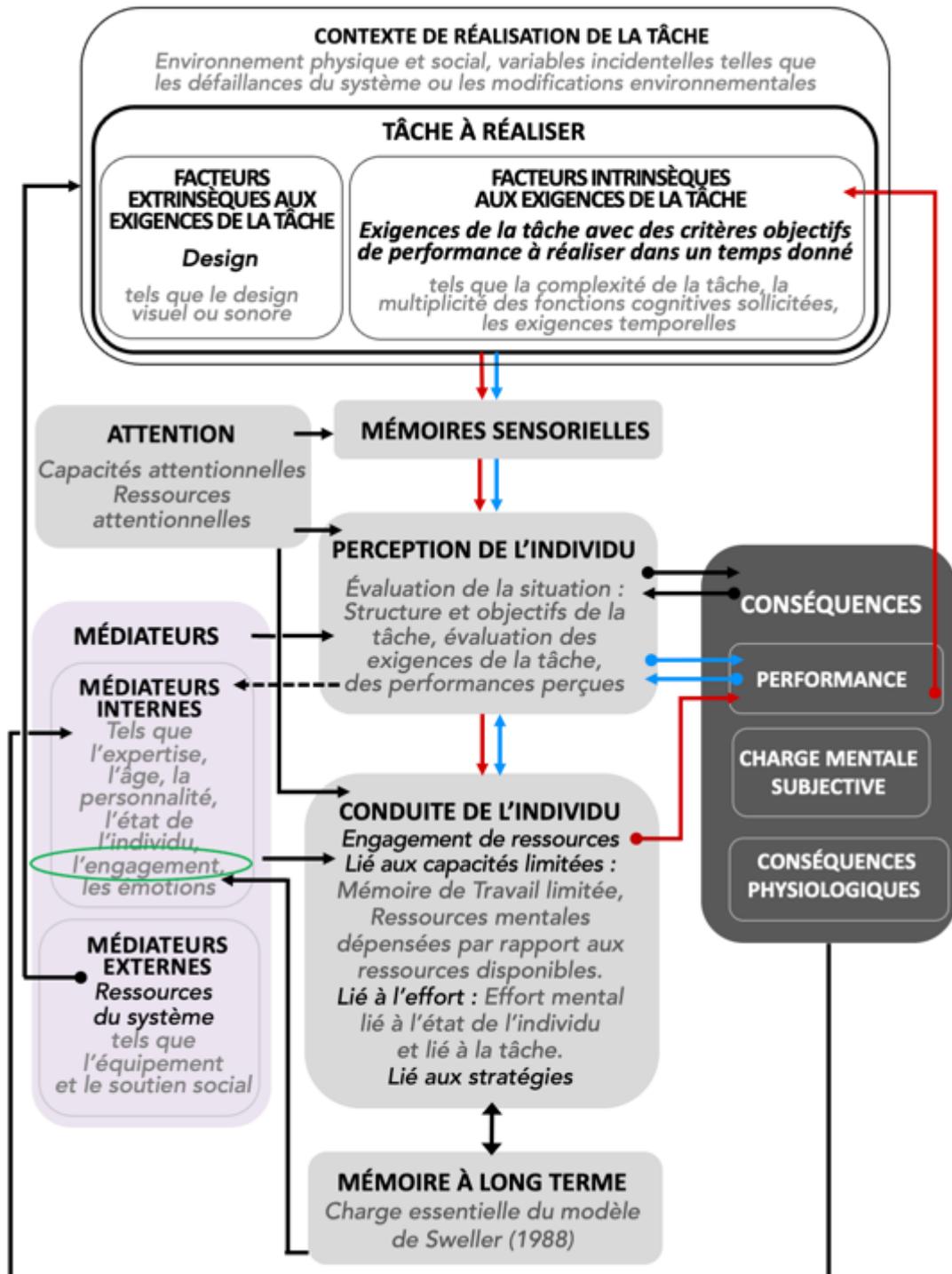


Figure 13 : Notre modèle conceptuel de la charge mentale combinant différents modèles théoriques. L'explication des flèches est donnée dans le corps du texte en amont de la Figure.

Sur la base de ce cadre conceptuel issu de la revue de littérature, nous pouvons définir la charge mentale comme *le niveau d'activation des ressources mentales limitées pour répondre aux critères de performance (objectifs et/ou subjectifs) d'une activité accomplie en un temps donné, par un individu qui y consacre de l'attention, de l'effort mental et déploie des stratégies de régulation pour maintenir une charge mentale dans des limites acceptables. La mobilisation délibérée de ressources mentales (effort mental) et des stratégies résulte de la mise en relation d'informations pertinentes issues de la mémoire à long terme ou du contexte avec les nouveaux éléments d'information issus de l'activité à réaliser. Par ailleurs, cet individu est influencé par des variables internes (expertise, âge, engagement ...) et externes (soutien social, équipement technologique ...) qui peuvent être des médiateurs de la perception et de la conduite de l'opérateur. En conséquence, l'individu a une expérience subjective de la charge mentale, ainsi que des répercussions comportementales et physiologiques qui, à leur tour, ont un impact sur la perception et les médiateurs internes de l'individu.*

Ci-après, un tableau avec les définitions phares de la charge mentale, leurs dimensions principales et leurs manques, que la nouvelle définition tente de combler.

Auteurs	Définitions de la charge mentale	Dimensions	Manques
Wickens (2002)	<i>Coûts encourus par un opérateur lors de l'exécution de plusieurs tâches qui utilisent un ensemble commun de ressources</i>	<ul style="list-style-type: none"> • Quantité de ressources limitées • Processus de traitement entraînant un coût 	<ul style="list-style-type: none"> • Exigences de la tâche • Mise en jeu de facteurs liés à l'environnement • Concept propre à chaque individu • Conséquences sur le traitement de la tâche et sur l'individu • Facteur temporel • Aspects biologiques • Stratégies de régulation

			<ul style="list-style-type: none"> • Médiateurs internes et externes à l'individu
Cain (2007)	<i>Construction mentale qui reflète la contrainte mentale résultant de l'exécution d'une tâche dans des conditions environnementales et opérationnelles spécifiques, associée à la capacité de l'opérateur à répondre à ces exigences.</i>	<ul style="list-style-type: none"> • Exigences de la tâche • Quantité de ressources limitées • Processus de traitement entraînant un coût • Concept propre à chaque individu • Mise en jeu de facteurs liés à l'environnement 	<ul style="list-style-type: none"> • Conséquences sur le traitement de la tâche et sur l'individu • Aspects biologiques • Stratégies de régulation • Médiateurs internes et externes à l'individu • Facteur temporel
Kostenko (2017)	<i>Concept abstrait (non observable), relatif (à la notion de capacité), et subjectif (par la singularité de chaque opérateur). Cependant, la charge mentale peut être vue comme un construit multidimensionnelle : causes (contraintes), effets (conséquences de l'astreinte) et médiateur (boucles de régulation).</i>	<ul style="list-style-type: none"> • Exigences de la tâche • Processus de traitement entraînant un coût • Concept propre à chaque individu • Mise en jeu de facteurs liés à l'environnement • Conséquences sur le traitement de la tâche et sur l'individu • Stratégies de régulation • Quantité de ressources limitées • Médiateurs internes et externes à l'individu 	<ul style="list-style-type: none"> • Aspects biologiques • Facteur temporel
Longo et al. (2022)	<i>Degré d'activation d'un pool fini de ressources, limité en capacité, pendant le traitement cognitif d'une tâche primaire dans le temps, médiée par des facteurs environnementaux et situationnels stochastiques externes, ainsi qu'affectée par des caractéristiques internes définies d'un opérateur humain, pour faire face à des demandes de tâches statiques, en consacrant des efforts et de l'attention.</i>	<ul style="list-style-type: none"> • Quantité de ressources limitées • Processus de traitement entraînant un coût • Exigences de la tâche • Concept propre à chaque individu • Mise en jeu de facteurs liés à l'environnement • Conséquences sur le traitement de la tâche et sur l'individu • Facteur temporel 	<ul style="list-style-type: none"> • Médiateurs internes à l'individu • Stratégies de régulation • Aspects biologiques
Nouvelle définition	<i>Cf. paragraphe en gras italique au-dessus de ce tableau.</i>	<ul style="list-style-type: none"> • Quantité de ressources limitées • Processus de traitement entraînant un coût • Exigences de la tâche • Concept propre à chaque individu • Mise en jeu de facteurs liés à l'environnement • Conséquences sur le traitement de la tâche et sur l'individu 	<ul style="list-style-type: none"> • Aspects biologiques

		<ul style="list-style-type: none"> • Facteur temporel • Médiateurs internes et externes à l'individu • Stratégies de régulation 	
--	--	--	--

Tableau 1 : Définitions phares de la charge mentale avec leurs dimensions, leurs manques et la nouvelle définition de la charge mentale proposée dans le cadre de la thèse avec leurs dimensions et leurs manques.

Cette nouvelle définition de la charge mentale pallie les manques des définitions phares de ce concept, sans être exhaustive. En effet, les soubassements biologiques, par exemple, ne sont pas précisés, notamment, pour ne pas alourdir davantage la définition qui est étendue et pour être centrée sur une vision psycho-ergonomique, discipline de la thèse. Le modèle conceptuel pourrait donc être amélioré par des études en neurosciences et en biologie.

Revenons sur l'engagement (encadré en vert sur la **Figure 13**) qui est intéressant à prendre en compte dans le cadre de départ où l'individu réalise une tâche cognitive. L'engagement correspondant, dans un contexte Interaction Humain-Machine (IHM), à un *état cognitif, affectif et comportemental qui donne envie à l'individu d'être impliqué* (O'Brien et Toms, 2010) durant l'activité. C'est donc une mesure de la qualité de l'expérience utilisateur (Lehmann et al., 2012). Dehais et al. (2020) se sont intéressés au lien entre l'engagement et la charge mentale en passant par le prisme des performances. En effet, selon ces auteurs, il y aurait un lien simple entre les performances et l'engagement. En outre, comme nous l'avons vu à la section 1.1.2 *Approches centrées sur les performances*, le modèle en U inversé relie directement les performances et la charge mentale (Cf. **Figure 3** de la section 1.1.2). De plus, il existerait une zone de confort où les performances à une tâche sont optimales entre l'engagement et l'éveil. Nous retrouvons également cette zone dans le concept de la charge mentale (Cf. la traduction française de Mallat (2019) du modèle temps/charge mentale ; Cf. **Figure 2** du chapitre 1). L'éveil représentant un état de préparation physiologique pour répondre à des contingences externes (Pribram et McGuinness, 1975 ; cité par Dehais et al., 2020). Par conséquent, nous pouvons supposer un lien non-linéaire entre l'engagement et la charge mentale. En effet, si un individu est dans de bonnes conditions

(par exemple d'éveil physiologique), même si sa charge mentale augmente (en termes d'exigences de la tâche), son engagement peut également augmenter (en termes de mobilisation de ressources attentionnelles), afin de maintenir de bonnes performances. Une relation que nous allons tenter d'appréhender lors de la phase expérimentale.

Pour finir, la nouvelle définition proposée nous guidera tout au long du manuscrit, à commencer par le chapitre suivant concernant les mesures de ce concept multidimensionnel.



Figure 1 du Chapitre 2 (chaque chapitre possède sa numérotation indépendante du reste du manuscrit) : Plan général du manuscrit et détail des sous-parties du chapitre 2.

Le second chapitre du manuscrit introduit les techniques d'évaluation de la charge mentale (Cf. **Figure 1**) qui reflètent la nature multidimensionnelle du concept. En effet, il existe un large éventail de techniques de mesures.

Tout d'abord, la mesure de la charge mentale est essentielle à plusieurs niveaux. En effet, Cain (2007) précise que la principale raison de mesurer la charge mentale est de quantifier le coût mental de l'exécution des tâches, afin de prédire la performance des opérateurs et des systèmes. De plus, selon Leduc (2019), la mesure de la charge mentale répond à un enjeu scientifique, plus précisément méthodologique de voir "objectiver" ce concept. La mesure répond également à un enjeu social pour repérer les seuils de la charge mentale, notamment la surcharge mentale où l'intégrité psychique et physique peuvent être détériorées. Pour évaluer la charge mentale, trois catégories de mesures existent (O'Donnell et Eggemeier, 1986 ; Reid et Nygren, 1988 ; Miller, 2001 ; Galy et al., 2012 ; Muñoz-de-Escalona et Cañas, 2019) : les mesures subjectives, les mesures de performance et les mesures psychophysiologiques. Dans ce manuscrit, chaque catégorie de mesure sera présentée dans une section dédiée, au regard des principaux critères pour expertiser un outil de mesure (Miller, 2001 ; Cegarra et Morgado, 2009 ; Moustafa et al., 2017 ; Verhulst, 2018 ; Butmee et al., 2018). En effet, la qualité de la mesure est d'autant meilleure qu'elle répond à des critères variés. Il s'agit des critères de **validité** nommée également "spécificité" (capacité de la mesure à évaluer la charge mentale), de **diagnostic** (possibilité d'identifier la source des modulations de la charge mentale), de **sensibilité** (capacité de la mesure à discriminer des changements dans les variations de la charge mentale), d'**intrusivité** (qui fait référence au degré d'interférence de la méthode avec l'exécution d'une tâche). On note également les critères de **fiabilité** (capacité à donner des mesures avec le moins d'incertitude possible), de **cohérence** nommée également "sélectivité" (Cegarra et Morgado, 2009) correspondant à la capacité de la mesure à rester inchangée lorsque la charge mentale n'évolue pas, mais aussi de **moindre équipement** (la mesure d'évaluation doit demander un minimum de matériel pour éviter d'interférer avec la performance du participant) et d'**acceptabilité** (la mesure doit être acceptée par les participants). Enfin, les critères de moindre équipement et d'acceptabilité entrent dans le critère de coût, tant de la mise en œuvre, de l'administration ou du traitement des données. Selon Miller (2001), si une mesure

nécessite un équipement coûteux ou des experts spécialisés pour superviser (en amont, pendant ou en aval) une partie du test, alors elle n'est pas totalement rentable.

2.1 Mesures subjectives

Les mesures subjectives sont les mesures de la charge mentale qui sont les plus proches de l'essence de ce concept et constituent l'indicateur le plus valide et le plus sensible (selon Hart et Staveland, 1988). L'évaluation de la charge mentale se fait par l'individu lui-même. Cela peut se réaliser par des



verbatim (propos spontanés de l'individu), des entretiens de recherche (des “interviews”) a posteriori de l'activité ou des auto-évaluations à travers des questionnaires.

En ce qui concerne la charge mentale, il existe une variété de techniques d'auto-évaluation à travers des questionnaires qu'on peut distinguer en deux catégories : les échelles unidimensionnelles et multidimensionnelles de la charge mentale. Selon Hancock et al. (2021), l'évaluation unidimensionnelle se distingue de l'évaluation multidimensionnelle dans l'utilisation d'un arbre de décision pour arriver à une évaluation unique et chiffrée de la charge mentale. Contrairement aux échelles multidimensionnelles qui consistent en un ensemble de sous-échelles qui peuvent être traitées individuellement pour diagnostiquer la ou les sources de la charge mentale, ou être combinées pour obtenir un score global de la charge mentale.

2.1.1 Échelles unidimensionnelles

Une échelle unidimensionnelle a pour objectif d'avoir une évaluation aboutissant à un seul et unique résultat. C'est le cas de l'échelle unidimensionnelle de Cooper-Harper (*Cooper-Harper aircraft handling Qualities Rating scale* ou HQR ; Cooper et Harper, 1969). Initialement développée au sein de la NASA pour l'évaluation des cockpits par les pilotes d'essai, l'échelle HQR s'appuie sur un arbre de décision basé sur la tâche, les caractéristiques de l'avion et la charge de travail. Un arbre de décision est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre. Les nœuds de l'arbre sont construits depuis sa racine vers sa base où les nœuds terminaux sont appelés “*en feuilles*”. Les différentes décisions possibles sont donc situées aux extrémités des branches (les “*feuilles*”), et sont atteintes en fonction de décisions prises

(Bertet et al., 2009). La principale limite de l'échelle HQR est qu'elle est spécifique au domaine aéronautique, ce qui a été surmonté par l'échelle de Cooper-Harper modifiée (*Modified Cooper-Harper Scale* ou MCHS ; Wierwille et Casali, 1983).

Plus tard, Tattersall et Foord (1996), ont proposé l'auto-évaluation instantanée (*Instantaneous Self Assessment* ; ISA), mise au point pour fournir des évaluations subjectives de la charge mentale durant la tâche, principale limite des mesures subjectives. Durant l'ISA, les individus répondent à un signal lumineux pour préciser leur niveau de charge mentale. Les réponses étaient données soit à l'aide d'un clavier à cinq touches, chacune étiquetée pour représenter un niveau de charge mentale différent (1 = sous-utilisé ; 2 = détendu ; 3 = confortable ; 4 = élevé ; 5 = excessif), soit en donnant les réponses orales équivalentes. Dans leur étude, Tattersall et Foord (1996) ont noté une corrélation significative entre l'ISA, le SWAT (une échelle multidimensionnelle de la charge mentale) et la fréquence cardiaque (qui est un indicateur psychophysologique de la charge mentale). De plus, Tattersall et Foord (1996) précisent que l'ISA est sensible aux variations de la complexité de la tâche. Cependant, la performance à la tâche principale s'est avérée moins bonne pendant les périodes où des réponses à l'ISA étaient requises, qu'il s'agisse de réponses orales ou manuelles. Il y a donc une limite liée au critère d'intrusivité de la mesure. Ce résultat suggère que la mesure de la charge mentale subjective durant la tâche (sans impacter les performances à la tâche principale), est difficile à mettre en place.

En conclusion, il existe plusieurs échelles unidimensionnelles qui sont pour la plupart non validées empiriquement (Mandrick, 2013). Cependant, elles peuvent être utiles pour augmenter la précision sur une dimension en particulier de la charge mentale (De Waard, 1996 ; Mallat, 2019), comme la difficulté perçue avec l'échelle unidimensionnelle d'Eccles et Wigfield (1995) ou l'effort perçu avec le *Rating Scale Mental Effort* (RSME) de Zijlstra (1993). Cependant, les échelles unidimensionnelles de la charge mentale doivent préférentiellement être utilisées en complément d'échelles multidimensionnelles. En effet, les échelles unidimensionnelles ne sont pas représentatives de la multidimensionnalité reconnue du concept de la charge mentale (Cf. Chapitre 1 - *Modèles et définitions de la charge mentale*).

2.1.2 Échelles multidimensionnelles

Parmi les échelles multidimensionnelles, trois ont été validées. Il s'agit de l'indice du *NASA Task Load Index* (NASA-TLX ; Hart et Staveland, 1988), du *Subjective Workload Assessment Technique* (SWAT ; Reid et Nygren, 1988) et du *Workload Profile* (WP ; Tsang et Velazquez, 1996). Ces questionnaires multidimensionnels comportent un nombre d'items différents, en fonction des dimensions jugées pertinentes pour l'évaluation de la charge mentale. Par ailleurs, chacun de ces questionnaires repose sur une méthodologie ou un modèle différents, ainsi, nous les considérerons dans des sous-sections distinctes.

2.1.2.1 NASA-TLX

Le *NASA Task Load Index* (NASA-TLX ; Hart et Staveland, 1988) est l'outil d'évaluation subjective de la charge mentale le plus couramment utilisé dans les études expérimentales. Selon Hancock et al. (2021), l'échelle unidimensionnelle de Cooper et Harper (1969) a contribué au développement du questionnaire NASA-TLX. Plus tard, Hart et Staveland (1988) se sont appuyés sur cette échelle unidimensionnelle et un cadre conceptuel multidimensionnel (Cf. section 1.2.2 *Approches incluant les stratégies de l'individu*) pour mesurer la charge mentale dans le cadre de l'aviation (Hart, 2006). En effet, le questionnaire NASA-TLX a été mis au point par le *Human Performance Group du Ames Research Center* de la NASA et testé initialement dans plus de quarante simulations en laboratoire. À la suite d'une vingtaine d'études scientifiques, Hart et Staveland (1988) ont identifié dix, puis six facteurs pouvant être pris en compte pour déterminer la charge mentale ressentie. Il s'agit de trois dimensions associées aux exigences de la tâche (les exigences mentales, les exigences physiques, la pression temporelle), deux dimensions associées aux stratégies de l'individu (performance, effort) et une dimension concernant l'état émotionnel de l'opérateur (frustration). Les questions associées à chacune des dimensions se retrouvent en Annexe 1 du manuscrit. Après chaque étape, niveau ou fin d'une activité, le participant note chacune des dimensions sur une échelle de Likert de 0 à 100, score désignant un niveau croissant d'intensité de la dimension citée. Cependant, en fonction de l'activité, certaines dimensions peuvent être sollicitées plus fortement que d'autres. Ainsi, Hart et Staveland (1988), proposent une procédure de comparaison des dimensions par paires. Par conséquent, dans un second temps, les participants doivent sélectionner la dimension

(présentée par paires) qui a le plus contribué à la charge globale. Cette deuxième étape permet de pondérer chaque dimension de ce concept. La somme totale des scores est ensuite moyennée pour obtenir une évaluation globale en la divisant par 15 (ce nombre étant la somme des poids attribués au préalable aux critères). La formule de la charge mentale (CM) issue du NASA-TLX est présentée dans l'équation ci-après où d_i représente la note fournie par un individu après l'exécution d'une tâche, w_i est le poids associé à cette dimension :

$$NASA_{CM} = \sum_{i=1}^6 (d_i \times w_i) \frac{1}{15}$$

Étant donné que cette seconde étape ajoute du temps dans la passation du questionnaire, un score global non pondéré peut être utilisé en calculant la moyenne des six dimensions pour obtenir un indice brut de charge mentale (Raw Task Load Index - RTLX). Cette méthode a été validée par Byers et al. (1989). Trente ans plus tard, Cegarra et Morgado (2009) ont démontré que la version française conserve une forte corrélation entre le score pondéré (TLX) et le score non pondéré (RTLX). Cependant, dans près de trente études comparant le RTLX au TLX, Hart (2006) a constaté que la pondération améliore la sensibilité, la fiabilité inter-évaluateurs et la reproductibilité.

2.1.2.2 Workload Profile

Les auteurs du *Workload Profile* (WP ; Tsang et Velazquez, 1996) font l'hypothèse que les dimensions de la charge mentale peuvent être définies par les réservoirs de ressources attentionnelles proposées dans le modèle des ressources multiples (Cf. section 1.1.3 *Approches centrées sur les capacités limitées de l'individu*) de Wickens (1984, 1987, 2002, 2008). Ainsi, un des avantages majeurs à utiliser le WP par rapport aux autres questionnaires multidimensionnels est qu'il repose sur des bases théoriques qui ont été soumises à de nombreux tests empiriques (Tsang et Velazquez, 1996). Les prédictions de la charge mentale fondées sur la théorie peuvent donc être faites a priori et testées a posteriori. Autrement dit, la capacité de diagnostic des dimensions du WP peut être jugée par rapport aux prédictions théoriques du modèle des ressources multiples. De plus, comparé aux autres échelles, le WP permet de fournir des informations sur la nature des exigences de la tâche. En effet, à travers ce questionnaire, nous pouvons savoir si la tâche sollicite davantage le traitement verbal ou le traitement auditif de l'information.

Concrètement, les dimensions de la charge mentale considérées dans le WP sont les suivantes : le traitement central (correspondant à des activités telles que la mémorisation ou la prise de décision), la sélection et l'exécution de la réponse appropriée (par exemple le choix entre une réponse vocale ou motrice), le traitement spatial, le traitement verbal, le traitement visuel, le traitement auditif, la sortie manuelle et la sortie vocale (Rubio et al., 2004). Pour chaque tâche, les participants doivent indiquer sur une échelle de Likert un nombre entre 0 (aucune demande) et 100 (demande maximale) qui représente la proportion de ressources attentionnelles utilisées dans chacune des huit dimensions suivantes de la charge mentale : le traitement central (WP1), la sélection et exécution des réponses (WP2), traitement spatial (WP3), traitement verbal (WP4), le traitement visuel (WP5), le traitement auditif (WP6), la production manuelle (WP7) et la production verbale (WP8). Les questions associées à chacune des dimensions se retrouvent en Annexe 2 du manuscrit.

Un chiffre de "0" signifie que la tâche n'exige pas de ressources dans la dimension évaluée ; un chiffre de "1" signifie que la tâche exige une attention maximale. L'indice de charge mentale (CM) final est une somme des huit dimensions où d_i représente la note fournie par un individu après l'exécution d'une tâche comme indiqué dans la formule suivante :

$$WP_{CM} = \sum_{i=1}^8 d_i$$

L'article principal du questionnaire (Tsang et Velazquez, 1996, p. 362) précise que ce questionnaire peut être évalué de manière unidimensionnelle. Ainsi, en fonction de l'activité, nous pouvons en amont sélectionner les dimensions du WP pertinentes.

2.1.2.3 SWAT

Le *Subjective Workload Assessment Technique* (SWAT ; Reid et Nygren, 1988) a été proposé à la suite d'une revue de la littérature sur la charge mentale où trois dimensions ont émergé : la pression temporelle (*Time Load*), la charge d'effort mental (*Mental Effort Load*) et la charge de stress psychologique (*Psychological Stress Load*). La *pression temporelle* est définie comme le temps disponible. La *charge d'effort mental* rend compte des capacités limitées de l'être humain (Cf. sections 1.1.3 *Approches centrées sur les capacités limitées de l'individu* et 1.2.1 *Approches centrées sur l'effort mental*

fourni par l'individu). Enfin, la dimension *stress psychologique* englobe d'autres variables liées à l'opérateur, telles que la motivation, l'entraînement, la fatigue, la santé et l'état émotionnel. Par ailleurs, selon Reid et Nygren (1988), pour ces trois dimensions, la charge mentale est composée de trois niveaux (faible, moyen et élevé).

De plus, lors de l'utilisation de l'échelle du SWAT, les participants doivent effectuer une procédure en amont de l'activité correspondant à un tri de cartes suivie d'une procédure de notation de la tâche. La première étape de l'analyse des cartes triées consiste à déterminer le niveau de concordance entre les participants à l'aide du coefficient de concordance de Kendall (W). Puis, une analyse est effectuée pour obtenir une échelle de charge mentale allant de 0 à 100.

Cependant, le SWAT présente un inconvénient majeur, la procédure chronophage. Ainsi, au début des années 2000, Luximon et Goonetilleke (2001) ont proposé une version simplifiée, le *Simplified-Subjective Workload Assessment Technique* (S-SWAT). La procédure conventionnelle de pré-traitement peut être remplacée par une simple moyenne non pondérée pour obtenir une échelle de haute sensibilité.

2.1.2.4 Comparaisons entre les mesures subjectives

Un des avantages des mesures multidimensionnelles de la charge mentale, par rapport aux mesures unidimensionnelles, est la plus grande diagnosticité qu'elles peuvent atteindre (Wierwille et Eggemeier, 1993).

Concernant la comparaison entre les mesures multidimensionnelles, Miyake et al. (2000) précisent que le NASA-TLX et le SWAT sont relativement similaires et plus particulièrement les dimensions charge temporelle (du SWAT) et demande temporelle (du NASA-TLX) ; charge d'effort mental (SWAT) et demande mentale (NASA-TLX) ; et la charge de stress psychologique (SWAT) et la Frustration (NASA-TLX). Par conséquent, avoir le NASA-TLX et le SWAT dans une même étude est redondant quant à l'information obtenue. Pour la dernière paire de comparaison stress psychologique-Frustration, les auteurs du SWAT eux-mêmes (Reid et Nygren, 1988, p.191) précisent que le stress psychologique peut être relié à « *la confusion, la frustration et/ou l'anxiété de l'opérateur* ».

Quelques années plus tard, Rubio et al. (2004) comparent le NASA-TLX, le WP et le SWAT. Le WP est préconisé pour une comparaison de la charge mentale entre plusieurs tâches de niveaux de difficulté différents. Le NASA-TLX, pour prédire la performance

d'un individu à une tâche. Le WP ou le SWAT sont plus indiqués pour analyser la demande cognitive ou les ressources attentionnelles allouées durant une tâche. De plus, l'étude a démontré la complémentarité entre les questionnaires du WP et du NASA-TLX, car ils compensent leurs limites respectives. Par exemple, le NASA-TLX n'est pas conçu pour évaluer le niveau de charge mentale dans un système multimodal, contrairement au questionnaire WP qui le réalise. Par ailleurs, le NASA-TLX est plus étroitement lié à la façon dont une tâche est mise en œuvre dans un contexte donné, via les trois dimensions associées à l'activité (les exigences mentales, les exigences physiques, la pression temporelle), contrairement au WP.

Enfin, selon le modèle d'Hart et Staveland (1988), la charge mentale imposée a un impact sur la perception de l'opérateur sur sa propre performance. Ainsi, le cinquième item du questionnaire NASA-TLX interroge sur la performance perçue. Le questionnaire NASA-TLX est donc plus sensible à la prédiction de la performance perçue. Concernant le WP, Rubio et al. (2004) précisent qu'il était le seul questionnaire à révéler des différences dues à la complexité des tâches. Le WP est alors, parmi les trois instruments, celui qui est le plus sensible. L'étude montre également le haut pouvoir diagnostique du WP. En effet, le questionnaire WP étant basé sur la théorie des ressources multiples (Wickens, 1984, 1987, 2002, 2008), il a permis de détecter précisément les différences dans le type de ressources attentionnelles demandées pour une activité donnée. Plus récemment, Hancock et al. (2021) précisent qu'il y a une prédominance de l'effort dans les évaluations subjectives de la charge mentale. Ainsi, il y a une possibilité d'utiliser les échelles unidimensionnelles questionnant l'effort mental (telles que le *Rating Scale Mental Effort* (RSME) de Zijlstra, 1993), au lieu d'échelles multidimensionnelles, plus chronophages dans la passation expérimentale.

2.1.3 Conclusion sur les mesures subjectives

Pour conclure, qu'elles soient unidimensionnelles ou multidimensionnelles, les mesures subjectives sont les plus proches de l'essence de la charge mentale, car elles constituent l'indicateur le plus valide et le plus sensible. De plus, elles constituent la seule source d'information sur l'impact subjectif d'une activité tout en intégrant les effets de nombreux facteurs contribuant à la charge mentale (stress, motivation, facteurs environnementaux, etc.). Cependant, l'utilisation de jugements intellectualisés à travers des questionnaires pose question de leur pertinence. En effet, l'obligation formelle de

fournir une note encourage les participants à exprimer leurs jugements dans un format standardisé et à adopter les critères d'évaluation imposés par l'expérimentateur (Piéron, 1959 ; Hart et Staveland, 1988). De plus, les mesures subjectives sont dépendantes des facteurs individuels de la charge mentale (Boyer, 2019) où les individus peuvent surestimer ou sous-estimer leur niveau de charge mentale en fonction par exemple de l'état motivationnel et de la personnalité de l'individu.

Outre ce problème, le principal inconvénient des mesures subjectives est qu'elles ne peuvent pas être utilisées pendant l'activité et ne sont donc pas sensibles aux changements d'état de charge mentale au cours de la tâche. De plus, des auteurs (Sperandio, 1988 ; Cain, 2007) soulignent qu'elles peuvent entraîner des biais, comme des biais cognitifs définis comme étant des « *erreurs mentales causées par des stratégies de traitement simplifié de l'information* » (Heuer, 1999). Par exemple, les individus peuvent ne plus se souvenir des détails intrinsèques de l'activité après l'avoir réalisée. Ainsi, selon Hart et Staveland (1988), seules les informations les plus récentes sont directement accessibles, car l'expérience de chaque moment est remplacée par celle du moment suivant. Les évaluations subjectives n'incluent donc pas nécessairement toutes les informations pertinentes et peuvent inclure des informations non pertinentes. Ce type de mesure est donc vulnérable aux effets de récence (Hancock et al., 2021). Ces effets peuvent être accentués pour certaines populations, telles que les personnes âgées ou atteintes de maladies (Leduc, 2019). Pour pallier la limite du *non-temps réel* des mesures subjectives, une mesure unidimensionnelle peut être proposée à chaque essai ou étape de l'activité en plus d'une échelle multidimensionnelle a posteriori de l'activité. Autrement dit, au lieu de proposer les questionnaires multidimensionnels en fin d'activité comme c'est couramment le cas, nous pouvons en amont identifier si l'activité est décomposable en différentes étapes où l'on pourrait, insérer à chacune de ces discontinuités un questionnaire unidimensionnel de la charge mentale. Cependant, cette possibilité dépend de l'activité. Elle ne peut se faire pour des activités continues où il n'y a pas de temps d'arrêt comme les tâches cognitives ayant un défilement continu d'items. De plus, en procédant ainsi, on ajoute une tâche supplémentaire à la tâche en cours de réalisation, ce qui augmente artificiellement le niveau de charge expérimentée par les individus. Par ailleurs, Schmeck et al. (2014) ont observé une différence significative des performances entre les participants remplissant un questionnaire en cours de réalisation ou a posteriori de l'activité.

Pour dépasser cette limite, Tattersall et Foord (1996) ont proposé l'ISA (Cf. *section 2.1.1 Échelles unidimensionnelles*) mise au point pour fournir une évaluation subjective de la charge mentale durant l'activité. Néanmoins, cette technique s'est avérée intrusive, car elle perturbait les performances. Il semble ainsi difficile d'envisager des mesures subjectives de la charge mentale pendant l'activité. Cependant, une autre solution envisageable, consiste à ne pas les utiliser seules, mais de les coupler à un autre indicateur de la charge mentale directement disponible durant l'activité, les performances.

2.2 Performances

La charge mentale peut être mesurée à travers les performances qui peuvent être appréhendées en deux catégories (Miller, 2001 ; Cain, 2007 ; Butmee et al., 2018) : les mesures directes (nommées également “mesures primaires”) et les mesures indirectes de performance (“mesures secondaires”).



2.2.1 Évaluation directe de la performance

L'évaluation directe de la performance est également nommée performance à une tâche primaire. Cette dernière signifie que l'opérateur réalise une tâche unique. Les mesures primaires pour évaluer la charge mentale supposent que les niveaux de charge mentale suivent les niveaux de performances (Cuvelier, 2012 ; Mandrick, 2013 ; Hancock et al., 2021) comme le propose le modèle conceptuel de Meister (1976 ; Cf. *section 1.1.2 Approche centrée sur les performances*). De plus, pour une activité donnée, il peut y avoir plusieurs dimensions de performances. La vitesse, la précision, les temps de réaction et les taux d'erreurs sont les plus couramment utilisés lors des tâches (Cain, 2007). Ce type de mesure représente un indicateur direct, non intrusif et il n'existe aucun obstacle réel à leur évaluation (Hancock et al., 2021). Néanmoins, prendre en compte les performances est plus ou moins accessible en fonction de l'activité (tâches cognitives VS course à pied ; Hancock et al., 2021). En outre, il n'y a pas forcément d'indicateurs de performances comparables et facilement transférables d'une tâche à une autre (Sirevaag et al., 1993 ; Raufaste et al. 2004).

Par ailleurs, si on se réfère aux modèles des ressources attentionnelles de la charge mentale (Cf. section 1.1.3 *Approches centrées sur les capacités limitées de l'individu*), une augmentation de la charge mentale n'entraîne pas nécessairement une diminution des performances si cette augmentation reste dans les limites de la capacité des ressources mentales de l'individu. Ainsi, les performances peuvent être dégradées si les exigences de la tâche commencent à dépasser cette capacité de l'individu. Mais les opérateurs peuvent ajuster leur stratégie pour éviter d'atteindre la limite supérieure des capacités de réserve (Wickens, 2015 ; Young et al., 2015). Par conséquent, pour Cain (2007), il est difficile d'évaluer le coût de l'exécution de la tâche primaire par les seules mesures de performance en raison des changements dans la "*réaffectation stratégique de la capacité mentale*". Ainsi, les performances à une tâche primaire sont particulièrement utiles lorsque les exigences excèdent la capacité de l'opérateur de sorte que la performance se dégrade par rapport aux niveaux de base (Cain, 2007). Autrement dit, dans les situations à faible exigence, les performances à la tâche primaire ne sont pas sensibles pour mesurer la charge mentale (Butmee et al., 2018).

De plus, le critère de validité de la mesure est mis à mal, lorsque l'on considère l'ensemble des facteurs susceptibles d'affecter les performances à une tâche. En effet, ces dernières peuvent être impactées par d'autres facteurs, comme le stress ou la fatigue (Wickens et al., 2015). De plus, selon Butmee et al. (2018), une autre limite des mesures de performances d'une tâche primaire, est qu'elles ne permettent pas d'identifier l'origine, autrement dit la source des modulations de la charge mentale. On est donc face à un problème lié au critère de diagnostic.

Les mesures de performance à la tâche primaire, bien qu'ayant de nombreuses limites, sont néanmoins pertinentes, en raison de la multiplicité des dimensions pouvant être considérées, permettant ainsi d'obtenir un critère de haute fiabilité (capacité à détecter les modulations de la charge mentale). De plus, parmi les critères de mesure de la charge mentale, la performance satisfait le critère d'équipement (la mesure d'évaluation doit nécessiter un équipement minimal). Prendre en compte les performances à la tâche primaire pour mesurer la charge mentale est donc tout à fait justifiée à condition qu'elles ne soient pas utilisées seules.

En résumé, la relation entre la charge mentale et la performance n'est pas linéaire. En effet, le niveau de performance ne fournit pas d'information sur les coûts impliqués dans

la réponse adaptative aux exigences d'une situation. Ainsi, la performance est peut-être insuffisante en soi pour évaluer la charge mentale, mais nécessaire (Wilson, 2004 ; Cain, 2007).

2.2.2 Évaluation indirecte de la performance

L'évaluation indirecte de la performance est également nommée, performance à une tâche secondaire (ou tâche ajoutée selon Leduc, 2019). Ce type de mesure fournit un indice de la capacité restante de l'individu lors de l'exécution des tâches primaires (Cain, 2007). Pour cela, les performances à une tâche secondaire sont couramment considérées lors d'un paradigme de double tâche, où l'on demande à l'individu d'effectuer une deuxième tâche exigeante sur le plan cognitif, parallèlement à une tâche principale (Marchand et al., 2021). Les performances à la tâche secondaire sont donc des mesures supplémentaires aux performances à la tâche primaire.

L'origine de cette méthode repose sur la mesure de la différence entre la capacité mentale consommée par la tâche primaire et la capacité totale disponible (Mulder, 1979). Il y a donc un lien avec la théorie des ressources multiples de Wickens (1984, 1987, 2002, 2008 ; Cf. section *1.1.3 Approches centrées sur les capacités limitées de l'individu*). En effet, l'exécution de la tâche primaire nécessite un certain nombre de ressources mentales. Les ressources restantes sont donc utilisées pour exécuter la tâche secondaire. Ainsi, le paradigme de la double tâche peut être utilisé pour déterminer s'il existe des capacités mentales inutilisées (Miller, 2001).

Dans cette méthode, les performances de la tâche secondaire sont prises en compte pour déduire la charge mentale requise par la tâche primaire. Par conséquent, lorsque la charge mentale de la tâche primaire augmente, la capacité disponible est en partie utilisée, et la performance de la tâche secondaire diminue, comme représentée sur la **Figure 2**. Sur cette dernière, deux doubles tâches sont représentées l'une à côté de l'autre et toutes deux se situent en-dessous des limites de la capacité totale de l'individu (100 % sur l'axe des ordonnées). Celle de gauche a une faible demande à la tâche primaire et donc des performances élevées à la tâche secondaire. Celle de droite a une forte demande à une tâche primaire et donc des performances faibles à la tâche secondaire (Cf. **Figure 2**). Par conséquent, selon Hancock et al. (2021), la performance de la tâche secondaire reflète les différences de capacité disponible de l'individu.

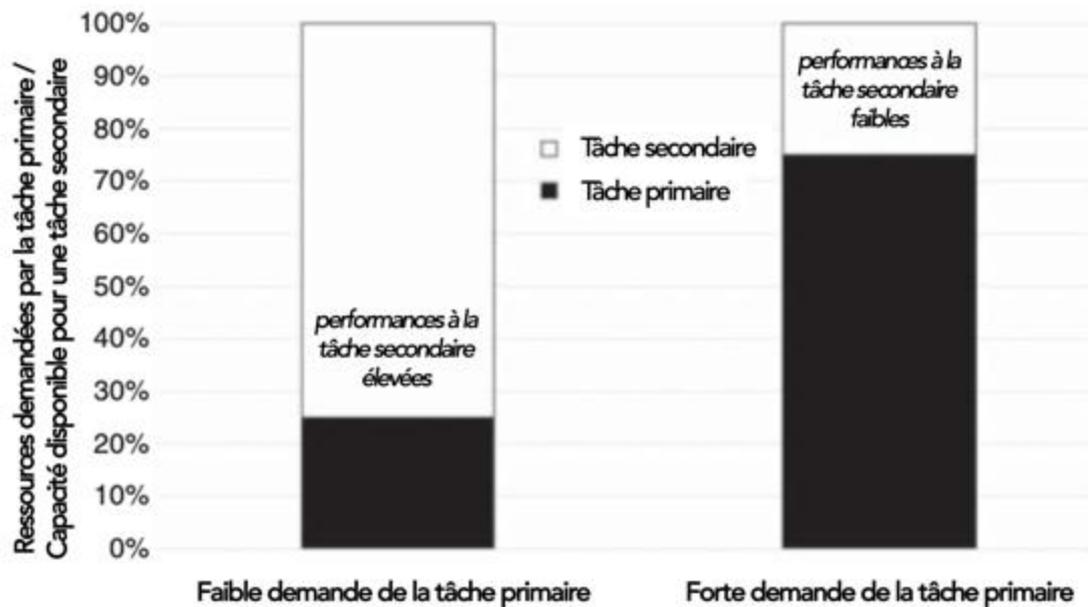


Figure 2 : Représentation du paradigme de la double tâche selon Hancock et al., 2021 (traduction française dans le cadre de cette thèse). L'ombrage noir représente les ressources demandées par la tâche primaire ; l'ombrage blanc représente la capacité disponible pour une tâche secondaire.

Pour que cette technique soit valable, il est impératif que les tâches primaires et secondaires utilisent les mêmes ressources (par exemple, une tâche primaire visuelle avec une tâche secondaire visuelle). Cela permet de s'assurer que la technique mesure réellement la capacité disponible (et donc la même dimension de la charge mentale) plutôt qu'un ensemble distinct de ressources (Hancock et al., 2021).

Un des inconvénients de l'évaluation indirecte de la performance est l'intrusivité qui a été soulignée par plusieurs auteurs (Miller, 2001 ; Cain, 2007 ; Butmee et al., 2018 ; Hancock et al., 2021 ; Longo et al. 2022). Ce caractère intrusif de deux tâches qui se disputent les mêmes ressources attentionnelles soulève un problème d'interférence. En effet, l'hypothèse de base du paradigme de la double tâche est que seule la capacité disponible est dirigée vers la tâche secondaire. Bien qu'il soit prouvé que la priorisation intentionnelle de la tâche primaire peut atténuer l'interférence, Kantowitz (2000 ; cité par Hancock et al., 2021) précise que l'hypothèse de l'absence d'interférence avec la tâche primaire ne peut être généralisable à tous les contextes. Par ailleurs, malgré les consignes mettant l'accent sur la nécessité de maintenir la priorité sur la tâche principale,

il a été démontré (Young et Stanton, 2007) que l'utilisation d'une tâche secondaire peut avoir des effets délétères sur la performance à la tâche principale.

Selon Hancock et al. (2021), une alternative possible au problème de l'interférence d'une tâche secondaire semble résider dans l'utilisation d'une modalité d'entrée différente ou de modes de réponse alternatifs. En partant de l'hypothèse que la capacité attentionnelle soit, en réalité, un mélange hybride de ressources unitaires et multiples, de sorte que les dimensions distinctes du modèle des ressources multiples de Wickens (1984, 1987, 2002, 2008) ne seraient pas exclusivement séparées. Au contraire, les ressources seraient alimentées par un réservoir commun (Hancock et al., 2021), comme proposé par les modèles à réservoir unique de Moray (1967), Kahneman (1973), Young et Stanton (2002) ou encore Wickens et al. (2015). Autre alternative, considérer une tâche secondaire qui est intégrée dans l'environnement global de la tâche (Schlegel, 1993). Une tâche intégrée dans l'environnement fait partie de la tâche primaire, mais n'est pas essentielle à sa réalisation comme lire des informations sur un tableau de bord pendant la conduite automobile. On peut supposer que l'intrusivité indésirable pose moins de problèmes lorsque la tâche secondaire est une tâche naturelle et intégrée qui fait partie de la routine de l'opérateur.

Enfin, Hancock et al. (2021) soulèvent une dernière limite au paradigme de la double tâche pour la mesure de la charge mentale, liée à la sensibilité, car ce paradigme n'est efficace que pour évaluer des changements aigus de la difficulté. Dans les cas où plusieurs mesures de tâches primaires sont accessibles, celles-ci peuvent être plus sensibles qu'une seule tâche secondaire (Wierwille et Gutmann, 1978). De plus, la tâche secondaire peut contribuer à la charge mentale indépendamment des exigences de la tâche primaire (Hancock et al., 2021). Il est donc difficile d'étudier la sous-charge mentale en utilisant une tâche secondaire.

Pour toutes ces raisons, Cain (2007) préconise de choisir une tâche secondaire qui n'interfère pas avec la tâche primaire (consommer des ressources similaires, mais ne pas interagir avec la tâche primaire) et qui est facile à apprendre. Mais l'objectif de non-interaction n'est pas chose aisée dans les situations expérimentales et encore moins de la vie réelle.

2.2.3 Conclusion sur les mesures de performances

En comparant les deux catégories de performances, nous observons une plus grande validité et sensibilité dans la mesure de la charge mentale pour les performances à une tâche secondaire par rapport à une tâche primaire. Cependant, cette dernière est plus facile à mesurer que la performance à une tâche secondaire, et sa précision a été étudiée de manière plus approfondie (Miller, 2001).

Dans tous les cas, il apparaît que la performance est un indicateur objectif de la charge mentale, mais peut être insuffisante, car elle n'est pas systématiquement représentative de ce concept. En fonction de différents facteurs internes et externes, les individus peuvent varier leurs efforts et subir davantage de charge mentale pour maintenir un niveau de performance constant (Meijman et O'Hanlon, 1984). En ce sens, les mesures de performance peuvent ne pas être à elles seules, suffisamment sensibles aux fluctuations et aux différentes dimensions de la charge mentale, alors que les évaluations subjectives et d'autres indicateurs comme, les indicateurs physiologiques peuvent l'être (Radüntz, 2017).

2.3 Mesures physiologiques

La troisième grande catégorie de mesure de la charge mentale concerne les mesures physiologiques. Ce terme fait référence à la réponse physiologique lors d'événements psychologiques. L'utilisation de ces méthodes pour la mesure de la charge mentale est basée sur l'idée qu'une augmentation des exigences mentales entraîne une augmentation de la réponse



physiologique (Moray, 1979 ; Miller, 2001 ; Butmee et al. 2018 ; Hancock et al., 2021). Par ailleurs, la plupart des mesures physiologiques sont fondées sur le modèle à ressource unique de la charge mentale, qui stipule que chaque individu a une capacité de traitement limitée, les mécanismes cognitifs nécessaires à l'exécution des tâches étant considérés comme un seul ensemble de ressources (Moray, 1967 ; Cf. *section 1.1.3 Approches centrées sur les capacités limitées de l'individu*). Lorsque les exigences de la tâche augmentent, le système nerveux central augmente l'offre de ressources. Les

mesures physiologiques reposent sur le concept selon lequel cette activité générale ou ses manifestations peuvent être mesurées (Moray, 1967).

Par ailleurs, ces mesures constituent un type d'indice naturel de la charge mentale, puisque l'activité mentale exige une activité physiologique (Young et al., 2015). Sur la base de la catégorisation du système nerveux, les mesures psychophysiologiques peuvent être divisées en deux catégories (Mallat, 2019 ; Hancock et al., 2021) : les mesures périphériques (liées au système nerveux périphérique) et les mesures centrales (liées au système nerveux central).

2.3.1 Les mesures périphériques

Les mesures périphériques évaluent l'activité nerveuse périphérique. Concernant l'évaluation de la charge mentale, les mesures périphériques peuvent se décliner en cinq catégories : les mesures cardiaques, les mesures respiratoires, les mesures oculaires, les mesures musculaires et les mesures de la peau.

Les mesures cardiaques

Nous commençons par les mesures cardiaques, car ce sont les mesures physiologiques les plus courantes pour mesurer la charge mentale (Miller, 2001 ; Charles et Nixon, 2019). Les techniques liées à l'activité cardiaque mesurent l'activité électrique du cœur, d'où la dénomination ECG pour ÉlectroCardioGraphie. Selon Miller (2001), trois méthodes existent pour mesurer la charge mentale via l'activité cardiaque : la fréquence cardiaque, la variabilité de la fréquence cardiaque et la pression artérielle. La fréquence cardiaque (FC) est généralement appréhendée par le nombre de battements dans une période de temps. La charge mentale augmente avec la FC (Wilson, 1993). Concernant la variabilité de la fréquence cardiaque (VFC), elle évalue les intervalles entre les battements du cœur au cours du temps. Plus la charge mentale augmente, plus la VFC diminue (Delliaux et al., 2019 ; Hancock et al., 2021). Enfin, concernant la pression artérielle, elle augmente avec la charge mentale (Veltman et Gaillard, 1996 ; Stuijver et al., 2014 ; Hancock et al. 2021). Par ailleurs, quelle que soit la manière de mesurer l'activité cardiaque, il est important de distinguer la charge physique de la charge mentale, car une augmentation de la première entraînera également une augmentation de la fréquence cardiaque (Jorna, 1993), de même que pour les mesures respiratoires (Roscoe 1992 ; Cain, 2007 ; Charles et Nixon, 2019 ; Marchand, 2021).

Les mesures respiratoires

Concernant la charge mentale, la fréquence respiratoire, correspondant au nombre de respirations par période de temps (Miller, 2001 ; Longo et al., 2022), est la mesure la plus utilisée (Roscoe, 1992 ; Miller, 2001 ; Charles et Nixon, 2019) parmi les mesures respiratoires. Plus la charge mentale augmente, plus la fréquence respiratoire augmente (Miller, 2001 ; Lean et Shan, 2012 ; Charles et Nixon, 2019 ; Tao et al., 2019 ; Marchand, 2021). En effet, les exigences mentales accrues sont assouvies par l'augmentation de la fréquence et de la profondeur de la respiration (Roscoe, 1992). Enfin, pour toutes les mesures respiratoires, elles ne peuvent s'utiliser lors de tâches nécessitant la production de paroles, car cette dernière peut interrompre et modifier les schémas respiratoires (Roscoe 1992 ; Miller, 2001 ; Charles et Nixon, 2019).

Les mesures oculaires

Selon Zagermann et al. (2016), les mesures oculaires peuvent être distinguées en deux catégories, selon si les mouvements sont volontaires ou pas (Cf. **Figure 3**). Parmi les mouvements oculaires volontaires, une fixation est un état de concentration où l'œil reste immobile pendant un temps donné (entre 200-300 millisecondes à plusieurs secondes, Zagermann et al., 2016). Chen et al. (2011) précisent qu'une augmentation des fixations est un indicateur d'une augmentation de l'attention nécessaire à mesure que la complexité de l'activité augmente. Une autre mesure volontaire concerne les saccades correspondant au déplacement entre deux fixations. Plus la charge mentale augmente, plus les saccades s'allongent et s'intensifient en vitesse (Chen et al., 2011 ; Zagermann et al., 2016).

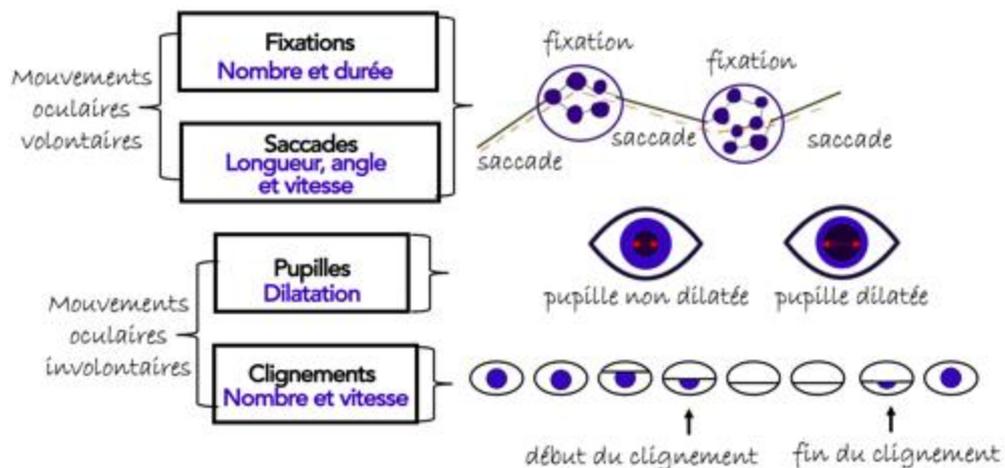


Figure 3 : Différents types de mesures oculaires pour la mesure de la charge mentale

Parmi les mouvements oculaires involontaires, plusieurs auteurs (May, 1990 ; Pflöging et al., 2016 ; Marinescu et al., 2018) précisent que le diamètre pupillaire augmente, autrement dit se dilate, lorsque la charge mentale augmente. Une autre mesure involontaire concerne les clignements des yeux qui correspondent aux moments où le diamètre de la pupille est bloqué par la paupière de près de la moitié, jusqu'à ce qu'il soit complètement fermé et rouvert au-dessus de la moitié (Cf. **Figure 3** ; Chen et al., 2011). Plus la charge mentale augmente, plus les clignements (nombre et vitesse) diminuent (Veltman et Gaillard, 1996 ; Chen et al., 2011). Pour conclure sur les mesures oculaires, si la tâche comporte principalement des stimuli visuels, c'est qu'elle sollicite d'une manière accrue l'attention visuelle, à ce moment-ci les mesures oculaires sont recommandées (Veltman et Gaillard, 1996 ; Miller, 2001). Enfin, bien qu'elles puissent être un bon indicateur de la charge mentale, la lumière, la qualité de l'air, ou la climatisation peuvent avoir des effets significatifs sur toutes les mesures oculaires rapportées (De Waard, 1996 ; Charles et Nixon, 2019).

Les mesures musculaires

Les mesures musculaires de la charge mentale se réalisent principalement à travers l'ÉlectroMyoGraphie (EMG) en supposant que la tension musculaire et la charge mentale soient liées (O'Donnell et Eggemeier, 1986). Lorsque la charge mentale augmente, l'amplitude de l'EMG augmente (Fallahi et al., 2016 ; Tao, 2019 ; Hancock et al., 2021). Différents muscles faciaux (tels que le frontalis et le corrugator) se sont avérés être sensibles aux changements de la charge mentale (De Waard, 1996). Cependant, les mesures musculaires sont intrusives et ne peuvent s'utiliser lors d'activité nécessitant une activité physique intense.

Les mesures de la peau

Concernant la charge mentale, l'Activité ÉlectroDermale (AED) a été la plus utilisée parmi les mesures de la peau (Charles et Nixon, 2019). Plus la charge mentale augmente, plus l'AED augmente (De Waard, 1996 ; Miller, 2001 ; Tao et al., 2019). Cependant, Hancock et al. (2021) pointent du doigt la résolution temporelle élevée de l'AED (de l'ordre de plusieurs secondes). De plus, cette mesure est sensible à la charge physique (Marchand et al., 2021). Pour toutes ces raisons, les mesures cutanées sont limitées dans

leur usage, sachant qu'elles sont sensibles à la température et à l'humidité de l'environnement, à l'âge, au sexe, ou encore à l'heure de la journée (Hancock et al., 2021).

2.3.2 Les mesures centrales

Les mesures physiologiques présentées jusqu'à présent fournissent une évaluation indirecte de la charge mentale. Effectivement, les mesures indirectes détaillées dans la section précédente sont toutes influencées par les signaux que le cerveau envoie lorsqu'il est soumis à différents niveaux de charge mentale (Miller, 2001). Pour une évaluation directe, il existe une possibilité : accéder au système nerveux central composé du cerveau et de la moelle épinière. Dans le cadre de la mesure de la charge mentale, seule l'activité du cerveau est étudiée. À ce jour, plusieurs techniques de mesure de l'activité cérébrale permettent de mesurer la charge mentale, à commencer par l'ÉlectroEncéphaloGraphie (EEG). Cette dernière est utilisée par la plupart des recherches dans le domaine, suivie par l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) et plus récemment, la Spectroscopie dans le Proche InfraRouge fonctionnelle (SPIRf).

La spectroscopie dans le proche infrarouge fonctionnelle

De l'anglais "*functional Near-Infrared Spectroscopy*", la SPIRf permet de mesurer le taux d'oxygénation de l'hémoglobine (on parle de l'oxy-Hb). Concernant la mesure de la charge mentale, lorsqu'elle augmente, l'oxy-Hb augmente et le désoxy-Hb diminue au niveau du cortex pré-frontal (Peck et al., 2014). Concernant les avantages de la SPIRf, le temps d'installation est court, elle peut être portable et est résistante à certains mouvements de l'utilisateur (Solovey et al., 2009 ; Peck et al., 2014 ; Aghajani et al. 2017). Par ailleurs, elle a une bonne résolution spatiale (5 mm à 1,5 cm), mais une résolution temporelle faible (Peck et al., 2014). Une limite qu'on retrouve également pour l'IRMf.

L'Imagerie par Résonance Magnétique fonctionnelle

L'Imagerie par Résonance Magnétique fonctionnelle (IRMf) et la SPIRf se basent sur le même principe technique, celui de la réponse hémodynamique. Elles partagent le même avantage lié à la résolution spatiale élevée (de l'ordre de 4 à 5 mm pour l'IRMf), mais également un inconvénient commun, celui de la résolution temporelle. Il faut un

délai de 1 à 5 secondes entre l'activation de la zone corticale et sa prise en compte par l'IRMf. Mais contrairement à la SPIRf, l'IRMf est une méthode intrusive, car le dispositif physique dans lequel le participant est positionné le contraint à rester allongé et immobile. De plus, l'IRMf est également coûteuse à l'achat et à l'entretien, nécessitant un personnel expert, des salles et un protocole spécifiques. Concernant la mesure de la charge mentale, plus cette dernière augmente, plus les signaux IRMf sont complexes sur le plan dynamique (Dhamala et al., 2002). Malgré la bonne résolution spatiale de l'IRMf, ses contraintes techniques ont abouti à un nombre d'études limité sur la charge mentale, contrairement à l'EEG qui est une mesure centrale phare pour évaluer ce concept (Barachant, 2013).

L'ÉlectroEncéphaloGraphie

L'ÉlectroEncéphaloGraphie (EEG) est une méthode qui enregistre, via des électrodes placées de manière précise sur le cuir chevelu, l'activité électrique du cerveau, principalement les neurones pyramidaux (Marchand et al., 2021).

Différents aspects des signaux EEG peuvent être évalués, notamment dans le domaine temporel et le domaine fréquentiel. Concernant ce dernier, l'EEG contient plusieurs bandes de fréquences allant de 0,1 à 60 Hz (Cain, 2007 ; Brookhuis et De Waard, 2010 ; Charles et Nixon, 2019 ; Hancock et al., 2021 ; Marchand et al., 2021 ; Longo et al., 2022). Il existe une classification de ces rythmes cérébraux selon leurs gammes de fréquence dont les plages associées à ces gammes peuvent varier en fonction des différentes normes et applications (Longo et al., 2022). Premièrement, il y a les ondes Delta comprises entre 0,1 à 5 Hz qui correspondent, entre autres, au sommeil profond sans rêve ; les ondes Thêta de 5 à 8 Hz correspondent à la somnolence, à la relaxation profonde, aux processus mnésiques et est liée à la régulation émotionnelle ; les ondes Alpha comprises entre 8 à 12 Hz correspondent à la détente tout en ayant un état d'attention ; les ondes Bêta de 12 à 39 Hz correspondent au traitement mental actif, à la réflexion, à la concentration ou à l'anxiété et les ondes Gamma au-delà des 40 Hz correspondent aux activités mentales intenses, telles que la résolution de problèmes et aux fonctions motrices. Concernant la charge mentale (Cf. **Figure 4**), il est convenu qu'une augmentation de la charge mentale induit une augmentation des ondes Thêta en zone frontale et une diminution des ondes Alpha en zone pariétale (Hankins et Wilson, 1998 ; Miller, 2001 ; Roy, 2015 ; Babiloni, 2019 ; Charles et Nixon, 2019 ; Tao et al.,

2019 ; Longo et al., 2022 ; Raufi et Longo, 2022). De plus, il se peut que les ondes Delta et Gamma augmentent avec la charge mentale (Tao et al., 2019).

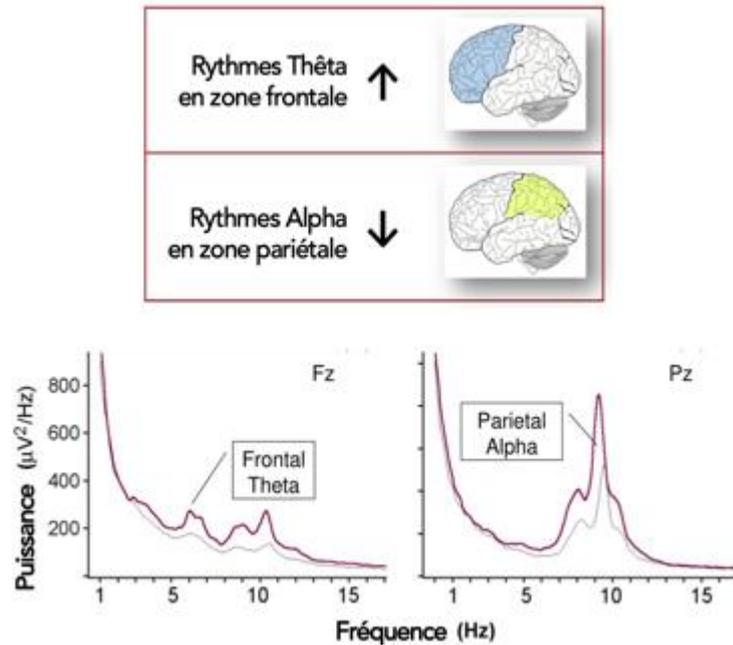


Figure 4 : Représentation schématique de l'augmentation des ondes ondes Thêta et de la diminution des ondes Alpha lors que la charge mentale augmente. Image issue de Babiloni (2019) pour celle du haut et de Trejo et al. (2007) pour celle du bas.

En plus de la mesure cérébrale à travers les bandes de fréquence (domaine fréquentiel), une deuxième approche de l'EEG, se fait dans le domaine temporel. Celle-ci est basée sur les "potentiels évoqués" liés aux événements (ERP-*Event Related Potential*), qui sont des informations au regard de la temporalité dans le signal EEG, synchronisés avec des stimuli (Hancock et al., 2021 ; Marchand et al., 2021). Concernant la mesure de la charge mentale à travers les mesures ERP, la composante P300 qui apparaît environ 300 millisecondes après un stimulus est la plus étudiée, car elle est une mesure fiable de la charge mentale (Charles et Nixon, 2019). Plusieurs auteurs (Hancock et al., 2021 ; Marchand et al., 2021) précisent que plus la charge mentale est élevée, plus l'amplitude de P300 est faible.

Pour finir, même si l'EEG a une résolution temporelle élevée (inférieure à 1 ms), sa résolution spatiale est faible (de l'ordre du centimètre), ce qui rend difficile la localisation de l'origine de l'activité neuronale (Peck et al., 2014). Autres inconvénients : l'existence de différences inter-individuelles (comme celles liées à l'âge ; McEvoy et

al., 2001), un temps de mise en place important, le besoin d'une expertise du personnel, les risques d'une activité musculaire pouvant perturber le signal EEG (Miller, 2001). Dans la dernière décennie, une piste a été explorée pour contrecarrer les limites de l'EEG, en couplant cette mesure à celle de la SPIRf. En effet, parmi toutes les mesures centrales de la charge mentale, la SPIRf est la plus robuste aux artefacts de mouvements et la plus portable (Peck et al., 2014). De plus, la SPIRf donne accès à des paramètres hémodynamiques qui ne sont pas accessibles avec l'EEG (qui est sensible aux signaux électriques et non à la réponse hémodynamique). Ces deux mesures sont donc complémentaires. Ainsi, la combinaison EEG/SPIRf a une précision significativement supérieure dans la détection des biomarqueurs de la charge mentale comparée à celle de l'EEG ou de la SPIRf prises individuellement (Aghajani et al., 2017).

2.3.3 Conclusion sur les mesures physiologiques

Les mesures physiologiques ont l'avantage d'être des mesures objectives de la charge mentale. De plus, elles sont quantitatives, continues et moins sujettes aux biais. Parmi les mesures physiologiques, les mesures centrales sont les plus sensibles à la charge mentale. En entonnoir, parmi toutes les mesures centrales, l'EEG est la plus efficace pour mesurer ce concept (Zhang et al., 2018), avec des perspectives encourageantes si on l'associe à la SPIRf. Cependant, si la tâche nécessite une activité ou un déplacement physique, les mesures cérébrales ne sont pas adaptées, contrairement à une mesure oculaire qui peut être portable. De même, si l'activité nécessite une prise de parole, les mesures respiratoires et cérébrales sont à écarter, car elles entrent en concurrence avec l'enregistrement. Par ailleurs, en se basant sur le contexte, si une mesure physiologique semble pertinente dans la mesure de la charge mentale, il est important de mettre les avantages de cette mesure en balance avec ses contraintes (comme le temps, les ressources, l'intrusivité, ou le rapport signal/bruit). Selon Miller (2001), si une mesure nécessite un équipement coûteux ou des experts pour superviser une partie du test, alors elle peut ne pas être rentable. En outre, les mesures physiologiques sont des indicateurs difficiles à interpréter, car elles ne sont pas spécifiques à la charge mentale et peuvent résulter de l'activation d'autres processus comme le stress (Katmah et al. 2021), l'attention ou la fatigue mentale (Tran et al., 2020). Il y a donc un problème de validité de la mesure. De plus, en raison de la grande variabilité de la manière et du degré de réaction du corps humain à la charge mentale, ainsi que des différences inter-

individuelles associées à ce processus, aucune mesure psychophysique prise indépendamment n'évalue de manière concluante la charge mentale (Charles et Nixon, 2019). Par conséquent, il est toujours recommandé d'incorporer des mesures subjectives et/ou de performances aux mesures physiologiques (Leduc, 2019 ; Hancock et al., 2021).

2.4 Conclusion sur l'ensemble des mesures

Pour conclure sur le chapitre 2 qui concerne l'ensemble des catégories de mesures de la charge mentale, elles ont toutes leurs avantages et leurs inconvénients. L'idéal étant de ne pas les utiliser isolément. Ainsi, pour évaluer la charge mentale avec le plus de justesse possible, la triangulation des trois catégories de mesures (subjectives, performances et psychophysiques) permet de dépasser les limites de chaque mesure (Charles et Nixon, 2019 ; Longo, 2022). Enfin, pour l'aide au choix de mesures, Miller (2001) a présenté un ensemble de critères de considération pour sélectionner les mesures de la charge mentale appropriées à travers un arbre de décision traduit en français et mis sous forme de tableau dans le cadre de cette thèse (Cf. **Table 1**).

Considérations	Type de données	Méthodes
Intervalle de collectes des données	En continu	Physiologiques
		Mesures des performances primaires
		Mesures des performances secondaires
	Pendant	Évaluation unidimensionnelle
	Après	Évaluation multidimensionnelle
Intrusivité	Intrusif	Mesures cérébrales
		Mesures respiratoires
		Mesures de performance indirecte
	Non intrusif	Mesures subjectives a posteriori
		Autres mesures physiologiques Mesures de performance directes (tâche primaire)
Forme de collecte des données	Verbal	Évaluation unidimensionnelle
	Écrit	Évaluation multidimensionnelle
	Numérique	Physiologiques
		Mesures des performances primaires Mesures des performances secondaires
Considération temporelle (uniquement pour les données subjectives)	Oui	Évaluation unidimensionnelle
	Non	Évaluation multidimensionnelle

Sensibilité	Haute	Évaluation unidimensionnelle
		Évaluation multidimensionnelle
		Mesures cérébrales
	Moyenne	Mesures de performances secondaires
		Mesures des performances primaires
		Mesures cardiaques
		Mesures oculaires
Faible	Autre mesures physiologiques	
Coût de l'implémentation	Haute	Mesures cérébrales
		Mesures respiratoires
		Mesures oculaires
	Moyenne	Évaluation multidimensionnelle
		La plupart des mesures des performances primaires
		La plupart des performances secondaires
		Autres mesures physiologiques
Faible	Évaluation unidimensionnelle	
Fiabilité	Haute	Évaluation unidimensionnelle
		Évaluation multidimensionnelle
		Mesures cérébrales
		Activité oculaire
	Moyenne	Mesures de performances secondaires
		Autre mesures physiologiques
	Faible	Mesures des performances primaires

Table 1 : Tableau pour le choix de la technique de mesure de la charge mentale (traduction française dans le cadre de cette thèse de l'arbre de décision de Miller, 2011).

Maintenant que nous avons une connaissance des différents types de mesures de la charge mentale, nous allons, dans le chapitre suivant, présenter les applications existantes ou manquantes, pour ensuite présenter dans le détail la problématique de la thèse.

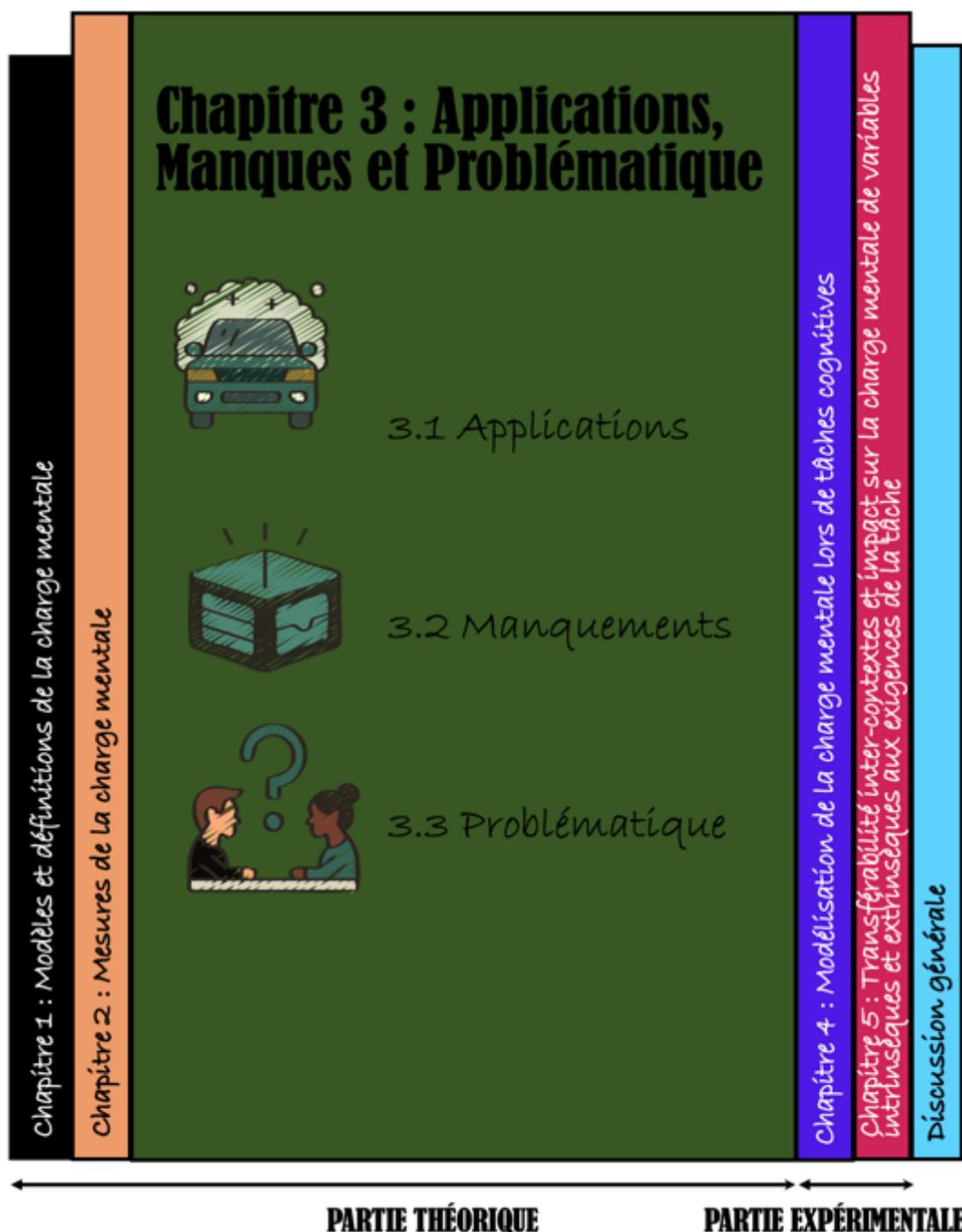


Figure 1 : Plan général du manuscrit et détail des sous-parties du chapitre 3

Comme précisé sur la **Figure 1**, après avoir décrit les modèles et les mesures de la charge mentale, ce troisième chapitre a pour objectif de présenter les applications existantes. Mais aussi de pointer les manques pour de potentiels usages basés sur la charge mentale, avant de clôturer le chapitre sur la problématique de la thèse.

3.1 Applications

Dans un monde de plus en plus technocentré qui impose des exigences mentales plus élevées, la charge mentale est devenue l'un des concepts les plus largement invoqués en ergonomie (Young et al., 2015). Dans ce contexte général, la charge mentale, en tant qu'indicateur du coût mental d'une activité, est souvent



étudiée lors de situations opérationnelles, telles que les transports ou le milieu médical (Young et al., 2015 ; Longo et al., 2022). Par ailleurs, Young et al. (2015) ont observé une évolution des domaines d'applications en fonction du temps. En effet, l'étude de la charge mentale dans l'ingénierie logicielle ou la Conception Assistée par Ordinateur (CAO) a eu un pic dans les années 80, avant de décroître à partir des années 90. En parallèle, c'est à partir des années 90, que les études sur la charge mentale appliquée aux transports (ferroviaire, automobile, aviation) se sont développées.

Parmi celles-ci, la conduite se distingue comme un domaine d'intérêt particulier, notamment à partir des années 2000. La conduite automobile est une activité complexe qui nécessite une bonne gestion des ressources attentionnelles, un contrôle cognitif efficace et des prises de décision (Gabaude et al., 2012). Ainsi, de nombreuses études sur la charge mentale ont eu lieu dans ce contexte (Recarte et Nunes, 2003 ; Young et Stanton, 2004 ; Gabaude et al., 2012 ; Borghini et al., 2014 ; Paxion et al., 2014 ; Milleville-Pennel et Charron, 2015 ; Foy et Chapman, 2018 ; Hidalgo-Muñoz et al., 2019). De plus, le secteur automobile propose de multiples innovations comme les systèmes d'assistance avancés, où l'impact sur la charge mentale nécessite d'être considéré (Ruscio et al., 2017). Par ailleurs, dans un contexte applicatif, la charge mentale est davantage étudiée comme un outil de compréhension d'une situation, que comme un concept à comprendre en lui-même. Ainsi, dans le secteur automobile, la charge mentale est essentiellement considérée comme un indicateur pour identifier les aspects qui perturbent la conduite, tels que l'ajout d'un système autonome (Ko et Ji, 2018 ; Kuribayashi et al., 2021 ; Figolová et al., 2022) ou l'impact de l'utilisation d'un téléphone portable (Törno et Bolling, 2006). Des recherches ont également étudié l'effet de l'expertise ou l'effet de l'âge sur la charge mentale, avec l'objectif d'identifier les médiateurs internes qui vont influencer la charge mentale dans un contexte donné,

l'automobile (Paxion et al., 2014 ; Milleville-Pennel et Marquez, 2020) ou l'aéronautique (Li et al., 2013 ; Rainieri et al., 2021).

Concernant le domaine aéronautique, la charge mentale a été étudiée à plusieurs reprises depuis les années 2000 (Young et al., 2015). Que ce soit la charge mentale mesurée lors du pilotage d'avion (Wilson, 2002 ; Wanyan et al., 2014 ; Schulte et al., 2015 ; Li et al., 2021) ou lors d'une tâche de contrôle du trafic aérien (Sperandio, 1971 ; Athènes et al., 2002 ; Averty et al., 2004 ; Mélan et Cascino, 2004 ; Loft et al., 2007 ; Djokic et al., 2010 ; Aricò et al., 2016). Dans ce dernier cas, connaître la charge mentale en cours d'activité est un enjeu important, car le contrôle du trafic aérien impose de multiples demandes simultanées, notamment la surveillance et l'intervention pour résoudre les conflits (Gopher et Donchin, 1986 ; Aricò et al. 2016). Prédire la charge mentale est donc un défi dans ce domaine, pour éviter les erreurs dues à des niveaux critiques de charge mentale. Par ailleurs, Loft et al. (2007) précisent la nécessité d'avoir de meilleurs modèles de la charge mentale des contrôleurs aériens basés, entre autres, sur les stratégies employées (comme la sélection des priorités, la gestion des ressources cognitives ou la régulation de leur propre performance). Cela, afin de pouvoir adapter la configuration du contrôle du trafic actuel, mais aussi créer des outils d'automatisation pour réduire la charge mentale des contrôleurs (Athènes et al., 2002).

Concernant le pilotage d'avions, c'est une tâche complexe qui fait appel à plusieurs capacités cognitives du pilote, telles que le traitement de flux d'informations en simultanée (Wilson, 2002 ; Li et al., 2021). Or, la charge mentale est l'un des principaux concepts qui émerge lors du traitement des tâches multiples (Li et al., 2021). Ainsi la mesurer durant l'activité du pilote d'avion (Wilson, 2002 ; Wanyan et al., 2014 ; Li et al., 2021) est importante, notamment pour améliorer les systèmes adaptatifs basés sur la charge mentale. Ces systèmes doivent adapter l'automatisation à la charge mentale, afin d'améliorer la performance du système. Selon Onken et Schulte (2010), un système adaptatif doit laisser l'opérateur humain effectuer son travail autant que possible sans intervenir. En cas de charge mentale critique, le système doit guider l'attention de l'humain vers la tâche objectivement la plus importante, ou transformer la situation en une situation que l'opérateur peut gérer. En dernier recours, le système adaptatif doit prendre en charge la tâche, si les coûts d'un échec, dû à une surcharge mentale, sont trop élevés. Cependant, l'introduction de systèmes automatiques dans les tâches de pilotage

d'avions, tels que le pilote automatique ou les systèmes de surveillance de vol peuvent à l'inverse réduire significativement la charge mentale des pilotes qui peuvent être en sous-charge mentale. Or, actuellement, la sous-charge mentale prédomine durant les phases de *croisière* et peut conduire les pilotes à ignorer certaines informations pertinentes en raison d'un manque d'engagement dans l'activité (Boyer, 2016). Étudier la charge mentale dans ce contexte permet donc d'éviter que les systèmes d'automatisation pour les pilotes accentuent les problématiques existantes.

En plus de ces thématiques phares du secteur automobile et de l'aéronautique, les intérêts se sont également tournés vers l'étude de la charge mentale lors de nos interactions avec nos portables ou nos ordinateurs (Iqbal et al., 2004 ; Törnros et Bolling, 2006), dans l'enseignement (Weinger et al., 2004), dans le secteur médical (Weinger et al., 2004 ; Byrne, 2011), dans les Interfaces Cerveau-Ordinateurs (Felton, et al., 2012 ; Aricò et al., 2016 ; Andreessen et al., 2021) ou encore dans le milieu militaire (Schulte et al., 2015). Par conséquent, la charge mentale a souvent été étudiée dans des cadres applicatifs définis ou lors d'usages réels avec une approche techno-centrée. De plus, pour tous ces domaines d'applications, la charge mentale est généralement mesurée durant l'activité ou en conséquence de l'activité. Il n'y a donc aucune certitude a priori de la valeur de la charge mentale dans laquelle se trouvera l'individu lorsqu'il effectuera la tâche.

3.2 Manques

Malgré la multiplicité des études sur la charge mentale, il subsiste des manques, dont trois principaux que nous avons identifiés et que nous présentons dans cette section. Le premier concerne le diagnostic des tests neuropsychologiques basés sur la sollicitation de fonctions cognitives (Cf. **Figure 2**). On parle également de tâches cognitives. Actuellement, l'analyse se base uniquement sur les performances pour établir des conclusions a posteriori de la réalisation de la tâche cognitive. Or, il serait pertinent d'analyser les performances et le coût mental généré par la réalisation de la tâche, à travers la mesure de la charge mentale. En effet, au cours d'une activité, un individu peut maintenir des performances élevées, mais elles peuvent conduire également à une surcharge mentale (Reid et Nygren, 1988 ; Raufaste et al., 2004 ; Cain,



2007 ; Radüntz, 2017). Ainsi, pour établir des conclusions qui ne soient pas biaisées par le maintien des performances au détriment de la charge mentale, il est important de pouvoir mesurer ces deux aspects. Par conséquent, lors de l'analyse des résultats aux tests neuropsychologiques, il y a un manque dans la possibilité de situer un individu sur la base de sa charge mentale et de sa performance par rapport à un groupe de référence partageant les mêmes caractéristiques démographiques et socio-professionnelles que lui (telles que l'âge, le sexe, le niveau scolaire ou le domaine professionnel). Un manque, car à notre connaissance, aucune tâche cognitive ne permet de le faire. La tâche de N-Back (nommée « NBT » dans le manuscrit de thèse) est réputée pour avoir des niveaux de complexité induisant des niveaux de charge mentale dans la zone intermédiaire de la charge mentale correspondant à une charge mentale intermédiaire faible, une charge mentale intermédiaire moyenne et une charge mentale intermédiaire élevée pour respectivement les niveaux 0-back, 1-back et 2-back (Arvaneh et al., 2015 ; Dimitrakopoulos et al., 2017 ; Ries et al., 2018). Cependant, aucune étude basée sur le NBT n'a contrôlé, ni étalonné de manière rigoureuse les résultats de charge mentale.

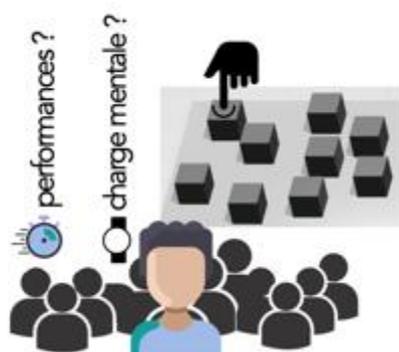


Figure 2 : Lors de la réalisation d'une tâche cognitive, pouvoir situer un individu par rapport à son groupe de référence sur la base de ses performances et de sa charge mentale.

Concernant le deuxième manque, il s'insère dans un contexte de recherche, où il pourrait être utile de maintenir une certaine charge mentale pendant une activité pour mesurer son impact sur différentes variables (comme le stress, le rythme cardiaque ou la présence en Réalité Virtuelle ; Cf. **Figure 3**). Ainsi, cela contribuerait à la compréhension de la charge mentale, de ses facteurs et de ses conséquences, afin d'améliorer les modélisations de ce concept.



Figure 3 : Connaître l'impact d'un niveau de charge mentale sur plusieurs variables dépendantes, telles que le stress ou le sentiment de présence.

Pour ce faire, nous devons moduler le niveau de complexité pour pouvoir maintenir constante la charge mentale souhaitée. L'adaptation du niveau de complexité en cours d'activité est primordiale, car l'habitué à la tâche, l'entraînement ou encore la fatigue peuvent influencer sur la charge mentale en l'augmentant ou en la diminuant.

Enfin, le troisième manque est de pouvoir mesurer la charge mentale en cours d'activité avec uniquement des données de performances. Actuellement, la mesure de la charge mentale se fait principalement a posteriori (avec des mesures subjectives) soit en cours d'activité avec des mesures physiologiques. L'utilité de mesurer la charge mentale durant l'activité est multiple. Un exemple est tiré du contexte des jeux vidéo, où l'enjeu est d'adapter pendant l'activité le niveau de complexité à la charge mentale pour pouvoir maximiser l'engagement du joueur (Ewing et al., 2016). D'autres situations, comme celle de la rééducation médicale, utilisent des Interfaces Cerveau-Ordinateurs (ICO) permettant de suivre l'activité cérébrale lors de la réalisation de tâches cognitives (Curran et Strokes, 2003 ; Roy, 2015 ; Carelli et al., 2017). Dans ce contexte, adapter le niveau de complexité de la rééducation au niveau de charge mentale souhaité permettrait une prise en charge médicale plus efficiente (Cf. **Figure 4**).



Figure 4 : Adaptation continue du niveau de complexité, afin d'obtenir une charge mentale optimale, dans la zone dite de confort.

Adapter le niveau de complexité en temps réel grâce à un indicateur physiologique (tel que l'EEG) est nommé boucle biocybernétique (Ewing et al., 2016). Une fois cette boucle construite, les exigences de la tâche peuvent être ajustées. Un processus qui implique l'exploration de l'interaction entre l'utilisateur et la réponse adaptative du jeu. Selon Ewing et al. (2016), le comportement de la boucle biocybernétique, l'interaction entre la psychophysiologie de l'utilisateur et le contrôle adaptatif constituent un objet d'étude en soi et n'est pas chose aisée. De plus, les ICO et les mesures physiologiques en général ont un coût (technique, expertise, intrusivité) que la mesure des performances n'a pas. Par conséquent, un des objectifs principaux de ce travail de thèse est de proposer une meilleure compréhension du lien entre la charge mentale subjective et la

performance pour pouvoir utiliser ces dernières dans la mesure de la charge mentale en cours d'activité.

Afin de répondre aux besoins énoncés dans les deux dernières applications (**Figures 3** et **4**), dans un premier temps, à travers les performances, il est nécessaire d'identifier l'évolution de la charge mentale, avec des transitions relativement proches entre les niveaux de charge mentale. L'objectif étant de pouvoir moduler les exigences de l'activité de manière subtile. Il serait donc judicieux d'étudier des niveaux "glissants" dans la plage intermédiaire de la charge mentale, car elle est essentielle dans de nombreux contextes, comme celui de la rééducation médicale où l'on ne cherche pas à sous-charger ou à surcharger le patient. Au contraire, des niveaux intermédiaires de charge mentale permettraient au patient d'être plus engagé dans l'activité, par rapport aux niveaux "extrêmes" de charge mentale. Néanmoins, de telles applications ne sont pas encore disponibles. En effet, il n'existe pas de tâche cognitive permettant d'obtenir avec certitude des intervalles de niveaux de charge mentale suffisamment distincts, mais également proches. Avec un défi : que ces niveaux de charge mentale soient directement prévisibles à partir du niveau de complexité et qu'un indicateur basé sur les performances soit disponible durant l'activité. Ainsi, l'objectif principal de notre recherche est de proposer plusieurs tâches candidates, qui ne dépendent pas d'une situation opérationnelle spécifique (pour être transférable et adaptable au plus grand nombre de contextes). Ces tâches doivent avoir des niveaux de complexité permettant d'induire des niveaux de charge mentale distincts qui se situent dans la zone intermédiaire.

3.3 Problématique

Plusieurs conditions sont nécessaires pour répondre aux manques évoqués précédemment. Comme vu à la section 3.2 *Manques*, la première condition consiste à avoir au moins une tâche avec différents niveaux de complexité. À noter que la complexité de la



tâche doit être considérée séparément de la difficulté de la tâche ; la première est considérée comme une propriété intrinsèque à la tâche, alors que la seconde est un phénomène perceptif (Longo et al., 2022). Pour mettre en place différentes tâches, nous considérons donc le niveau de complexité.

Deuxièmement, comme nous l'avons vu à la section 3.2, la tâche ne doit pas être rattachée à un contexte donné, c'est-à-dire qu'elle ne doit pas dépendre d'une situation opérationnelle spécifique (comme la conduite automobile), mais qu'elle doit refléter la sollicitation d'une ou de plusieurs fonctions cognitives de manière plus générique pour être transférable et adaptable au plus grand nombre de contextes. En effet, plusieurs auteurs (Berka et al., 2007 ; Radüntz, 2017 ; Guan et al., 2021) recommandent d'étudier la charge mentale à travers des tâches cognitives, car les situations réelles peuvent être décomposées en plusieurs fonctions cognitives. Ainsi, nous souhaitons avoir plusieurs tâches sollicitant chacune une fonction cognitive distincte. Actuellement, les études se basent principalement sur la tâche de N-back (nommée « NBT » dans le manuscrit de thèse) et en second lieu la tâche de Sternberg pour impacter la charge mentale lors de tâches cognitives. Cependant, ce sont deux tâches dont le matériel est un ensemble de lettres de l'alphabet latin à retenir, sollicitant ainsi un seul type de fonction cognitive, la mémoire de travail verbale, alors que nous aimerions disposer de plusieurs tâches sollicitant plusieurs fonctions cognitives distinctes. De plus, le NBT et la tâche de Sternberg sont socialement et culturellement marquées, car les analphabètes et les personnes utilisant d'autres alphabets (comme les caractères cyrilliques) en sont exclus. Par conséquent, notre objectif est d'avoir plusieurs tâches indépendantes d'un contexte et qui pourraient être utilisées par le plus grand nombre.

Par ailleurs, une question importante pour notre objectif concerne la mesure de la charge mentale. Comme nous l'avons vu au *Chapitre 2 - Mesures de la charge mentale*, différentes méthodes d'évaluation de la charge mentale existent et peuvent être divisées

en trois catégories : les mesures subjectives, les mesures de performance et les mesures psychophysiologiques. Les mesures les plus pertinentes sont les mesures subjectives, car elles sont directement basées sur les modèles explicatifs de la charge mentale. Parmi les trois questionnaires multidimensionnels reconnus pour mesurer la charge mentale, nous sélectionnons le WP et le NASA-TLX, car ils compensent les limites l'un de l'autre et sont donc complémentaires (Rubio et al., 2004 ; Paxion, 2014). En effet, ces questionnaires se basent sur différentes approches méthodologiques ou théoriques. Le WP s'appuie sur la *Multiple Resource Theory* (MRT) de Wickens (1984, 1987, 2002, 2008 ; Cf. section 1.1.3 *Approches centrées sur les capacités limitées de l'individu*). Le NASA-TLX, quant à lui, ne s'appuie pas directement sur un modèle théorique, mais sur une méthodologie (Cf. section 2.1.2 *Échelles multidimensionnelles*). Donc le WP s'intéresse à la saturation des multiples réservoirs de ressources attentionnelles, alors que le NASA-TLX s'intéresse aux variables externes (les trois premiers items) et internes (les trois derniers) à l'individu impactant la charge mentale.

Cependant, étant donné que les mesures subjectives peuvent conduire à des biais et ne sont pas disponibles durant l'activité, nous prenons en compte également les performances, comme seconde catégorie de mesure de la charge mentale. Ces dernières, bien que moins sensibles que les mesures subjectives, peuvent être pertinentes, du fait de leur multiplicité de variations possibles (temps de réalisation, erreurs, bonnes réponses, etc.), permettant ainsi d'obtenir un critère de fiabilité élevé (capacité à détecter les modulations de la charge mentale). Concernant les mesures physiologiques, une des plus efficaces pour mesurer la charge mentale, est l'EEG (Zhang et al., 2018). Or, cette mesure est une des plus bruitées (faible rapport signal sur bruit) à cause de la sensibilité élevée aux artefacts de mouvements et nécessite un temps de mise en place et une expertise importants. Ainsi, dans un premier temps, pour obtenir une bonne indication de la charge mentale durant l'activité, il semble pertinent de proposer des tâches avec suffisamment d'indicateurs de performance. Cela assurerait un modèle prédictif plus fiable de la charge mentale subjective, grâce à la richesse et à la diversité des dimensions de performance. De ce fait, nous pourrions déterminer si nous pouvons nous passer de données physiologiques, dont le rapport bénéfice/contrainte est peut être délétère.

Par ailleurs, pour étudier la possibilité de connaître a priori la valeur de la charge mentale dans laquelle se trouvera l'individu lorsqu'il effectuera la tâche, nous souhaitons

moduler un facteur intrinsèque aux exigences de la tâche, le niveau de complexité, pour étudier son impact sur la charge mentale (en s'appuyant sur diverses méthodes statistiques, telles que la classification supervisée). De plus, en nous basant sur le modèle proposé en section 1.2.3, nous souhaitons étudier le lien entre une variable interne à l'individu et la charge mentale. La variable interne a été identifiée lors de la section 3.1 *Applications*, et correspond à l'engagement du participant. Plusieurs modèles de l'engagement (Bucy et Gregson, 2001 ; Zichermann et Linder, 2013 ; Bonenfant et Philippette, 2018 ; Deci et Ryan, 2000) considèrent la motivation comme un point clé de ce concept. En outre, Bonenfant et Philippette (2018) proposent qu'il y ait un plus fort engagement dans une action qui présente des caractéristiques ludiques. Désormais, nous nous demandons si les caractéristiques ludiques, donc le design, sous-entend un ressenti ludique de l'individu. À savoir que le design d'une activité a souvent été étudié à travers la *gamification* correspondant à l'utilisation d'éléments de conception de jeu dans un contexte non ludique (Brigham, 2015). Par ailleurs, plusieurs auteurs (Auray, 2003 ; Brigham, 2015 ; Harwood et Garry, 2015 ; Looyestyn et al. 2017) précisent que l'utilisation de la *gamification* permet d'augmenter l'engagement. L'engagement pouvant être mesuré par des données comportementales, physiologiques, ou subjectives telles que le ressenti ludique, la motivation ou des échelles standardisées comme la *User Engagement Scale* (UES) short form (O'Brien et al., 2018). Ainsi, dans la volonté de comprendre davantage la charge mentale pour compléter les modèles actuels et d'identifier comment l'engagement à travers le ressenti ludique s'y intègre, il nous semble pertinent d'intégrer le ressenti *ludique* dans notre problématique de recherche.

D'autre part, il y a un défi lié aux modèles de classification de la charge mentale, celui d'avoir des modèles plus précis, plus fiables, plus génériques et donc des modèles inter-tâches. En conséquence, dans un premier temps, notre problématique de recherche est axée sur ces cinq objectifs qui seront opérationnalisés à la section 4.1.1 *Hypothèses opérationnelles de l'étude 1* :

- **Objectif 1** : Toutes les tâches cognitives développées pour nos recherches doivent fournir différents niveaux de charge mentale subjective par rapport aux niveaux de complexité.

- **Objectif 2** : Un modèle de classification (spécifique à chaque tâche) doit être capable de prédire la charge mentale subjective en se basant sur le niveau de complexité et la performance si cette dernière a suffisamment de variations possibles.
- **Objectif 3** : Un modèle de classification doit être capable de prédire la charge mentale subjective en ne considérant que la performance pendant la tâche si cette dernière a suffisamment de variations possibles.
- **Objectif 4** : Pour certaines tâches cognitives, un modèle de classification pour une tâche A doit prédire les niveaux de charge mentale d'une tâche B (modèle inter-tâches). Celles partageant une base commune (matériel commun par exemple) ou des types d'interactions communs (car cela permet d'avoir davantage de dimensions de performances similaires).
- **Objectif 5** : Nous souhaitons vérifier qu'il existe un lien entre la charge mentale et le ressenti ludique et en identifier la nature exacte.

Pour atteindre ces cinq objectifs, nous avons mis en place une **première étude** (présentée dans le Chapitre 4 de ce document). Cependant, lors de cette première étude, le ressenti ludique est uniquement mesuré. Ainsi, nous nous demandons s'il est possible d'augmenter le ressenti ludique d'une tâche en manipulant son design et notamment en passant d'un design classique à un design plus ludique. Autrement dit, le ressenti ludique d'une tâche peut-il être directement relié à son design ? Par ailleurs, qu'en est-il de l'impact d'un design ludique sur la charge mentale ? À la suite d'une revue de la littérature, nous n'avons identifié aucune étude qui analysait l'impact du design ludique sur la charge mentale lors de tâches cognitives.

Une **deuxième étude** est donc nécessaire. Ainsi, nous désirons étudier une variable extrinsèque aux exigences de la tâche, un design ludique (comparé à un design classique), pouvant exercer une influence sur l'engagement (via le ressenti ludique et également la motivation) et la charge mentale. De plus, nous souhaitons savoir si l'accentuation d'un critère du design ludique, comme la multiplicité de la couleur des items impacterait l'engagement (via le ressenti ludique et la motivation) et la charge mentale. En effet, nous souhaitons identifier les limites du design ludique sur une dimension donnée, la couleur. En nous demandant si la couleur des items permet un plus grand engagement dans l'activité sans induire une surcharge mentale liée aux ressources attentionnelles mobilisées. Par ailleurs, en fonction des applications identifiées à la

section 3.2 *Manques*, il se peut que l'utilisateur sollicite non pas une fonction cognitive individuellement, mais plusieurs. Ainsi, la multiplicité des fonctions cognitives sollicitées, qui est une condition se rapprochant davantage de situations réelles, est une variable intrinsèque aux exigences de la tâche, qui nous semble pertinente à prendre en compte. Nous nous demandons donc si l'augmentation du nombre de fonctions cognitives correspond à une augmentation de la charge mentale basée sur la somme des fonctions cognitives sollicitées. De plus, nous nous demandons s'il existe des différences entre une tâche simple et la même tâche, qui cette fois, est suivie d'une autre tâche cognitive.

En outre, pour aller plus loin sur la question de la transférabilité de modèles de classification, nous souhaitons savoir si un modèle de classification réalisé pour une tâche donnée dans un contexte expérimental A (en ligne par exemple) peut prédire la charge mentale subjective pour une tâche identique, mais réalisée dans un contexte expérimental B (en présentiel par exemple). Par conséquent, lors de l'étude 1, nous testons la transférabilité d'une tâche A vers une tâche B dans un même contexte opérationnel. Alors que dans l'étude 2 nous testons la transférabilité d'une tâche A dans un contexte A vers la même tâche A dans un contexte B. Nous déclinons donc la question de la transférabilité d'un modèle de classification selon la tâche et selon le contexte expérimental.

Enfin, comme vu au *Chapitre 2*, pour évaluer la charge mentale avec le plus de justesse, la triangulation des trois catégories de mesures (subjectives, de performance et psychophysologiques) peut être envisagée. Notre objectif à terme est donc de définir si le modèle établi dans la première partie de nos recherches et basé sur les mesures subjectives et de performances peut être amélioré par la prise en compte des données physiologiques. On souhaiterait ajouter l'EEG, car c'est la mesure cérébrale la plus efficace pour mesurer la charge mentale (Zhang et al., 2018). Ainsi, nous pourrions atteindre l'objectif d'avoir une mesure continue de la charge mentale via l'EEG qui est une mesure déjà présente dans le cadre de départ de l'essai clinique *GHOST*.

Comme il s'agit ici d'un travail d'envergure en termes de traitement des données (impliquant des compétences supplémentaires en traitement du signal), il ne pourra pas être mené à terme durant cette thèse, mais nous en préparerons les prémises. En conséquence, dans un second temps, notre problématique de recherche est axée sur trois

objectifs qui seront opérationnalisés à la section *5.1.1 Hypothèses opérationnelles de l'étude 2* :

- **Objectif 1** : Nous souhaitons savoir si un modèle de classification de la charge mentale subjective est capable de faire de la classification inter-contextes. Ainsi, pour une même tâche, un modèle de classification réalisé dans un contexte donné, serait capable de prédire les classes de charge mentale d'un autre contexte.
- **Objectif 2** : Nous souhaitons connaître la différence, de ressenti ludique et de charge mentale, entre un design classique et un design ludique avec des items unicolores, puis entre ce dernier et un design ludique avec des items multicolores.
- **Objectif 3** : Nous souhaitons connaître la différence, de charge mentale et de ressenti ludique, entre des tâches sollicitant de manière différente plus ou moins de fonctions cognitives.

Nous mettrons en place une seconde étude, présentée au chapitre 5, pour répondre à ces objectifs. Mais avant toute chose, nous présenterons l'étude 1, au chapitre suivant.

PARTIE EXPÉRIMENTALE

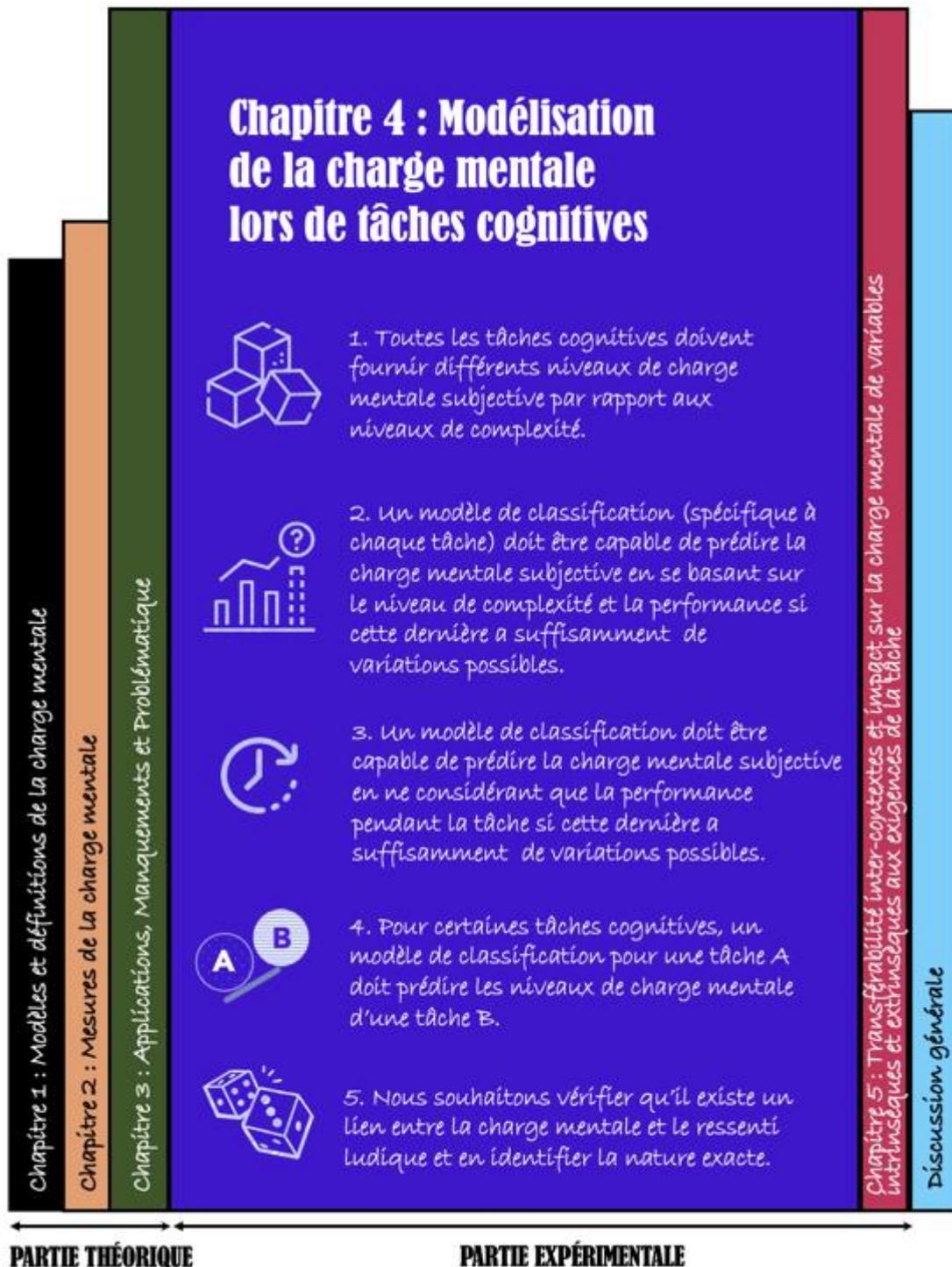


Figure 1 : Plan général du manuscrit de thèse et détail des objectifs du chapitre 4 présentant l'étude 1.

Les objectifs de cette première étude sont d'avoir plusieurs tâches cognitives candidates pour pouvoir prédire le niveau de charge mentale subjective associé à un niveau de complexité. Puis, d'avoir un indicateur temps réel basé sur les performances, qui est

capable de mesurer la charge mentale en cours d'activité. De plus, nous souhaitons savoir si les modèles de classification sont inter-tâches, donc transférables d'une tâche cognitive A à une tâche cognitive B. Enfin, dans cette première étude, nous nous interrogeons sur le lien entre la charge mentale subjective et le ressenti ludique de la tâche cognitive. Ainsi, nous avons plusieurs objectifs, présentés au niveau de la **Figure 1**, que nous avons opérationnalisés à la section suivante.

4.1 Hypothèses de l'étude et choix des tâches

4.1.1 Hypothèses opérationnelles de l'étude 1

En se basant sur les objectifs de notre problématique (Cf. section 3.3 *Problématique*), nous avons choisi, à travers cette première étude expérimentale de vérifier les hypothèses opérationnelles suivantes :

- **Hypothèse H1** : Toutes les tâches cognitives développées pour nos recherches fournissent différents niveaux de charge mentale subjective par rapport aux niveaux de complexité qui les caractérisent.
- **Hypothèse H2** : Un modèle de classification (spécifique à chaque tâche) est capable de prédire la charge mentale subjective en se basant sur le niveau de complexité et la performance si cette dernière a suffisamment de variations possibles.
- **Hypothèse H3** : Un modèle de classification est capable de prédire la charge mentale subjective en ne considérant que la performance pendant la tâche si cette dernière a suffisamment de variations possibles.
- **Hypothèse H4** : Pour certaines tâches cognitives, un modèle de classification pour une tâche A peut prédire les niveaux de charge mentale d'une tâche B. Celles partageant une base commune (matériel commun par exemple) ou des types d'interactions communs (car cela permet d'avoir davantage de dimensions de performances similaires).
- **Hypothèse H5** : Il y a une corrélation négative entre la charge mentale et le ressenti ludique.

4.1.2 Choix des tâches cognitives

Pour tester nos hypothèses opérationnelles, il nous faut un certain nombre de tâches cognitives candidates. Dans un premier temps, nous avons sélectionné le test de N-back

(NBT), car c'est une tâche cognitive reconnue pour mesurer la charge mentale. Elle nous permettra donc de comparer nos résultats avec la littérature scientifique (Cf. 3.3 *Problématique*). Pour le NBT de notre étude, nous nous sommes appuyés sur cinq variantes basées sur plusieurs études (Dimitrakopoulos et al., 2017 ; Shin et al., 2018 ; Zhang et al., 2018 ; Tremmel et al., 2019) dont la princeps, celle de Kirchner (1958). Un tableau comparatif (avec entre autres, les participants, le nombre d'essais et les mesures) entre le NBT de notre étude et les différentes variantes du NBT est proposé en Annexe 3.

De plus, par rapport à la problématique énoncée au chapitre précédent, nous avons comme objectif d'avoir plusieurs tâches sollicitant chacune une fonction cognitive distincte. Étant donné que le NBT sollicite la mise à jour de la mémoire verbale, nous souhaitons d'autres tâches basées sur des fonctions cognitives différentes.

Dans un second temps, nous avons donc identifié les fonctions cognitives fréquemment sollicitées dans les trois applications que nous avons énoncées à la section 3.2 *Manques*. Puis, nous avons réalisé une revue de la littérature sur toutes les tâches cognitives rattachées à chaque fonction cognitive identifiée. Ensuite, nous avons comparé les différentes tâches qui étaient les plus appropriées aux applications mentionnées à la section 3.2 *Manques* et pour une expérimentation en ligne. En effet, cette première étude a eu lieu lors de la pandémie de la Covid-19. Ainsi, nous avons dû nous adapter et proposer une étude distancielle. Par conséquent, nous étions limités dans le choix des tâches dont les modalités de réponse devaient impliquer principalement des interactions motrices avec l'ordinateur (pointer la souris de l'ordinateur ou appuyer sur le clavier). Effectivement, dans le contexte de l'expérience en ligne, la production orale était difficile à enregistrer.

La **Table 1** présente trois fonctions cognitives identifiées, les principales tâches cognitives affiliées et les tâches sélectionnées pour notre étude.

Fonctions cognitives (FC)	Principales tâches cognitives liées à la FC (listes non exhaustives)	Tâches conservées
Mémoire de travail via l'Empan mnésique	L'épreuve des matrices (Brooks, 1967) ; Le <i>Block Tapping Test</i> (Corsi, 1972) ; La tâche d'Empan Visuel (Wilson et al., 1987) ; Le <i>Visual Pattern Test</i> (Della Sala et al., 1997).	Test de Corsi

Inhibition mentale	Le Stroop (Stroop, 1935) ; Les tâches de “ <i>tapping</i> ” ; Le paradigme d’amorçage négatif (Tipper, 1985) ; La tâche de signal de stop (Logan, 1994) ; Le test de Go/No-Go (Vidal et al. (2012) ; Le test de Hayling (Burgess et Shallice, 1996).	Test de Go/No-Go
Flexibilité mentale	Cf. Table 2	WCST

Table 1 : Lien entre les fonctions cognitives identifiées, les tâches cognitives affiliées et les tâches sélectionnées pour notre seconde étude.

Pour la seconde fonction cognitive, nous souhaitons nous intéresser à la Mémoire de Travail (MDT) liée aux tâches d’Empan mnésique. En effet, dans les applications identifiées en 3.2 *Manques*, les tâches d’Empan mnésique sont couramment utilisées par les praticiens lors d’évaluation basées sur des tâches cognitives. Les tâches d’Empan mnésique permettent d’évaluer les capacités mnésiques résiduelles et dysfonctionnelles de diverses populations neuropathologiques. Puisque certaines fonctions mnésiques peuvent être touchées et d’autres non à la suite d’une atteinte cérébrale, plusieurs outils ont vu le jour pour l’évaluation de différentes composantes mnésiques. Nous avons donc réalisé une revue de la littérature pour identifier parmi plusieurs (Cf. **Table 1**), la tâche d’Empan mnésique la plus adaptée à notre contexte d’étude. Notre choix s’est porté sur le test de Corsi. Il est adapté à un test en ligne (pas de modalité verbale évitant ainsi les contraintes liées aux limites techniques), les consignes sont simples et il est adapté à des analphabètes et à différentes cultures. De plus, le test de Corsi sollicite la mémoire visuo-spatiale, contrairement au NBT sollicitant la mémoire verbale. Pour le test de Corsi de notre étude, nous nous sommes appuyés sur cinq variantes (Kessels et al., 2000 ; Fournier et Albaret, 2013 ; Brunetti et al., 2014 ; Björngrim, 2019) dont la princeps, celle de Corsi (1972). Un tableau comparatif entre le Corsi de notre étude et les différentes variantes du Corsi est proposé en Annexe 4.

Pour la troisième fonction cognitive, nous nous intéressons à l’inhibition mentale qui est également impliquée dans de nombreux tests neuropsychologiques (Miyake et al., 2000) qui entrent en jeu dans les applications identifiées dans la partie 3.2 *Manques*. Lors de l’inhibition mentale, le focus attentionnel reste fixé sur un type de stimuli sans être perturbé par la survenue intempestive de stimuli différents (Goddefroy et GREFEX,

2008). Dans cette logique, l'inhibition mentale peut être considérée comme une des fonctions de contrôle remplies par le système attentionnel superviseur (Norman et Shallice, 1986) ou par l'administrateur central de la MDT (Baddeley, 1986). Afin de sélectionner la tâche la plus appropriée pour notre contexte de recherche, la **Table 1** indique les principales épreuves d'inhibition mentale recensées dans la littérature scientifique. Après une comparaison des différentes tâches, notre choix s'est porté sur le test de Go/No-Go (GNG), car il est adapté à une expérimentation en ligne (pas de modalité verbale). De plus, il est adapté à des analphabètes et à différentes cultures. Enfin, plusieurs études (Falkenstein et al., 1999 ; Vidal et al., 2012) l'ont utilisé dans un contexte EEG. Ces études utilisant le GNG avec l'EEG peuvent donc nous servir de base pour un de nos objectifs issu de notre problématique 3.3 *Problématique*. Il existe cependant différents types de tâches de GNG. Nous avons choisi de nous baser sur celle de Vidal et al. (2012), car la présentation des stimuli est simple. Ce sont des formes classiques (cercles, carrées) identifiables par le plus grand nombre, peu importe la culture, contrairement au GNG d'Harmony et al. (2009) par exemple, se basant sur des lettres. Un tableau comparatif du GNG de Vidal et al. (2012) et du GNG de notre étude est proposé en Annexe 5.

Pour la quatrième fonction cognitive, nous nous intéressons à la flexibilité mentale qui, comme l'inhibition mentale, est impliquée dans de nombreux tests neuropsychologiques (Miyake et al., 2000). La flexibilité mentale correspond à un déplacement volontaire du foyer attentionnel d'une catégorie de stimuli à une autre (Miyake et al., 2000 ; Collette, 2004). Nous avons donc réalisé une revue de la littérature pour identifier toutes les tâches cognitives sollicitant la flexibilité mentale (Cf. **Table 2**). Nous nous sommes rendu compte qu'il existait trois types de flexibilité mentale (Eslinger et Grattan, 1993). La flexibilité réactive (les contraintes de l'environnement exigent un changement de réponse), la flexibilité spontanée (le participant présente des réponses variées dans un environnement stable), la flexibilité auto-générée (produire des mots, des pensées, des réponses diversifiées dans un environnement stable). La **Table 2** met au regard les trois types de flexibilité mentale avec les tâches cognitives associées à chacune d'entre elles.

Types	Différentes tâches					
Flexibilité réactive	<i>Wisconsin Card Sorting Test</i> (WCST ; Heaton, 1981)	<i>Trail Making Test</i> (TMT ; Reitan, 1958)	La tâche de nombre-lettre adaptée (Rogers et Monsell, 1995)	Tâche de plus-moins (Jersil, 1927)	Tâche <i>globale-locale</i> avec figures (Navon, 1977)	Test de Brixton (Burgess et Shallice, 1996)
Flexibilité spontanée	<i>Wisconsin Card Sorting Test</i> (Heaton, 1981)	<i>Dimensional Change Card Sort</i> (Zelazo et al., 1996)	<i>Modified Card Sorting Test</i> (Nelson, 1976) puis traduction française (Godefroy et Grefex, 2008)			
Flexibilité auto-générée	Tâche verbale phonémique	Tâche de production aléatoire de chiffres	Tâche de fluence verbale sémantique	Tâche de fluence graphique	Test de flexibilité (Eslinger et Grattan, 1993)	

Table 2 : Les tâches mesurant les différentes composantes de la flexibilité mentale

Comme nous pouvons l’observer au niveau de la **Table 2**, le *Wisconsin Card Sorting Test* de Heaton (WCST ; Heaton, 1981) a l’avantage de mêler la composante spontanée et la composante réactive de la flexibilité mentale (Salicé et Juhel, 2005). Ainsi, avec le WCST, nous manipulons la flexibilité mentale dans ses deux plus grandes composantes qu’on retrouve également dans les applications identifiées dans notre problématique de recherche. Pour mettre en place le WCST de notre étude, nous nous sommes appuyés sur deux variantes (Godefroy et GREFEX, 2008 ; Laurent, 2013). Un tableau comparatif entre le WCST de notre étude et les différentes variantes du WCST est proposé en Annexe 6 du manuscrit.

Enfin, comme nous l’avons vu au *Chapitre 2 - Mesures de la charge mentale*, une des mesures de ce concept correspond aux performances à une tâche secondaire. Ce type de mesure est couramment considéré lors d’un paradigme de la double tâche, où l’on demande à l’individu d’effectuer une deuxième tâche parallèlement à une tâche principale (Marchand et al., 2021). La réalisation de deux tâches simultanées est considérée comme une épreuve exécutive, reposant sur l’administrateur central de la MDT du modèle de Baddeley (1986). L’administrateur central qui sélectionne les actions les plus efficaces en fonction de la tâche à effectuer et qui les coordonne. Ainsi, la dernière tâche cognitive de notre étude est une double tâche où nous souhaitons avoir comme tâche principale un test que nous avons déjà identifié au préalable. Notre

choix s'est porté sur le test de Corsi, car ses consignes simples et rapides sont propices à insérer la consigne d'une double tâche. La tâche secondaire s'est portée sur une tâche de calcul mental, donc sollicitant la mémoire de travail verbale, qui est une fonction cognitive distincte de la mémoire visuo-spatiale du Corsi. Un choix pour éviter l'effet d'interférence d'une tâche secondaire. En effet, selon Hancock et al. (2021), l'utilisation d'une modalité d'entrée différente permet d'éviter le problème d'interférence, en partant de l'hypothèse que les ressources attentionnelles sont alimentées par un réservoir commun (Hancock et al., 2021). De plus, le calcul mental, sous forme écrite, permet d'être implanté dans une étude en ligne sans pour autant avoir de modalité auditive, et ainsi éviter les problématiques techniques (pas de retour sonore du côté du participant). Pour le choix des calculs mentaux, nous nous sommes appuyés sur l'étude de So et al. (2017), ayant utilisé le calcul mental avec différents niveaux de complexité. Un aspect qu'on voulait avoir, en maintenant le niveau du Corsi stable, et en modulant les niveaux de complexité à travers le calcul mental.

4.2 Matériel et méthodes

4.2.1 Participants

Cinquante-trois participants francophones en bonne santé, résidant à Nantes (et sa métropole) et répondant aux critères d'inclusion (parlant couramment le Français, droitiers, ayant une vision normale ou corrigée à la normale et une audition normale ou corrigée à la normale), ont participé à cette expérience en ligne (28 femmes, 24 hommes et 1 personne non-binaire correspondant à une "*identification à des identités de genre en dehors des catégories masculines ou féminines*" (Poirier et al., 2019)). Leur âge moyen était de 28,64 ans (écart-type = 6,89 ans). Ils étaient volontaires et ont été informés du but réel de l'expérience. 96,23 % d'entre eux (51 participants) avaient au moins un niveau Bac + 2, dont 43,4 % avaient un Master. Le reste des participants avait le niveau Baccalauréat. Par ailleurs, 66,26 % travaillaient pour l'entreprise onepoint, 9,43 % venaient de l'Université de Nantes et les autres étaient des connaissances de l'équipe de recherche.

4.2.2 Tâches et procédures

Pour avoir de nombreux participants en un temps donné et en temps de restrictions pandémiques, nous avons opté pour une étude en ligne en deux sessions expérimentales,

afin d'éviter les effets de fatigue. Ainsi, les participants pouvaient avoir deux ou trois tâches cognitives maximum pour une session donnée. Par conséquent, les instructions pour le test ont été conçues pour une expérience optimale en ligne (comme la nécessité d'être assis dans une pièce calme, ou de mettre le téléphone en mode silencieux ; Cf. Annexe 11 - questionnaire pré-test de l'étude 1). Enfin, pour chaque action des cinq tâches cognitives, aucun retour visuel n'est fourni aux participants, quelle que soit la précision de la réponse. Cela, afin d'éviter d'influencer les résultats aux questionnaires, notamment la dimension *Performance ressentie* du NASA-TLX.

Tâches cognitives

Toutes les tâches ont été créées grâce au logiciel de développement *Unity3D* (Version 2019.4.30f1) par Sébastien Maire, développeur onepoint sous la supervision de la doctorante qui s'est appuyée sur le protocole expérimental. Par ailleurs, les participants ont entrepris cinq tâches cognitives de manière aléatoire. Comme vu à la section 4.2.1, les tâches étaient les suivantes : la tâche de N-back (NBT), le test de Corsi, le WCST, le test Go/No-Go (GNG) et une double tâche (DT). Avant de détailler la structure des tâches, la **Figure 2** présente une vue générale du design expérimental de l'étude. Au sein d'une même session expérimentale, la durée entre chaque tâche cognitive est régulée par le participant lui-même. S'il voulait enchaîner ou faire une pause, il pouvait le faire.

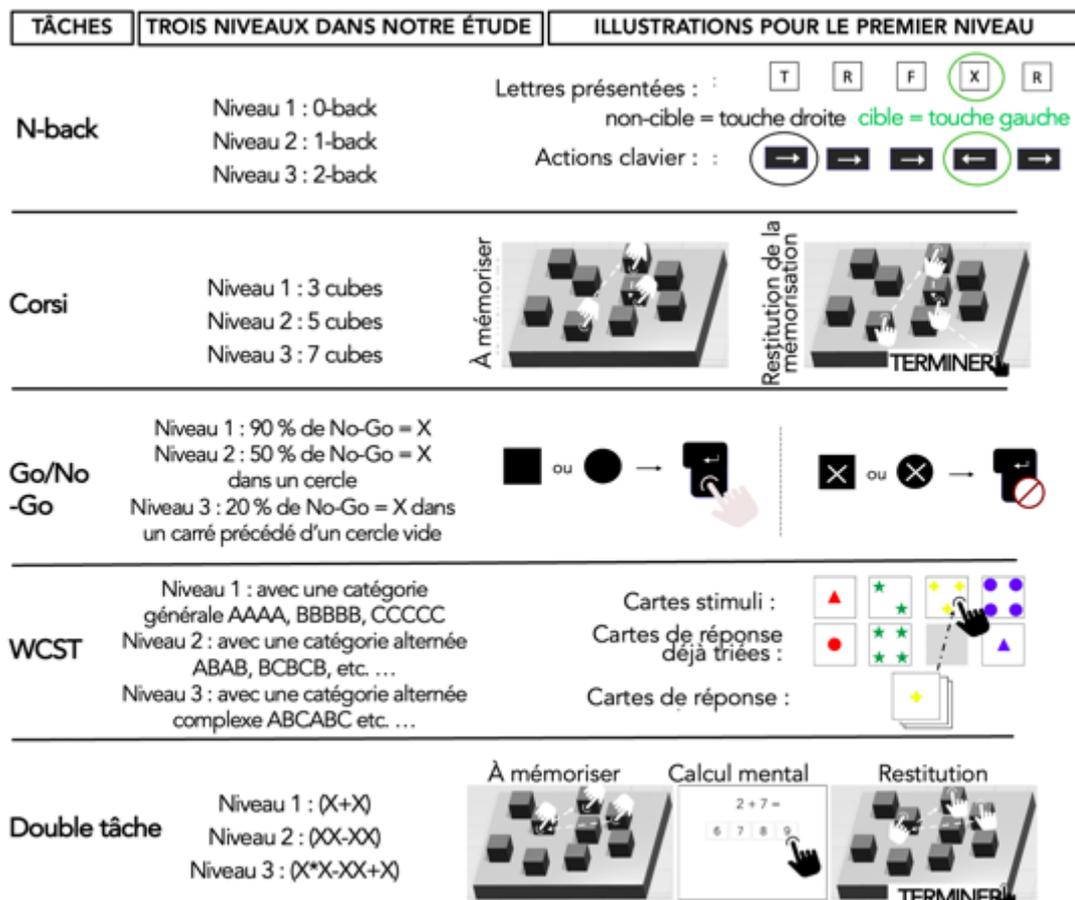


Figure 2 : Vue d'ensemble du design expérimental de l'étude 1 où les tâches sont présentées de manière aléatoire aux participants.

Le NBT est réputé pour avoir des niveaux de complexité induisant des niveaux de charge mentale dans la zone intermédiaire de la charge mentale correspondant à une charge mentale intermédiaire faible, une charge mentale intermédiaire moyenne et une charge mentale intermédiaire élevée pour respectivement les niveaux 0-back, 1-back et 2-back (Arvaneh et al., 2015 ; Dimitrakopoulos et al., 2017 ; Ries et al., 2018). Ainsi, le NBT a servi de référence pour calibrer les niveaux de complexité des autres tâches lors de pré-tests où nous avons testé plusieurs niveaux de complexité avec au moins cinq participants. Les mesures de performance ont été l'indicateur pour sélectionner les trois niveaux de complexité. Cependant, avec ces mesures, l'inconvénient était que nous ne pouvions pas obtenir des indicateurs comparables, car ils étaient différents d'une tâche à l'autre et difficilement transférables (Sirevaag et al., 1993 ; Raufaste et al., 2004). Nous avons donc défini un indicateur commun à toutes les tâches, correspondant au pourcentage de réponses attendues. Par exemple, si le participant avait toutes ses

réponses correctes, les réponses attendues étaient de 100 %. Ainsi, lors des pré-tests à plusieurs niveaux de complexité, nous en avons sélectionné trois, en fonction des résultats aux réponses attendues.

4.2.2.1 Tâche de N-back (NBT)

Proposée par Kirchner (1958), la tâche de N-back (NBT) implique la capacité de mise à jour de la Mémoire de Travail. En effet, le NBT demande aux participants de réagir différemment lorsqu'un stimulus est le même que le stimulus présenté avant lui. Dans un premier temps, les participants lisent des instructions écrites (Annexe 7) accompagnées de schémas explicatifs et d'une consigne-vidéo. Puis, ils effectuent une tâche d'entraînement correspondant au niveau de complexité "1-rang" (Cf. **Figure 3**), où ils réalisent 10 essais ayant 3 cibles.

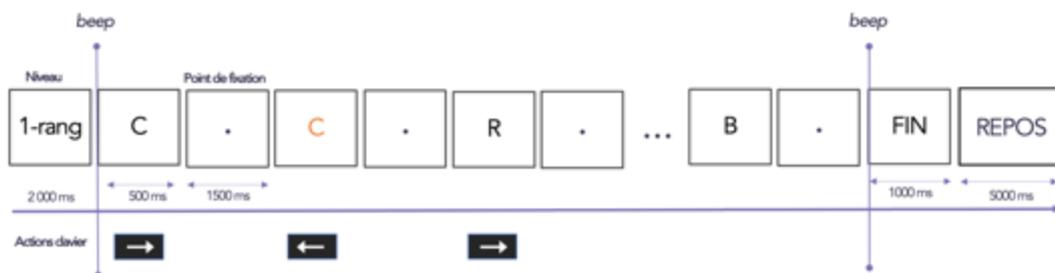


Figure 3 : Structure détaillée de la tâche d'entraînement du N-back. La lettre orangée sert à titre d'exemple pour mettre en évidence la lettre cible. Dans les faits, toutes les lettres présentées aux participants sont noires.

Puis, chaque participant réalise les trois niveaux du NBT (0-rang, 1-rang et 2-rang) de manière aléatoire. Nous n'allons pas au-delà de 2-rang, car dans les résultats de l'étude de Scharinger et al. (2017), les performances du NBT n'ont indiqué aucune différence significative entre les niveaux 3-rang et 4-rang. De plus, lors de nos pré-tests, nous avons observé que le niveau 3-rang dégradait les performances au-delà de la moyenne des autres tâches cognitives.

Plus précisément, le NBT de notre étude présente quinze lettres capitales individuelles : B, C, D, F, G, H, J, L, N, P, Q, R, S, T et V. Les participants doivent répondre en indiquant si la lettre est une cible ou une non-cible. Lorsqu'une cible apparaît, le participant appuie avec son index droit sur la touche directionnelle gauche du clavier. Inversement, en cas de non-cible, le participant appuie avec son majeur droit sur la touche directionnelle droite du clavier.

Concernant la tâche 0-back, la cible est toujours la lettre “X”. Pour les autres conditions, la cible est la lettre présentée “N” rangs plus tôt dans la séquence. Dans ces deux autres conditions, la lettre “X” n'est pas présentée dans la séquence. L'essai continue par la présentation d'une consonne toutes les 1500 ms. Chaque consonne est présentée durant 500 ms, suivie d'un point de fixation pendant l'intervalle inter-stimulus de 1500 ms (Cf. **Figure 3**).

Un son *beep* (250 ms) marque le début et la fin de la présentation des stimuli de la tâche de NBT. De plus, le mot « FIN » est affiché durant 1000 ms à la suite du déroulement de l'ensemble des stimuli, suivi du mot “REPOS” pendant 5000 ms. Chaque condition consiste en des blocs de 20 lettres, dont 30 % sont des stimuli cibles. Nous avons donc un défilement de 20 lettres avec 6 cibles. La durée d'un niveau (hors consignes écrites et questionnaires) est donc de 48 s.

Les dimensions de performance lors du NBT sont les *Fausses alarmes* (appui sur la touche gauche/cible lors de l'affichage d'une lettre non-cible (où le participant doit appuyer sur la touche droite). Sur 20 lettres qui défilent, on peut avoir maximum 14 *Fausses alarmes*), les *Omissions* (ne pas appuyer sur la touche gauche/cible alors qu'il s'agissait d'une lettre cible et d'appuyer sur la touche droite. On peut avoir maximum 6 *Omissions*), les *Non-réponses* (aucune réponse entre deux items. On peut avoir maximum 20 *Non-réponses*), les *Réponses attendues* (appui sur la touche gauche en cas de cible et sur la touche droite en cas de non-cible), le *Temps de réaction* pour l'ensemble des items (temps entre la présentation des stimuli et l'appui sur une touche de réponse) et le *Temps de réaction* uniquement pour les bonnes réponses.

Enfin, la durée totale du NBT est d'environ 13 minutes.

4.2.2.2 Test de Corsi

Le *block tapping task* dit le test de Corsi (Corsi, 1972) se base sur la mémoire visuo-spatiale qui consiste à retenir une séquence de pointage de cubes, puis à reproduire la séquence dans le même ordre d'apparition. Ainsi, il s'agit d'une tâche d'empan mnésique qui nécessite de se souvenir des différents emplacements des cubes pointés qui constituent la séquence (dimension spatiale) et de l'ordre dans lequel ils apparaissent (dimension temporelle).

Dans notre étude, le matériel (Cf. **Figure 4**) est constitué de neuf cubes gris foncé en perspectives 3D montés de manière non symétrique sur une planche de couleur gris clair. Le pointage des séquences visuo-spatiales se fait par l'expérimentatrice virtuelle au rythme d'un cube par seconde.

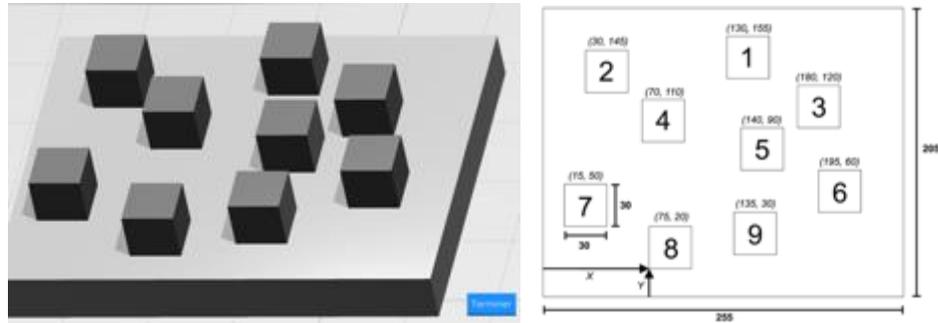


Figure 4 : Matériel du test de Corsi avec neuf cubes en perspectives 3D de la présente étude (figure gauche) avec des emplacements (figure droite) basés sur l'étude de Kessels et al. (2000). Les coordonnées (en mm) sont mesurées du coin inférieur gauche du plateau au coin inférieur gauche de chaque cube. Les chiffres (de 1 à 9) ne sont pas visibles par les participants.

Après la lecture des consignes (Annexe 7), le participant réalise une phase d'entraînement qui consiste à refaire la séquence des deux cubes "2" et "6". Lorsque le participant doit réaliser sa séquence, une consigne "À VOUS" s'affiche. Lorsque les cubes sont pointés par le participant, la face supérieure du cube s'allume en bleu pour confirmer que le logiciel a bien détecté la réponse. Après avoir donné sa réponse, le participant doit indiquer lorsque son tour est terminé en appuyant sur le bouton "TERMINER" qui s'affiche dès le début du tour du participant (Cf. **Figure 5**).

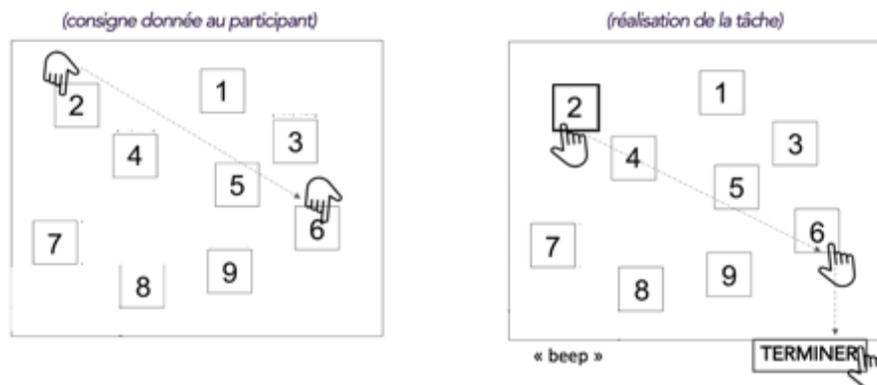


Figure 5 : Exemple de consigne et d'interactions lors du test de Corsi.

Une fois l'entraînement terminé, le participant réalise trois niveaux d'empan mnésique (à deux séquences chacun) dans un ordre aléatoire. Après avoir pré-testé huit niveaux de complexité, nous avons pu déterminer les trois qui génèrent trois niveaux de performances bien distincts. À la suite de cinq pré-tests, les niveaux de complexité suivants : 3 cubes, 5 cubes et 7 cubes, ayant chacun 2 essais ont été gardés et sont surlignés en vert sur la **Figure 6**.

8-5	
6-4	
4-7-2	3-7-8-2-9-4
8-1-5	5-9-1-7-4-2-8
3-4-1-7	5-7-9-2-8-4-6
6-1-5-8	5-8-1-9-2-6-4-7
5-2-1-8-6	5-9-3-6-7-2-4-3
4-2-7-3-1	5-3-8-7-1-2-4-6-9
3-9-2-4-8-7	4-2-6-8-1-7-9-3-5

Figure 6 : Séquences pré-définies d'empan mnésique du test de Corsi par Kessels et al. (2000) et les séquences conservées dans notre étude surlignées en vert.

Concernant les dimensions de performances du Corsi, nous avons les *Cubes pointés*, les *Cubes faux*, les *Cubes exacts* (somme des cubes correctement rappelés), les *Séquences exactes* (nombre de cubes pointés dans l'ordre et la position attendue), les *Réponses attendues* (similarité entre la séquence attendue et la séquence réalisée), les *Omissions* (ne pas pointer sur un cube alors qu'il faisait partie de la séquence attendue) et le *Temps total* (temps entre l'apparition du bouton « TERMINER » et l'appui sur ce bouton). Enfin, la durée totale du test de Corsi est d'environ 11 minutes.

4.2.2.3 Test de Go/No-Go (GNG)

Le Go/No-Go (GNG ; Vidal et al., 2012) est une tâche d'inhibition mentale de la réponse motrice. Pour notre étude, elle se présente avec un enchaînement aléatoire de deux formes : des cercles et des carrés noirs sous fond blanc. La lettre X apparaît aléatoirement sur ces formes. Lorsque la forme est vide, les participants doivent cliquer sur la touche « ENTRÉE ». Lorsque le X est présent, les participants doivent inhiber leur réponse motrice. Après la lecture des consignes (Annexe 7), les participants effectuent une tâche d'entraînement (Cf. **Figure 7**) avec 10 essais (5 cibles).

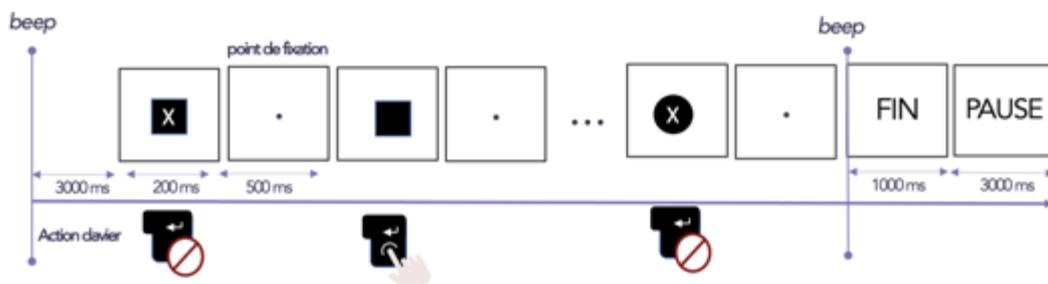


Figure 7 : Structure détaillée de la tâche d'entraînement du Go/No-Go.

Une fois l'entraînement terminé, le participant réalise trois niveaux de complexité. Tous présentent 10 items défilant avec une fréquence de 700 ms. Chaque item est présenté durant 200 ms., suivi d'un point de fixation pendant l'intervalle inter-stimulus de 500 ms (Cf. **Figure 7**). Un son *beep* (250 ms) est présenté au début et à la fin de la tâche. De plus, le mot « FIN » est affiché à la fin de la tâche durant 1000 ms, suivi du mot « PAUSE » affiché durant 3000 ms. Pour les pré-tests, nous avons testé neuf niveaux de complexité (Cf. **Table 3**).

Niveau	% de No-Go	Types de No-Go	Autres items du niveau
Niveau simple	90 %	No-Go = X	10% de présentation aléatoire de cercles et de carrés vides
	70%	No-Go = X	30% de présentation aléatoire de cercles et de carrés vides
	50%	No-Go = X	50% de présentation aléatoire de cercles et de carrés vides
Niveau moyen	50%	No-Go = "X" dans cercle	30% de X dans carré et 20% de cercles/carrés vides
	30%	No-Go = X	70% de cercle/carrés vides aléatoires
Niveau difficile	30%	No-Go = "X" dans cercle	30% de X dans carrés et 40% de cercles/carrés vides
	30%	No-Go = "X" dans cercle	50% de X dans carrés et 20% de cercles/carrés vides
Niveau très difficile	30%	No-Go = "X" dans carrés précédés de cercles vides	30% de No-Go = X dans carrés précédés de cercles vides 10% de X dans carrés précédés de carrés vides 10% de X dans cercles précédés de cercles vides
	20%	No-Go = "X" dans carrés précédés de cercles vides	20% de No-Go = X dans carrés précédés de cercles vides 10% de X dans carrés précédés de carrés vides 20% de X dans cercles précédés de cercles vides

Table 3 : Différents niveaux de Go/No-Go pré-testés, ceux en couleurs sont ceux qui sont conservés pour l'étude finale.

À la suite des pré-tests, on identifie ces trois niveaux de complexité (sous fond coloré sur la **Table 3**), comme permettant d'avoir des niveaux de performances distincts :

- Niveau *simple* avec 90 % de No-Go = X
- Niveau *moyen* avec 50 % de No-Go = "X" dans cercle
- Niveau *très difficile* avec 20 % de No-Go = "X" dans carrés précédés de cercles vides.

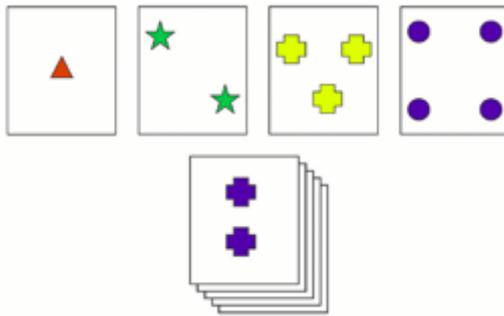
Les performances mesurées à la tâche de GNG sont les *Omissions* (pas d'appui sur la touche alors qu'il s'agissait d'une cible Go. Tous les niveaux ont 10 items. Ainsi, pour le niveau 1, nous avons au maximum 1 *Omission*, pour le niveau 2, au maximum 5 *Omissions* et pour le niveau 3, au maximum 8 *Omissions*), les *Fausse alarmes* (FA ; correspondant à l'appui sur la touche lors de l'affichage d'une cible No-Go où normalement le participant n'appuie pas. Pour le niveau 1, nous avons maximum 9 FA, pour le niveau 2, maximum 5 FA et pour le niveau 3, maximum 2 FA), les *Réponses*

attendues (le calcul est : $100 * (10 - (FA + Omissions) / 10)$), le *Temps de réaction* uniquement pour réponses correctes (de l'affichage d'un stimulus Go jusqu'à l'appui sur ENTRÉE), le *Temps de réaction pour l'ensemble des items confondus*.

Enfin, la durée totale du GNG est d'environ 11 minutes.

4.2.2.4 *Wisconsin Card Sorting Test (WCST)*

Le *Wisconsin Card Sorting Test* (WCST ; Heaton, 1981) est une tâche de catégorisation



au cours de laquelle le participant doit s'adapter aux changements de règles. Cette tâche mesure principalement la flexibilité mentale correspondant au déplacement volontaire du foyer attentionnel d'une catégorie de stimuli à une autre (Miyake et al., 2000 ; Collette, 2004).

Figure 8 : Configuration du Wisconsin Card Sorting Test (WCST).

Dans notre version du WCST, la partie supérieure de l'écran comporte quatre cartes cibles (une carte avec un triangle rouge, une carte avec un deux étoiles vertes, une carte avec trois croix jaunes et une carte avec un quatre ronds bleus) comme le montre la **Figure 8**. À chaque changement de consigne, la position des cartes dans la rangée change de manière aléatoire.

Quant à la partie inférieure, elle se compose d'un jeu de plusieurs cartes qui varie de 16 à 48 cartes (en fonction du niveau de complexité). Les items présents sur les cartes sont définis selon trois dimensions, chacune déclinée en quatre modalités : les items peuvent être des triangles, des étoiles, des croix ou des ronds, être au nombre de 1, 2, 3 ou 4, ou de couleur jaune, rouge, vert ou bleu. Une carte comportera toujours des items de forme et de couleur identiques. Pour effectuer le tri des cartes, les participants doivent utiliser la souris d'ordinateur ou le pavé tactile. Puis, ils doivent cliquer sur la carte cible correspondante qui se trouve en haut de l'écran. Les dimensions à prendre en compte pour le tri apparaissent à l'écran de manière écrite durant 5 secondes.

Pour nos pré-tests, nous avons identifié quatre niveaux de complexité. Un niveau par critère spécifique, un autre par catégorie générale, un niveau par catégorie alternée et un autre niveau par catégorie alternée complexe qui sont détaillés dans la **Table 4**.

Niveaux	Règles	Dimensions	ABC Nombre de cartes	B Nombre de cartes
Niveau 0 – par critère spécifique	Règle A	Couleur bleue	4 cartes	–
	Règle B	En fonction des triangles	5 cartes	4 cartes
	Règle C	Trois éléments	3 cartes	–
Niveau 1 – par catégorie générale	Règle A	Forme	4	–
	Règle B	Couleur	5	4
	Règle C	Nombre	3	–
Niveau 2 – par catégorie alternée	Règle A	Forme/Nombre	4	–
	Règle B	Nombre/Couleur	5	4
	Règle C	Forme/Couleur	3	–
Niveau 3 – par catégorie alternée complexe	Règle A	Forme/Nombre/Couleur	4	–
	Règle B	Nombre/Forme/Couleur	5	4
	Règle C	Forme/Nombre/Couleur	3	–

Table 4 : Différents niveaux du WCST lors des pré-tests

Niveau 0 - par catégorie spécifique

Dans ce niveau, les participants doivent associer chaque carte du bas avec une des cartes du haut en fonction d'une règle basée sur une catégorie spécifique. Par exemple, pour la consigne « Associez en fonction de la COULEUR BLEUE », les participants associent toutes les cartes du bas aux cartes du haut en fonction de cette couleur.

Niveau 1 - par catégorie générale

Dans ce niveau, les participants doivent se focaliser sur une seule dimension pour effectuer le tri des cartes. Ainsi, les participants doivent trier les cartes du bas en fonction de leur forme, de leur nombre et de leur couleur uniquement (Cf. **Figure 9**).

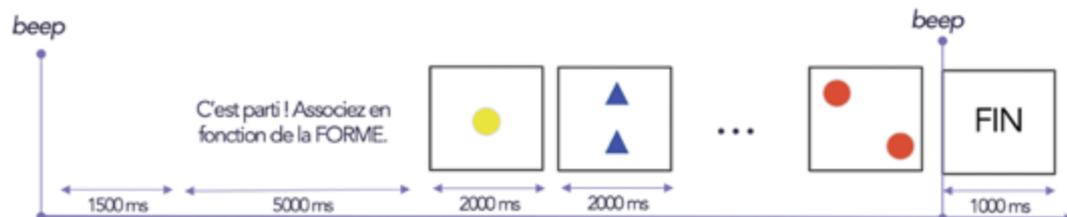


Figure 9 : Structure détaillée de la tâche de WCST pour le niveau 1 - par catégorie générale

Niveau 2 - par catégorie alternée

Dans ce niveau, les participants doivent alterner entre deux dimensions distinctes (alternance de type ABABAB), telles que les dimensions FORME et NOMBRE.

Niveau 3 - par catégorie alternée complexe

Concernant, le niveau alterné complexe, les participants doivent alterner entre trois dimensions (alternance de type ABCABC), comme la FORME, le NOMBRE et la COULEUR.

À la suite de nos pré-tests, nous avons gardé trois niveaux de complexité, ceux qui sont dans des cellules colorées au niveau de la **Table 4**. Nous avons écarté le niveau 0 qui générait des problèmes de compréhension récurrents.

Concernant les performances recueillies au WCST, nous avons : le *Nombre d'erreurs* (réponses qui ne correspondent pas à ce qui est attendu. On peut en avoir au maximum 16 pour le niveau 1, maximum 32 pour le niveau 2 et maximum 44 pour le niveau 3), les *Non-réponses* (nombre de non-réponses entre deux cartes présentées. Pour cette dimension, à la suite de problèmes techniques, les données ont été recueillies pour une partie seulement des participants), le *Nombre d'erreurs persévératrices* (qui s'observent lorsqu'une réponse erronée correspond à la catégorie utilisée par le participant lors de sa réponse précédente. Autrement dit, lorsque le participant persiste/persévère dans une catégorie de réponse), le *Temps de réaction de tous les items confondus* et les *Réponses attendues* (calcul : $100 * (\text{nombre de cartes totales} - \text{Erreurs}) / \text{nombre total de cartes}$). Les *Non-réponses* ne sont pas comptées, car elles ne sont pas disponibles pour l'ensemble des participants (problématique technique).

Enfin, la durée totale du WCST est d'environ 15 minutes.

4.2.2.5 Double tâche (DT)

La double tâche (DT) reprend en partie le test de Corsi basé sur la mémoire visuo-spatiale avec un niveau fixe, celui avec trois cubes à mémoriser. En parallèle, une tâche de calcul mental est insérée entre le pointage des cubes par l'expérimentatrice et la restitution du pointage par les participants (Cf. **Figure 10**). Ainsi, la tâche de calcul mental perturbe le rappel de la mémorisation visuo-spatiale du test de Corsi.

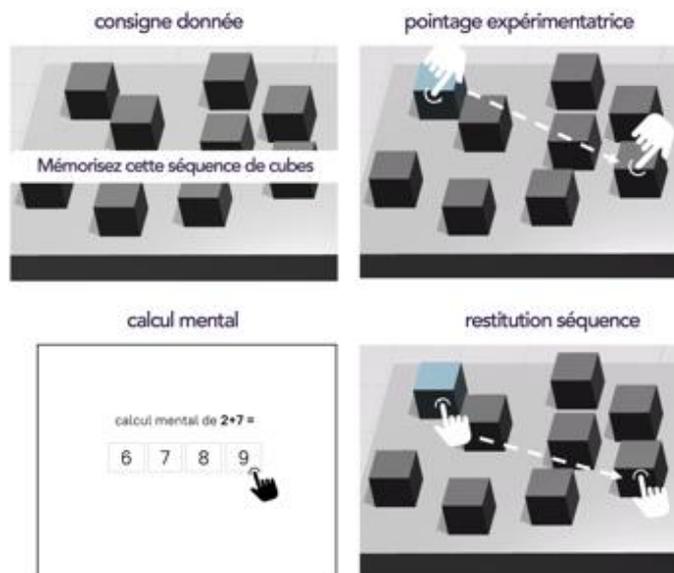


Figure 10 : Matériel et déroulement de la double tâche durant l'entraînement.

Les calculs mentaux ont trois niveaux de complexité, correspondant à un niveau simple (une addition à un chiffre), un niveau intermédiaire (une addition ou une soustraction à deux chiffres) et un niveau difficile avec, selon les termes des auteurs, un « *mix d'opérations* » de multiplications ou de divisions avec une addition ou une soustraction. Trois niveaux de complexité ayant chacun deux essais (présentés dans le même ordre indépendamment des passations). Les niveaux de complexité sont quant à eux réalisés dans un ordre aléatoire. En outre, pour les deux essais par niveau, nous avons veillé à ce qu'il y ait le même niveau de difficulté perçue. Pour ce faire, les pré-tests ont également eu cette utilité d'ajustement qui nous ont permis d'avoir trois niveaux de complexité distincts (Cf. **Table 5**).

Essais	Entraînement			Niveau simple		
	Séquence Corsi	Calcul mental	Choix proposés	Séquence Corsi	Calcul mental	Choix proposés
1er essai	2-6	2+7	6,7,8,9	5-8-1	5+2	6,7,8,9
2ème essai	8-5	3+5	6,7,8,9	3-4-1	4+2	6,7,8,9
Essais	Niveau intermédiaire			Niveau difficile		
	Séquence Corsi	Calcul mental	Choix proposés	Séquence Corsi	Calcul mental	Choix proposés
1er essai	5-9-3	33-12	19, 20, 21, 22	5-3-8	7*3-10+3	12, 13, 14 et 15
2ème essai	6-1-5	28-11	16,17,18, 19	3-9-2	4*6-13+2	12, 13, 14 et 15

Table 5 : Différents niveaux de complexité de la double tâche.

Concernant le matériel de la double tâche (Cf. **Figure 10**), il est identique au matériel du test de Corsi. Concernant le calcul mental, il se réalise dans un fond blanc composé de quatre choix de réponses cliquables. Par ailleurs, avant le pointage de cubes, la consigne « Mémorisez cette séquence de cubes » apparaît au-dessus du plateau virtuel pendant trois secondes. Puis, le premier cube est pointé par l'expérimentatrice. 1 seconde après la fin du dernier cube pointé, on affiche sur un fond blanc le calcul mental. Dès que le participant appuie sur une des quatre réponses, le plateau de Corsi réapparaît (1000 ms plus tard) avec la consigne : « Répétez la séquence précédemment mémorisée ». Le bouton "TERMINER" s'affiche dès le début de cette étape.

Dans un premier temps, le participant réalise une phase d'entraînement. Puis, il réalise les trois niveaux de complexité dans un ordre aléatoire (Cf. **Table 5**).

Concernant les performances à la DT, nous avons les *Cubes faux*, les *Cubes exacts*, les *Cubes pointés*, les *Omissions* (correspondant au fait de ne pas pointer sur un cube alors qu'il faisait partie de la séquence attendue), le *Temps total pour le Corsi*, le *Temps total pour le calcul mental*, les *Réponses attendues* (correspondant à la similarité entre la séquence pointée et la séquence attendue), les *Réponses correctes au calcul mental* parmi les quatre choix possibles et les *Séquences exactes* (correspondant au nombre de cubes pointés dans l'ordre et la position attendus).

Enfin, la durée totale de la DT est d'environ 13 minutes.

La **Table 6** est un récapitulatif des performances avec la ligne bleue "Réponses attendues" commune à toutes les tâches. Par ailleurs, les dimensions de performances sont classées pour permettre une comparaison entre les tâches.

N-back (NBT)	Corsi	Wisconsin (WCST)	Go/No-Go (GNG)	Double Tâche (DT)
Fausses alarmes	Cubes faux	Nombre d'erreurs	Fausses alarmes	Cubes faux Bonne réponse calcul
Omissions	Omissions Cubes pointés		Omissions	Omissions Cubes pointés
Non réponses		Non réponses Erreurs persévératrices		
Le pourcentage de réponses attendues	Le pourcentage de réponses attendues	Le pourcentage de réponses attendues	Le pourcentage de réponses attendues	Le pourcentage de réponses attendues
	Séquences exactes			Séquences exactes
	Cubes exacts			Cubes exacts
Temps de réaction tous items confondus		Temps de réaction tous items confondus	Temps de réaction tous items confondus	
Temps de réaction pour les bonnes réponses			Temps de réaction pour bonnes réponses.	
Temps total		Temps total	Temps total	Temps total
	Temps total			Temps total Corsi

Table 6 : Dimensions de performances pour toutes les tâches cognitives de l'étude.

La Figure 11 propose un autre aperçu de notre plan expérimental qui résume l'ensemble des données communes aux cinq tâches cognitives.

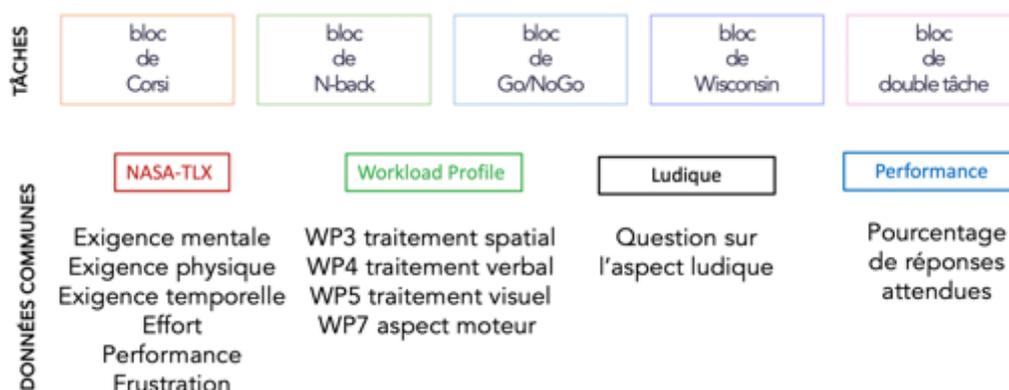


Figure 11 : Vue générale de l'ensemble des données communes aux tâches cognitives.

4.2.3 Questionnaires

Comme nous l'avons vu au Chapitre 2 concernant les mesures de la charge mentale, les questionnaires NASA-TLX et WP sont multidimensionnels et complémentaires. Ils ont donc été utilisés dans la présente étude dans l'objectif de mesurer la charge mentale subjective. Par ailleurs, en s'appuyant sur notre problématique de recherche, nous mesurons le ressenti ludique à travers une question sur une échelle de Likert de 0 à 100

que nous avons mise en place : « *Dans quelle mesure la tâche que vous venez de faire a été ludique, amusante ?* ».

Questionnaire NASA-TLX

Comme nous l'avons vu au Chapitre 2, le NASA-TLX (Hart et Staveland, 1988 ; Cegarra et Morgado, 2009 ; Annexe 1) est un outil d'évaluation subjectif et multidimensionnel qui évalue la charge mentale perçue associée aux tâches à travers six dimensions : *Exigences mentales, Exigences physiques, Exigences temporelles, Performance, Effort et Frustration*. Après chaque niveau d'une tâche, le participant note chaque dimension de 0 à 100. Dans notre étude, nous avons considéré chaque dimension du questionnaire NASA-TLX et son score global. Pour ce dernier, afin de gagner du temps d'expérimentation, nous utilisons la version non pondérée en faisant la moyenne des six dimensions (RTLX ; Cf. section 2.1.2.1 NASA-TLX).

Questionnaire WP

Le questionnaire *Workload Profile* (WP ; Tsang et Velazquez, 1996 ; Annexe 2) est une autre technique d'évaluation de la charge mentale qui demande aux participants de fournir la proportion des ressources attentionnelles utilisées à travers huit dimensions : *Traitement central perceptif, Sélection et exécution de la réponse, Traitement spatial, Traitement verbal, Traitement visuel, Traitement auditif, Sortie manuelle et Sortie vocale* (Rubio et al., 2004). La définition de chaque dimension est donnée aux participants. Pour chaque tâche, ils doivent fournir un nombre entre 0 (aucune demande) et 100 (demande maximale) qui représente la proportion de ressources attentionnelles utilisées dans chacune des huit dimensions. Avant la sélection des items, nous avons procédé à une traduction française (Cf. Annexe 2) de ce questionnaire à l'aide de deux personnes anglo-saxonnes et en pré-testant auprès de cinq individus. Par ailleurs, comme il n'y avait pas de tâches de résolution, de sélection, d'audition et de parole dans notre étude, seules quatre des huit dimensions ont été proposées aux participants. En effet, en se référant à l'article princeps du questionnaire WP (Tsang et Velazquez, 1996, p. 362), il est précisé que les résultats à ce questionnaire peuvent être évalués de manière unidimensionnelle. Ainsi, pour notre étude, nous avons considéré uniquement les

échelles correspondant au traitement spatial (WP3), au traitement verbal (WP4), au traitement visuel (WP5) et à la production manuelle (WP7).

La durée de la passation de l'ensemble de ces questionnaires est d'environ six minutes.

4.2.4 Procédure expérimentale

Après avoir obtenu la validation éthique du Comité Éthique pour les Recherches Non Interventionnelles (CERNI ; Annexe 8), nous avons lancé la campagne de recrutement des participants. Chaque participant a été clairement informé des objectifs et du déroulement de l'étude. Cette expérience a adopté un plan intra-sujet ($P53 < T5 * N3 >$) ; avec P = participants, T = Tâches et N = Niveaux). La phase expérimentale s'est déroulée comme suit :

- Présentation d'un résumé de l'étude (Annexe 9) par courrier électronique ou via les réseaux sociaux professionnels.
- Lorsque le participant manifestait son intérêt, l'expérimentatrice lui envoyait un lien avec un mot de passe personnel.
- Le participant se connectait à la session expérimentale en ligne après avoir saisi le mot de passe. De plus, une signature électronique de la lettre de consentement (Annexe 10) était demandée avant le questionnaire pré-test (avec des questions sur l'âge, le genre, le niveau d'éducation par exemple ; Annexe 11).
- Une fois ces étapes achevées, le participant réalisait une des deux sessions expérimentales sur deux jours distincts séparés a minima d'une semaine. Toutes les tâches comprenaient des instructions (Annexe 7) sur la manière de les réaliser, suivies d'une séance d'entraînement pour se familiariser avec la présentation des stimuli et avec les interactions requises (Annexe 7).
- Après chaque niveau de complexité de chaque tâche, il y avait l'administration d'une version électronique des questionnaires standardisés du NASA-TLX et du WP. Ainsi que la question sur le ressenti ludique. Tous les participants étaient exposés aux tâches et aux niveaux de complexité dans un ordre aléatoire. De plus, après chaque niveau de complexité de chaque tâche, une question sur la possibilité d'avoir été distrait par l'environnement extérieur a été posée (Annexe 12).
- Par ailleurs, lorsque l'ensemble des tâches ont été réalisées, pour une session donnée, un questionnaire post-test était proposé aux participants (Annexe 12).

- À la fin de l'étude, tous les participants recevaient 15 euros en bons d'achat.

4.3 Résultats de l'étude 1

L'analyse de nos résultats commence par la vérification que les niveaux de complexité de nos cinq tâches génèrent trois niveaux de performance distincts. La suite de la section *Résultats* est composée de cinq parties principales correspondant aux cinq hypothèses liées à notre problématique de l'étude 1 présentées ci-après.



Pour tester l'ensemble de ces hypothèses opérationnelles, nous avons procédé à plusieurs analyses statistiques que nous présentons dans un plan illustré à la **Figure 12**.

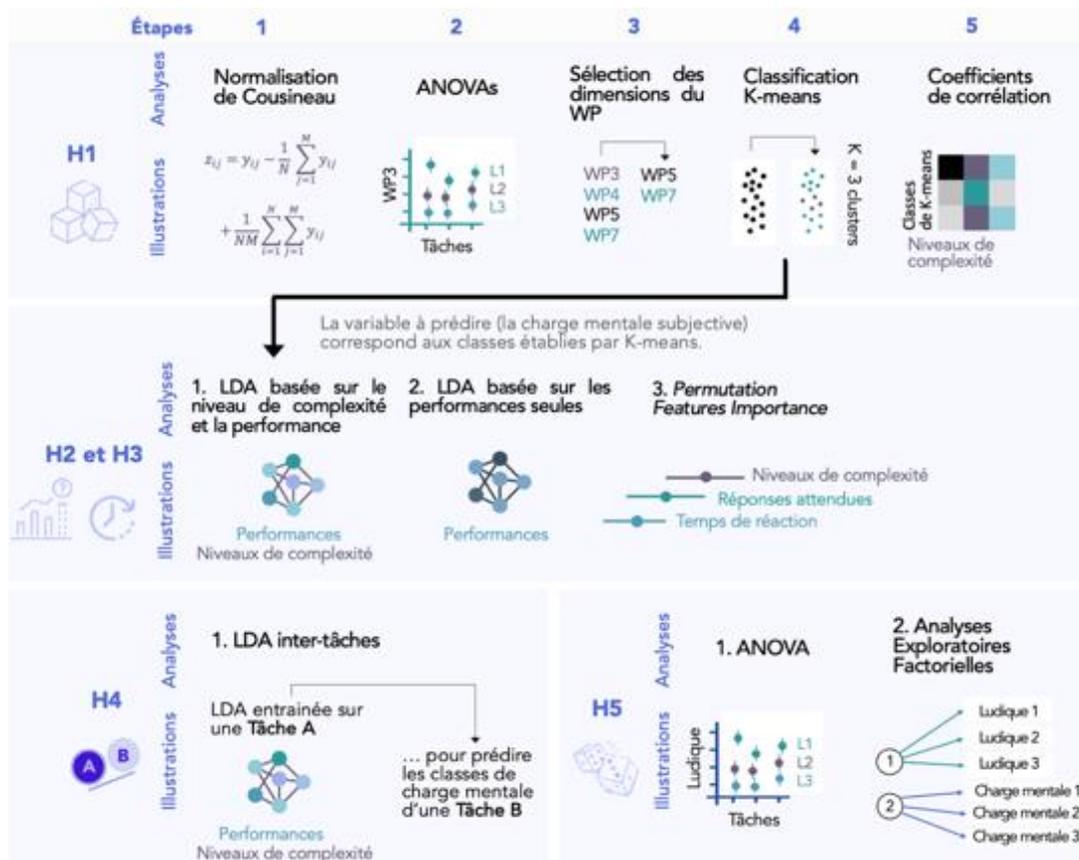


Figure 12 : Plan d'analyse statistique des données de l'étude 1 en lien avec les hypothèses opérationnelles issue de la problématique générale.

Correspondance entre les classes de performances et le niveau de complexité

Avant de répondre à ces cinq hypothèses, nous vérifions si les niveaux de complexité de nos cinq tâches généraient trois niveaux de performance distincts. Pour ces analyses, nous avons utilisé le logiciel JASP (version JASP 0.16). Pour comparer les performances entre les tâches, nous avons établi un critère de performance commun aux cinq tâches, les *Réponses attendues* (Cf. section 4.2.2 *Tâches et procédures*). En préambule, nous avons vérifié l'homogénéité de nos distributions et de nos variances et avons conclu qu'elles ne l'étaient pas. Pour cette raison, nous avons utilisé des statistiques non paramétriques. Une ANOVA à mesures répétées non paramétrique (avec le test de Friedman) a été réalisée, avec le niveau de complexité comme variable indépendante et la performance comme variable dépendante. De plus, des tests de rang signé de Wilcoxon et des comparaisons post hoc de Conover ont ensuite été utilisés.

Ainsi, nous avons comparé les performances en termes de *Réponses attendues* entre les cinq tâches (**Figure 13**) et nous avons observé un effet principal significatif du niveau de complexité (Table 1 partie A en Annexe 13 ; $X^2(2) = 145,255$; $p < 0,001$). Les niveaux 1 et 2 étaient significativement différents (Table 1, partie B en Annexe 13 ; $T(740) = 2,887$; $p = 0,004$), ainsi que les niveaux 2 et 3 ($T(740) = 2,592$; $p = 0,010$). Par ailleurs, le niveau de complexité 1 était le niveau le plus faible, et le niveau 3 le plus complexe. De plus, considérés tâche par tâche, tous les niveaux de complexité étaient significativement différents (Table 1, partie C en Annexe 13), sauf concernant les niveaux 2 et 3 du GNG ($W = 561,000$; $p = 0,086$) et de la DT ($W = 460,000$; $p = 0,102$). Par conséquent, les trois niveaux de complexité ont généré trois niveaux distincts de *Réponses attendues* pour le NBT, le Corsi et le WCST, mais pas pour le GNG et la DT où les niveaux 2 et 3 n'étaient pas significativement différents.

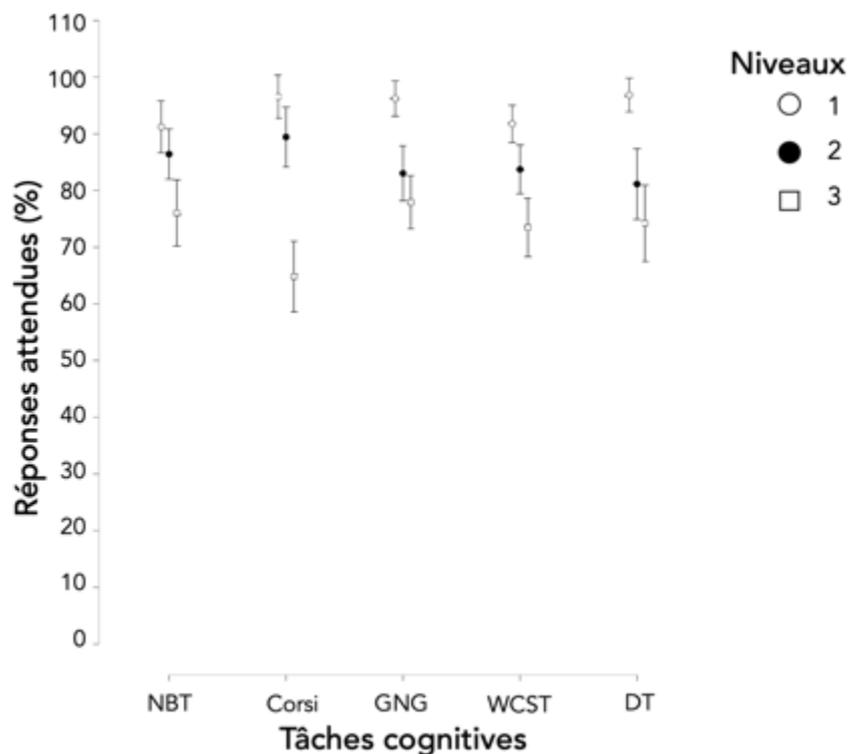


Figure 13 : Les Performances en termes de *Réponses attendues* selon le type de tâche cognitive et le niveau de complexité.

De plus, il y avait un effet principal significatif du type de tâche (Table 1, partie A en Annexe 13 ; $X^2(4) = 16,735$; $p = 0,002$). Considérées niveau par niveau, les *Réponses attendues* n'étaient pas significativement différentes entre le NBT et le WCST, de même pour le GNG et la DT (Table 2, partie D en Annexe 13). Enfin, le Corsi et le WCST

étaient les deux tâches dont la charge mentale globale était significativement différente entre tous les niveaux de complexité (Table 1, partie D en Annexe 13).

En conclusion, les niveaux de complexité du NBT, du Corsi et du WCST généraient plusieurs niveaux de *Réponses attendues*, contrairement aux niveaux de complexité du GNG et du WCST. Les résultats de ces deux dernières tâches peuvent s'expliquer par des niveaux d'exigences de la tâche demandés qui étaient trop proches entre les niveaux 2 et 3 qui n'étaient pas significativement différents. Une piste d'explication qui sera davantage discutée lors de la Discussion de l'étude 1.

4.3.1 Correspondance entre les niveaux de charge mentale subjective et les niveaux de complexité

Dans cette section d'analyse, nos résultats permettront de tester l'hypothèse 1 (**H1 en violet**) de l'étude 1.



Pour cette partie, nous avons utilisé des statistiques non paramétriques (tests de Friedman avec le niveau de complexité comme variable indépendante et les mesures subjectives comme variable dépendante, de Wilcoxon et des comparaisons post hoc de Conover), toujours avec le logiciel JASP (version JASP 0.16).

Normalisation des données subjectives

Tout d'abord, pour toutes les mesures subjectives, telles que les questionnaires NASA-TLX et WP, nous avons normalisé ces données grâce à la correction proposée par Cousineau (Morey, 2008). Cette normalisation était basée sur des intervalles de confiance intra-sujets, et elle était recommandée pour les données subjectives telles que celles des questionnaires. La méthode de Cousineau pourrait être décrite comme suit. Soit y_{ij} le score du i e participant dans la $j^{\text{ème}}$ condition ($i = 1, \dots, N ; j = 1, \dots, M$). Ensuite, les observations normalisées z_{ij} ont été définies comme :

$$z_{ij} = y_{ij} - \frac{1}{N} \sum_{j=1}^M y_{ij} + \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M y_{ij}$$

Dans notre cas, nous avons soustrait la moyenne de toutes les conditions (les cinq tâches) et ajouté la moyenne de tous les participants à toutes les conditions expérimentales. Pour se rendre visuellement compte de l'impact de la normalisation Cousineau sur les données subjectives, nous avons réalisé des graphiques de distributions (**Figure 14**) des scores de charge mentale globale (sur 600) en fonction de la tâche et du niveau de complexité.

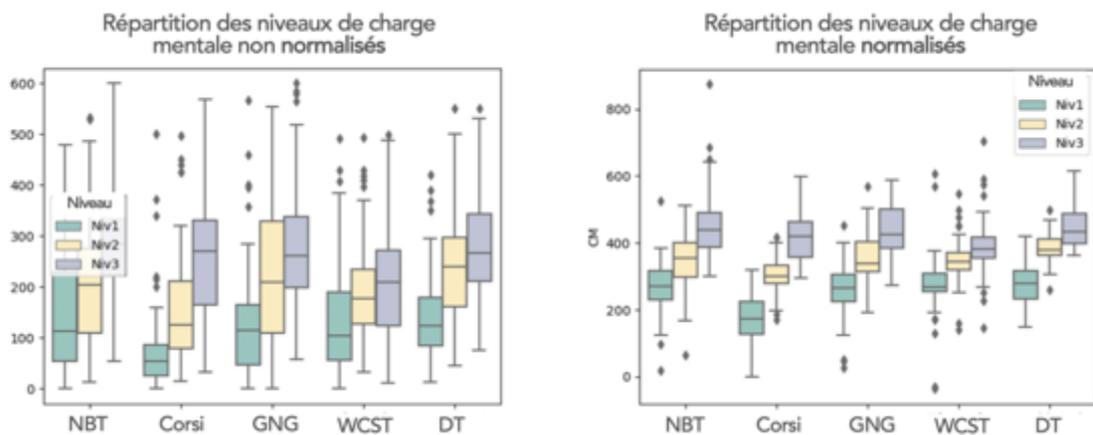


Figure 14 : Les scores du NASA-TLX global sans normalisation de Cousineau (à gauche) et avec (à droite).

On remarque qu'avec la normalisation de Cousineau, les différents scores du NASA-TLX global sont plus distincts d'un niveau de complexité à un autre. Les recouvrements dus aux variabilités individuelles (telles que le niveau de fatigue ou de concentration) ont donc diminué. C'était exactement ce qu'on recherchait à faire. La normalisation Cousineau était donc adaptée à notre contexte d'étude.

Questionnaire du NASA-TLX

Score global

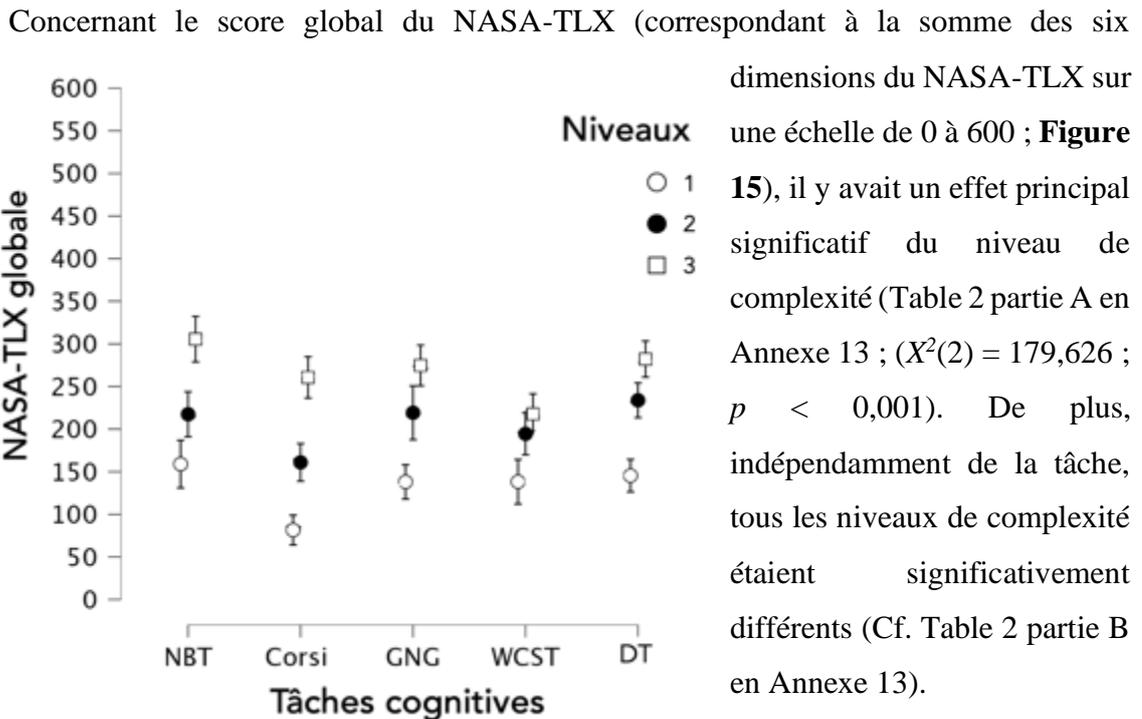


Figure 15 : La charge mentale basée sur le score du NASA-TLX global selon le type de tâche et le niveau de complexité.

Considérés tâche par tâche, tous les niveaux de complexité étaient également significativement différents (Table 2 partie C en Annexe 13). Par conséquent, les trois niveaux de complexité ont généré trois niveaux distincts de charge mentale globale.

De plus, il y avait un effet principal significatif du type de tâche (Table 2, partie A en Annexe 13 ; $X^2(4) = 33,605$; $p < 0,001$). Pour tous les niveaux de complexité, la charge mentale globale n'était pas significativement différente entre le NBT et le GNG, de même pour le NBT et la DT, ainsi que le GNG et la DT (Table 2, partie D en Annexe 13). Enfin, entre le NBT et le Corsi, la charge mentale globale était significativement différente entre tous les niveaux de complexité (Table 2, partie D en Annexe 13).

Analyse dimension par dimension

Concernant le score des dimensions du questionnaire NASA-TLX, nous considérerons d'abord les **Exigences mentales** (Cf. **Figure 16**). Il y avait un effet principal significatif du niveau de complexité (Table 3, part A en Annexe 13 ; $X^2(2) = 164,416$; $p < 0,001$). Indépendamment de la tâche, tous les niveaux de complexité étaient significativement différents (Table 3, part B en Annexe 13). De plus, considéré tâche par tâche (Table 3,

part C en Annexe 13), tous les niveaux de complexité étaient significativement différents à l'exception des niveaux 2 et 3 du WCST ($W = 546,000$; $p = 0,135$). Ainsi, les trois niveaux de complexité ont généré trois niveaux distincts d'*Exigences mentales* à l'exception du WCST.

De plus, il y avait un effet principal significatif du type de tâche ($X^2(4) = 30,873$; $p < 0,001$). Pour tous les niveaux de complexité, les *Exigences mentales* n'étaient pas significativement différentes entre le NBT et le GNG, de même pour le NBT et la DT, ainsi que pour le GNG et la DT (Table 3, partie D en Annexe 13). Par ailleurs, entre le NBT et le Corsi, les *Exigences mentales* étaient significativement différentes entre tous les niveaux de complexité (Table 3, partie D en Annexe 13).

Concernant les *Exigences physiques*, premièrement, par rapport aux autres dimensions, il s'agissait de celle ayant la valeur la plus faible pour toutes les tâches considérées, avec une moyenne inférieure à 30 sur 100 sur l'échelle de Likert (**Figure 16**). De plus, malgré qu'il y avait un effet principal significatif du niveau de complexité ($X^2(2) = 14,769$; $p < 0,001$; Table 3, part A en Annexe 13), indépendamment de la tâche, les niveaux de complexité n'étaient pas significativement différents (Table 3, part B en Annexe 13). Ce qui pouvait être expliqué par la taille de l'effet, le coefficient de concordance de Kendall (Kendall's W) qui est utilisé pour évaluer l'accord entre les évaluateurs. Le Kendall's W varie entre 0 (pas d'accord) à 1 (complètement en accord). Concernant l'effet principal significatif du niveau de complexité pour les *Exigences physiques*, il était relativement faible ($W = 0,061$). De plus, considérés tâche par tâche, les trois niveaux de complexité n'étaient pas différents (Table 3, part C en Annexe 13). Par conséquent, nos trois niveaux de complexité n'ont pas généré trois niveaux distincts d'*Exigences physiques*.

Par ailleurs, il n'y avait aucun effet principal significatif du type de tâche ($X^2(4) = 8,776$; $p = 0,067$; Table 3, part A en Annexe 13).

Concernant les *Exigences temporelles* (Cf. **Figure 16**), il y avait un effet principal significatif du niveau de complexité (Table 3 en Annexe 13 ; $X^2(2) = 49,422$; $p < 0,001$). Indépendamment de la tâche, seuls les niveaux 1 et 2 ($T(740) = 2,008$; $p = 0,045$) étaient significativement différents, ainsi que les niveaux 1 et 3 ($T(740) = 3,286$; $p = 0,001$). Considérés tâche par tâche, tous les niveaux de complexité étaient significativement

différents, sauf pour les niveaux 2 et 3 du GNG ($W = 419,000$; $p = 0,689$). Ainsi, nos trois niveaux de complexité ont généré trois niveaux distincts d'*Exigences temporelles* à l'exception du GNG.

De plus, il y avait un effet principal significatif du type de tâche ($X^2(4) = 51,179$; $p < 0,001$). Pour tous les niveaux de complexité, les *Exigences temporelles* n'étaient pas significativement différentes entre le NBT et le GNG (Table 3, partie D en Annexe 13). Par ailleurs, entre le NBT et le Corsi, les *Exigences temporelles* étaient significativement différentes entre tous les niveaux de complexité, de même entre le Corsi et le GNG, ainsi que le GNG et le WCST (Table 3, partie D en Annexe 13).

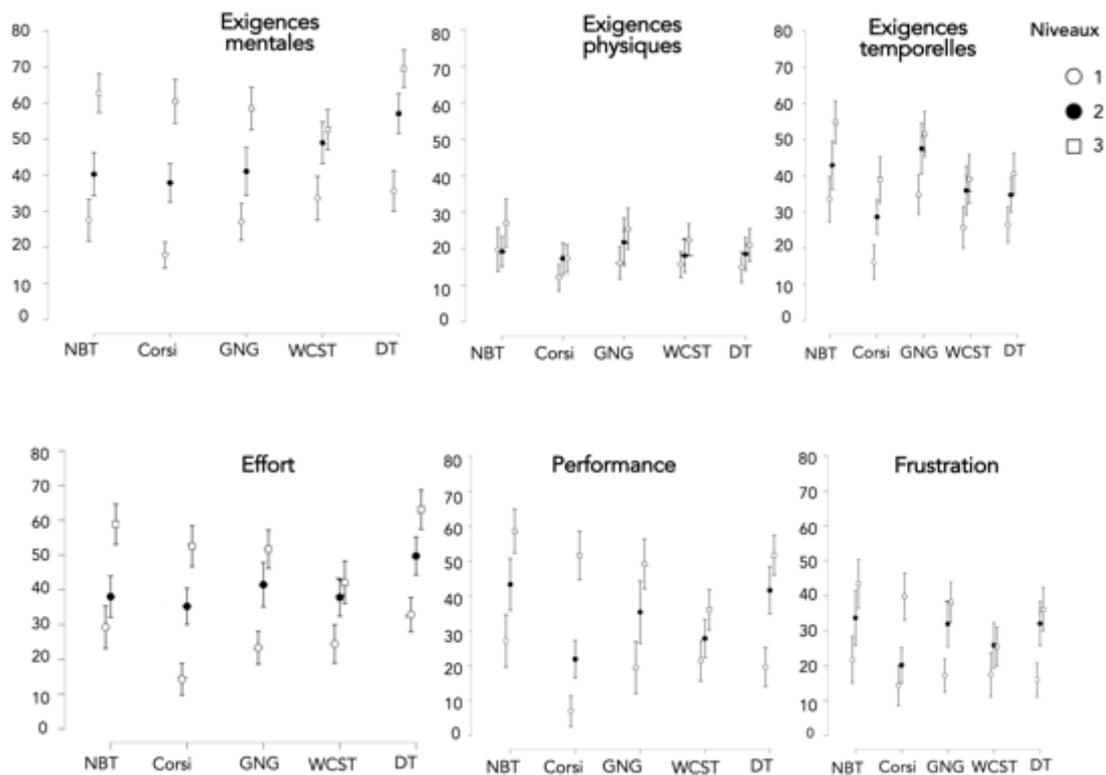


Figure 16 : Scores des six dimensions du NASA-TLX en fonction du type de tâche et du niveau de complexité.

Concernant l'*Effort mental* (Cf. **Figure 16**), il y avait un effet principal significatif du niveau de complexité (Table 3 en Annexe 13 ; $X^2(2) = 143,925$; $p < 0,001$). Indépendamment de la tâche, tous les niveaux étaient significativement différents. De plus, considéré tâche par tâche, tous les niveaux de complexité étaient significativement différents à l'exception des niveaux 2 et 3 du WCST ($X^2 = 518,000$; $p = 0,081$). Par conséquent, les trois niveaux de complexité ont généré trois niveaux d'*Effort* distincts,

à l'exception du WCST. De plus, il y avait un effet du type de tâche ($X^2(4) = 37,273 ; p < 0,001$). Pour tous les niveaux de complexité, l'*Effort* n'était pas significativement différent entre le NBT et le GNG (Table 3, partie D en Annexe 13). Par ailleurs, entre le Corsi et la DT, l'*Effort* était significativement différent entre tous les niveaux de complexité, de même entre le WCST et la DT (Table 3, partie D en Annexe 13).

Concernant le ressenti de la *Performance* (Cf. **Figure 16**), il y avait un effet principal significatif du niveau de complexité ($X^2(2) = 137,616 ; p < 0,001 ;$ Table 3, partie A en Annexe 13). Indépendamment de la tâche, tous les niveaux étaient significativement différents. De plus, considéré tâche par tâche, tous les niveaux de complexité étaient significativement différents, même pour les niveaux 2 et 3 du WCST ($W = 409,500 ; p = 0,028$). Ainsi, les trois niveaux de complexité ont généré trois niveaux distincts de *Performance*.

En outre, il y avait un effet du type de tâche ($X^2(4) = 31,676 ; p < 0,001$). Cependant, il n'y avait aucune paire de tâches où la *Performance* était significativement différente entre tous les niveaux de complexité (Table 3, partie D en Annexe 13). Ce qui pouvait être expliqué par la taille de l'effet, le coefficient de concordance de Kendall (Kendall's W) utilisé pour évaluer l'accord entre les évaluateurs. Le Kendall's W varie entre 0 (pas d'accord) à 1 (complètement en accord). Pour l'effet du type de tâche, il était relativement faible ($W = 0,064 ;$ Table 3, partie A en Annexe 13). Par ailleurs, pour tous les niveaux de complexité, la *Performance* n'était pas significativement différente entre le NBT et le GNG (Table 3, partie D en Annexe 13), de même pour le NBT et la DT, ainsi que le GNG et la DT.

Enfin, concernant la *Frustration* (Cf. **Figure 16**), il y avait un effet principal significatif du niveau de complexité (Table 3 en Annexe 13 ; $X^2(2) = 81,364 ; p < 0,001$). Indépendamment de la tâche, tous les niveaux étaient significativement différents, excepté pour les niveaux 2 et 3 ($T(740) = 1,662 ; p = 0,097$). De plus, considérées tâche par tâche, les comparaisons par paires ont montré que cela n'était vrai que pour le Corsi et la DT.

Par ailleurs, il y avait un effet du type de tâche ($X^2(4) = 19,209 ; p < 0,001$). Cependant, il n'y avait aucune paire de tâches qui était significativement différente entre tous les niveaux de complexité (Table 3, partie D en Annexe 13). Cela est dû à l'indice de

Kendall's relativement faible ($W = 0,036$). Enfin, pour tous les niveaux de complexité, la *Frustration* n'était pas significativement différente entre le NBT et le GNG (Table 3, partie D en Annexe 13), de même entre le NBT et la DT.

En conclusion, les trois niveaux de complexité ont généré trois niveaux distincts de charge mentale globale, mais aussi concernant la dimension *Performance* pour toutes les tâches. Le Corsi et la DT étaient les tâches les plus discriminantes et le WCST était la moins discriminante. À travers le WCST, nous avons observé la limite d'utiliser uniquement les performances pour la mesure de charge mentale. Les performances peuvent se dégrader à mesure que le niveau de complexité augmente, mais la charge mentale n'augmente pas de manière linéaire. Par ailleurs, les *Exigences physiques* était la seule dimension qui n'était pas influencée par le niveau de complexité quelle que soit la tâche et par rapport aux autres dimensions, la charge mentale des *Exigences physiques* était la dimension la plus faible pour toutes les tâches considérées. Ce dernier résultat était cohérent avec les interactions physiques lors de nos tâches, qui se limitaient aux interactions avec le clavier ou aux mouvements souris/pavé numérique. Une piste d'explication qui sera davantage discutée lors de la Discussion de l'étude 1.

Questionnaire WP

Pour cette partie, nous détaillons les résultats de chaque dimension du WP que nous avons précédemment sélectionnée, à savoir *WP3-Traitement spatial*, *WP4-Traitement verbal*, *WP5-Traitement visuel*, et *WP7-Production manuelle*.

Concernant le WP5, il s'agit de la dimension présentant le niveau de charge le plus élevé par rapport aux autres dimensions (**Figure 17**), et cela pour presque toutes les tâches considérées, avec une moyenne supérieure à 30 sur 100 sur l'échelle de Likert.

De plus, malgré qu'il y avait un effet principal significatif de la tâche (Table 4, partie A en Annexe 13 ; $X^2(4) = 13,893$; $p = 0,008$), les comparaisons par paires entre les tâches n'indiquaient aucune différence significative (Table 4, partie C en Annexe 13). Ce qui pourrait être expliqué par la taille de l'effet qui était faible ($W = 0,026$). Par ailleurs, il y avait un effet principal significatif du niveau de complexité ($X^2(2) = 43,682$; $p < 0,001$), mais seuls les niveaux 1 et 3 ($T(740) = 3,056$; $p = 0,002$). Enfin, considérés

tâche par tâche, tous les niveaux de complexité étaient significativement différents pour le NBT et le Corsi uniquement.

Concernant le WP3 (Figure 17), il y avait un effet principal significatif de la tâche (Table 4, partie A en Annexe 13 ; $X^2(4) = 86,422$; $p < 0,001$). Le WP3 pour la DT et le Corsi étaient significativement plus élevés que pour le GNG. De plus, le WP3 pour le NBT était significativement plus faible que pour toutes les tâches (excepté pour le GNG). De plus, il y avait un effet principal significatif du niveau de complexité (Table 4, partie B en Annexe 13 ; $X^2(2) = 27,797$; $p < 0,001$), mais seuls les niveaux 1 et 3 étaient statistiquement différents ($T(740) = 2,447$; $p = 0,015$). Enfin, considérés tâche par tâche, tous les niveaux de complexité étaient significativement différents pour le Corsi uniquement.

Concernant le WP4 (Cf. Figure 17), il y avait un effet principal significatif de la tâche (Table 4 partie A en Annexe 13 ; $X^2(4) = 55,564$; $p < 0,001$). Le WP4 du NBT et du WCST a généré la charge mentale la plus élevée par rapport aux autres tâches. De plus, il y avait un effet principal significatif du niveau de complexité sur toutes les tâches ($X^2(2) = 13,809$; $p = 0,001$), mais aucun niveau n'était statistiquement différent avec un autre, même les niveaux 1 et 3 ($T(740) = 1,699$; $p = 0,090$). Ainsi, considéré tâche par tâche, tous les niveaux de complexité étaient significativement différents pour le WCST uniquement.

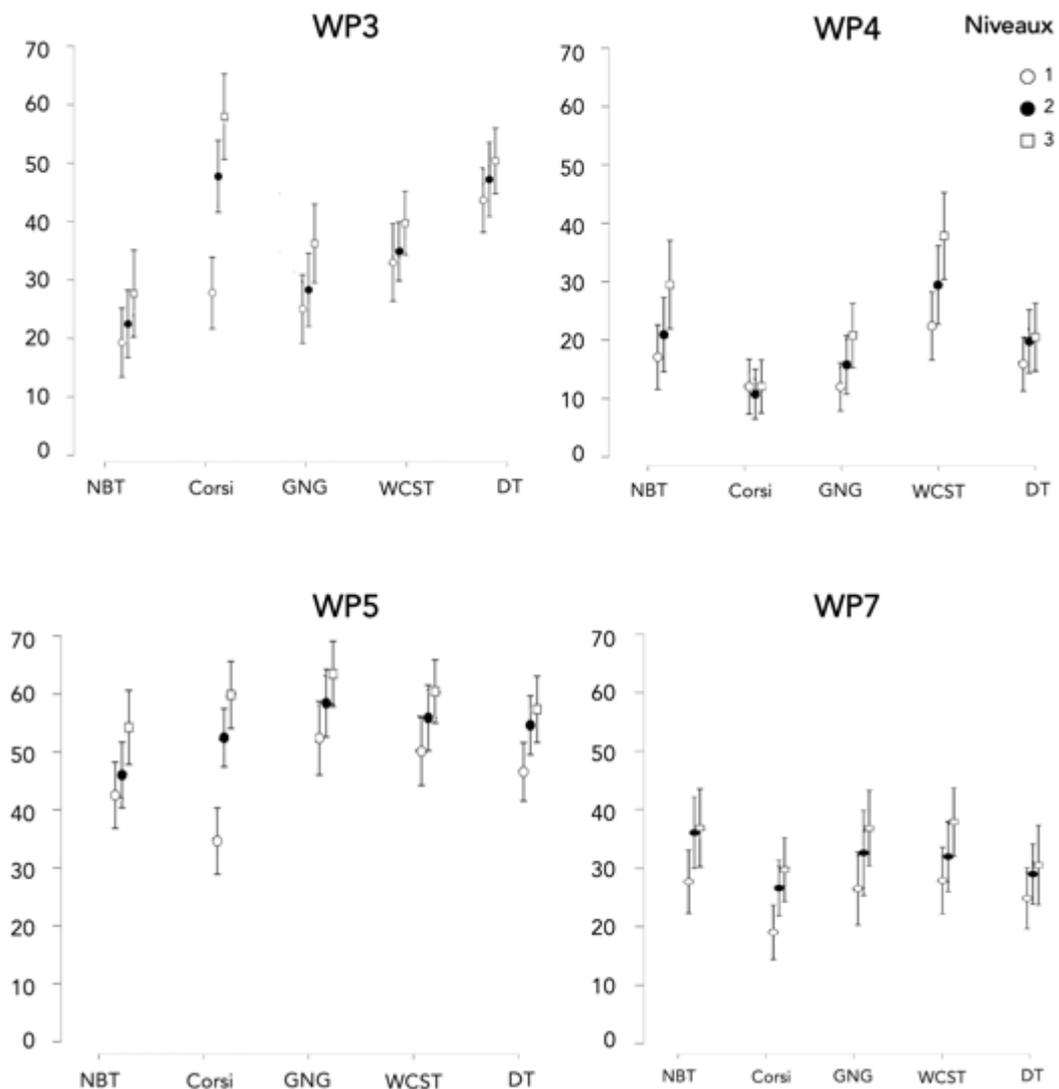


Figure 17 : Le score de charge mentale des quatre dimensions du questionnaire WP (le WP3, le WP4, le WP5 et le WP7) selon le type de tâche et le niveau de complexité. Enfin, concernant le WP7 (Cf. Figure 17), il n'y avait pas d'effet principal significatif de la tâche (Table 4, partie A en Annexe 13 ; $X^2(4) = 5,944$; $p = 0,203$). De plus, malgré qu'il y avait un effet principal significatif du niveau de complexité sur ($X^2(2) = 15,296$; $p < 0,001$), les niveaux n'étaient pas significativement différents, même pour les niveaux 1 et 3 ($T(740) = 1,809$; $p = 0,071$).

En conclusion, les dimensions du WP ont une sensibilité différente à la complexité selon la tâche concernée. Dans notre contexte d'étude, le WP5 était une dimension importante, car quelle que soit la tâche, la charge mentale du WP5 était la plus élevée par rapport aux autres dimensions de ce questionnaire. Ce résultat s'expliquait par le fait que les tâches étaient réalisées sur écran avec un traitement visuel de l'information. De

plus, notre protocole à trois niveaux de complexité a généré trois niveaux distincts de WP5-visuel pour le NBT et le Corsi. C'était également le cas concernant le WP3-spatial pour le Corsi uniquement et concernant le WP4-verbal pour le WCST. Par ailleurs, le WP7-manuel était la seule dimension qui n'a pas été influencée par le niveau de complexité quel que soit le type de tâche. Un résultat à mettre au regard de la dimension *Exigences physiques* du NASA-TLX qui était la moins sollicitée des dimensions de ce questionnaire. Ces résultats s'expliquaient par le fait que les interactions manuelles étaient limitées lors de la réalisation des tâches.

De plus, concernant le WP3, la DT et le Corsi étaient significativement supérieurs au GNG. Ce résultat s'expliquait par le fait que ces deux tâches sollicitaient la mémoire visuo-spatiale. Concernant le WP4, le NBT et le WCST ont généré les charges mentales les plus élevées par rapport aux autres tâches. Ce résultat s'expliquait par le fait que ces deux tâches amenaient implicitement le participant à dire mentalement ou à voix haute l'information pour la garder "*en tête*", c'est-à-dire en mémoire de travail verbale. Enfin, concernant le WP7, c'était la seule tâche sans effet principal significatif de la tâche. Cela s'explique par le fait que toutes avaient un rendement manuel similaire (interactions limitées au clavier ou à la souris).

Toutes ces pistes d'explication seront davantage discutées lors de la Discussion de l'étude 1.

4.3.2 Clustering de la charge mentale subjective et vérification de la corrélation entre les classes subjectives et le niveau de complexité

Dans cette section d'analyse, nos résultats permettront de vérifier la seconde partie de l'hypothèse 1 de l'étude 1. Pour ces analyses, nous avons utilisé le logiciel JASP (version JASP 0.16).

Raisonnement de notre approche

La comparaison des moyennes de charge mentale à travers un regroupement basé sur les niveaux de complexité pourrait suggérer artificiellement la présence de différents niveaux de charge mentale en masquant le chevauchement des classes. Nous avons donc utilisé une autre méthode pour observer des regroupements (couramment nommés par l'anglais *clusterings*) de valeurs de charge mentale, non pas en fonction du niveau de complexité, mais en fonction de la proximité entre les valeurs observées. Les classes

obtenues ont ensuite été comparées à celles observées en considérant les niveaux de complexité.

Par conséquent, l'autre étape de l'analyse de nos résultats consistait à déterminer si nous pouvions obtenir trois niveaux de charge mentale sans tenir compte du niveau de complexité. Pour cela, nous avons utilisé la technique des K-means pour effectuer le *clustering*. Les études existantes (Al-Mohair et al., 2015 ; Shaheen et al., 2020) utilisant le *clustering* ont fait référence à K-means comme une technique simple, fiable et robuste de *clustering*. Le principe était que le *clustering* K-means divisait les données en K ensembles. Le nombre d'ensembles possibles dépendait de la nature des données ou de la possibilité plausible du nombre d'ensembles qu'une donnée peut avoir. Par conséquent, nous avons utilisé cette classification non supervisée, avec l'algorithme Hartigan-Wong, un paramètre par défaut dans le logiciel JASP. Nous avons observé le contrôle qualité des clusters avec trois critères. Les deux premiers étaient le Critère d'Information d'Akaike (AIC) et le Critère d'Information Bayésien (BIC) qui sont des méthodes analytiques pour estimer la qualité de l'ajustement des modèles statistiques les uns aux autres pour un ensemble de données (plus les valeurs sont basses, meilleurs sont les résultats du *clustering*). De plus, nous avons ajouté le score de Silhouette qui comparait la distance moyenne intra-cluster à la distance moyenne du cluster le plus proche (variant de -1 à 1, où 1 représentait un score parfait). Il était important d'avoir plusieurs critères de qualité, car cela nuancait les résultats. Enfin, pour s'émanciper des niveaux de complexité, on fait désormais l'hypothèse que l'on peut ici se défaire de la variable "individu" en considérant que la normalisation de Cousineau a suffisamment réduit les variabilités intra-individuelles. Par conséquent, nous n'avons pas 53 vecteurs associés à chaque niveau de complexité, mais 159 vecteurs liés à une tâche, sans information sur le niveau de complexité.

Sélection de variables pour le clustering K-means

Pour toutes les tâches, nous avons conservé l'ensemble des dimensions du NASA-TLX, car ce questionnaire doit être considéré dans son intégralité (Hart et Staveland, 1988) contrairement au questionnaire WP. En effet, en se référant à l'article princeps de ce questionnaire (Tsang et Velazquez, 1996, p. 362), il a été précisé que le WP peut être côté de manière unidimensionnelle. Par conséquent, avant l'étape de *clustering*, pour chaque tâche, nous avons identifié les dimensions du WP les plus pertinentes en fonction

de la logique de la tâche, c'est-à-dire des exigences cognitives qu'elle implique. Nous avons également considéré les résultats de l'analyse précédente précisant quelles dimensions du WP sont les plus affectées pour une tâche donnée. La sélection des dimensions subjectives pour chaque tâche est résumée dans la **Table 7**.

	WP3-spatiale	WP4-verbale	WP5-visuel	WP7-manuel
NBT		X	X	
Corsi	X		X	X
GNG			X	X
WCST	X	X	X	
DT	X	X	X	

Table 7 : Sélection des dimensions du WP qui seront utilisées par le clustering K-means.

Premièrement, pour toutes les tâches, le WP5 a été sélectionné, car il s'agissait d'une dimension importante. En effet, quelle que soit la tâche, la charge mentale basée sur le WP5 était la plus élevée par rapport aux autres dimensions du WP (Cf. *Questionnaire WP dans la section Résultats*).

Pour le NBT, les dimensions du WP retenues étaient le WP4 et le WP5. En effet, les comparaisons post hoc de Conover ont indiqué que la charge mentale basée sur le WP3 pour le NBT était significativement plus faible que pour toutes les tâches (excepté pour le GNG). De plus, le NBT présentait un ensemble de lettres au centre de l'écran sans mouvement spatial des stimuli et les interactions manuelles se limitaient au pavé directionnel (flèche gauche et droite uniquement) et ne faisaient pas partie de la logique de la tâche (le NBT sollicite uniquement de la mémoire de travail verbale).

Pour le Corsi, les dimensions du WP retenues étaient le WP3, le WP5 et le WP7. En effet, le Corsi avait un score de WP4 inférieur à toutes les autres tâches. De plus, la dimension verbale n'était pas censée entrer en jeu, car la tâche était basée sur la mémoire visuo-spatiale.

Pour le GNG, les dimensions du WP retenues étaient le WP5 et le WP7. Nous avons écarté le WP3, car cette tâche présentait un ensemble de stimuli au centre de l'écran (disques noirs et carrés noirs) sans mouvement spatial de ces stimuli. Nous avons également écarté le WP4, car le score pour cette dimension au GNG était significativement inférieur à celui du NBT et du WCST (tâches impliquant un traitement verbal).

Pour le WCST, les dimensions WP retenues étaient WP3, WP4 et WP5. Nous avons écarté WP7 car les interactions manuelles ne faisaient pas partie de la logique de la tâche (ce qui implique principalement une flexibilité mentale).

Enfin, pour la DT, les dimensions du WP retenues étaient le WP3, le WP4 et le WP5. En effet, la DT sollicitait la mémoire visuo-spatiale (WP3) lors du test de Corsi et la dimension verbale lors de la tâche de calcul. Nous avons écarté le WP7, car tous ses niveaux n'étaient pas significativement différents (Cf. Table 4 en Annexe 13).

Résultats

Après l'étape de sélection des dimensions du WP, nous avons réalisé un *clustering* K-means des données avec toutes les variables subjectives sélectionnées (Cf. **Table 7**) en plus des variables du NASA-TLX pour déterminer le nombre de clusters proposés associés aux scores des critères de qualité. Nous avons donc effectué le *clustering* K-means de trois manières. La première consistait à réaliser le modèle K-means en l'optimisant par le score de Silhouette, puis la valeur BIC et AIC avec une limite de 10 clusters (paramètre par défaut sur le logiciel JASP). Ensuite, nous avons comparé ces scores avec une autre façon de regrouper les données : en fixant $K = 3$ clusters.

À titre d'exemple, la **Figure 18** présente visuellement les chevauchements entre les trois clusters de charge mentale subjective (avec le NASA-TLX/WP5) pour le Corsi. Les résultats sont présentés dans un espace ACP (Analyses en Composantes Principales), car la représentation dans l'espace ACP permet de diminuer le nombre de dimensions des données. Effectivement, il y avait entre sept et neuf dimensions de charge mentale, ce qui était impossible à représenter tel quel sur un plan hors ACP. Ainsi, la partie A présentait les trois classes de charge mentale basées sur le K-means. La partie B présentait la même distribution de points que la partie A, mais à travers le niveau de complexité. Nous avons pu constater qu'entre les deux techniques, il y avait une bonne concordance dans le positionnement des individus statistiques dans les classes de charge mentale. Cependant, il y a peu ou pas de chevauchement de points dans la partie A. Concernant la partie B, nous avons observé quelques chevauchements, notamment aux limites des classes 1 et 3, qui étaient proches du niveau 2. Pour finir, le *clustering* en classes de charge mentale correspond bien à ce qui est attendu à travers les niveaux de complexité, malgré quelques divergences dans le niveau intermédiaire.

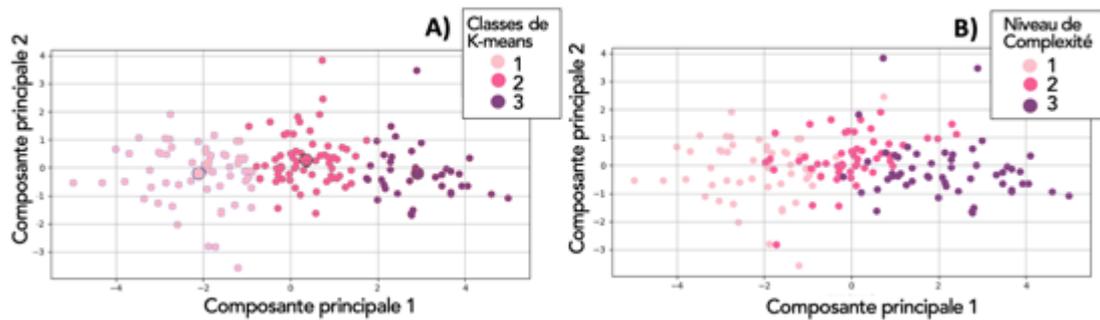


Figure 18 : Trois classes de charge mentale (basées sur K-means) dans l'espace de l'analyse en composantes principales (ACP) pour le test Corsi. Dans la partie A, visualisation indépendante du niveau de complexité. Dans la partie B, la même distribution de points (que dans la partie A), mais à travers le niveau de complexité.

Ainsi, pour chacune des variables subjectives sélectionnées, nous disposons de 4 X 3 scores avec 4 modes de *clustering* et 3 critères de qualité de *clustering* (Cf. **Table 8**).

Pour l'analyse des résultats, dans un premier temps, pour chaque critère de qualité, par exemple la valeur AIC, nous avons identifié parmi les variables subjectives retenues (Cf. **Table 7**) celle qui donnait le meilleur score de valeur AIC (cellules grises sur la **Table 8**). Puis, dans un second temps, nous avons identifié les variables subjectives sélectionnées, fixées à $K = 3$, qui donnaient les deux meilleures valeurs AIC. Les cellules vertes et bleu clair correspondaient respectivement à la meilleure variable subjective sélectionnée (basée sur l'AIC) fixées à 3 et à la seconde. Cette méthode a permis de sélectionner les variables subjectives qui donnaient les meilleurs scores pour $K = 3$ par rapport aux scores obtenus dans les mêmes conditions, mais avec un K optimisé.

Ensuite, nous souhaitons déterminer si les valeurs de charge mentale classées dans chacune des trois classes de *clustering* ont bien été obtenues dans les conditions censées induire ces niveaux de charge mentale. Ainsi, nous avons examiné s'il y avait une forte corrélation (allant de "0" - aucune corrélation à "100" - meilleure corrélation) entre l'affectation d'une donnée charge mentale dans une classe charge mentale par K-means et l'affectation des données par niveaux de complexité qui devait lui correspondre.

Concernant le NBT, nous avons comparé la fiabilité des classements sur la base des différentes combinaisons de variables subjectives sélectionnées sur la base du NASA-TLX, du WP4 et du WP5 (Cf. **Table 8**). Pour chaque sélection de variables, nous avons

considéré les critères de qualité de *clustering* pour chaque technique : lorsqu'elle était optimisée par le score de Silhouette, la valeur AIC et la valeur BIC. En comparant les valeurs AIC (plus les valeurs sont basses, meilleurs sont les résultats de clustering) entre une classification fixée à trois groupes ou optimisée par l'AIC, les résultats variaient de 513,520 (cellule ombrée dans le **Table 8**) à 790,680.

For NBT	Clusters	N	R ²	Clustering quality criteria			Subjective dimensions
				AIC value	BIC value	Silhouette score	
Optimize by AIC	9	159	0.604	644.050	865.010	0.120	NASA + WP4 + WP5
Optimize by BIC	5	159	0.510	699.340	822.100	0.160	NASA + WP4 + WP5
Optimize by Silhouette	2	159	0.311	902.460	951.560	0.270	NASA + WP4 + WP5
Fixed K=3	3	159	0.412	790.680	864.340	0.180	NASA + WP4 + WP5
Optimize by AIC	10	159	0.662	513.520	728.340	0.140	NASA + WP4
Optimize by BIC	6	159	0.581	547.360	676.260	0.160	NASA + WP4
Optimize by Silhouette	2	159	0.326	773.310	816.270	0.270	NASA + WP4
Fixed K=3	3	159	0.439	662.860	727.300	0.230	NASA + WP4
Optimize by AIC	10	159	0.662	513.540	728.360	0.160	NASA + WP5
Optimize by BIC	5	159	0.557	560.200	667.610	0.180	NASA + WP5
Optimize by Silhouette	2	159	0.346	751.810	794.770	0.290	NASA + WP5
Fixed K=3	3	159	0.456	644.160	708.600	0.210	NASA + WP5

Table 8 : Regroupement K-means pour le NBT. Les cellules grisées correspondent aux meilleurs scores pour chaque colonne de critères de qualité de clustering (valeur AIC, valeur BIC et score Silhouette). Pour chaque K=3 d'un critère de qualité de clustering, nous avons le vert (le meilleur score de qualité de clustering) et le bleu clair (le second). De plus, parmi les K = 3 fixés, la sélection de variables NASA/WP5, a obtenu le score AIC le plus bas (donc le meilleur score) avec 644,160 (ce qui était relativement proche de 513,520, le meilleur score avec optimisation AIC). Concernant la valeur BIC (plus les valeurs sont basses, meilleurs sont les résultats de clustering) avec le K = 3 fixé, la sélection de variables NASA/WP5, a obtenu le score BIC le plus bas avec 708,600 (ce qui était relativement proche de 667,610, le meilleur score avec optimisation BIC). Concernant le score de Silhouette (allant de « -1 » - mauvais clustering à « 1 » - très bon clustering) avec le K = 3 fixé, la sélection de variables NASA/WP4, a obtenu le score de Silhouette le plus élevé (donc le meilleur score) avec 0,230 (ce qui était relativement proche de 0,290, le meilleur score avec l'optimisation de Silhouette). Ainsi, nous avons pu obtenir trois classes de niveaux charge mentale avec des indicateurs de haute qualité pour les sélections de variables NASA-TLX/WP4 et NASA/WP5. Concernant le

coefficient de corrélation (**Table 13**) entre les classes charge mentale avec K-means et les classes charge mentale avec niveaux de complexité, nous avons obtenu respectivement 0,64 et 0,65 pour les sélections NASA/WP4 et NASA/WP5 respectivement.

Concernant le Corsi (Cf. **Table 9**), nous avons comparé la fiabilité des classements sur la base des différentes combinaisons de variables subjectives sélectionnées sur la base du NASA-TLX, du WP3, du WP5 et du WP7. Concernant la valeur de l'AIC avec $K = 3$ fixé (**Table 9**), la sélection de variables NASA/WP3, a obtenu le score AIC le plus bas avec 573,400 (ce qui était relativement proche de 457,160, le meilleur score avec optimisation AIC). Concernant la valeur BIC avec le $K = 3$ fixé, la sélection de variables NASA/WP3, a obtenu le score BIC le plus bas avec 637,840 (ce qui était relativement proche de 609,490, le meilleur score avec optimisation BIC). Concernant le score de Silhouette avec le $K = 3$ fixé, la sélection de variables NASA/WP3/WP5, a obtenu le score de Silhouette le plus élevé avec 0,260 (qui était relativement proche de 0,340, le meilleur score avec l'optimisation de Silhouette). Il a également été observé que la sélection de variables NASA/WP5 arrivait systématiquement à la deuxième place, tout en étant très proche des résultats de NASA/WP3. Ainsi, nous avons pu obtenir trois classes de niveaux charge mentale avec des indicateurs de haute qualité pour les sélections de variables NASA/WP3, NASA/WP5 et NASA/WP3/WP5. Concernant le coefficient de corrélation (**Table 13**) entre les classes charge mentale avec K-means et les classes charge mentale avec les niveaux de complexité, nous avons obtenu respectivement 0,79 ; 0,82 et 0,80 pour les sélections de variables NASA/WP3, NASA/WP5 et NASA/WP3/WP5 respectivement.

For Corsi	Clusters	N	R ²	Clustering quality criteria			Subjective dimensions
				AIC value	BIC value	Silhouette score	
Optimize by AIC	10	159	0.645	684.160	960.360	0.110	NASA + WP3 + WP5 + WP7
Optimize by BIC	4	159	0.515	761.130	871.610	0.170	NASA + WP3 + WP5 + WP7
Optimize by Silhouette	2	159	0.350	959.920	1015.160	0.280	NASA + WP3 + WP5 + WP7
Fixed K=3	3	159	0.471	806.110	888.970	0.230	NASA + WP3 + WP5 + WP7
Optimize by AIC	10	159	0.695	545.560	791.070	0.150	NASA + WP3 + WP5
Optimize by BIC	5	159	0.592	595.200	717.960	0.190	NASA + WP3 + WP5
Optimize by Silhouette	2	159	0.376	821.300	870.400	0.300	NASA + WP3 + WP5
Fixed K=3	3	159	0.510	667.420	741.080	0.260	NASA + WP3 + WP5
Optimize by AIC	9	159	0.653	583.060	804.020	0.160	NASA + WP3 + WP7
Optimize by BIC	5	159	0.558	638.470	761.220	0.180	NASA + WP3 + WP7
Optimize by Silhouette	2	159	0.358	843.770	892.870	0.290	NASA + WP3 + WP7
Fixed K=3	3	159	0.478	707.760	781.410	0.230	NASA + WP3 + WP7
Optimize by AIC	10	159	0.659	590.640	836.150	0.140	NASA + WP5 + WP7
Optimize by BIC	5	159	0.556	641.820	764.580	0.170	NASA + WP5 + WP7
Optimize by Silhouette	2	159	0.355	846.920	896.030	0.290	NASA + WP5 + WP7
Fixed K=3	3	159	0.476	710.040	783.700	0.230	NASA + WP5 + WP7
Optimize by AIC	9	159	0.701	457.160	650.500	0.150	NASA + WP3
Optimize by BIC	5	159	0.609	502.070	609.490	0.210	NASA + WP3
Optimize by Silhouette	2	159	0.395	697.340	740.300	0.340	NASA + WP3
Fixed K=3	3	159	0.520	573.400	637.840	0.250	NASA + WP3
Optimize by AIC	10	159	0.698	473.520	688.340	0.160	NASA + WP5
Optimize by BIC	5	159	0.607	504.290	611.700	0.190	NASA + WP5
Optimize by Silhouette	2	159	0.390	702.570	745.530	0.330	NASA + WP5
Fixed K=3	3	159	0.518	575.340	639.790	0.250	NASA + WP5
Optimize by AIC	10	159	0.689	483.950	698.770	0.160	NASA + WP7
Optimize by BIC	6	159	0.608	517.440	646.330	0.210	NASA + WP7
Optimize by Silhouette	2	159	0.371	723.850	766.810	0.310	NASA + WP7
Fixed K=3	3	159	0.491	605.250	669.690	0.240	NASA + WP7

Table 9 : Regroupement K-means pour le Corsi. Les cellules grisées correspondent aux meilleurs scores pour chaque colonne de critères de qualité de clustering (valeur AIC, valeur BIC et score Silhouette).

Concernant le GNG (Cf. Table 10), nous avons comparé la fiabilité des classements sur la base des différentes combinaisons de variables subjectives sélectionnées sur la base du NASA-TLX, du WP5 et du WP7. Concernant la valeur AIC et la valeur BIC avec K = 3 fixé (Table 10), la sélection de variables NASA/WP5, a obtenu les scores AIC et BIC les plus bas. Concernant le score de Silhouette avec le K = 3 fixé, la sélection

de variables de variables NASA/WP7, a obtenu le score de Silhouette le plus élevé. En ce qui concerne le coefficient de corrélation (**Table 13**) entre les classes de charge mentale avec K-means et les classes charge mentale avec les niveaux de complexité, nous avons obtenu 0,65 pour les sélections de variables NASA/WP5 et NASA/WP7.

For GNG	Clusters	N	R ²	Clustering quality criteria			Subjective dimensions
				AIC value	BIC value	Silhouette score	
Optimize by AIC	9	159	0.608	639.420	860.380	0.120	NASA + WP5 + WP7
Optimize by BIC	5	159	0.505	706.140	828.900	0.140	NASA + WP5 + WP7
Optimize by Silhouette	2	159	0.327	882.710	931.820	0.280	NASA + WP5 + WP7
Fixed K=3	3	159	0.426	773.510	847.170	0.260	NASA + WP5 + WP7
Optimize by AIC	9	159	0.632	532.560	725.900	0.130	NASA + WP5
Optimize by BIC	6	159	0.559	572.010	700.910	0.160	NASA + WP5
Optimize by Silhouette	2	159	0.356	740.450	783.410	0.310	NASA + WP5
Fixed K=3	3	159	0.453	647.420	711.870	0.250	NASA + WP5
Optimize by AIC	10	159	0.660	516.580	731.400	0.120	NASA + WP7
Optimize by BIC	6	159	0.573	556.680	685.570	0.180	NASA + WP7
Optimize by Silhouette	2	159	0.352	744.800	787.770	0.300	NASA + WP7
Fixed K=3	3	159	0.451	649.250	713.700	0.270	NASA + WP7

Table 10 : Regroupement K-means pour le GNG. Les cellules grisées correspondent aux meilleurs scores pour chaque colonne de critères de qualité de clustering (valeur AIC, valeur BIC et score Silhouette).

Concernant le WCST (Cf. **Table 11**), nous avons comparé la fiabilité des classements sur la base des différentes combinaisons de variables subjectives sélectionnées sur la base du NASA-TLX, du WP3, du WP4 et du WP5. Pour tous les critères de qualité de clustering, la sélection de variables NASA/WP5 a obtenu les meilleurs scores avec K = 3 fixé, suivie de près par la sélection de variables NASA/WP3 (**Table 11**). De plus, lorsque le clustering était obtenu par le score de Silhouette, K-means proposait K = 3 pour la sélection de ces deux variables. Par conséquent, nous avons pu disposer de trois classes de niveaux de charge mentale avec des indicateurs de haute qualité pour les variables NASA/WP3 et les variables NASA/WP5. Concernant le coefficient de corrélation (**Table 13**) entre les classes de charge mentale avec K-means et les classes de charge mentale avec des niveaux de complexité, nous avons obtenu respectivement 0,50 et 0,51 pour les sélections de variables NASA/WP3 et NASA/WP5.

For WCST	Clusters	N	R ²	Clustering quality criteria			Subjective dimensions
				AIC value	BIC value	Silhouette score	
Optimize by AIC	9	159	0.548	805.420	1054.000	0.120	NASA + WP3 + WP4 + WP5
Optimize by BIC	5	159	0.450	871.640	1009.740	0.140	NASA + WP3 + WP4 + WP5
Optimize by Silhouette	2	159	0.234	1124.910	1180.150	0.210	NASA + WP3 + WP4 + WP5
Fixed K=3	3	159	0.344	987.330	1070.200	0.190	NASA + WP3 + WP4 + WP5
Optimize by AIC	10	159	0.577	694.240	939.750	0.120	NASA + WP4 + WP5
Optimize by BIC	6	159	0.501	726.270	873.580	0.140	NASA + WP4 + WP5
Optimize by Silhouette	2	159	0.252	977.410	1026.510	0.210	NASA + WP4 + WP5
Fixed K=3	3	159	0.363	853.550	927.210	0.200	NASA + WP4 + WP5
Optimize by AIC	10	159	0.579	692.440	937.960	0.130	NASA + WP3 + WP4
Optimize by BIC	5	159	0.468	752.440	875.200	0.140	NASA + WP3 + WP4
Optimize by Silhouette	3	159	0.361	855.560	929.220	0.210	NASA + WP3 + WP4
Fixed K=3	3	159	0.361	855.470	929.130	0.210	NASA + WP3 + WP4
Optimize by AIC	10	159	0.585	685.160	930.670	0.110	NASA + WP3 + WP5
Optimize by BIC	5	159	0.485	731.320	854.080	0.150	NASA + WP3 + WP5
Optimize by Silhouette	3	159	0.374	839.180	912.830	0.210	NASA + WP3 + WP5
Fixed K=3	3	159	0.374	839.180	912.830	0.210	NASA + WP3 + WP5
Optimize by AIC	10	159	0.609	572.300	787.120	0.110	NASA + WP3
Optimize by BIC	5	159	0.508	614.170	721.580	0.150	NASA + WP3
Optimize by Silhouette	3	159	0.399	707.070	771.510	0.230	NASA + WP3
Fixed K=3	3	159	0.399	707.030	771.480	0.230	NASA + WP3
Optimize by AIC	10	159	0.622	558.200	773.020	0.130	NASA + WP4
Optimize by BIC	7	159	0.565	579.150	729.530	0.130	NASA + WP4
Optimize by Silhouette	2	159	0.274	830.920	873.880	0.230	NASA + WP4
Fixed K=3	3	159	0.387	720.100	784.550	0.220	NASA + WP4
Optimize by AIC	10	159	0.609	572.150	786.980	0.120	NASA + WP5
Optimize by BIC	5	159	0.513	608.460	715.870	0.170	NASA + WP5
Optimize by Silhouette	3	159	0.400	705.550	769.990	0.230	NASA + WP5
Fixed K=3	3	159	0.400	705.550	769.990	0.230	NASA + WP5

Table 11 : Regroupement K-means pour le WCST. Les cellules grisées correspondent aux meilleurs scores pour chaque colonne de critères de qualité de clustering (valeur AIC, valeur BIC et score Silhouette).

Enfin, *concernant la DT* (Cf. Table 12), nous avons comparé la fiabilité des classements sur la base des différentes combinaisons de variables subjectives sélectionnées sur la base du NASA-TLX, du WP3, du WP4 et du WP5. Pour tous les critères de qualité de *clustering*, la sélection de variables NASA/WP5 a obtenu les meilleurs scores avec K = 3 fixé, suivie de près par la sélection de variables NASA/WP3

(Table 12). Ainsi, en se passant des niveaux de complexité, nous avons pu disposer de trois niveaux de charge mentale avec des indicateurs de haute qualité pour les sélections de variables NASA/WP3 et NASA/WP5. Concernant le coefficient de corrélation (Table 13) entre les classes charge mentale avec K-means et les classes de charge mentale avec niveaux de complexité, nous avons obtenu 0,75 et 0,76 pour respectivement les sélections de variables NASA/WP3 et NASA/WP5.

For DT	Clusters	N	R ²	Clustering quality criteria			Subjective dimensions
				AIC value	BIC value	Silhouette score	
Optimize by AIC	8	159	0.523	822.590	1043.550	0.100	NASA + WP3 + WP4 + WP5
Optimize by BIC	6	159	0.470	861.100	1026.820	0.110	NASA + WP3 + WP4 + WP5
Optimize by Silhouette	2	159	0.282	1056.380	1111.620	0.260	NASA + WP3 + WP4 + WP5
Fixed K=3	3	159	0.358	967.420	1050.280	0.150	NASA + WP3 + WP4 + WP5
Optimize by AIC	10	159	0.586	683.090	928.600	0.130	NASA + WP4 + WP5
Optimize by BIC	6	159	0.496	733.390	880.700	0.120	NASA + WP4 + WP5
Optimize by Silhouette	2	159	0.301	915.520	964.620	0.280	NASA + WP4 + WP5
Fixed K=3	3	159	0.379	832.920	906.580	0.160	NASA + WP4 + WP5
Optimize by AIC	10	159	0.585	684.650	930.160	0.100	NASA + WP3 + WP4
Optimize by BIC	6	159	0.506	720.500	867.810	0.140	NASA + WP3 + WP4
Optimize by Silhouette	2	159	0.294	923.880	972.980	0.270	NASA + WP3 + WP4
Fixed K=3	3	159	0.377	835.610	909.260	0.160	NASA + WP3 + WP4
Optimize by AIC	9	159	0.562	697.550	918.510	0.130	NASA + WP3 + WP5
Optimize by BIC	5	159	0.472	747.210	869.970	0.140	NASA + WP3 + WP5
Optimize by Silhouette	2	159	0.309	905.340	954.440	0.280	NASA + WP3 + WP5
Fixed K=3	3	159	0.392	816.760	890.410	0.170	NASA + WP3 + WP5
Optimize by AIC	10	159	0.615	565.860	780.690	0.140	NASA + WP3
Optimize by BIC	6	159	0.532	602.000	730.890	0.120	NASA + WP3
Optimize by Silhouette	2	159	0.327	772.680	815.640	0.290	NASA + WP3
Fixed K=3	3	159	0.416	687.370	751.810	0.180	NASA + WP3
Optimize by AIC	10	159	0.634	544.880	759.700	0.150	NASA + WP4
Optimize by BIC	6	159	0.538	594.570	723.470	0.160	NASA + WP4
Optimize by Silhouette	2	159	0.318	782.500	825.470	0.280	NASA + WP4
Fixed K=3	3	159	0.406	699.340	763.790	0.180	NASA + WP4
Optimize by AIC	10	159	0.618	562.400	777.220	0.140	NASA + WP5
Optimize by BIC	5	159	0.503	619.570	726.980	0.150	NASA + WP5
Optimize by Silhouette	2	159	0.335	764.030	806.990	0.300	NASA + WP5
Fixed K=3	3	159	0.420	683.250	747.690	0.180	NASA + WP5

Table 12 : Regroupement K-means pour NBT. Les cellules grisées correspondaient aux meilleurs scores pour chaque colonne de critères de qualité de clustering (valeur AIC, valeur BIC et score Silhouette).

Combinaison subjective	NBT					
	NASA + WP5			NASA + WP4		
Niveaux de complexité	1	2	3	1	2	3
CM0	37	16	1	25	8	0
CM1	15	32	26	27	37	20
CM2	1	5	26	1	8	33
Corrélation de coefficient	0,65			0,64		

Combinaison subjective	Corsi								
	NASA + WP3			NASA + WP5			NASA + WP3 + WP5		
Niveaux de complexité	1	2	3	1	2	3	1	2	3
CM0	41	6	0	46	6	0	43	5	0
CM1	12	43	20	7	44	22	10	45	22
CM2	0	4	33	0	3	31	0	3	31
Corrélation de coefficient	0,79			0,82			0,8		

Combinaison subjective	WCST					
	NASA + WP3			NASA + WP5		
Niveaux de complexité	1	2	3	1	2	3
CM0	34	6	4	34	6	4
CM1	17	41	37	17	42	37
CM2	2	6	12	2	5	12
Corrélation de coefficient	0,5			0,51		

Combinaison subjective	GNG					
	NASA + WP5			NASA + WP7		
Niveaux de complexité	1	2	3	1	2	3
CM0	20	4	0	20	3	0
CM1	31	33	13	31	34	14
CM2	2	16	40	2	16	39
Corrélation de coefficient	0,65			0,65		

Combinaison subjective	DT					
	NASA + WP3			NASA + WP5		
Niveaux de complexité	1	2	3	1	2	3
CM0	39	2	0	39	1	0
CM1	13	36	17	14	37	18
CM2	1	15	36	0	15	35
Corrélation de coefficient	0,75			0,76		

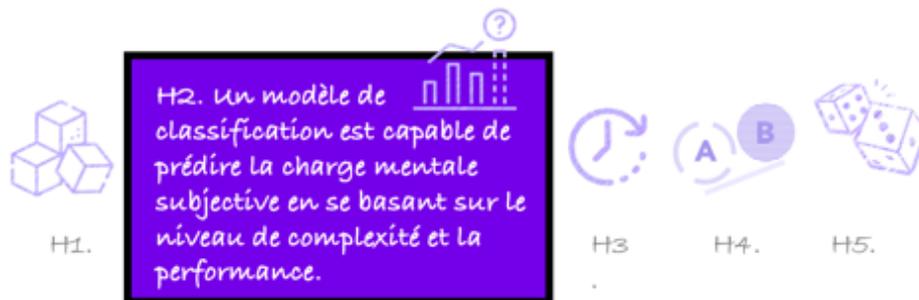
Table 13 : Coefficient de corrélation (Spearman) entre les classes K-means de charge mentale subjective et les niveaux de complexité pour toutes les tâches cognitives.

En conclusion, sans tenir compte des niveaux de complexité, nous avons pu obtenir trois niveaux de charge mentale subjective pour chacune des cinq tâches tout en ayant des critères de qualité de *clustering* élevés, c'est-à-dire des classes de charge mentale distinctes. Lorsque nous avons comparé chaque critère de qualité entre les tâches, nous avons observé que les scores variaient relativement peu. Cependant, pour les trois

critères de qualité de *clustering* (AIC, BIC et le score de Silhouette), le Corsi était systématiquement en tête des tâches et le WCST souvent en dernière position. De plus, si on comparait les classes de charge mentale obtenues via les labels de niveau de complexité et via le *clustering*, on observait une bonne correspondance, surtout pour le Corsi. Le *clustering* en classes de charge mentale correspondait bel et bien à ce qui est attendu à travers les niveaux de complexité, malgré quelques divergences dans le niveau intermédiaire. Une explication était que le Corsi était une tâche pour laquelle le plus de dimensions de charge mentale subjective permettaient cette distinction entre les classes (Cf. section 4.3.1). Une piste d'explication qui sera davantage discutée lors de la Discussion de l'étude 1.

4.3.3 Classification supervisée de la charge mentale subjective basée sur les niveaux de complexité et les performances

Dans cette section d'analyse, nos résultats permettront de vérifier la deuxième hypothèse (H2) de l'étude 1 présentée ci-après.



Pour ces analyses, nous avons utilisé le langage de programmation *Python* (version 3.9.7). Le code et une partie des analyses ont été réalisés par le stagiaire Thomas LeJan sous la supervision d'Isabelle Milleville-Pennel et de la doctorante.

Pour tester notre deuxième hypothèse opérationnelle, notre objectif était de proposer un modèle de classification supervisée de la charge mentale subjective, spécifique à chaque tâche cognitive. Pour chaque analyse, nous avons d'abord considéré le niveau de complexité comme un prédicteur du modèle.

Méthodes statistiques utilisées

Concernant la technique de classification supervisée, nous avons retenu l'Analyse Discriminante Linéaire (de l'anglais *Linear Discriminant Analysis* - LDA), car des études antérieures sur la charge mentale (via des données physiologiques), telles que

Cinaz et al., 2013 ou Mohanavelu et al. (2022) ont montré que cette méthode fournissait de meilleurs résultats de classification corrects par rapport à l'algorithme K-plus proche voisin (*K-Nearest Neighbor algorithm* - K-NN) ou la machine à vecteurs de support (*Support Vector Machine* - SVM). Par ailleurs, en plus de son faible coût de calcul (par rapport au SVM par exemple), la LDA a été utilisée dans le domaine des tâches cognitives (Abibullaev et An, 2012 ; Yoo et al., 2020). Enfin, l'hyperplan utilisé par la LDA, pour séparer les différentes classes, permettait d'identifier les caractéristiques qui maximisaient la variance inter-classes, tout en minimisant la variance intra-classes (Mohanavelu et al., 2022). C'était l'un des objectifs de notre étude.

Ainsi, dans notre modèle LDA, les variables prédictives (X) sont les différentes données de performance, précisées au niveau de la section 4.2.2 *Tâches et procédures* et le niveau de complexité. La variable à prédire (Y) concerne la classe de charge mentale dont les trois niveaux ont été établies précédemment par K-means. Les hyper-paramètres de la LDA sont déterminés par GridSearch. Le solveur utilisé peut être SVD (*Singular Value Decomposition*), LSQR (*Least Squares Solution*) ou Eigen (*Eigenvalues decomposition*). Le second hyper-paramètre à déterminer est le seuil "tol" permettant de dire si une valeur singulière est déterminante. De plus, pour évaluer les performances du modèle de classification LDA, nous avons utilisé la méthode de validation croisée. Dans le processus de validation croisée, le dataset global est sous-divisé en un dataset d'entraînement et un dataset de test. On entraîne donc le modèle sur le dataset d'entraînement et on le teste sur le dataset de test. On recommence ensuite le processus en établissant un nouveau dataset d'entraînement ainsi qu'un nouveau dataset de test. Le score de validation croisée global est alors la moyenne des scores de validation croisée obtenus à chaque itération. Ainsi, pour cette section, l'estimation de la charge mentale était faite en inter-sujets. Car, l'entraînement de la classification (*training test*) a été faite sur plusieurs sujets et le test (*testing test*) a été effectué sur un sujet non pris en compte dans l'entraînement.

De plus, afin de prévenir les problèmes qui pourraient être liés à des classes non-équilibrées en nombre et donc de mésestimer l'efficacité du modèle de classification, on

$$BalancedAccuracy = \frac{1}{K} \sum_{k=1}^K Recall_k$$

$$Recall_k = \frac{TP_k}{TP_k + FN_k}$$

décide d'utiliser le score de précision de *Balanced Accuracy* dont la formule, en multiclasse ($K \geq 2$) est donnée par la formule ci-après :

Notons “K” le nombre de classes. Cette métrique a également l'avantage de communiquer une information sur la bonne prédiction de chaque classe, en calculant, pour tout $k \leq K$, $Recall_k$, le rapport des vrais positifs sur la totalité des positifs, pour la classe k.

Nous considérons ensuite le score de précision obtenu avec la combinaison niveaux de complexité et performances et celui obtenu avec les combinaisons de performances seules. Concernant la sélection des variables de charge mentale, nous avons utilisé la meilleure sélection pour chaque tâche en fonction des résultats précédents (ANOVAs, K-means et coefficients de corrélation entre les classes de charge mentale avec K-means et les classes de charge mentale avec les niveaux de complexité). Il est important de noter que la variable à prédire, la charge mentale subjective, correspondait aux classes établies par K-means.

Nous avons ensuite calculé la contribution de chaque variable au modèle de classification. Pour cela, nous avons utilisé la méthode de *Permutation Feature Importance* (PFI). La PFI est définie comme la diminution du score du modèle lorsque la valeur d'une variable prédictive est mélangée de manière aléatoire. En effet, la PFI permute les fonctionnalités à chaque tour, puis supprime une fonctionnalité de la liste et associe l'augmentation de l'erreur en tant que rang à la fonctionnalité qui a été supprimée. L'inconvénient de cette approche est que la PFI ne nous indique pas le nombre de fonctionnalités à utiliser, elle indique simplement les mieux classées (Bel, 2020).

Résultats

Modèle de classification du NBT

En ce qui concerne cette tâche, les meilleures sélections subjectives étaient NASA-TLX/WP4 et NASA-TLX/WP5. Nous avons donc comparé ces deux sélections pour la LDA. Pour la première sélection, notre modèle a pu déterminer la classe de charge mentale subjective en se basant sur le *Niveau de complexité* et la performance avec le meilleur score de validation croisée égal à 57,6 % ($\pm 3,64$ %). Dans ce modèle, les mesures de performance avaient trois dimensions : *Fausses alarmes*, *Omissions* et *Temps de réaction pour tous les items*. Pour le NASA-TLX/WP5, le résultat était à peu près similaire avec 55,53 % ($\pm 4,53$ %) pour le meilleur score de validation croisée. Pour vérifier que notre modèle de classification de la charge mentale n'était pas biaisé par le fait d'être principalement basé sur une seule variable, nous avons considéré l'importance des différentes variables à travers la PFI (toutes les variables de performance combinées). Nous avons donc calculé l'importance des différentes variables sur le score de précision pour le NASA-TLX/WP4 (**Figure 19**). Nous avons observé que le modèle LDA était principalement basé sur le *Niveau de complexité* et les *Réponses attendues* pour la prédiction.

Modèle de classification du Corsi

Concernant cette tâche, les meilleures sélections étaient NASA-TLX/WP3, NASA-TLX/WP5 et NASA-TLX/WP3/WP5. Pour la seconde sélection, notre modèle a pu déterminer la classe de charge mentale subjective en se basant uniquement sur le *Niveau de complexité* avec le meilleur score de validation croisée égal à 79,97 % ($\pm 0,0$ %). Pour le NASA-TLX/WP3 et NASA-TLX/WP3/WP5, le résultat était à peu près similaire avec respectivement 77,92 % ($\pm 0,0$ %) et 79,73 % ($\pm 0,0$ %) dans le meilleur score de validation croisée. Au niveau de la **Figure 19**, nous avons observé que le modèle LDA pour NASA-TLX/WP5 était principalement basé sur le *Niveau de complexité* pour la prédiction.

Modèle de classification du GNG

Concernant cette tâche, les meilleures sélections ont été NASA-TLX/WP5 et NASA-TLX/WP7. Pour la première sélection, notre modèle a pu déterminer la classe de charge mentale subjective en se basant sur le *Niveau de complexité* avec le meilleur score de validation croisée égal à 51,35 % ($\pm 1,75$ %). Dans ce modèle, les mesures de performance avaient deux dimensions, les *Réponses attendues* et les *Omissions*. Pour la

sélection de variables NASA-TLX/WP7, le résultat était à peu près similaire avec 50,65 % ($\pm 1,71$ %) pour le meilleur score de validation croisée.

Au niveau de la **Figure 19**, nous avons observé que le modèle LDA pour NASA-TLX/WP5 était principalement basé sur le *Niveau de complexité*, les *Réponses attendues* et le *Temps de réaction pour tous les items* pour la prédiction.

Modèle de classification du WCST

Concernant cette tâche, les meilleures sélections étaient NASA-TLX/WP3 et NASA-TLX/WP5. Pour la première sélection, notre modèle a pu déterminer la classe de charge mentale subjective en se basant sur le *Niveau de complexité* avec le meilleur score de validation croisée égal à 53,63 % ($\pm 1,65$ %). La performance avait une dimension : le *Nombre d'erreurs*. Pour la sélection de variables NASA-TLX/WP5, le résultat était à peu près similaire avec 53,86 % ($\pm 1,64$ %) dans le meilleur score de validation croisée avec le Niveau de complexité et le *Nombre d'erreurs* également. Au niveau de la **Figure 19**, nous avons observé que le modèle LDA pour NASA-TLX/WP5 était principalement basé sur le *Niveau de complexité* et les *Réponses attendues* pour la prédiction.

Modèle de classification de la DT

Enfin, pour la DT, les meilleures sélections étaient NASA-TLX/WP3 et NASA-TLX/WP5. Pour la seconde sélection, notre modèle a pu déterminer la classe de charge mentale subjective en se basant sur le *Niveau de complexité* avec le meilleur score de validation croisée égal à 73,71 % ($\pm 0,0$ %). Pour la sélection de variables NASA-TLX/WP3, le résultat était à peu près similaire avec 72,97 % ($\pm 0,0$ %).

Au niveau de la **Figure 19**, nous avons observé que le modèle LDA pour NASA-TLX/WP5 était principalement basé sur le *Niveau de complexité* pour la prédiction.

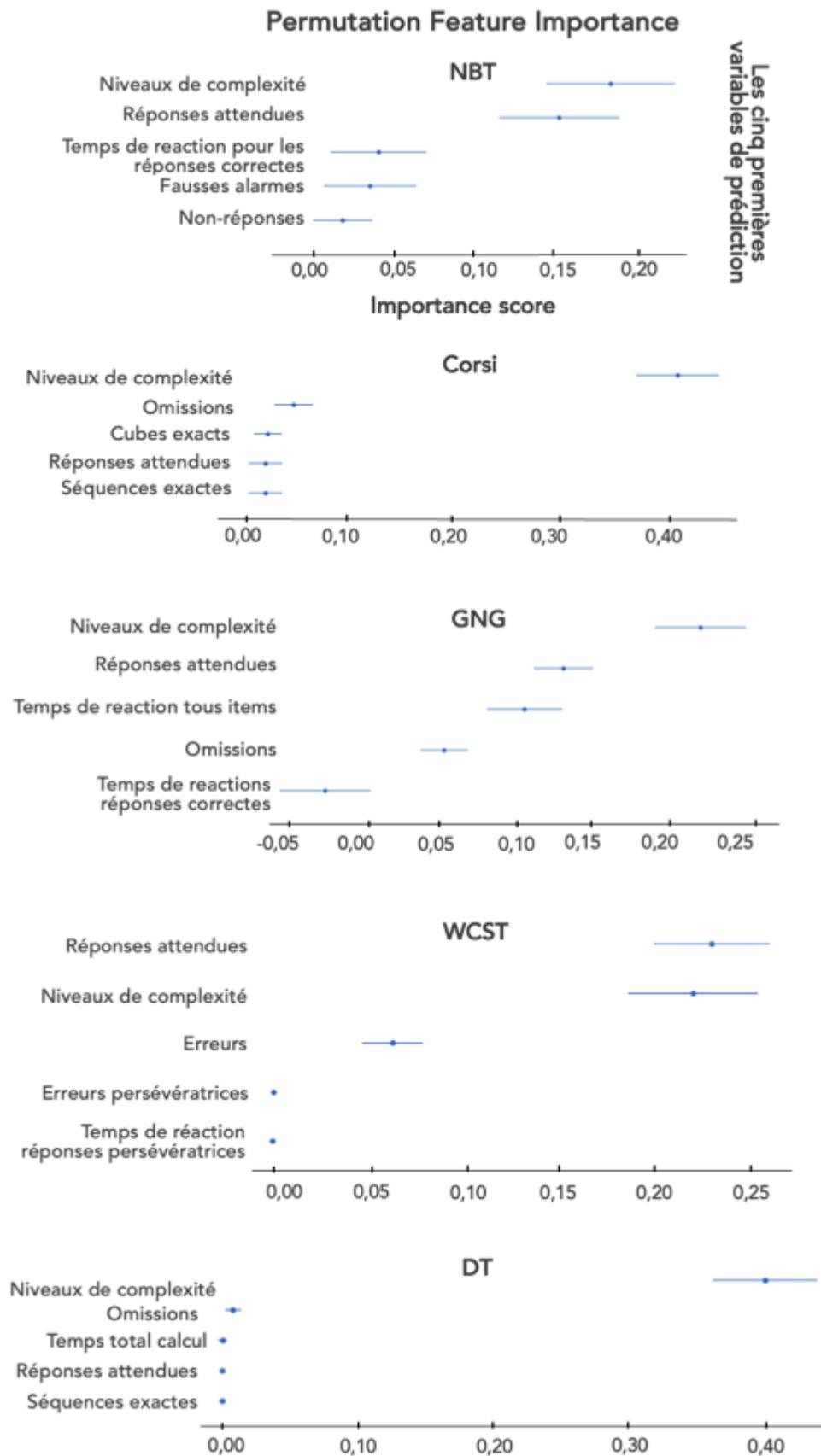


Figure 19 : Graphique à barres sur l'importance des cinq premières variables de performance pour chaque tâche cognitive.

En conclusion, nous avons pu prédire la charge mentale subjective bien au-delà du seuil de chance (égal à 33,3 %) en considérant le niveau de complexité et la performance comme prédicteurs des modèles de classification pour toutes les tâches cognitives. Notons que les meilleurs résultats de prédiction étaient basés sur les variables subjectives sélectionnées NASA-TLX/WP5 pour toutes les tâches, sauf pour le NBT. Le WP5-visuel semble donc bien être la variable du questionnaire WP la plus pertinente à considérer pour établir le modèle de classification. De plus, l'aspect visuel est prédominant dans la logique des tâches cognitives se déroulant dans un environnement numérique.

En outre, le Corsi et la DT étaient les tâches pour lesquelles on obtenait le meilleur résultat de classification, avec un score de validation croisée supérieur à 73 % ($\pm 0,0$ %). Cependant, pour ces deux tâches, lorsque le niveau de complexité était considéré, le modèle de classification de la charge mentale subjective reposait principalement sur lui (selon les PFI). À l'inverse, les autres tâches ont de moins bons résultats, mais l'importance des variables prédictives n'était pas principalement basée sur le niveau de complexité. Ces derniers résultats soutiennent les modèles théoriques mettant l'accent sur les exigences de la tâche dans la modélisation de la charge mentale. Une piste d'explication qui sera davantage discutée lors de la Discussion de l'étude 1.

4.3.4 Classification de la charge mentale subjective basée sur les performances uniquement

Procédure

Dans cette section d'analyse, nos résultats permettront de vérifier la troisième hypothèse (**H3**) de l'étude 1 précisée ci-après.



Pour ces analyses, nous avons utilisé le langage de programmation de *Python* (version 3.9.7). Le code et une partie des analyses ont été réalisés par le stagiaire Thomas LeJan

sous la supervision d'Isabelle Milleville-Pennel et de la doctorante. La procédure était la même que la section précédente, sauf que le modèle de classification était basé uniquement sur les différentes dimensions de performances.

Résultats

Concernant le NBT, sans le *niveau de complexité*, la précision de notre modèle pour NASA-TLX/WP4 était égale à 36,12 % ($\pm 2,64$ %). Dans ce cas, le modèle LDA a considéré un plus grand nombre de variables de performance comme étant utiles pour la prédiction (par rapport au modèle de classification prenant en compte également le niveau de complexité), qui sont les suivantes : *Non-réponses, Réponses attendues, Fausses alarmes, Omissions, Temps de réaction pour tous les items* et *Temps de réaction pour les réponses correctes*.

Concernant le Corsi, sans le *niveau de complexité*, la précision de notre modèle pour NASA-TLX/WP5 était égale à 53,28 % ($\pm 3,48$ %) avec quatre variables de performances du Corsi : *Faux cubes, Omissions, Séquences exactes, et Temps total*.

Concernant le GNG, sans le *niveau de complexité*, la précision de notre modèle NASA-TLX/WP5 était égale à 47,74 % ($\pm 1,9$ %) avec trois performances du GNG : *Temps de réaction pour les bonnes réponses, Temps de réaction pour tous les items* et *Réponses attendues*.

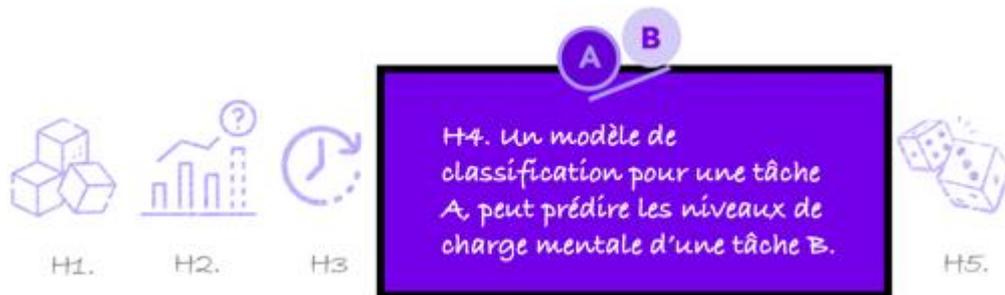
Concernant le WCST, sans le *niveau de complexité*, la précision de notre modèle NASA-TLX/WP5 était égale à 42,88 % ($\pm 3,43$ %) avec quatre performances du WCST : *Nombre d'erreurs, Erreurs persévératives, Temps de réaction pour tous les items* et *Réponses attendues*.

Concernant la DT, sans le *niveau de complexité*, la précision du modèle pour NASA-TLX/WP5 était égale à 44,74 % ($\pm 5,66$ %) avec six dimensions de performances de la DT : *Temps du calcul mental, Réponses attendues, Temps du Corsi, Cubes faux, Omissions* et *Séquences exactes*.

En conclusion, lorsque l'on considérait des modèles de classification basés uniquement sur des mesures de performance pour prédire la charge mentale subjective, le score de précision diminuait pour le Corsi et la DT, mais il restait au-dessus du seuil du hasard. Un résultat qui soutenait les modèles théoriques mettant en lien la charge mentale et les performances. Enfin, parmi les cinq tests cognitifs, le Corsi était la meilleure tâche pour prédire la charge mentale subjective en se basant sur le niveau de difficulté et/ou les performances. L'intérêt de ce résultat pour notre objectif de prédiction de la charge mentale en lien avec les manques énoncés au niveau de la section 3.2 sera discuté à la discussion de l'étude 2.

4.3.5 Transférabilité d'un modèle de classification inter-tâches

Dans cette section d'analyse, nos résultats permettront de vérifier la quatrième hypothèse (**H4**) de l'étude 1 présentée ci-dessous.



Nous nous sommes donc interrogés sur la transférabilité des modèles de classification d'une tâche à l'autre. Pour ces analyses, nous avons utilisé le langage de programmation de *Python* (version 3.9.7). Le code et une partie des analyses ont été réalisés par le stagiaire de l'entreprise onepoint Thomas LeJan sous la supervision d'Isabelle Milleville-Pennel et de la doctorante.

Pour tester l'hypothèse **H4**, nous avons dans un premier temps classer les tâches dans un tableau (Cf. **Table 14**) à deux entrées pour pouvoir réaliser des comparaisons par paires. Puis, en fonction de la logique de chaque tâche, nous avons identifié les dimensions communes en termes de charge mentale subjective et de performances. Le niveau de complexité faisait partie *de facto* d'une donnée commune à l'ensemble des tâches. Une fois les données communes identifiées, nous souhaitions savoir si un modèle entraîné sur une tâche A (en ligne sur la **Table 14**) pouvait déduire avec suffisamment

de précision les classes de charge mentale d'une tâche B (en colonne). C'était ce que nous avons appelé un modèle de classification inter-tâches.

De plus, pour cette section, l'estimation de la charge mentale était faite en inter-sujets. Car, l'entraînement de la classification (*training test*) a été faite sur plusieurs sujets et le test (*testing test*) a été effectué sur un sujet non pris en compte dans l'entraînement.

Index	Corsi	GNG	NBT	WCST	DT
Corsi	NASA-TLX; WP3; WP5; WP7; Cubes pointés; SeqExactes; Réponses Attendues; Niveau	NASA-TLX; WP5 Cubes Faux/Fausses Alarmes; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; WP3; WP5; WP7; Cubes faux; Séquences exactes; Cubes exacts; Temps total du Corsi; Niveau
GNG	NASA-TLX; WP5 Cubes Faux/Fausses Alarmes; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Fausses alarmes/Cubes Faux; Réponses attendues; Niveau
NBT	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; WP5; WP4 Nb d'erreurs/ Fausses Alarmes; Non réponses; Temps tous items cofondus; Niveau	NASA-TLX; Fausses Alarmes/Cubes Faux; Niveau
WCST	NASA-TLX; WP5; WP3; WP7; Fausses Alarmes/Nb d'erreurs; Réponses attendues; Niveau	NASA-TLX; WP5 Tps de réaction tous items cofondus; Nombre d'erreurs/ Fausses Alarmes; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; WP4; WP5; WP7; Nb d'erreurs/Cubes faux; Réponses attendues; Niveau
DT	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Fausses Alarmes/Cubes Faux; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX; Réponses attendues; Niveau	NASA-TLX-TLX; Réponses attendues; Niveau

Table 14 : Ensemble des données communes par paires de tâches cognitives.

Concernant les résultats, la **Table 15** s'appuyait sur les données entrantes de la **Table 14**. Chaque cellule de la **Table 15** correspond au score de précision des prédictions.

Index	Corsi	GNG	NBT	WCST	DT
Corsi	79.87	50.94	63.52	62.26	63.52
GNG	47.17	69.18	44.03	42.77	54.09
NBT	76.73	51.57	64.78	60.38	69.18
WCST	71.7	44.65	59.75	69.81	63.52
DT	73.58	57.23	61.01	55.97	71.7

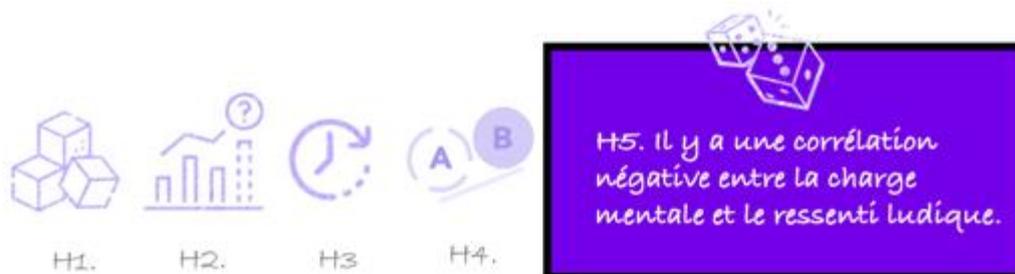
Table 15 : Score de précision des modèles de classifications. En ligne, modèle entraîné sur une tâche qui prédit les classes de charge mentale d'une autre tâche (en colonne).

Les résultats précisent que nous pouvons envisager la transférabilité d'un modèle de prédiction *inter-tâches*. On avait a minima un score de précision des prédictions égal à 44,03 pour un modèle entraîné sur le GNG qui prédit la charge mentale du NBT, ce qui est au-dessus du seuil du hasard (égal à 33,33). Les résultats les plus prometteurs, encadrés en bleu dans la **Table 15**, correspondent à des modèles entraînés sur le NBT, le WCST et la DT pouvant prédire avec suffisamment de précision (respectivement 76,73 ; 71,7 ; 73,58) les classes de charge mentale du test de Corsi, avec des résultats encourageant pour la première paire de tâches cognitives NBT/Corsi.

En conclusion, à partir d'autres modèles de classification, seul le Corsi permettait d'avoir des résultats encourageants dans la transférabilité inter-tâches. En effet, lorsque nous entraînions un modèle de classification sur une autre tâche cognitive (NBT, WCST, DT), nous pouvions prédire les classes de charge mentale subjective du Corsi. Ainsi, nous pouvons envisager d'entraîner un unique modèle de classification basé sur le NBT, par exemple, pouvant prédire la charge mentale du NBT et du Corsi. Les potentialités de ce résultat seront discutées davantage dans la discussion de l'étude.

4.3.6 Lien entre le ressenti ludique, la performance et la charge mentale subjective

Dans cette section d'analyse, nos résultats permettront de vérifier la dernière hypothèse de l'étude 1, la **H5**.



Pour cette sous-partie, nous avons utilisé des statistiques non paramétriques (tests de Friedman, de Wilcoxon et des comparaisons post hoc de Conover) avec le logiciel JASP (version JASP 0.16).

Tout d'abord, concernant le lien entre le ressenti *Ludique* et le *Niveau de complexité*, les résultats précisent qu'il n'y avait pas d'effet principal significatif du *Niveau de complexité* (Cf. **Figure 20** ; Table 5 en Annexe 13 ; $X^2(2) = 161,566$; $p < 0,001$). Ainsi, indépendamment de la tâche, aucun niveau n'était significativement différent. À l'inverse, il y avait un effet principal significatif du type de tâche (Table X en Annexe 13 ; $X^2(2) = 0,523$; $p = 0,770$). Le ressenti *Ludique* pour le WCST était significativement plus élevé que pour le NBT, le Corsi et le GNG. De plus, le ressenti *Ludique* pour le NBT était significativement plus faible que pour toutes les tâches, mise à part pour le GNG (Table X en Annexe 13).

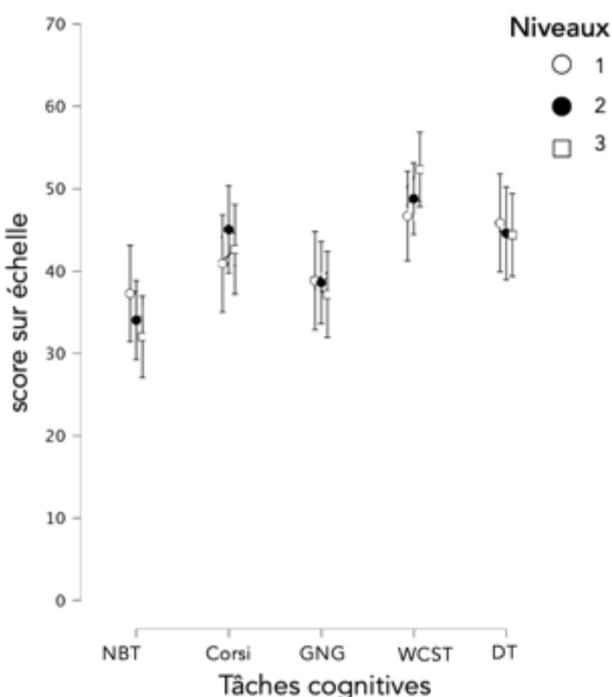
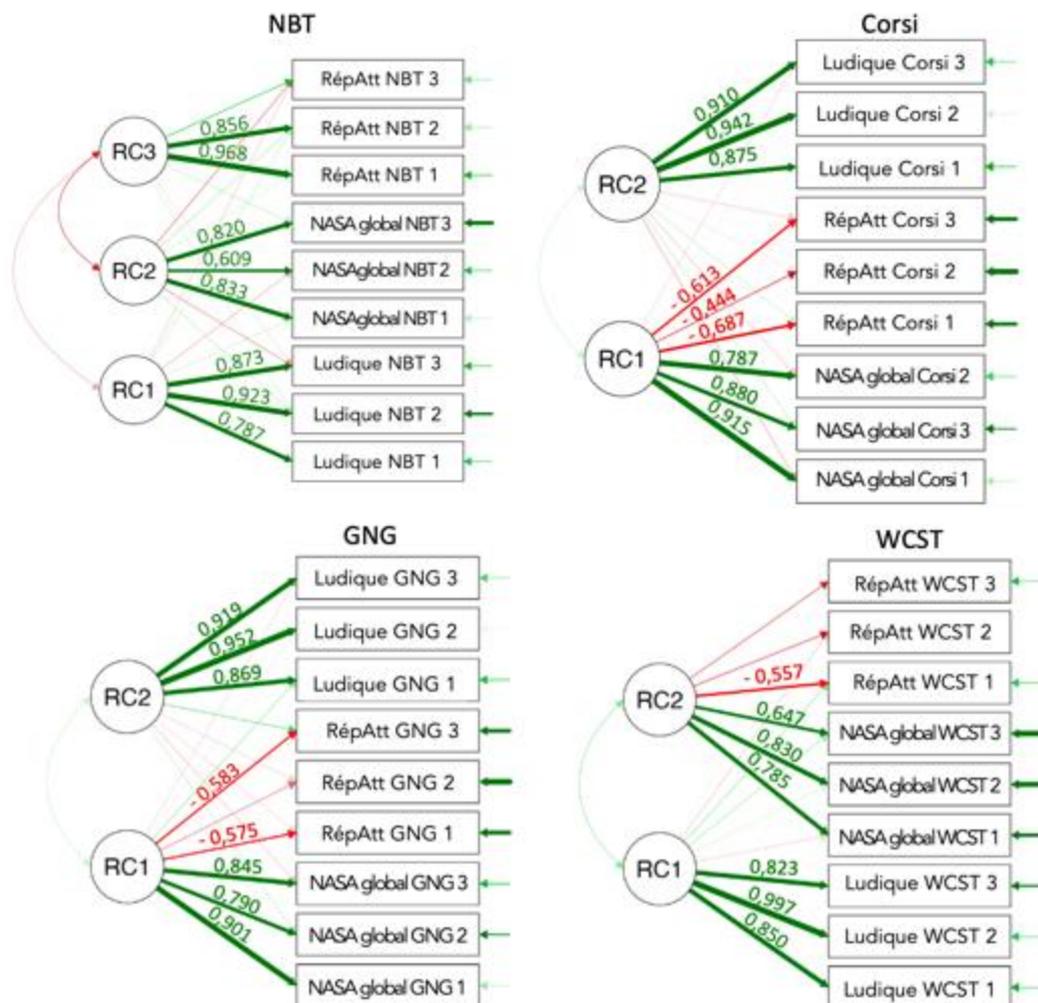


Figure 20 : Le ressenti ludique de la tâche en fonction du type de tâche et du niveau de difficulté.

De plus, pour vérifier plus précisément notre hypothèse **H4** sur le lien entre le ressenti *Ludique*, la charge mentale (en termes de charge mentale globale) et la performance (en termes de *Réponses attendues*), nous avons réalisé des Analyses Exploratoires Factorielles (AEF) avec le logiciel JASP (version JASP 0.16). L'AEF est une technique statistique permettant d'identifier les facteurs latents sous-jacents dans un ensemble de

données, en plus d'identifier les corrélations entre les variables. Plus précisément, l'AEF permet de construire un intervalle de confiance pour une expérience à mesures répétées. On peut alors comparer les moyennes entre les conditions, ce qui revient à déterminer un intervalle de confiance. Pour ces statistiques, nous avons utilisé les données non-normalisées, car l'AEF doit porter sur des scores bruts. Le cas contraire, on perd la valeur absolue de la variable qui est significative pour l'AEF.

Dans notre étude, l'AEF a révélé un schéma similaire pour le Corsi, le GNG, le WCST et la DT avec les variables *Réponses attendues* et le NASA-TLX qui étaient expliquées par le même facteur sous-jacent (**Figure 21**).



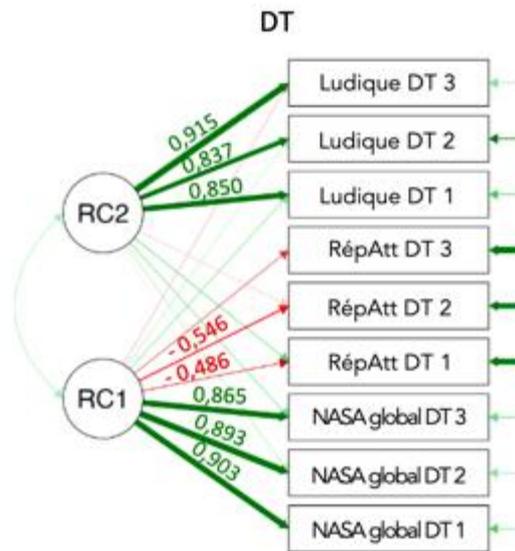


Figure 21 : Analyses Exploratoires Factorielles entre le ressenti ludique, le NASA-TLX global et la performance (en termes de Réponses attendues) pour chaque tâche cognitive. Les lignes vertes représentent les corrélations positives et les lignes rouges, les corrélations négatives.

De plus, pour chacune de ces quatre tâches, la variable *Réponses attendues* était corrélée négativement à ce facteur explicatif, alors que la variable NASA-TLX est corrélée positivement à ce facteur (**Figure 21**). Concernant la variable *Ludique*, elle était expliquée par un second facteur. Par conséquent, le ressenti *Ludique* était indépendant de la charge mentale globale et de la performance. Concernant le NBT, nous avons obtenu un profil différent avec trois facteurs sous-jacents pour chacune des trois variables. Ainsi, il n’y avait pas de corrélation négative entre la charge mentale globale et les *Réponses attendues*. Cependant, le ressenti *Ludique* était toujours une dimension indépendante.

En conclusion, pour les cinq tâches cognitives, relativement similaires en termes de stimuli et d’interactions, le ressenti ludique différait d’une tâche à une autre. Cependant, au sein d’une même tâche, il n’y avait pas de différence entre les niveaux de complexité. Une piste d’explication est que le ressenti ludique est une caractéristique intrinsèque à la tâche. De plus, le ressenti *Ludique* était une dimension indépendante de la charge mentale globale et des *Réponses attendues*. Ainsi, peu importe les performances ou la charge mentale subjective d’un opérateur, le ressenti qu’il a sur l’aspect *Ludique* d’une interface ne sera pas impacté. Un exemple concret de ce résultat est un participant qui réalise un jeu de cartes tout en étant en surcharge mentale avec des performances faibles

(corrélation négative entre ces deux concepts), il va tout de même considérer le jeu de cartes comme ludique. Les potentialités de ce résultat seront discutées davantage dans la discussion de l'étude.

4.4 Discussion de l'étude 1

Dans cette étude, nous avons plusieurs tâches cognitives avec des niveaux de complexité distincts. Pour rappel, les hypothèses suivantes ont été posées :

				
H1. Toutes les tâches cognitives fournissent différents niveaux de charge mentale subjective par rapport aux niveaux de complexité.	H2. Un modèle de classification est capable de prédire la charge mentale subjective en se basant sur le niveau de complexité et la performance.	H3. Un modèle de classification est capable de prédire la charge mentale subjective en ne considérant que la performance pendant la tâche.	H4. Un modèle de classification pour une tâche A peut prédire les niveaux de charge mentale d'une tâche B.	H5. Il y a une corrélation négative entre la charge mentale et le ressenti ludique.

Conception de tâches cognitives à trois niveaux de complexité

Tout d'abord, avant de discuter ces cinq hypothèses, il importe de vérifier si les trois niveaux de complexité génèrent trois niveaux de performances en termes de *Réponses attendues*. Comme prévu, c'est le cas pour le NBT, ce qui est conforme aux études précédentes (Arvaneh et al., 2015 ; Dimitrakopoulos et al., 2017 ; Ries et al., 2018). Dans ce contexte, dans un premier temps, le NBT sert de référence pour calibrer les niveaux de complexité des autres tâches. Pour au moins deux tâches (Corsi et WCST), plusieurs niveaux de complexité génèrent plusieurs niveaux de *Réponses attendues*. Concernant le GNG et la DT, les derniers niveaux de complexité ne sont pas différents en termes de *Réponses attendues*. Ces résultats peuvent s'expliquer par des niveaux d'exigences de la tâche demandés qui sont trop proches entre les derniers niveaux pour ces deux tâches. Par exemple, dans la DT, ajouter une multiplication (niveau de complexité le plus élevé) à une soustraction (niveau intermédiaire) n'a pas d'impact sur les performances. Ce résultat n'est pas en accord avec l'étude de So et al. (2017) qui a obtenu des différences significatives dans les mesures de performance entre un niveau intermédiaire de soustraction et un niveau supérieur de multiplication. Cependant, la limite avec l'étude de So et al. (2017) est qu'elle mêle au dernier niveau de complexité des opérations de multiplication et de division, ce qui rend difficile la distinction entre

les deux types d'opérations. Peut-être qu'une division aurait donné des différences plus importantes dans notre étude. Néanmoins, sur la base de notre phase de pré-test, nous avons observé que la division aurait conduit à une trop grande complexité par rapport aux autres tâches cognitives.

Néanmoins, concernant notre premier objectif, la priorité était d'obtenir trois niveaux de complexité capables de générer trois niveaux de *Réponses attendues*, ce qui est le cas pour le NBT, le Corsi et le WCST. Le GNG et la DT sont néanmoins intéressants à étudier pour vérifier la relation entre la performance et la charge mentale. Pour ces deux tâches, on pourrait observer une relation non-linéaire entre la charge mentale et les performances, comme soutenu par plusieurs auteurs (Reid et Nygren, 1988 ; Cain, 2007 ; Raufaste et al., 2004 ; Radüntz, 2017).

Discussion sur l'hypothèse 1

L'hypothèse **H1** porte sur la possibilité que les niveaux de complexité induisent des niveaux de charge mentale distincts. Par rapport à la définition de la charge mentale, qui est un concept multidimensionnel, nous considérons systématiquement deux questionnaires : le WP et le NASA-TLX.



La charge mentale basée sur le WP

Concernant les dimensions du WP, pour certaines tâches, tous les niveaux de complexité génèrent des classes de charge mentale distinctes, comme le traitement spatial WP3 pour le Corsi. Cependant, aucune dimension du WP ne peut discriminer tous les niveaux de charge mentale pour le GNG et la DT. Ces derniers résultats sont en contradiction avec l'étude de Rubio et al. (2004) qui observait que le WP était le seul questionnaire à révéler des différences dues à la complexité de la tâche. Néanmoins, bien qu'il soit moins discriminant au niveau de la complexité que prévu, le questionnaire WP permet de mesurer la proportion de ressources attentionnelles utilisées comme spécifié dans Tsang et Velazquez (1996). Par exemple, dans notre étude, le Corsi et la DT sont les tâches cognitives qui génèrent la charge mentale la plus élevée basée sur le traitement spatial. Ces résultats sont cohérents avec la logique des tâches qui sont basées sur la mémoire visuo-spatiale.

La charge mentale basée sur le NASA-TLX

La charge mentale globale

Concernant le questionnaire NASA-TLX, tous les niveaux de complexité ont généré des niveaux distincts de charge mentale globale pour toutes les tâches. Cependant, il y avait une non-équivalence entre les niveaux globaux de charge mentale d'une tâche à l'autre. Ainsi, les niveaux obtenus étaient relatifs, et non absolus et dépendaient donc de la tâche, même si les "ordres" de valeurs pour un niveau donné sont assez proches. Par ailleurs, toutes tâches confondues, les moyennes de charge mentale globale variaient entre environ 100 et 300 sur une échelle de Likert de 600. Ainsi, les trois niveaux de charge mentale globale ne se situaient pas dans la zone de charge mentale intermédiaire, mais dans la première moitié de la charge mentale, à l'intérieur d'une zone basse et intermédiaire de charge mentale. Ce n'est pas exactement ce que nous voulions obtenir au départ, mais ce résultat montre néanmoins que nous disposons de tâches permettant d'atteindre des niveaux de charge mentale assez proches, mais néanmoins distinguables, ce qui est souvent difficile à obtenir avec des échelles subjectives. Une fois que nous savons que nous pouvons atteindre ce niveau de précision dans la discrimination des niveaux de charge mentale, nous pouvons ajouter un niveau de complexité aux tâches. À cet effet, le Corsi est le meilleur candidat car, parmi toutes les tâches, il a permis d'obtenir les niveaux de charge mentale globale les plus distincts. C'est un premier argument pour ne pas considérer le NBT comme une tâche de référence pour la mesure et la modélisation de la charge mentale.

Les dimensions du NASA-TLX

En ce qui concerne la mesure de chacune des dimensions du NASA-TLX, on observe que les dimensions contribuent différemment au score de la charge mentale globale. Ce résultat est conforme aux conclusions précédentes (Rubio et al., 2004 ; Fallahi et al., 2016 ; Longo, 2017). Ainsi, les dimensions du NASA-TLX peuvent être utilisées pour déterminer quelles dimensions tirent vers le haut ou vers le bas la charge mentale globale. Par exemple, pour tous les niveaux de complexité du NBT, la dimension *Exigences physiques* est la plus faible par rapport aux autres dimensions du NASA-TLX. Ce résultat est en accord avec l'étude de Malakoutikhah et al. (2021) qui avait également comparé les six dimensions du NASA-TLX pour le NBT et a permis d'observer que la dimension *Exigences physiques* était la plus faible par rapport aux autres dimensions du

NASA-TLX. Ce résultat est cohérent avec les sollicitations de notre tâche, car les interactions manuelles se limitent aux interactions avec le clavier ou aux mouvements souris/pavé numérique. Concernant les interactions avec la souris, Besançon et al. (2017) indiquent que la souris consomme moins d'*Exigences physiques* qu'une technique d'interaction réelle, mais plus d'*Exigences mentales*. Ce constat est corroboré par notre résultat précisant que la dimension *Exigences physiques* est la plus faible par rapport aux autres et également le résultat lié aux *Exigences mentales*, qui est l'une des dimensions les plus élevées (avec l'*Effort mental*).

Concernant notre première hypothèse **H1** - Toutes les tâches cognitives fournissent différents niveaux de charge mentale subjective par rapport aux niveaux de complexité - elle est confirmée si l'on considère la charge mentale globale. Le Corsi et la DT sont les tâches les plus discriminantes et le WCST est la moins discriminante. Une piste d'explication peut se retrouver au niveau du questionnaire NASA-TLX, qui est connu pour montrer une forte corrélation avec la performance (Rubio et al., 2004). Cela peut expliquer pourquoi la charge mentale basée sur le NASA-TLX ne permet pas de faire ressortir trois niveaux distincts pour toutes les dimensions du GNG pour lesquelles les performances ne sont pas impactées par le troisième niveau de complexité. Par ailleurs, à travers le WCST qui induit trois niveaux de performance distincts, mais pas trois niveaux de charge mentale globale, nous observons la limite d'utiliser uniquement les performances pour la mesure de charge mentale. Les performances peuvent se dégrader à mesure que le niveau de complexité augmente, mais la charge mentale n'augmente pas de manière linéaire et inversement si l'on considère le GNG. Le WCST dans notre étude est un exemple de la relation non-linéaire entre la charge mentale subjective et celle issue de la performance. Une piste d'explication peut se trouver au niveau du caractère ludique de la tâche. En effet, le WCST est la tâche cognitive ayant un ressenti ludique le plus élevé par rapport aux autres tâches cognitives. On peut supposer qu'un ressenti ludique élevé diminue l'impact de la complexité de la tâche sur la charge mentale ressentie. Par ailleurs, le niveau 3 de la charge mentale globale du WCST est le plus faible par rapport aux niveaux 3 des autres tâches cognitives. En parallèle, le niveau 3 du ressenti ludique du WCST est plus élevé que les deux autres niveaux de cette tâche. On peut ainsi supposer qu'un ressenti ludique élevé peut éviter la survenue de la surcharge mentale ressentie.

Discussion sur les classes de charge mentale et les niveaux de complexité

Par la suite, nous avons observé des regroupements (couramment nommés par l'anglais *clusterings*) de valeurs de charge mentale, non pas en fonction du niveau de complexité, mais en fonction de la proximité entre les valeurs observées. Les classes obtenues ont ensuite été comparées à celles observées en considérant les niveaux de complexité. Nous avons ainsi considéré la corrélation entre les classes de charge mentale de valeurs obtenues avec une méthode de *clustering* (considérant uniquement les mesures subjectives de charge mentale comme variables) et les classes de charge mentale de valeurs obtenues avec les niveaux de complexité.

Premièrement, parmi les tâches, le Corsi permet le meilleur *clustering*, c'est-à-dire les classes de charge mentale les plus distinctes. Une explication est que cette tâche est celle pour laquelle le plus de dimensions de charge mentale permettent cette distinction entre les classes. D'où l'intérêt de considérer toutes les dimensions de la charge mentale indépendamment l'une de l'autre, plutôt que la valeur globale dans un modèle de classification, comme l'ont fait Radüntz (2017) ainsi que Guan et al. (2021). De plus, si on compare les classes de charge mentale obtenues via les labels de niveau de complexité ou via le *clustering*, on observe une bonne correspondance, surtout pour le Corsi. Ainsi, le Corsi permet le meilleur *clustering* en classes de charge mentale et ces classes correspondent bien à ce qui est attendu à travers les niveaux de complexité. Ces résultats constituent le deuxième argument pour ne pas considérer le NBT comme une tâche de référence. De plus, à notre connaissance, aucune étude antérieure n'avait confirmé l'observation de trois classes distinctes de charge mentale en combinant deux méthodes d'analyse comme nous l'avons fait en considérant la correspondance entre les classes de charge mentale avec K-means et les classes de charge mentale avec des niveaux de complexité. Les résultats obtenus montrent l'intérêt de procéder de cette manière pour valider l'observation de classes de valeurs distinctes.

Discussion sur l'hypothèse 2

L'hypothèse **H2** concerne la prédiction de la charge mentale subjective en se basant sur le niveau de complexité et la performance. Pour toutes les tâches, la précision du modèle de classification de la charge mentale subjective était supérieure à 50 %, donc au-dessus du seuil du hasard égal à 33 % pour trois classes de charge mentale. Avec une précision supérieure à 70 % pour le Corsi et



la double tâche dont le modèle de classification de la charge mentale subjective reposait principalement sur le niveau de complexité (selon le résultat des PFI). Ce résultat soutient les modèles théoriques qui mettent l'accent sur la contribution des exigences de la tâche comme le modèle *contraintes/astreintes* (Karasek, 1979 ; Schlegel, 1993 ; Raufaste et al., 2004 ; vu au Chapitre 1).

Par ailleurs, le Corsi est la meilleure tâche pour prédire la charge mentale en se basant sur les niveaux de complexité et les performances. Par conséquent, cette tâche est bien dimensionnée en termes de niveaux de complexité. Ainsi, le Corsi a un haut niveau de fiabilité pour induire a priori (au moins au début de l'activité) le niveau de charge mentale dans lequel l'individu doit se trouver. Ce résultat est le troisième argument pour ne pas considérer le NBT comme une tâche de référence.

Discussion sur l'hypothèse 3

Désormais, nous souhaitons savoir si un modèle de classification peut prédire la charge mentale subjective uniquement sur la base des données de performance. Nous avons donc essayé de contourner les limites de la correspondance entre la charge mentale et les données de performance en privilégiant le nombre et donc la sensibilité des mesures de performance. Néanmoins, lorsque les modèles prédictifs de la charge mentale sont basés uniquement sur des mesures de performance, sans inclure le niveau de complexité, les résultats diminuent pour toutes les tâches, même si le résultat reste assez bon pour le Corsi et la DT. Notons que le Corsi est la seule tâche avec une précision supérieure à 50 %, bien au-dessus du seuil de hasard. Les performances peuvent ainsi être des indicateurs de la charge mentale. Un résultat qui soutient les modèles théoriques mettant en lien la charge mentale et les performances comme celui de la représentation en U inversé de la fonction de Hebb (1955) appliquée au concept de la charge mentale (Boyer, 2016 ; Cf. Chapitre 1). Cependant, nous nuancions ces modèles théoriques, car les dimensions de performances doivent être nombreuses et variées pour pouvoir correctement être liées aux changements de la charge mentale subjective. En effet, le Corsi et la DT ont un nombre de dimensions de performance supérieures aux autres tâches (7 et 9 pour le Corsi et la DT respectivement). Ainsi, multiplier les données de performance est un bon moyen de rendre cet indicateur plus fiable et plus sensible.



Néanmoins, même si, pour le Corsi et la DT, le niveau de prédiction est encore bien supérieur au niveau de hasard, on ne peut pas envisager de monitorer et adapter le niveau de complexité uniquement sur la base de ces modèles de classification. Ces résultats tendent à soutenir les études indiquant que les mesures de performance ne peuvent pas, à elles seules, décrire précisément la charge mentale, car les participants peuvent varier leurs efforts pour maintenir un niveau de performance constant (Reid et Nygren, 1988 ; Raufaste et al., 2004 ; Cain, 2007 ; Radüntz, 2017). Bien que les mesures de performance puissent rendre compte de la charge mentale en cours d'activité, il faut être conscient de leur limite. Ainsi, pour améliorer notre modèle, nous devrions le compléter par d'autres mesures de la charge mentale. Bien que les mesures psychophysiologiques aient plusieurs contraintes, elles peuvent être utiles pour améliorer nos modèles de classification. En effet, il pourrait être possible de trianguler les trois catégories de mesures (subjectives, de performance et psychophysiologiques) pour comprendre davantage la charge mentale d'un opérateur (Cf. *Chapitre 2 - Mesures de la charge mentale* ; Charles et Nixon, 2019 ; Longo, 2022). De la sorte, considérer la charge mentale à travers le prisme de ces trois catégories de mesures permettrait de dépasser les limites de chaque mesure. Une triangulation de mesures que nous avons tenté de mettre en place lors de l'étude 2, où nous avons recueilli les données issues de la charge mentale subjectives (le NASA-TLX et le WP), de performances et d'une mesure physiologique (l'EEG).

Discussion sur l'hypothèse 4

L'hypothèse **H4** concerne la transférabilité d'un modèle de classification d'une tâche A pour prédire la charge mentale subjective d'une tâche B. Les résultats précisent que nous pouvons envisager la transférabilité d'un modèle de classification inter-tâches uniquement pour certaines tâches cognitives. Ce résultat confirme de précédentes études (Walter et al., 2013 ; Appel et al., 2021) qui observaient une transférabilité d'un modèle de classification de la charge mentale inter-tâches (dont une tâche cognitive comme le NBT). Cependant, ces deux études (Walter et al., 2013 ; Appel et al., 2021) concernaient la classification de la charge mentale physiologique (issue de données EEG ou oculaires) et non subjective. Après une revue de la littérature, nous n'avons pas identifié d'étude ayant



réalisé de la classification de la charge mentale subjective (à trois classes) inter-tâches cognitives.

Concernant notre étude, à partir d'autres modèles de classification, seul le Corsi permet d'avoir des résultats encourageants dans la transférabilité inter-tâches. En effet, lorsque nous entraînons un modèle de classification sur une autre tâche cognitive (NBT, WCST, DT), nous pouvons prédire les classes de charge mentale subjective du Corsi. Ainsi, nous pouvons envisager d'entraîner un unique modèle de classification basé sur le NBT, par exemple, pouvant prédire la charge mentale du NBT et du Corsi. Effectivement, le meilleur résultat concerne le NBT qui peut prédire avec suffisamment de précision les classes de charge mentale du Corsi. Un résultat étonnant, car ces deux tâches ont des structures différentes. La première est une tâche séquentielle où les stimuli et le temps de réponse sont fixés, contrairement au Corsi où le participant peut prendre le temps qu'il souhaite pour répondre (même si la consigne précise qu'il doit aller le plus vite possible). Une piste d'explication est que ce sont des tâches sollicitant toutes deux la mémoire de travail contrairement au GNG (inhibition mentale) ou au WCST (flexibilité mentale) par exemple. Une autre piste d'explication se trouve au niveau des dimensions du NASA-TLX qui sont sollicitées pour les deux tâches. En effet, l'*Exigence mentale* et l'*Effort mental* sont significativement différents pour le NBT et le Corsi. Or, comme précisé plus haut dans la *Discussion*, ces deux dimensions sont les plus sollicitées dans la réalisation de tâches cognitives. Au vu de ce résultat encourageant pour plusieurs tâches cognitives, une potentialité est de pouvoir entraîner un modèle de classification unique avec une tâche cognitive A sur un échantillon de participants. Puis, de prédire la charge mentale subjective d'un participant donné lorsqu'il réalise la tâche A et la tâche B. On pourrait donc avoir un modèle de classification "générique" de la charge mentale subjective pour plusieurs tâches. Cette perspective permettrait d'économiser le temps de mise en place, ou encore le temps d'entraînement d'un modèle de classification supervisé. Cependant, la transférabilité reste pour l'instant au stade d'une perspective envisageable. Il faut encore améliorer la fiabilité des modèles de classification inter-tâches, avant d'envisager des modèles génériques. Pour finir, afin d'étudier un autre aspect de la transférabilité d'un modèle de classification, il nous reste à tester la transférabilité inter-contextes permettant de prédire les classes de charge mentale d'une tâche A d'un contexte B à partir d'un modèle de classification de d'une tâche A d'un contexte A. Ce qui sera réalisé lors de la seconde étude expérimentale de la thèse.

Discussion sur l'hypothèse 5



Tout d'abord, cette étude permet d'observer que le ressenti ludique est une dimension indépendante de la charge mentale globale et de la performance en termes de *Réponses attendues*. Notre hypothèse de départ concernant la corrélation négative entre la charge mentale et le ressenti ludique est donc à nuancer. En effet, l'individu peut avoir un ressenti ludique fort, tout en ayant une charge mentale élevée. Cependant, nous supposons que l'effet de la complexité de la tâche sur la charge mentale est minimisé, voire annihilé lorsque le ressenti ludique est élevé. Effectivement, par rapport aux autres tâches cognitives, le WCST a le ressenti ludique le plus élevé, tout en ayant les niveaux 2 et 3 de la charge mentale globale subjective qui n'étaient pas significativement différents (et bien qu'ayant des performances différentes entre les deux niveaux de complexité).

De plus, pour cinq tâches cognitives, relativement similaires en termes de stimuli et d'interactions, le ressenti ludique diffère. Le WCST est la tâche considérée comme la plus ludique en raison de l'aspect "jeu de cartes" ayant plusieurs couleurs et formes qu'on retrouve dans les jeux d'enfants. À l'inverse, la tâche de NBT est jugée la moins ludique, en raison de la mémorisation de lettres qui n'ont pas de sens entre elles. Cela est aussi dû à l'aspect "tâche expérimentale" que le NBT peut avoir, et non "jeu", que peut avoir le WCST. Une piste d'explication est que l'apparence de la tâche, autrement dit le design, impacte le ressenti ludique. Une autre piste d'explication est que le ressenti ludique ne provient pas du design de l'activité (variable extrinsèque aux exigences de la tâche), mais des mécaniques de l'activité (variable intrinsèque aux exigences de la tâche). À la suite de nos résultats, le ressenti ludique ne peut provenir du niveau de complexité (variable intrinsèque aux exigences de la tâche). En effet, nous observons qu'au sein d'une même tâche cognitive, il n'y a pas de différence au niveau du ressenti ludique entre les niveaux de complexité. Le ressenti ludique peut donc provenir d'une autre variable intrinsèque aux exigences de la tâche, impactant directement les mécanismes de résolution de l'activité en lien avec les processus cognitifs sollicités. Une manière de tester cette hypothèse pourrait être d'imbriquer entre elles deux tâches cognitives pour lesquelles le ressenti ludique est connu. Ainsi, les structures des deux tâches seraient modifiées tout en permettant d'évaluer l'impact de ce changement sur le

ressenti ludique de la nouvelle tâche combinée. Une hypothèse que nous essayerons de vérifier dans l'étude 2 où nous avons mesuré le ressenti ludique, mais également manipulé le design ludique (variable extrinsèque aux exigences de la tâche), et la multiplicité des fonctions cognitives (variable intrinsèque aux exigences de la tâche).

Apport de l'étude et applications possibles

Enfin, l'ensemble de ces résultats permettent de contribuer à de futures applications existantes, comme celles identifiées au *Chapitre 3* que nous présentons de nouveau sur la **Figure 22**. À l'appui de nos résultats, seul le Corsi pourrait être une tâche cognitive pertinente pour pouvoir tester la faisabilité de l'application 1 correspondant à la possibilité de situer un individu par rapport à son groupe de référence. Pour ce faire, une étude d'étalonnage est nécessaire pour confirmer cette hypothèse applicative. Or, ce n'est pas le corps du questionnement de la thèse. Ainsi, la doctorante ne le fera pas durant son doctorat. Cependant, c'est une application qu'il serait intéressant de tester par la suite.

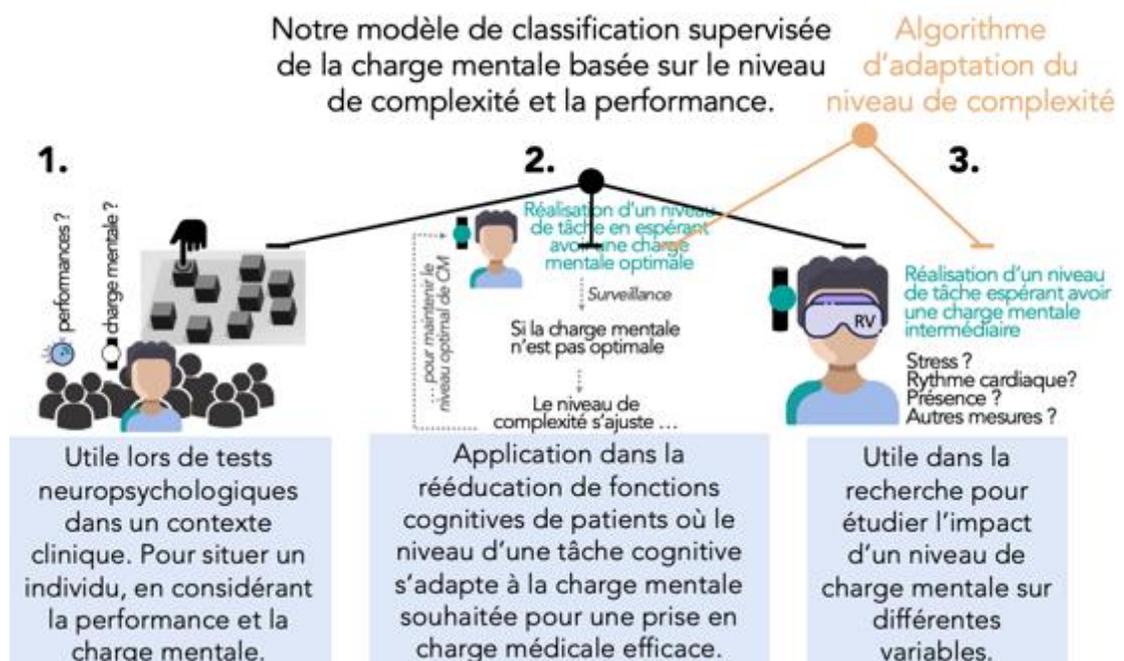


Figure 22 : Applications possibles liées aux méthodes statistiques pour y arriver.

Concernant les deux autres applications, la deux et la trois (Cf. **Figure 22**), le Corsi serait également une bonne tâche candidate. Pour le confirmer, il faudrait réaliser plusieurs algorithmes de modulation du niveau de complexité, les pré-tester, puis mener

une étude avec le meilleur algorithme de modulation du niveau de complexité. Étant donné que la thèse n'est pas centrée sur le développement algorithmique, nous n'avons pas lancé une étude sur ce point précis, mais nous avons tout de même réfléchi à une piste à explorer. À titre d'exemple, la **Figure 23** illustre un premier algorithme de modulation du niveau de complexité qui prend en entrée le niveau de complexité, la charge mentale estimée et la précision de la prédiction pour déterminer dans quel niveau de complexité la tâche cognitive continue à évoluer.

Le début du code présenté en **Figure 23** signifie de manière littérale ceci : *“Si la précision de la prédiction de la charge mentale est supérieure à 80 % et si la charge mentale est faible : 0 (ce qui est minimum) et aussi si le niveau de complexité est inférieur à 3, alors le niveau de complexité augmente (+1).”*

```
def LVL_MODULATION(MWL, PROBA, LVL):  
    if PROBA[MWL] > 0.8:  
        if MWL == 0:  
            if LVL < 3:  
                LVL = LVL + 1  
        if MWL == 2:  
            if LVL > 1:  
                LVL = LVL - 1  
    else:  
        l = PROBA.copy()  
        l[MWL] = 0  
        new_MWL = np.argmax(l)  
        if new_MWL == 0:  
            if LVL < 3:  
                LVL = LVL + 1  
        if new_MWL == 2:  
            if LVL > 1:  
                LVL = LVL - 1  
    return(LVL)
```

Figure 23 : Version 1 de l'algorithme d'adaptation du niveau de complexité.

Limites

Avant de conclure, il est important de préciser que notre expérience a été menée dans un environnement en ligne. Les résultats pourraient donc être différents avec des tâches réalisées dans un environnement plus standardisé de type laboratoire. Ainsi, lors de notre seconde étude (Cf. *Chapitre 5*), nous avons testé la transférabilité de modèles de classification de l'étude 1 réalisée en ligne et de l'étude 2 réalisée en présentiel. Autre limite, le modèle de charge mentale pourrait être enrichi par d'autres variables qui pourraient l'influencer, telles que des variables externes (comme le design de la tâche) à l'individu. L'impact du design (à travers l'aspect ludique) sur la charge mentale est une question de recherche que nous avons étudiée lors de la seconde étude. De même, pour une autre variable externe à l'individu, correspondant à la multiplicité des fonctions cognitives, que nous avons étudiée lors de la deuxième étude.

Conclusion

Pour conclure, nous avons testé cinq tâches cognitives sans contexte applicatif qui, toutes, fournissaient trois niveaux distincts de charge mentale globale. De plus, grâce à des méthodes statistiques combinées, nous avons pu valider l'obtention de trois classes de charge mentale associés à trois niveaux de complexité. Parmi nos cinq tâches, le Corsi a obtenu différents niveaux de charge mentale de manière suffisamment fiable pour que nous puissions faire des prédictions sur la charge mentale attendue. Par conséquent, cette étude fournit une base pour les recherches futures qui souhaiteraient prédire le niveau de charge mentale avant l'activité et ajuster le niveau de complexité pour conserver le niveau de charge mentale souhaité. En outre, grâce à nos analyses, nous avons pointé les limites du NBT qui ne devrait pas être une tâche de référence pour modéliser la charge mentale, comme l'ont fait plusieurs auteurs (Dimitrakopoulos et al., 2017 ; Beh et al., 2021 ; Malakoutikhah et al., 2021). Par ailleurs, parmi nos cinq tâches cognitives, le Corsi était la meilleure candidate pour la prédiction de la charge mentale subjective, lorsque le modèle de classification se base sur le niveau de complexité et/ou les performances. Le modèle que nous avons proposé pour le Corsi est encore faible pour permettre le suivi de la charge mentale en ligne, mais il reste néanmoins très prometteur et peut être amélioré.

Ainsi, notre prochain objectif est de permettre d'améliorer ce modèle avec des mesures physiologiques (comme les données EEG) de charge mentale disponibles pendant une tâche. Un ajout que nous avons réalisé lors de l'étude 2 (Cf. *Chapitre 5*), bien que l'intégration de cette mesure physiologique ne soit pas considérée dans le cadre de cette thèse, car elle nécessite de fortes compétences en traitement du signal et entraînant des délais de traitement trop élevés.



Figure 1 : Plan général du manuscrit de thèse et détail des objectifs du chapitre 5 présentant l'étude 2.

Un certain nombre de questionnements ont découlé des résultats de l'étude 1 (Cf. section 4.4 Discussion). Les objectifs (Cf. **Figure 1**) de cette seconde étude sont d'y répondre.

Après avoir montré la faisabilité de la transférabilité inter-tâches lors de l'étude 1, nous souhaitons décliner la question de la transférabilité d'un modèle de classification selon le contexte expérimental. Ainsi, le premier objectif de l'étude 2 est de tester la transférabilité d'un modèle de classification de la charge mentale subjective entre une passation en ligne et une passation dans un contexte en laboratoire (classification inter-contextes).

Le second objectif concerne l'approfondissement de la modélisation de la charge mentale, en se basant sur le modèle théorique proposé au Chapitre 1, notamment sur les variables internes à l'individu comme le ressenti ludique. La prise en compte du ressenti ludique fait également partie du souhait de mesurer une variable interne à l'individu liée à l'engagement (notion centrale dans les applications identifiées au Chapitre 3). Une mesure de l'engagement qui se fera à travers le ressenti ludique et également à travers la motivation de l'individu à réaliser la tâche.

Lors de la première étude, nous avons conclu qu'en fonction de la tâche, le ressenti ludique différait. Le WCST (jeu de cartes avec des couleurs) était la tâche considérée comme la plus ludique. À l'inverse, le NBT (défilement de lettres noires) était jugée la moins ludique. Une piste d'explication est que l'apparence de la tâche, autrement dit le design, impacte le ressenti ludique. Une hypothèse que nous essayerons de vérifier lors de cette seconde étude, en comparant un design classique avec un design ludique. En outre, nous souhaitons identifier les limites du design ludique sur une dimension donnée, la couleur. Autrement dit, nous souhaitons connaître l'impact de l'accentuation d'un critère du design ludique, comme la multitude de couleurs des items. Ainsi, nous comparerons un design ludique avec des items unicolores et un design ludique avec des items multicolores. Nous posons l'hypothèse qu'il y aurait plus de ressenti ludique pour la situation avec des items multicolores. Effectivement, Goethe (2019) a indiqué que la couleur a un effet sur nos émotions. Ainsi, la couleur est une première étape vers un design ludique. Donc, en augmentant les couleurs, on devrait accentuer les émotions et donc le ressenti ludique.

Une autre piste d'explication sur la différence de ressenti ludique entre des tâches cognitives (un des résultats de l'étude 1) est que le ressenti ludique ne provienne pas du design de l'activité (variable extrinsèque aux exigences de la tâche), mais des mécaniques de l'activité (variable intrinsèque aux exigences de la tâche). À la suite des

résultats de l'étude 1, le ressenti ludique ne peut provenir du niveau de complexité (variable intrinsèque aux exigences de la tâche). En effet, nous avons observé qu'au sein d'une même tâche cognitive, il n'y avait pas de différence au niveau du ressenti ludique entre les niveaux de complexité. Le ressenti ludique peut donc provenir d'une autre variable intrinsèque aux exigences de la tâche impactant directement les mécanismes de résolution de l'activité en lien avec les processus cognitifs sollicités. Une manière de tester cette hypothèse pourrait être d'imbriquer entre elles deux tâches cognitives pour lesquelles le ressenti ludique est connu. Ainsi, les structures des deux tâches seraient modifiées tout en permettant d'évaluer l'impact de ce changement sur le ressenti ludique de la nouvelle tâche combinée. Par ailleurs, en se basant sur les trois applications identifiées à la section 3.2 *Manques*, il se peut que l'utilisateur sollicite non pas une fonction cognitive individuellement, mais plusieurs. Ainsi, la multiplicité des fonctions cognitives sollicitées, qui est une condition se rapprochant davantage de situations réelles est le second argument de la prise en compte de cette variable lors de cette seconde étude.

En outre, une fois que nous saurons si le ressenti ludique est influencé par le design ludique ou la multiplicité des fonctions cognitives, nous nous demanderons comment évolue la charge mentale dans ces deux situations (impact du design ludique ou impact de la multiplicité de fonctions cognitives).

Nous supposons que la charge mentale liée à la mobilisation des ressources attentionnelles sera plus élevée avec une tâche au design ludique, par rapport à une tâche classique sans design. Effectivement, il y aura plus d'éléments d'informations (couleurs, stimuli contextuels) à prendre en compte. Par ailleurs, nous supposons que la charge mentale liée à la mobilisation des ressources attentionnelles augmentera par la multiplication de la couleur des items, par rapport au design ludique unicolore, car il y aura plus d'informations à prendre en compte. Concernant les autres dimensions de la charge mentale, liées notamment aux caractéristiques de la tâche, telles que les exigences mentales, ou liées aux stratégies de l'individu, comme l'effort mental, nous pouvons supposer une évolution différente. En effet, si le design ludique impacte le ressenti ludique, alors cette variable (le ressenti ludique) pourrait agir comme un médiateur minimisant la charge mentale ressentie pour ces dimensions.

De plus, nous supposons que la charge mentale liée aux exigences de la tâche sera plus élevée lors d'une tâche avec de multiples fonctions cognitives sollicitées, par rapport à une tâche simple. En effet, une tâche combinée demande plus d'exigences mentales qu'une tâche simple. De surcroît, nous nous demandons si une augmentation du nombre de fonctions cognitives correspond à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées) ou à une augmentation proportionnelle de la charge mentale. Cependant si le ressenti ludique est impacté positivement ou négativement par la multiplicité des fonctions cognitives, il pourrait minimiser ou à l'inverse augmenter la charge mentale ressentie dans cette condition. Cependant, la structure même de la tâche combinée ne permettra pas de déterminer la part exacte d'action du ressenti ludique et de la multiplication des fonctions cognitives sur la charge mentale.

Lors de cette seconde étude, nous allons donc étudier l'impact sur la charge mentale et sur l'engagement (via le ressenti ludique et la motivation) d'une variable intrinsèque aux exigences de la tâche, la multiplicité des fonctions cognitives sollicitées, et d'une variable extrinsèque aux exigences de la tâche, le design ludique.

Enfin, nous proposerons également de tester la transférabilité de modèles de classification (pour déduire les classes de charge mentale subjective) inter-contextes pour une même tâche cognitive. Pour finir, à travers cette seconde étude, nous souhaitons mettre en place un protocole permettant d'obtenir les données nécessaires afin de préparer un futur travail de triangulation de la mesure de la charge mentale, se basant sur des données EEG, sur de multiples dimensions de performances et sur des données subjectives.

5.1 Hypothèses de l'étude et choix des tâches

5.1.1 Hypothèses opérationnelles de l'étude 2

Nous avons choisi, à travers cette seconde étude expérimentale, de vérifier les hypothèses opérationnelles suivantes :

- **Hypothèse H1** : Un modèle de classification réalisé dans une situation expérimentale A (en ligne), peut prédire la charge mentale subjective d'une situation expérimentale B (en présentiel).

- **Hypothèse H2 :**

- a) Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, alors le design ludique, par rapport à un design classique, augmente le ressenti ludique.
- b) Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, la charge mentale subjective liée aux caractéristiques de la tâche (comme les exigences mentales) ou aux stratégies de l'individu (comme l'effort mental) devrait diminuer. En parallèle le design ludique devrait augmenter la charge mentale subjective en lien avec la mobilisation des ressources attentionnelles.
- c) Si le design ludique est accentué par la couleur et que le ressenti ludique est lié au design, alors il devrait augmenter et les effets consécutifs sur la charge mentale devraient être accentués.

- **Hypothèse H3 :**

- a) Si le ressenti ludique provient de la multiplicité des fonctions cognitives, alors une tâche combinée modifie le ressenti ludique, par rapport à une tâche simple.
- b) Une augmentation du nombre de fonctions cognitives entraîne une augmentation de la charge mentale. Cette augmentation ne correspond pas à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées), mais plutôt à une augmentation proportionnelle de la charge mentale.

5.1.2 Choix des tâches cognitives

Cette seconde étude est dans la continuité de la première. Ainsi, en nous appuyant sur les résultats de l'étude 1 (Cf. section 4.3 *Résultats de l'étude 1*) et les hypothèses opérationnelles de l'étude 2, nous conservons le test de N-Back (NBT), le test de Corsi et le test de Go/No-Go (GNG) pour cette seconde étude. Le Corsi, car parmi les cinq tâches, il a obtenu différents niveaux de charge mentale de manière suffisamment fiable pour que nous puissions faire des prédictions sur la charge mentale attendue. De plus, ses consignes simples et rapides en font un bon candidat pour comprendre et modéliser la charge mentale à travers une tâche cognitive. Au contraire du WCST, dont les trois niveaux de complexité n'ont pas pu générer trois classes de charge mentale en termes d'*Exigences mentales* et d'*Effort mental* (contrairement aux autres tâches de l'étude 1).

Or, lors de la réalisation de tâches cognitives, les dimensions *Exigences mentales* et *Effort mental*, sont celles contribuant le plus à la charge mentale subjective (Besançon et al., 2017 ; Louis et al., 2023 *In press*). En outre, des études ont utilisé le NBT (Ke et al., 2015 ; Roy, 2015 ; Pergher et al., 2019 ; Kutafina et al., 2021) et le GNG (Falkenstein et al., 1999 ; Vidal et al., 2012) pour mesurer des concepts mentaux à travers l'EEG. Ainsi, ces études peuvent servir de base dans le choix de l'étiquetage des données EEG lors de notre seconde étude. Enfin, dans une visée de multiplication de fonctions cognitives lors d'un même essai, nous nous basons sur le Corsi (tâche la plus simple au niveau de la compréhension des consignes) avec un ajout d'inhibition mentale (à travers le GNG). En effet, après plusieurs pré-tests, nous avons observé que le GNG était le plus propice pour une imbrication avec le Corsi, par rapport au NBT, au WCST ou à la DT. Le GNG proposant un enchaînement séquentiel d'items simples (basés sur des formes) spatialisés au même endroit, contrairement au NBT où les items (des lettres à mémoriser en sollicitant la mise à jour de la MdT) nécessitent plus de ressources mentales, ou au WCST où les items sont spatialisés à différents endroits de l'interface. Concernant la DT, cette tâche est déjà basée sur le Corsi, donc on ne peut pas l'imbriquer davantage au test de Corsi. Enfin, en se basant sur les résultats de la transférabilité inter-tâches de l'étude 1 (Cf. section 4.3.5), le NBT et le Corsi s'étaient distingués, par le résultat de la précision de la prédiction de la charge mentale subjective. Par conséquent, nous conservons trois tâches cognitives issues de l'étude 1, le NBT, le Corsi et le GNG. À cela s'ajoutent plusieurs tâches pour pouvoir vérifier l'hypothèse concernant l'impact d'un design ludique sur l'engagement et la charge mentale. Ainsi, ce sont des tâches cognitives (sollicitant les mêmes fonctions cognitives que les tâches classiques) avec un design ludique. Il y a donc un NBT ludique, deux versions différentes pour le Corsi ludique et un GNG ludique. Enfin, pour tester l'hypothèse sur l'impact de la multiplicité de fonctions cognitives sur l'engagement et la charge mentale, nous avons mis en place deux tâches dites "combinées", car elles sollicitent en un unique essai plusieurs fonctions cognitives. Ces différentes tâches et leur intérêt pour répondre aux différentes hypothèses de l'étude seront détaillés dans la section 5.2.2.

5.2 Matériel et méthodes

5.2.1 Participants

Vingt volontaires (9 femmes, 11 hommes) francophones, résidant dans la région nantaise et répondant aux critères d'inclusion (parlant couramment le Français, droitiers, ayant une vision et une audition normale ou corrigée à la normale), ont participé à cette étude expérimentale en laboratoire. Leur âge moyen était de 33,55 ans (écart-type = 8,46 ans). Ils ont été informés de l'objectif réel de l'expérience. Tous avaient au moins le baccalauréat et 85 % d'entre eux avaient au moins une Licence, dont 65 % un Master. Par ailleurs, 70 % travaillaient pour l'entreprise onepoint, 20 % venaient du réseau social *LinkedIn* et les autres provenaient du Centre Hospitalier Universitaire (CHU) de Nantes, du Laboratoire des Sciences du Numérique de Nantes (LS2N) et de l'École Centrale de Nantes (ECN).

5.2.2 Tâches et procédures

Au total, les participants réalisaient neuf tâches expérimentales dans un ordre aléatoire (Cf. **Figure 2**) sur un écran placé en face d'eux au CHU de Nantes. Toutes les tâches ont été créées à travers le logiciel de développement *Unity3D* (Version 2019.4.30f1) par Sébastien Maire, développeur au sein de l'entreprise onepoint, sous la supervision de la doctorante qui s'était appuyée sur le protocole expérimental que nous détaillons dans les sections suivantes.



Figure 2 : Vue d'ensemble du design expérimental de l'étude 2

5.2.2.1 Tâches cognitives classiques

Lors de l'étude 2, les tâches cognitives classiques permettront de tester l'hypothèse **H1** (concernant la transférabilité inter-contextes d'un modèle de classification) et

permettront de mettre en place un protocole pour la triangulation des catégories de mesures de la charge mentale.

Nous reprenons donc le NBT, le Corsi et le GNG de l'étude 1 (que nous avons détaillé en section 4.2.2). Elles contiennent toujours trois niveaux de complexité pour pouvoir comparer les trois classes de charge mentale issues des trois niveaux de complexité de l'étude 1 avec les classes de charge mentale issues de l'étude 2. Mais aussi, de vérifier la correspondance entre les trois niveaux de complexité et les trois niveaux de charge mentale issus de l'EEG. Des niveaux de charge mentale via l'EEG qui seront pris en compte dans un travail ultérieur à la thèse, au regard des mesures subjectives de la charge mentale et de performances. D'ailleurs, par rapport à l'EEG, l'étude 2 contient des modifications liées à cette technique de mesure. En effet, il est préconisé d'avoir une durée d'essai suffisante, car les premières secondes de données EEG risquent d'être exclues pour réduire les artefacts dus aux mouvements parasites (comme la déglutition ou les mouvements de la tête), souvent observés en début d'enregistrement (Lim et al., 2018). Par conséquent, nous avons déterminé qu'il nous fallait a minima des essais d'une longueur de 40 secondes, correspondant à la longueur d'essai du NBT de l'étude 1 avec 20 items (comme l'étude de Ke et al., 2015). Or, le GNG de l'étude 1 avait une durée inférieure à 10 secondes (Cf. section 4.2.2.3). Concernant le Corsi, avec ses deux essais par niveau, il avait une durée inférieure à 20 secondes (pour le niveau de complexité le plus élevé). Ainsi, pour cette seconde étude, nous avons pris comme référence la longueur de l'essai du NBT équivalente à 40 secondes et nous avons ajusté la longueur du Corsi et du GNG. Concernant ce dernier, nous sommes passés de 10 items par niveau dans la première étude, à 58 items pour cette seconde étude. Concernant le Corsi, après avoir comparé plusieurs possibilités lors de pré-tests, nous avons choisi d'ajouter trois essais à chacun des niveaux de complexité. Pour chaque essai, nous présentons au niveau de la **Table 1** les différents cubes pointés par l'ordinateur.

Niveau	Essai 1	Essai 2	Essai 3	Essai 4	Essai 5
3 cubes	4-7-2	8-1-5	3-4-1	6-1-5	9-3-5
5 cubes	5-2-1-8-6	4-2-7-3-1	3-9-2-4-8	7-8-2-9-4	4-2-6-8-1

7 cubes	5-9-1-7-4-2-8	5-7-9-2-8-4-6	8-1-9-2-6-4-7	9-3-6-7-2-4-3	8-7-1-2-4-6-9
---------	---------------	---------------	---------------	---------------	---------------

Table 1 : Les cubes pointés dans chacun des cinq essais des trois niveaux de complexité du Corsi de l'étude 2.

Enfin, concernant les dimensions de performance pour ses trois tâches cognitives classiques, ce sont les mêmes que la tâche de NBT (Cf. section 4.2.2.1), du Corsi (Cf. section 4.2.2.2) et du GNG (Cf. section 4.2.2.3) de l'étude 1.

Pour finir, concernant la durée totale du NBT classique de l'étude 2, elle est de 13 minutes, telle que le NBT de l'étude 1 (Cf. section 4.2.2.1). Cependant, pour le Corsi et le GNG, la durée totale est différente de leurs homologues de l'étude 1, car la durée de l'essai est plus longue. Par conséquent, pour le Corsi, nous passons d'une durée de 11 minutes à 13 minutes. Pour le GNG, nous passons de 11 minutes à 16 minutes.

5.2.2.2 Tâches cognitives ludiques

Les tâches cognitives ludiques permettront de tester l'hypothèse **H2** (concernant l'impact du design ludique sur la charge mentale et sur l'engagement). Afin de tester l'hypothèse, nous nous sommes appuyés sur les mêmes mécaniques de jeu (nombre d'items et interactions similaires) que pour les tâches cognitives classiques. Par conséquent, les dimensions de performance sont les mêmes entre une tâche classique et sa version ludique.

De plus, en vue de répondre à la question sur l'impact du design ludique sur la charge mentale et l'engagement, inutile d'avoir les trois niveaux de complexité pour chaque tâche. Étant donné que les enjeux de notre problématique (Cf. section 3.3) se situent principalement au niveau de la charge mentale intermédiaire, nous avons conservé les niveaux intermédiaires des tâches cognitives classiques (le niveau 1-rang pour le NBT, le niveau à 5 cubes pour le Corsi et le niveau de complexité intermédiaire pour le GNG). Par ailleurs, afin de pouvoir mettre en place un design ludique, nous avons utilisé la narration (autrement dit le changement du style linguistique des consignes) et les couleurs. En effet, Oyshi et Tumpa (2018) ont analysé les jeux populaires et ont observé que tous avaient une narration et étaient colorés. De plus, Goethe (2019) a indiqué que

la couleur a un effet sur nos émotions, ce qui pourrait être considéré comme un premier pas vers un design ludique.

Après plusieurs semaines d'échanges avec l'équipe de recherche et les équipes créatives de l'entreprise onepoint (designers, concepteurs 3D, développeurs), nous avons



contextualisé les tâches cognitives dans un environnement aquatique où neuf bouées colorées sont réparties avec les mêmes positions que les cubes du test de Corsi classique (Cf. **Figure 3**).

Figure 3 : Environnement des tâches cognitives ludiques.

Lors des consignes, par la narration, nous invitons le participant à jouer à un jeu virtuel où son ami et lui étaient partis en vacances au bord de mer. Dans ce contexte, plusieurs jeux avaient lieu, comme un échange de lancer de ballon ou repérer un poisson qui change de couleur lorsqu'il passe dans des bouées flottantes. Les premiers jeux ludiques correspondent à quatre tâches cognitives : le NBT ludique, deux Corsi ludiques (avec des bouées de couleurs similaires ou différentes) et le GNG ludique. Il y a deux Corsi, car le premier, l'unicolore (Cf. **Figure 4**) est uniforme au niveau de la couleur des bouées comme le test de Corsi classique l'est pour les cubes. Concernant le second, le multicolore (Cf. **Figure 5**), les bouées peuvent prendre des couleurs différentes (comme le rouge, le bleu ou le vert) pour s'approcher de la condition "tâches combinées" où l'alternance des couleurs était nécessaire. En effet, pour les "tâches combinées", nous avons comme objectif d'insérer le GNG au Corsi. Pour ce faire, il fallait une modalité visuelle (les bouées de couleurs) en plus de la modalité visuo-spatiale (les différentes positions des bouées).

Dans les prochains paragraphes, nous détaillerons chacune des tâches ludiques.

Le N-back (NBT) ludique

Cette tâche permettra de tester l'hypothèse **H3** (impact de la multiplicité des fonctions cognitives sur la charge mentale et sur l'engagement).

Principe et matériel

Les participants lisent dans un premier temps des consignes écrites (Annexe 14) et visualisent une consigne-vidéo sur la façon d'accomplir le NBT ludique. Puis, ils effectuent une phase d'entraînement, où ils réalisent 10 essais (3 cibles), afin de se familiariser avec la présentation des stimuli (des poissons de différentes couleurs) et avec les boutons de réponse (touches directionnelles). Le niveau d'entraînement a une durée de 28 secondes et correspond au niveau de difficulté « 1-rang ».

Les durées et les interactions sont les mêmes que le NBT classique. Sauf que le NBT ludique contient un unique niveau de complexité correspondant au niveau intermédiaire du NBT classique, le N = 1 rang. Plus précisément, dans le NBT ludique, un poisson (Cf. poisson rouge indiqué par des flèches directionnelles rouges sur la **Figure 3**) saute toutes les 1,5 secondes de manière aléatoire sur une des 9 bouées jaunes et cela durant 20 sauts. Dans ce contexte ludique, le poisson peut prendre sept couleurs différentes (rouge, jaune, bleu, vert, marron, gris et violet). De plus, le poisson reste durant 500 millisecondes (ms.) au niveau d'une bouée, suivi d'un intervalle de 1500 ms. La durée d'un essai est donc de 48 secondes. Enfin, parmi les 20 sauts du poisson sur les bouées, 30 % sont des cibles. L'objectif des participants est de comparer, à chaque saut du poisson, la couleur du poisson à la précédente. Si la couleur est identique à la précédente, le participant doit appuyer sur la touche directionnelle gauche du clavier. Concernant tous les autres cas, il doit appuyer sur la touche directionnelle de droite (interactions similaires avec le NBT classique).

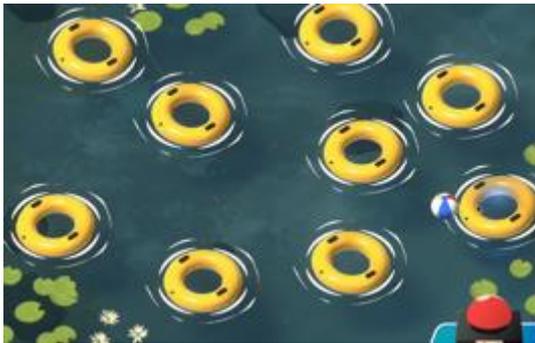
Pour finir, les dimensions de performance sont les mêmes que le NBT de l'étude 1 (Cf. section 4.2.2.1). Enfin, la durée totale du NBT ludique est d'environ 13 minutes.

Le Corsi ludique unicolore

Cette tâche permettra de vérifier les hypothèses **H2** (impact du design ludique sur la charge mentale et sur l'engagement) et **H3** (impact de la multiplicité des fonctions cognitives sur la charge mentale et sur l'engagement).

Principe et matériel

Le Corsi ludique unicolore est une tâche de mémoire visuo-spatiale qui consiste à retenir une séquence de lancers d'un ballon sur des bouées jaunes, puis à reproduire la séquence dans le même ordre de lancer. Les consignes données aux participants sont situées au niveau de l'Annexe 14. Le Corsi unicolore se déroule dans un environnement où l'ensemble des neuf bouées sont jaunes et suivent le même positionnement que la planche de Corsi classique. Lorsque les participants passent la souris au-dessus des



bouées jaunes, leurs phases supérieures s'allument. De plus, lorsque les participants cliquent à l'aide de la souris sur les bouées, le ballon se dirige vers elles (Cf. **Figure 4**).

Figure 4 : Environnement du Corsi ludique unicolore

Une fois l'entraînement terminé (même séquence que le Corsi classique ; bouée 2 - bouée 6), les participants réalisent les cinq mêmes séquences du niveau intermédiaire du Corsi classique dans un ordre aléatoire. Le ballon est lancé par l'ami virtuel (en réalité l'ordinateur) sur les bouées jaunes à une vitesse d'une bouée par seconde. Lorsque les participants doivent réaliser leur séquence, une consigne « À VOUS » s'affiche. Immédiatement après, les participants cliquent avec la souris sur les séquences de bouées dans le même ordre de lancer que l'ami virtuel. Après avoir donné leur réponse, les participants doivent indiquer quand leur tour est terminé en appuyant sur le buzzer rouge "TERMINER" qui s'affiche en bas à droite de l'écran dès le début du tour des participants (Cf. **Figure 4**).

Pour finir, les dimensions de performance sont les mêmes que celles du Corsi classique de l'étude 1 (Cf. section 4.2.2.2). Enfin, la durée totale du Corsi ludique unicolore est d'environ 11 minutes.

Le Corsi multicolore

Cette tâche permettra de vérifier les hypothèses **H2** (impact du design ludique sur la charge mentale et sur l'engagement) et **H3** (impact de la multiplicité des fonctions cognitives sur la charge mentale et sur l'engagement).

Principe et matériel

Le test de Corsi multicolore est similaire au Corsi unicolore, sauf qu'il se déroule dans un environnement où les neuf bouées sont de couleurs différentes (jaune, rouge, bleu, vert, marron, gris et violet) avec la possibilité que deux ou trois bouées soient de même couleur en simultanée (Cf. **Figure 5**). De plus, entre les cinq lancers de ballon à mémoriser et la restitution par les participants, les bouées changent de couleur pour



éviter que les participants mémorisent la couleur au lieu de la position visuo-spatiale.

Figure 5 : Environnement du Corsi ludique multicolore.

Au niveau de la consigne (Annexe 14), il est donc précisé de se focaliser sur les positions des bouées et non sur leurs couleurs qui vont changer entre la mémorisation et leur restitution.

Pour finir, les dimensions de performance sont les mêmes que le Corsi classique de l'étude 1 (Cf. section 4.2.2.2). Enfin, la durée totale du Corsi ludique multicolore (11 minutes) est la même que le Corsi ludique unicolore précisée au-dessus.

Le Go/No-Go (GNG) ludique

Cette tâche permettra de tester l'hypothèse **H3** (impact de la multiplicité des fonctions cognitives sur la charge mentale et sur l'engagement).

Principe et matériel

Le GNG ludique est une tâche d'inhibition mentale de la réponse motrice. Cette tâche se déroule dans le même environnement que le Corsi multicolore (détailé au paragraphe précédent) avec des interactions différentes. Lors du GNG ludique, lorsque les participants appuient sur la touche ENTRÉE, le ballon se dirige vers les bouées.

Dans un premier temps, les participants reçoivent des consignes écrites (Annexe 14) et sous forme de vidéo sur la façon d'accomplir le GNG ludique, puis ils effectuent une tâche d'entraînement avec 10 essais (5 cibles), afin de se familiariser avec la présentation des stimuli et avec le bouton de réponse (ENTRÉE). Le GNG ludique est un enchaînement aléatoire de bouées qui s'allument durant 4550 ms toutes les 9775 ms durant 58 fois. Les temporalités sont issues de pré-tests pour prendre en compte l'aspect visuo-spatial du GNG ludique qui impacte le temps de réaction par rapport au GNG classique (avec des items qui ne sont pas spatialisés). L'objectif du participant est de lancer le ballon (en appuyant sur le bouton ENTRÉE) sur toutes les bouées qui



s'allument, sauf sur les bouées grises contenant une fleur (Cf. **Figure 6**). Il y a un seul niveau de complexité correspondant à 50 % de No-Go qui sont des bouées grises contenant une fleur.

Figure 6 : Environnement du GNG ludique

Pour finir, les dimensions de performance sont les mêmes que le GNG de l'étude 1 (Cf. section 4.2.2.3). Enfin, la durée totale du GNG ludique est d'environ 12 minutes.

5.2.2.3 Tâches cognitives combinées

L'objectif de ces deux dernières tâches du protocole de l'étude 2 est de reprendre les trois fonctions cognitives de nos tâches ludiques (mémoire visuo-spatiale, mise à jour de la mémoire verbale et inhibition mentale) dans un même essai et d'étudier l'impact sur la charge mentale et sur l'engagement. Ces tâches comparées à celles qui sollicitent une fonction cognitive principale au cours du même essai (comme le Corsi multicolore) permettent de tester l'hypothèse **H3** (impact de la multiplicité des fonctions cognitives sur la charge mentale et sur l'engagement).

Principe et matériel

Afin d'atteindre cet objectif, nous procédons en deux parties (partie A et partie B) qui se présentent de manière aléatoire entre les participants. La partie A, vise à tester l'effet de la combinaison de deux fonctions cognitives. Elle comprend le Corsi (séquence de

lancer de ballon à mémoriser) combiné au GNG intermédiaire (inhibition mentale avec les fleurs sur les bouées grises). Elle se décompose donc comme ceci :

Partie A = Corsi 5 bouées combiné au GNG intermédiaire au cours d'un même essai.

La partie B vise, quant à elle, à tester l'impact de la réalisation d'une première tâche simple sur la réalisation consécutive des deux tâches combinées. Ainsi, elle commence par le NBT (poisson qui change de couleur 10 fois de suite), puis continue par le Corsi combiné au GNG. Ce qui donne une décomposition de l'essai en deux étapes, comme ci-après :

Partie B = NBT (1 rang) suivi par le Corsi 5 bouées combiné au GNG intermédiaire.

De plus, entre la 1^{ère} partie (NBT) et la seconde (Corsi + GNG), il y a une phrase de transition (« 2^{ème} partie du jeu, votre ami va lancer le ballon ») de 3000 ms.

Concernant le matériel et les interactions de la partie NBT, ils sont similaires à ceux du NBT ludique seul (Cf. **Figure 3**). Pour la partie Corsi/GNG, nous avons le même matériel que le Corsi ludique multicolore (Cf. **Figure 5**). Cependant, les participants doivent mémoriser une séquence de cinq lancers de ballon, puis restituer la mémorisation en cliquant deux fois sur les bouées. Lors de la restitution de la séquence de lancer de ballon, des fleurs peuvent se positionner sur des bouées qui doivent être pointées. Dans ce cas, les participants ne doivent pas faire le second clic sur la bouée grise avec une fleur (No-Go) et la réponse sera validée avec un seul clic.

Pour les deux parties A et B, nous avons quatre séquences de pointage avec 2 cibles No-Go (bouée grise avec fleur). Dans toutes les séquences, il y a une bouée « piège » où l'on retrouve une fleur sur une bouée non grise à cliquer. Le reste de la séquence n'a pas de fleur.

Les bouées à mémoriser pour chaque séquence sont les suivantes et une représentation graphique est proposée en Annexe 15 :

Partie A : 3-2-8-4-9 ; 8-7-4-9-6 ; 6-8-1-2-7 et 1-8-6-2-4

Partie B : 6-4-8-2-9 ; 3-4-2-7-6 ; 8-1-9-2-6 et 9-2-4-6-1

Les mesures de la performance lors de la tâche combinée peuvent se décomposer en deux parties, celles concernant la partie du NBT correspondant aux mêmes dimensions de performance que le NBT classique (Cf. section 4.2.2.1). Nous avons également les

mesures de performances lors du Corsi/GNG qui sont détaillées ci-après. Il y a les *Bouées fausses*, les *Fausse alarmes* (deux clics au lieu d'un seul lors de l'affichage d'une cible No-Go), les *Omissions* (pas de deuxième clic sur la bouée), les *Bouées pointées*, les *Bouées exactes*, les *Réponses attendues* (correspondant au pourcentage de similarité entre la séquence attendue et la séquence réalisée du Corsi, à cela s'ajoute le résultat du GNG : $100 * (25 - (\text{Fausse Alarmes} + \text{Omissions}) / 25)$), les *Séquences exactes* (nombre de bouées pointées dans l'ordre et la position attendue), le *Temps de réaction du GNG pour les bonnes réponses*, le *Temps de réaction pour tous les items confondus*, et le *Temps total* (entre l'apparition du bouton « TERMINER » et l'appui sur ce bouton). Enfin, la durée totale de la tâche combinée est d'environ 16 minutes.

5.2.3 Mesures subjectives et physiologiques

Mesures subjectives

Les questionnaires de la charge mentale de l'étude 2 sont les mêmes que ceux de l'étude 1. Pour un rappel succinct, nous avons le NASA-TLX (Hart et Staveland, 1988 ; Cegarra et Morgado, 2009 ; Annexe 1), le *Workload Profile* (WP ; Tsang et Velazquez, 1996 ; Annexe 2) avec les mêmes échelles considérées, correspondant au traitement spatial (WP3), au traitement verbal (WP4), au traitement visuel (WP5) et à la production manuelle (WP7).

Comme vu en introduction du Chapitre 5, en plus de la mesure subjective de la charge mentale, nous souhaitons mesurer une variable interne à l'individu liée à l'engagement au travers du ressenti ludique et de la motivation. Ainsi, nous mesurons le ressenti *Ludique* à travers une question proposée sur une échelle de Likert de 0 à 100 que nous avons mise en place : « *Dans quelle mesure la tâche que vous venez de faire a été ludique, amusante ?* ». De plus, spécificité de l'étude 2, nous ajoutons une question concernant l'aspect motivationnel de la tâche. La dimension *Motivation* est donc une question sur une échelle de Likert de 0 à 100 que nous avons mise en place et qui correspond à : « *Quel était votre niveau de motivation pour réaliser la tâche ?* ».

La durée de la passation de l'ensemble de ces questionnaires est d'environ six minutes.

Mesures physiologiques

Enfin, la principale différence avec l'étude 1 est la prise en compte des données ÉlectroEncéphaloGraphiques (EEG) réalisées à travers seize électrodes dites "humides"

(liées au gel conducteur qu'on ajoute au niveau de l'électrode et du cuir chevelu). Comme vu à la section 2.3.2 *Les mesures centrales*, il est convenu qu'une augmentation de la charge mentale correspond à une augmentation des ondes Thêta en zone frontale et à une diminution des ondes Alpha en zone pariétale (Hankins et Wilson, 1998 ; Miller, 2001 ; Roy, 2015 ; Charles et Nixon, 2019 ; Tao et al., 2019 ; Longo et al., 2022 ; Raufi et Longo, 2022). En se basant sur l'étude de Raufi et Longo (2022), nous identifions plusieurs électrodes pour pouvoir mesurer les ondes Thêta (Cf. **Figure 7**) au niveau des zones frontales. Nous avons ainsi un groupe d'électrodes pour la bande Thêta. De même, pour la bande Alpha des zones corticales pariétales.

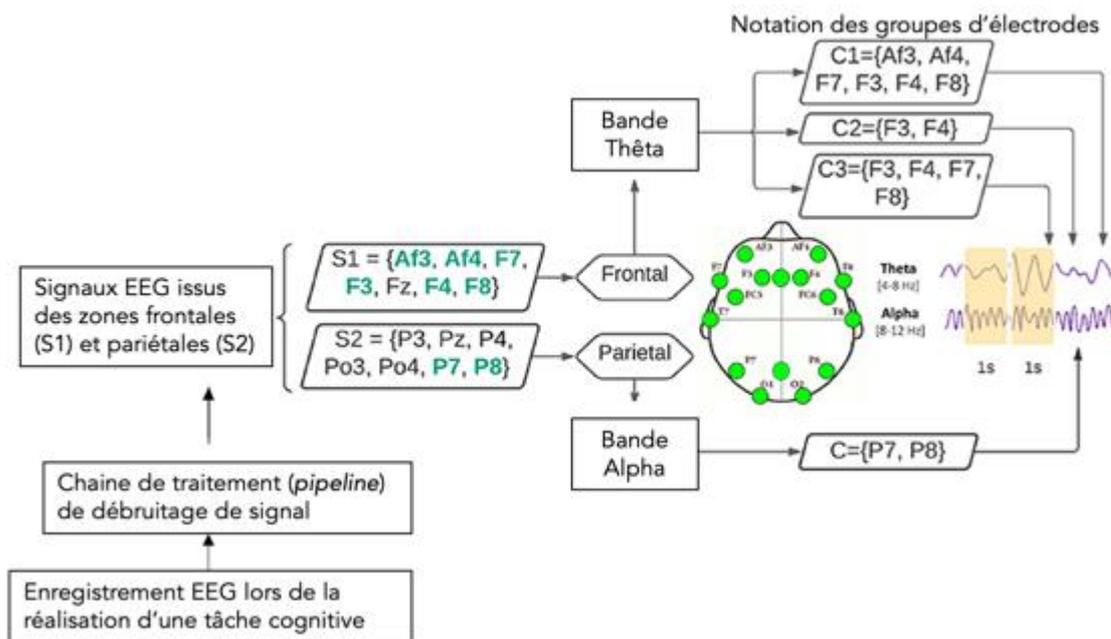


Figure 7 : Illustration du choix des électrodes EEG pour la mesure de la charge mentale. Illustration inspirée de Raufi et Longo (2022).

5.2.4 Procédure expérimentale

Après avoir obtenu un avis favorable du Comité de Protection des Personnes (CPP) Nord-Ouest III (Annexe 16), nous avons lancé la campagne de recrutement. Nous avons envoyé un résumé de l'étude (Annexe 17) par courrier électronique aux potentiels participants. Si nous avons un retour, nous proposons une réunion individuelle où chaque individu a été clairement informé des objectifs et du déroulement de l'étude lors d'une *visite d'information* individuelle à J-7 (Cf. **Figure 8**) avant la passation au CHU de Nantes. Cette visite d'information, durant 45 minutes, avait également pour objectif

de vérifier les critères d'inclusion et de non-inclusion, mais aussi de permettre un délai de réflexion de sept jours lié au recueil de la non-opposition du participant.

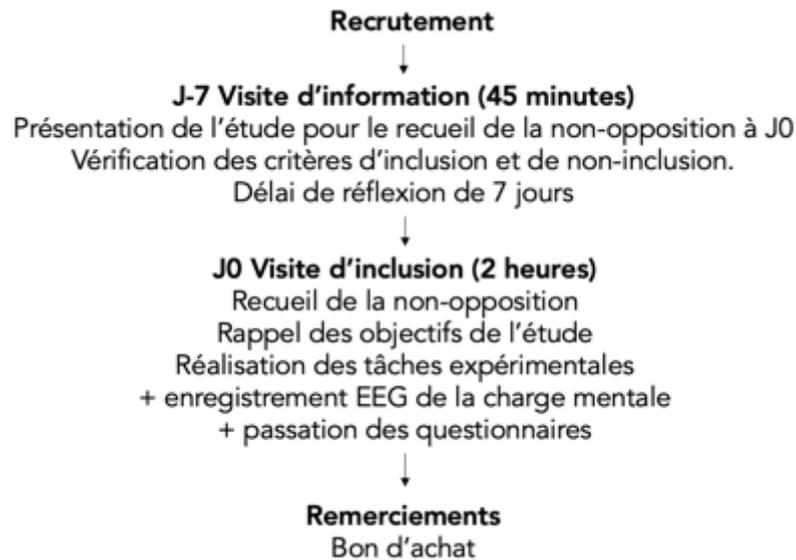


Figure 8 : Schéma général de l'organisation de l'étude 2

Avant d'établir une date pour la prochaine visite d'inclusion, le participant était invité à signer la lettre de consentement, nommée *Note d'information pour la participation à la recherche* (Annexe 18). Après avoir établi une date de passation au CHU de Nantes, la phase expérimentale se déroulait comme suit :

- Accueil au CHU de Nantes (Hôpital Guillaume et René Laennec, Boulevard Jacques Monod - St Herblain, 44093 NANTES au laboratoire ICONIC au niveau -2) par la doctorante, puis accompagnement au laboratoire expérimental.
- Vérification de l'éligibilité du participant et sa non-opposition à l'étude.
- Rappel de l'étude à l'oral par la doctorante et l'ingénieur spécialisé en EEG (Aurélien Van Langenhove).
- Un questionnaire pré-test (Annexe 19) était proposé aux participants.
- Mise en place du dispositif d'enregistrement EEG contenant seize électrodes *humides* (Cf. **Figure 9**). La mise en place a une durée de 20 minutes, car elle comprend le choix de la taille du bonnet en fonction de la morphologie de la tête du participant, la mise en place des seize électrodes avec une pipette contenant du gel conducteur, la vérification des signaux et au besoin l'ajustement des électrodes.



Figure 9 : Mise en place du dispositif d'enregistrement EEG par la doctorante.

- Enregistrement de deux lignes de base (*baselines*) EEG d'une minute chacune au minimum. La première *baseline* consistait en un enregistrement EEG les yeux ouverts fixant un point fixe devant soi avec le moins de tension et mouvements musculaires. La seconde *baseline* était la même chose avec les yeux fermés. Les *baselines* EEG avaient plusieurs objectifs, à commencer par l'étalonnage des algorithmes de nettoyage d'artefacts. Elles servaient également de références statistiques pour comparer les signaux EEG entre les tâches cognitives.
- Le participant réalisait toutes les tâches (Cf. **Figure 10**) qui comprenaient des instructions, sous forme de textes, d'images et de vidéos, sur la manière de les réaliser, suivies d'une phase d'entraînement pour se familiariser avec la présentation des stimuli et avec les interactions requises.



Figure 10 : Dispositif expérimental de l'étude 2.

- Les expérimentateurs étaient attentifs aux signaux EEG présentés sur l'écran 2 (Cf. **Figure 10**), car des électrodes pouvaient s'assécher ou bouger durant l'enregistrement. Si c'était le cas, il y avait un ajustement au moment de la phase de questionnaires se déroulant à la fin des essais de chaque niveau de complexité.
- Tous les participants étaient exposés aux tâches et aux niveaux de complexité dans un ordre aléatoire. Puis, après chaque niveau de complexité de chaque tâche, il y avait l'administration de questionnaires (Cf. section 5.2.3 et Annexe 20).
- Par ailleurs, lorsque l'ensemble des tâches a été réalisé, un questionnaire post-test était proposé aux participants (Annexe 20).
- À la fin de la passation de toutes les tâches expérimentales, le dispositif EEG était retiré et une serviette était proposée au participant pour retirer le gel de ses cheveux.
- À la fin de l'étude, tous les participants recevaient 50 euros en bons d'achat et étaient accompagnés à l'accueil du CHU de Nantes.

5.3 Résultats de l'étude 2

La section *Résultats* est composée de trois parties correspondant aux trois hypothèses présentées ci-dessous qui sont liées à la problématique de l'étude 2.



<p>H1. Un modèle de classification réalisé dans une situation expérimentale A (en ligne), peut prédire la charge mentale subjective d'une situation expérimentale B (en présentiel).</p>	<p>H2a. Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, alors le design ludique, par rapport à un design classique, augmente le ressenti ludique.</p> <p>H2b. Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, la charge mentale subjective liée aux caractéristiques de la tâche (comme les exigences mentales) ou aux stratégies de l'individu (comme l'effort mental) devrait diminuer. En parallèle le design ludique devrait augmenter la charge mentale subjective en lien avec la mobilisation des ressources attentionnelles.</p> <p>H2c. Si le design ludique est accentué par la couleur et que le ressenti ludique est lié au design, alors il devrait augmenter et les effets consécutifs sur la charge mentale devraient être accentués.</p>	<p>H3a. Si le ressenti ludique provient de la multiplicité des fonctions cognitives, alors une tâche combinée modifie le ressenti ludique, par rapport à une tâche simple.</p> <p>H3b. Une augmentation du nombre de fonctions cognitives entraîne une augmentation de la charge mentale. Cette augmentation ne correspond pas à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées), mais plutôt à une augmentation proportionnelle de la charge mentale.</p>
---	---	---

Pour tester l'ensemble de ces hypothèses opérationnelles, nous avons procédé à plusieurs analyses statistiques que nous présentons dans un plan illustré à la **Figure 11**.

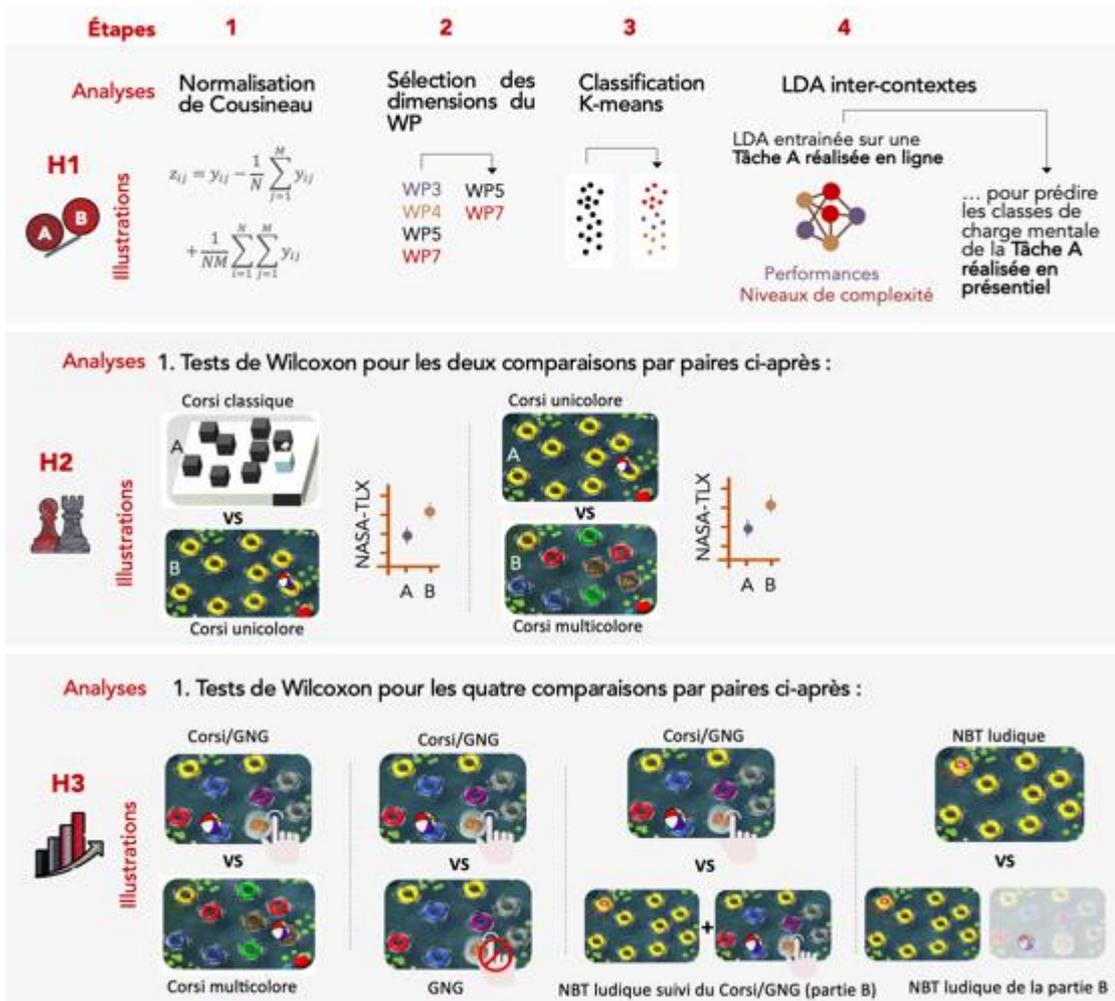
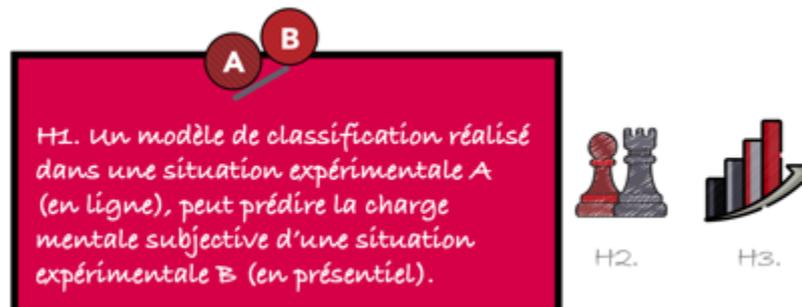


Figure 11 : Plan d'analyse statistique des données de l'étude 2 en lien avec les hypothèses opérationnelles issue de la problématique générale.

5.3.1 Transférabilité d'un modèle de classification inter-contextes

Dans cette section d'analyse, nos résultats permettront de vérifier la première hypothèse de l'étude 2, la **H1** précisée ci-dessous.



Par conséquent, l'objectif de cette section est de définir le score de précision de la prédiction de la charge mentale subjective de l'étude 2 avec un modèle de classification entraîné sur les données de l'étude 1. Pour ces analyses, nous avons utilisé le langage de

programmation *Python* (version 3.9.13). Le code et une partie des analyses ont été réalisés par une ingénieure de l'entreprise onepoint, Bérengère Mathieu, sous la supervision de la doctorante.

Tâches cognitives sélectionnées

Pour tester l'hypothèse **H1** de l'étude 2, nous nous sommes intéressés aux trois tâches cognitives communes aux deux études expérimentales. Les tâches sont donc le NBT, le Corsi et le GNG (Cf. **Figure 12**).

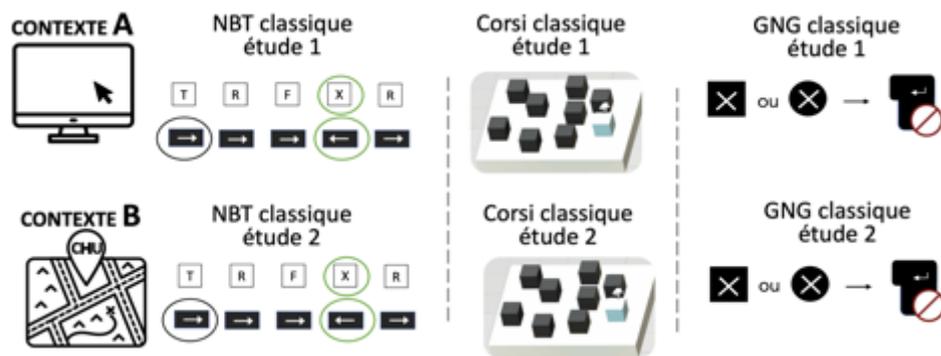


Figure 12 : Illustrations des tâches cognitives sélectionnées pour répondre à l'hypothèse 1 sur la classification supervisée inter-contextes.

Méthodes statistiques utilisées

Tout d'abord, pour chaque tâche cognitive, en considérant qu'une donnée correspond à un participant associé à un niveau de complexité, nous avons $53 \times 3 = 159$ données pour l'étude 1 et $20 \times 3 = 60$ données pour l'étude 2.

Par ailleurs, nous avons pu mettre en place un nouveau modèle LDA (méthode détaillée à la section 4.3.3) qui se base sur les classes de charge mentale subjective issues de K-means. Pour évaluer les performances du modèle de classification LDA, nous avons utilisé la méthode de validation croisée. Dans le processus de validation croisée, le dataset global est sous-divisé en un dataset d'entraînement et un dataset de test. Nous entraînons donc le modèle sur le dataset d'entraînement et nous le testons sur le dataset de test. Nous recommençons ensuite le processus en établissant un nouveau dataset d'entraînement, ainsi qu'un nouveau dataset de test. Le score de validation croisée global est alors la moyenne des scores de validation croisée obtenus à chaque itération. De plus, pour cette section, l'estimation de la charge mentale était faite en inter-session.

Car, l'entraînement de la classification (*training test*) a été faite sur la première session et le test (*testing test*) a été effectué sur la seconde.

Concernant les données, le modèle s'entraînait sur les données de performance (*X train*) et de charge mentale (*Y train*) de l'étude 1. Puis, ce modèle de classification testait sur le dataset de test, correspondant aux performances (*X test*) et à la charge mentale (*Y test*) de l'étude 2. Comme pour l'étude 1, nous avons proposé un modèle de classification supervisée de la charge mentale subjective, spécifique à chaque tâche cognitive.

Précisons que l'entraînement de la classification supervisée a été réalisé à partir des données de l'étude 1, avec 70 % des données pour l'apprentissage et 30 % des données pour l'évaluation de la qualité de cet apprentissage. Nous avons réalisé une validation croisée avec vingt itérations, en calculant le *Balanced accuracy score* moyen de l'ensemble de ces itérations.

Paramètres des modèles de classification

Pour les données subjectives de la charge mentale, nous nous sommes basés sur le questionnaire du NASA-TLX et les mêmes dimensions du WP que la section 4.3.3. Par conséquent, pour le NBT, la sélection de variables était le NASA-TLX/WP4. Pour le Corsi, la sélection de variables NASA-TLX/WP5 et pour le GNG, la sélection de variables NASA-TLX/WP7.

Concernant les prédicteurs de la charge mentale subjectives, nous avons le *Niveau de complexité* pour tous les modèles de classification de chaque tâche. Pour les performances, nous avons repris celles qui ont obtenu de meilleurs résultats lors des LDA de l'étude 1 présentés aux sections 4.3.3 et 4.3.4. Par conséquent, pour le NBT, les performances correspondaient aux *Réponses attendues*, aux *Fausses alarmes*, aux *Omissions* et au *Temps de réaction pour tous les items*. Pour le Corsi, les *Réponses attendues*, les *Omissions*, les *Séquences exactes* et le *Temps total* étaient sélectionnés. Enfin, pour le GNG, les performances correspondaient aux *Réponses attendues*, au *Temps de réaction tous items confondus* et au *Temps de réaction pour les bonnes réponses*.

Résultats de la classification supervisée

Avant la présentation des résultats, nous tenons à préciser que nous pensions initialement réutiliser directement les modèles de classification établis lors de l'étude 1. Cependant, nous en avons réalisé de nouveaux, car en cours de thèse, l'équipe a subi le remaniement de certains de ses membres. Ainsi, les nouvelles analyses ont finalement dû être réalisées par une autre personne. Malgré plusieurs itérations, nous n'avons pas réussi à avoir des résultats similaires entre les deux séries d'analyse, en termes de classification supervisée de la charge mentale subjective. Les analyses se basaient cependant sur les mêmes données de performances et de niveaux de complexité. Plusieurs pistes peuvent expliquer cette différence de résultats. Tout d'abord, nous n'avons pas pu reproduire les mêmes paramètres de K-means et de LDA (ces informations n'ayant malheureusement pas été sauvegardées), qui étaient donc différents d'un collègue à un autre. Enfin, pour le Corsi, spécifiquement, pour cette seconde étude, il y a eu un oubli lié à la mise en pourcentage des données de performances. Ainsi, les résultats sont différents de la première à la seconde étude, avec une différence plus marquée pour le test de Corsi. Malgré ces différences, pour une tâche donnée, dans la classification de la charge mentale subjective basée sur la performance et le niveau de complexité, ce qui nous intéressait, dans cette section, c'était le résultat de transférabilité inter-contextes.

Modèle de classification inter-contextes du NBT

Tout d'abord, concernant le NBT, le modèle de classification entraîné sur les données de l'étude 1 (70 % de *train* et 30 % de *test*) avait un score de précision égal à 48,5 % (± 19 %). Par la suite, le score de précision de la prédiction de la charge mentale subjective de l'étude 2 (100 % des données subjectives de l'étude 2 sont en *test*) avec un modèle de classification entraîné sur les données de l'étude 1, était de 34 %. Le second résultat n'avait pas d'écart-type, puisqu'il y avait un seul test sur l'ensemble des données.

Modèle de classification inter-contextes du Corsi

Concernant le Corsi, le modèle de classification entraîné sur les données de l'étude 1 avait un score de précision égal à 51,6 % (± 19 %). Par la suite, le score de précision de la prédiction de la charge mentale subjective de l'étude 2 avec un modèle de classification entraîné sur les données de l'étude 1 était de 47 %.

Modèle de classification inter-contextes du GNG

Enfin, concernant le GNG, le modèle de classification entraîné sur les données de l'étude 1 avait un score de précision égal à 62,8 % (± 21 %). Par la suite, le score de précision de la prédiction de la charge mentale subjective de l'étude 2 avec un modèle de classification entraîné sur les données de l'étude 1 était de 45 %.

En conclusion, à l'appui de ces résultats, nous pouvons envisager la transférabilité d'un modèle de prédiction *inter-contextes* pour le test de Corsi et le GNG. En effet, nous n'observons pas de chute du score de précision à la limite du niveau de hasard (égal à 33 % pour trois classes de charge mentale) d'une classification à une autre pour ces tâches d'un contexte A à un contexte B. Contrairement au NBT, qui pour le second score de précision de la prédiction (34 %), est proche du niveau du hasard. Ainsi, nous pouvons songer à entraîner un unique modèle de classification basé sur le test de Corsi réalisé en ligne, par exemple, pouvant prédire la charge mentale subjective du Corsi lorsqu'il est réalisé dans une situation expérimentale en laboratoire. Les potentialités de ce résultat seront discutées davantage dans la discussion de l'étude 2.

5.3.2 Impact du design ludique

Dans cette section d'analyse, nos résultats permettront de vérifier la deuxième hypothèse de l'étude 2, la **H2** précisée ci-après.



H2a. Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, alors le design ludique, par rapport à un design classique, augmente le ressenti ludique.

H2b. Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, la charge mentale subjective liée aux caractéristiques de la tâche (comme les exigences mentales) ou aux stratégies de l'individu (comme l'effort mental) devrait diminuer. En parallèle le design ludique devrait augmenter la charge mentale subjective en lien avec la mobilisation des ressources attentionnelles.

H2c. Si le design ludique est accentué par la couleur et que le ressenti ludique est lié au design, alors il devrait augmenter et les effets consécutifs sur la charge mentale devraient être accentués.

Dans cette section, nous ne nous intéressons qu'aux trois versions du test de Corsi, car les mécaniques de jeu sont similaires, contrairement aux versions classiques/ludiques

du NBT et du GNG. En effet, les versions ludiques du NBT et du GNG ont servi pour le questionnement suivant concernant l'impact de la multiplicité de fonctions cognitives sur la charge mentale et sur l'engagement (Cf. section 5.3.3).

Afin de connaître l'impact du design ludique sur la charge mentale, nous avons décomposé cette question de recherche en deux sous-questions. Ainsi, dans un premier temps, nos résultats portaient sur l'impact du design ludique en prenant en compte le test de Corsi classique et sa version ludique unicolore (Cf. **Figure 13**). Puis, l'impact d'un aspect spécifique du design, la pluralité de couleurs, en comparant le Corsi unicolore et le Corsi multicolore (Cf. **Figure 13**).

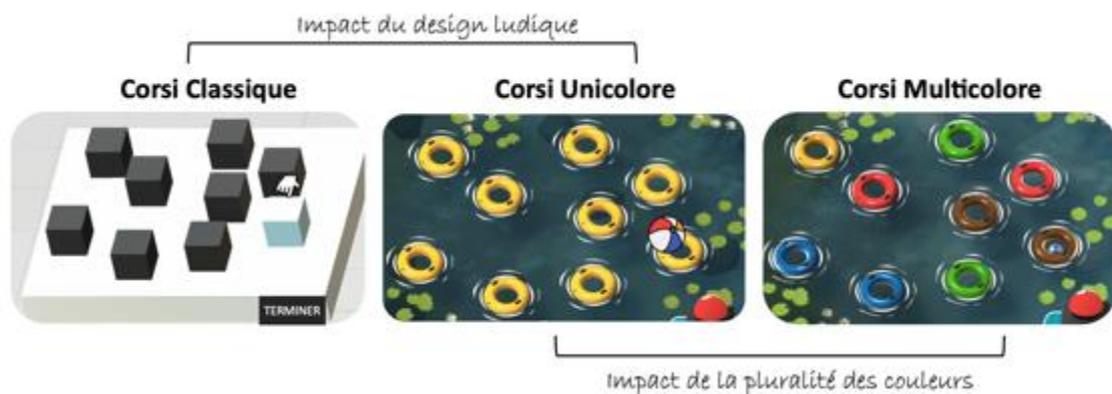


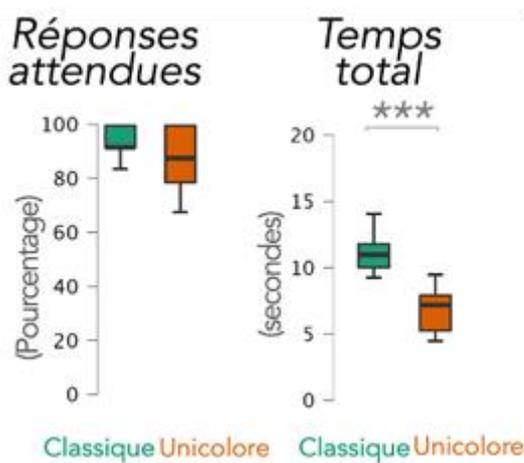
Figure 13 : Matériel du Corsi classique (à gauche), du Corsi ludique unicolore (au centre) et du Corsi ludique multicolore (à droite).

Concernant ces analyses, nous avons utilisé le logiciel JASP (version JASP 0.16). En préambule, nous avons vérifié l'homogénéité de nos distributions et de nos variances et nous nous sommes rendu compte qu'elles ne l'étaient pas. Pour cette raison, nous avons utilisé des statistiques non paramétriques. Ainsi, des tests de rang signé de Wilcoxon ont été utilisés. De plus, toutes les mesures subjectives, telles que les questionnaires NASA-TLX et WP, ont été normalisées grâce à la correction de Cousineau (Morey, 2008) que nous avons détaillée à la section 4.3 de l'étude 1.

5.3.2.1 Impact d'un design ludique avec des items unicolores

Dans cette section d'analyse, nos résultats permettront de vérifier *H2a* et *H2b* de la deuxième hypothèse **H2** portant sur l'impact d'un design ludique sur l'engagement et la charge mentale. Pour cette section, nous avons donc comparé le Corsi classique et le Corsi ludique unicolore. Ainsi, les deux tâches sont de complexité équivalente et la seule différence porte sur le design. Avant de comparer l'engagement et la charge mentale

subjective entre ces deux Corsi, nous voulions comparer un critère objectif lié à la charge mentale, les performances. Il y a deux dimensions de performance du Corsi pertinentes à prendre en compte, les *Réponses attendues* (concernant la précision de la réponse) et le *Temps total* (correspondant au temps entre l'apparition du *buzzer* rouge « TERMINER » et l'appui sur ce bouton). Les autres dimensions de performance du Corsi (*Cubes pointés*, *Cubes faux*, *Cubes exacts*, *Séquences exactes* et *Omissions*) sont englobées dans la dimension *Réponses attendues*. Puis, nous comparerons les données subjectives mesurées via le questionnaire NASA-TLX, le WP, la question *Ludique* et la question *Motivation*.



Mesures de performances

Dans un premier temps, nous avons donc comparé les performances en termes de *Réponses attendues* entre les deux Corsi (graphique de gauche **Figure 14**).

Figure 14 : Performances en termes de *Réponses attendues* (à gauche) et de *Temps Total* (à droite) selon le Corsi classique et le Corsi ludique unicolore.

Considérés deux à deux, ils avaient des *Réponses attendues* qui n'étaient pas significativement différentes (Table 1 en Annexe 21 ; $W = 86,000$; $p = 0,145$). À l'inverse, les deux Corsi avaient un *Temps total* significativement différent et plus élevé pour le Corsi classique par rapport au Corsi ludique (Table 1 en Annexe 21 ; $W = 208,000$; $p < 0,001$).

Questionnaire du NASA-TLX

Score global

Désormais, nous nous intéressons aux mesures subjectives, à commencer par la charge mentale globale. Entre les deux Corsi, la charge mentale globale du NASA-TLX (correspondant à la somme des six dimensions du NASA-TLX sur une échelle de 0 à 600 ; **Figure 15**) n'était pas significativement différente (Table 2 en Annexe 21 ; $W = 101,000$; $p = 0,896$).

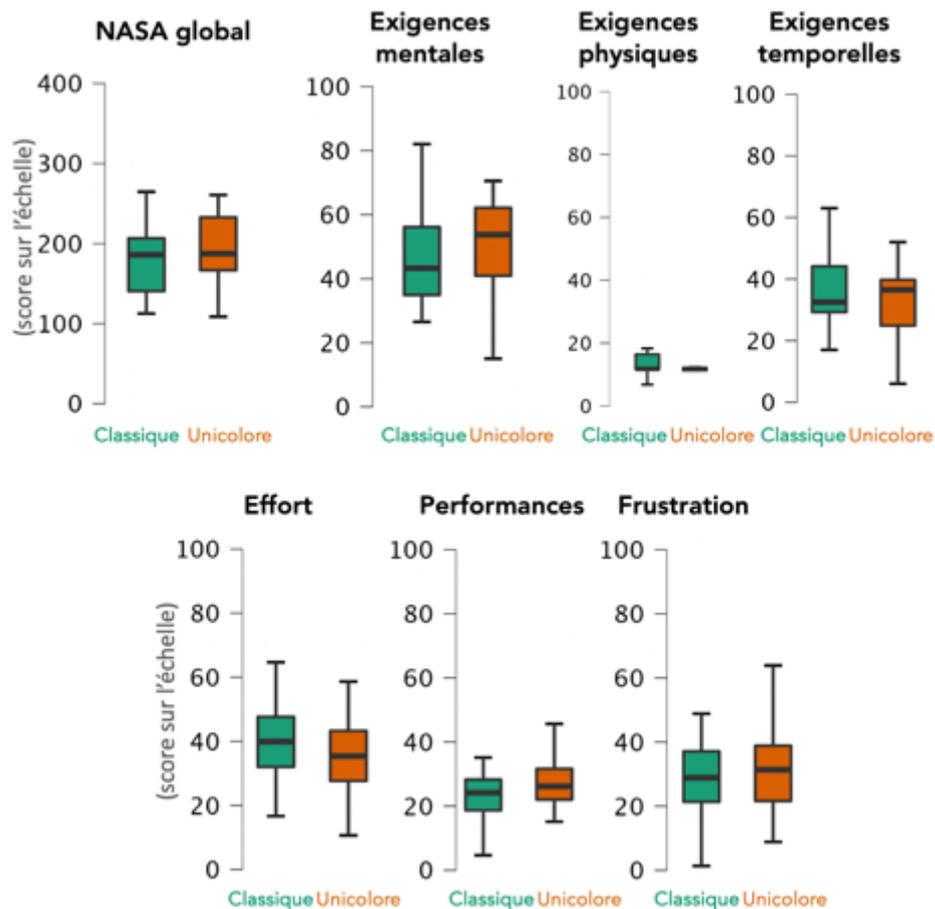


Figure 15 : La charge mentale basée sur le questionnaire NASA-TLX selon le Corsi classique et le Corsi ludique unicolore.

Analyse dimension par dimension du NASA-TLX

Concernant les dimensions du questionnaire NASA-TLX, nous considérerons d'abord les *Exigences mentales* (Cf. **Figure 15**). Entre les deux Corsi, la dimension *Exigences mentales* n'était pas significativement différente (Table 2 en Annexe 21 ; $W = 81,00$; $p = 0,380$). De même, pour la dimension *Exigences physiques* (Table 2 en Annexe 21 ; $W = 63,00$; $p = 0,233$). Nous observons les mêmes tendances pour toutes les autres dimensions du NASA-TLX. Les *Exigences temporelles* (Table 2 en Annexe 21 ; $W = 86,00$; $p = 1,000$), l'*Effort* (Table 2 en Annexe 21 ; $W = 101,00$; $p = 0,513$), les *Performances* (Table 2 en Annexe 21 ; $W = 62,500$; $p = 0,522$) et enfin, la *Frustration* (Table 2 en Annexe 21 ; $W = 71,000$; $p = 0,813$) n'étaient pas significativement différentes entre les deux Corsi.

Questionnaire du WP

Nous n'avons pas gardé la dimension WP4, car ce n'était pas une dimension sollicitée par le Corsi, que ce soit à l'appui des résultats de l'étude 1 (Cf. section 4.3) ou à l'appui de la logique du test de Corsi, qui ne sollicite pas de modalité verbale, mais la modalité visuo-spatiale. Ainsi, concernant le score des dimensions du questionnaire WP, nous considérerons d'abord le WP3 (Cf. **Figure 16**). Entre les deux Corsi, la dimension *WP3-spatial* était significativement différente et plus élevée pour le Corsi ludique par rapport au Corsi classique (Table 3 en Annexe 21 ; $W = 121,500$; $p = 0,035$).

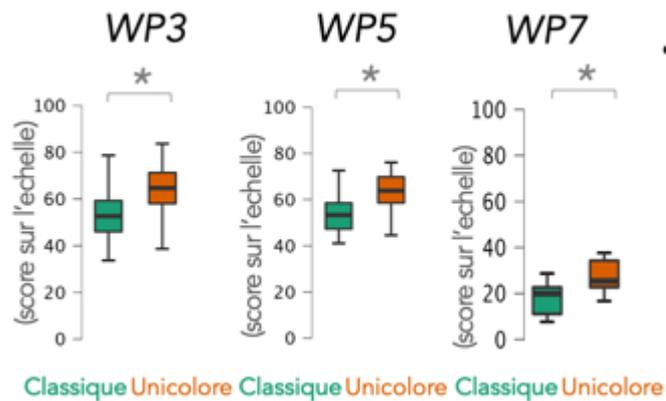


Figure 16 : La charge mentale à travers trois dimensions du questionnaire WP selon le Corsi classique et le Corsi ludique unicolore.

De plus, entre les deux Corsi, la dimension *WP5-visuel* était significativement différente et plus élevée pour le Corsi ludique par rapport au Corsi classique (Table 3 en Annexe 21 ; $W = 33,500$; $p = 0,025$), de même pour la dimension *WP7-manuel* (Table 3 en Annexe 21 ; $W = 26,000$; $p = 0,010$).

Questions sur les ressentis ludique et motivation

Concernant le ressenti ludique (Cf. **Figure 17**), entre les deux Corsi, il n'était pas significativement différent (Table 4 en Annexe 21 ; $W = 78,500$; $p = 0,519$). De même, concernant la dimension *Motivation* (**Figure 17** ; Table 5 en Annexe 21 ; $W = 51,000$; $p = 0,133$).

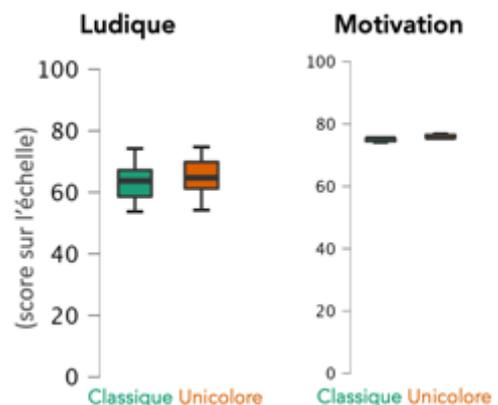


Figure 17 : Le ressenti Ludique selon le Corsi classique et le Corsi ludique unicolore.

En conclusion, le design ludique du Corsi augmente la charge mentale liée aux différentes ressources attentionnelles (visuelles, spatiales, motrices) mobilisées pour réaliser la tâche (mesurée à travers le WP) par rapport à un design classique. Ce qui n'est pas le cas de la charge mentale liée aux exigences de la tâche ou aux stratégies de l'individu (mesurée via le NASA-TLX) qui n'augmente pas avec le design ludique. De plus, avec un design ludique, les participants réalisent le Corsi plus rapidement (*Temps total* moins élevé) par rapport à un Corsi classique. Une piste d'explication est la nature intrinsèquement *challengeante* des activités avec un design ludique. Enfin, peu importe le design ludique ou non du Corsi, le ressenti ludique et la motivation n'avaient pas de différence, ce qui laisse penser que ces deux ressentis dépendent de la nature intrinsèque de la tâche. Des pistes d'explication que l'on discutera davantage lors de la discussion de l'étude 2.

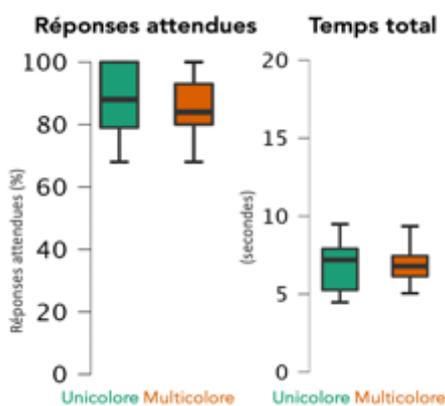
5.3.2.2 Impact d'un design ludique avec des items multicolores

Dans cette section d'analyse, nos résultats permettront de vérifier *H2c* de la deuxième hypothèse **H2** qui porte sur l'impact de la pluralité de la couleur des items sur la charge mentale et sur l'engagement. Pour cette section, nous avons donc comparé le Corsi ludique unicolore et le Corsi ludique multicolore.

Avant de comparer la charge mentale subjective entre ces deux Corsi, nous voulions comparer les performances. Comme pour la section précédente (Cf. 5.3.2.1), y a deux dimensions de performance du Corsi pertinentes à prendre en compte, les *Réponses attendues* et le *Temps total*. Par la suite, nous comparerons les données subjectives mesurées par le questionnaire NASA-TLX, le WP, la question *Ludique* et la question *Motivation*.

Mesures de performances

Dans un premier temps, nous avons donc comparé les performances en termes de *Réponses attendues* entre les deux Corsi ludiques (graphique de gauche **Figure 18**). Considérés deux à deux, ils avaient des *Réponses attendues* qui n'étaient pas significativement différentes (Table 1 en Annexe 22 ; $W = 97,000$; $p = 0,342$). De



même, les deux Corsi avaient un *Temps total* qui n'était significativement pas différent (Table 1 en Annexe 22 ; $W = 101,000$; $p = 0,898$).

Figure 18 : Performances en termes de Réponses attendues (à gauche) et de Temps Total (à droite) selon le Corsi unicolore et le Corsi multicolore.

Questionnaire du NASA-TLX

Score global

Désormais, nous nous intéressons aux mesures subjectives, à commencer par la charge mentale globale. Entre les deux Corsi, la charge mentale globale du NASA-TLX (Cf. **Figure 19**) n'était pas significativement différente (Table 2 en Annexe 22 ; $W = 55,000$; $p = 0,065$).

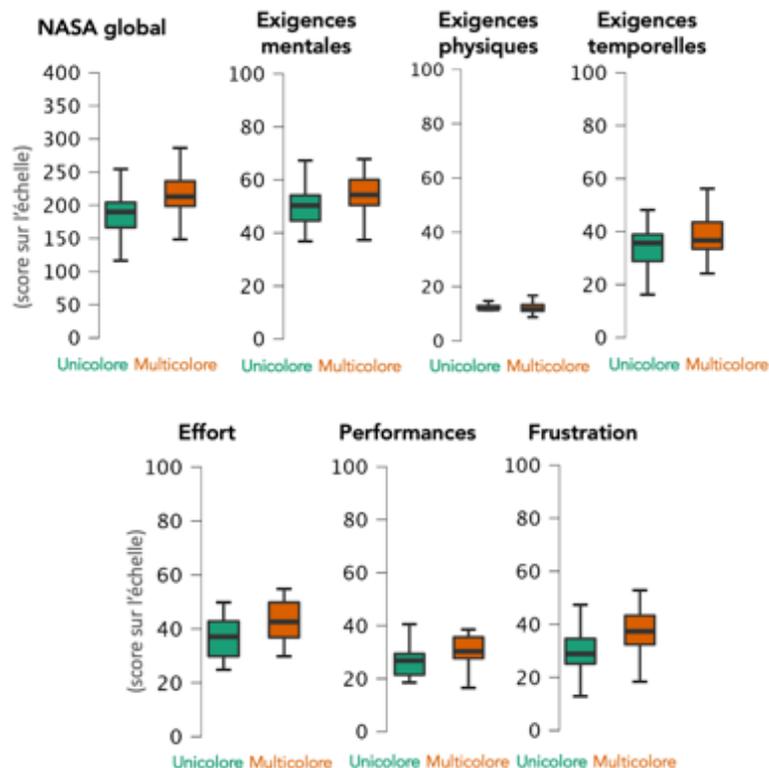


Figure 19 : La charge mentale basée sur le questionnaire du NASA-TLX selon le Corsi ludique unicolore et le Corsi ludique multicolore.

Analyse dimension par dimension

Concernant le score des dimensions du questionnaire NASA-TLX, nous considérerons d'abord les *Exigences mentales* (Cf. **Figure 19**). Entre les deux Corsi, la dimension *Exigences mentales* n'était pas significativement différente (Table 2 en Annexe 22 ; $W = 61,00$; $p = 0,296$), de même, pour la dimension *Exigences physiques* (Table 2 en Annexe 22 ; $W = 28,00$; $p = 0,408$). Nous avons les mêmes tendances pour toutes les autres dimensions du NASA-TLX. Les *Exigences temporelles* (Table 2 en Annexe 22 ; $W = 62,00$; $p = 0,316$), l'*Effort* (Table 2 en Annexe 22 ; $W = 53,500$; $p = 0,098$), les *Performances* (Table 2 en Annexe 22 ; $W = 47,000$; $p = 0,288$) et enfin, la *Frustration* (Table 2 en Annexe 22 ; $W = 58,000$; $p = 0,239$) n'étaient pas significativement différentes entre les deux Corsi.

Questionnaire du WP

Pour les mêmes raisons qu'à la section 5.3.2.1 *Impact d'un design ludique avec des items unicolores*, nous n'avons pas gardé la dimension WP4. Ainsi, concernant le score des dimensions du questionnaire WP, nous considérerons d'abord le WP3 (Cf. **Figure 20**). Entre les deux Corsi, la dimension *WP3-spatial* n'était pas significativement différente (Table 3 en Annexe 22 ; $W = 106,000$; $p = 0,383$).

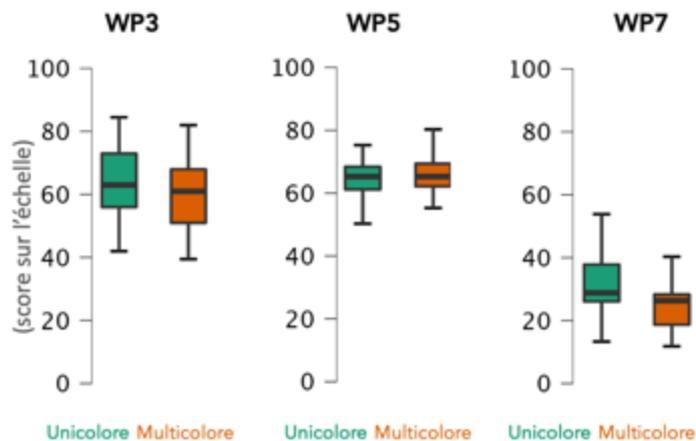


Figure 20 : La charge mentale à travers trois dimensions du questionnaire WP selon le Corsi ludique unicolore et le Corsi ludique multicolore.

Concernant le *WP5-visuel*, entre les deux Corsi, il n'était pas significativement différent (Table 3 en Annexe 22 ; $W = 60,000$; $p = 0,697$), de même pour le *WP7-manuel* (Table 3 en Annexe 22 ; $W = 105,000$; $p = 0,185$).

Question Ludique et Motivation

Concernant la dimension *Ludique* (Figure 21), entre les deux Corsi, elle n'était pas significativement différente (Table 4 en Annexe 22 ; $W = 71,500$; $p = 0,831$), de même pour la dimension *Motivation* (Figure 21 ; Table 5 en Annexe 21 ; $W = 29,000$; $p = 0,919$).

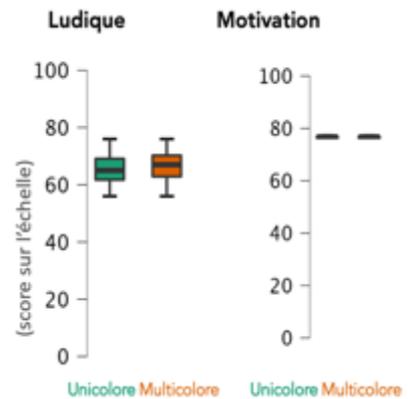


Figure 21 : Le ressenti Ludique (à gauche) et de Motivation (à droite) selon le Corsi ludique unicolore et le Corsi ludique multicolore.

En conclusion, un environnement ludique avec des items unicolores (uniquement des bouées jaunes) ou un environnement ludique avec des items multicolores (des bouées de différentes couleurs) n'a aucun impact significatif sur la charge mentale, qu'elle soit mesurée à travers le NASA-TLX, le WP, ou via les performances. De même, concernant le ressenti *Ludique* et le ressenti de *Motivation*. Une explication est que la couleur concerne une dimension restreinte du design ludique qui comprend d'autres éléments, comme ceux du contexte (vagues, nénuphars) ou de la consigne qui ont plus d'impact sur les ressources attentionnelles que la couleur. Une piste d'explication que l'on discutera davantage lors de la discussion de l'étude 2.

5.3.3 Impact de la multiplicité de fonctions cognitives

Dans cette section d'analyse, nos résultats permettront de vérifier la troisième hypothèse de l'étude 2, la **H3**.

H3a. Si le ressenti ludique provient de la multiplicité des fonctions cognitives, alors une tâche combinée modifie le ressenti ludique, par rapport à une tâche simple.

H3b. Une augmentation du nombre de fonctions cognitives entraîne une augmentation de la charge mentale. Cette augmentation ne correspond pas à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées), mais plutôt à une augmentation proportionnelle de la charge mentale.

Afin de connaître l'impact de la multiplicité de fonctions cognitives sollicitées par une tâche sur la charge mentale et sur l'engagement (via le ressenti ludique et la motivation), nous décomposons la question en trois sous-questions en réalisant quatre comparaisons inter-tâches (deux comparaisons pour la première sous-question et une comparaison pour chacune des deux autres sous-questions ; Cf. **Figure 22**).

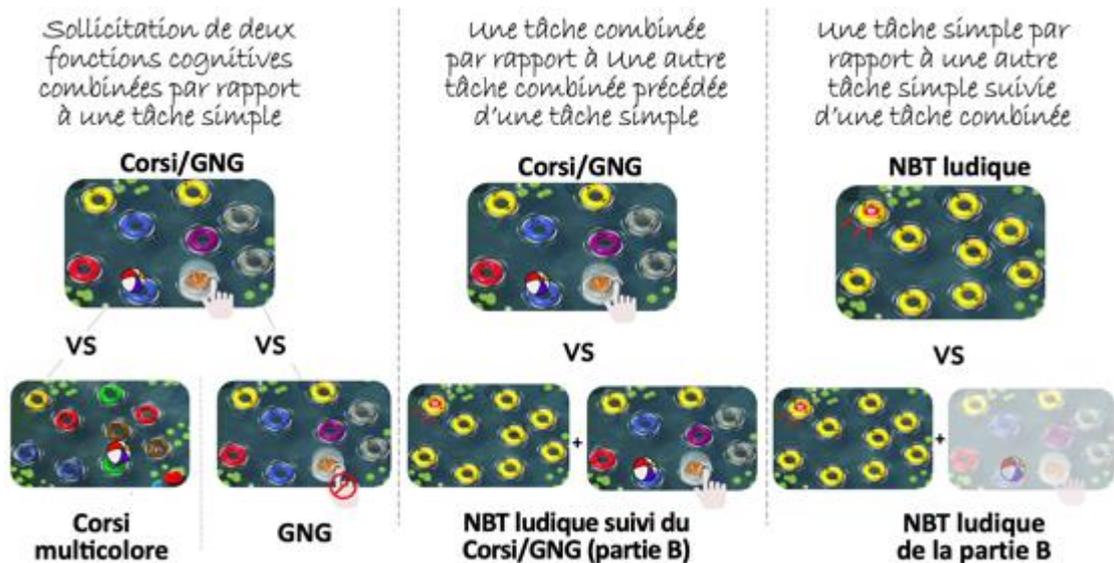


Figure 22 : Présentation du matériel des quatre comparaisons inter-tâches pour répondre à l'hypothèse 3 sur l'impact de la multiplicité de fonctions cognitives sur la charge mentale et sur l'engagement.

Ainsi, dans un premier temps, nos résultats porteront sur la comparaison entre la tâche de Corsi multicolore seul et la tâche de Corsi/GNG. Puis, nous réaliserons la comparaison entre la tâche de GNG ludique seul et la tâche de Corsi/GNG. Ces deux premières comparaisons permettront d'étudier si une tâche sollicitant deux fonctions cognitives combinées impacte de manière différente la charge mentale et l'engagement par rapport à une tâche sollicitant une fonction cognitive. En supposant qu'une augmentation du nombre de fonctions cognitives ne correspond pas à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées), mais plutôt à une augmentation proportionnelle de la charge mentale.

Dans un second temps, nous comparerons la tâche de Corsi/GNG seule avec la tâche de Corsi/GNG qui est précédée au cours du même essai du NBT ludique. Cette comparaison permettra d'étudier si une tâche réalisée en amont influence la charge mentale et l'engagement d'une tâche combinée, par rapport à une tâche combinée seule

dans l'essai. En posant l'hypothèse que la charge mentale augmente lorsqu'on sollicite une fonction cognitive en amont de la tâche combinée, par rapport à la tâche combinée seule. En effet, ajouter une fonction cognitive en amont peut générer l'impression que les ressources mentales sont déjà impactées par ce qu'on vient de faire et donc sont moins disponibles au moment de faire la tâche combinée.

Dans un troisième temps, nous comparerons le NBT ludique seul par rapport au NBT ludique lorsqu'il est suivi au cours d'un même essai d'une tâche combinée. Cette comparaison permettra d'étudier si la réalisation d'une tâche combinée précédée d'une tâche simple (non combinée) au cours d'un même essai, influence (par anticipation) la réalisation de la tâche simple. En effet, nous nous attendons à avoir une différence de performances entre les deux tâches, car les participants peuvent anticiper la seconde partie de la tâche lors du NBT, suivi du Corsi/GNG.

Pour ces analyses, nous avons utilisé le logiciel JASP (version JASP 0.16). En préambule, nous avons vérifié l'homogénéité de nos distributions et de nos variances et nous nous sommes rendu compte qu'elles ne l'étaient pas. Pour cette raison, nous avons utilisé des statistiques non paramétriques. Ainsi, des tests de rang signé de Wilcoxon ont été utilisés. De plus, pour toutes les mesures subjectives, nous les avons normalisées grâce à la correction de Cousineau (Morey, 2008) que nous avons détaillée à la section *4.3 Résultats de l'étude 1*.

Corsi multicolore vs Corsi multicolore combiné au Go/No-Go

Pour cette section, nous avons comparé le Corsi ludique multicolore avec la tâche combinée du Corsi/GNG (Cf. **Figure 22**). Avant de comparer la charge mentale subjective entre ces deux tâches, nous voulions comparer les performances en termes de *Réponses attendues*. Dans cette section, le *Temps total* n'est pas pris en compte, car le nombre d'essais entre les deux tâches est différent (cinq pour le Corsi unique par rapport à quatre pour le Corsi/GNG). Puis, nous comparerons les données subjectives mesurées via le questionnaire NASA-TLX, le WP, la question *Ludique* et la question *Motivation*.

Mesures de performances

Dans un premier temps, nous avons donc comparé les performances en termes de *Réponses attendues* entre les deux tâches (Cf. **Figure 23**). Considérées deux à deux, elles avaient des *Réponses attendues* qui étaient significativement différentes (Annexe 23 ; $W = 27,000$; $p = 0,007$). Une meilleure performance était observée pour le Corsi/GNG.

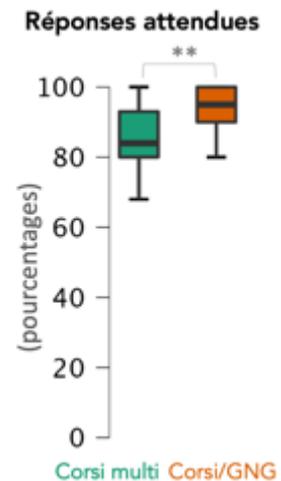


Figure 23 : Performances en termes de *Réponses attendues* selon le Corsi ludique multicolore et le Corsi/GNG.

Questionnaire du NASA-TLX

Score global

Désormais, nous nous intéressons aux mesures subjectives, à commencer par la charge mentale globale. Entre le Corsi ludique multicolore et la partie A (Corsi/GNG), la charge mentale globale du NASA-TLX (Cf. **Figure 24**) était significativement différente (Annexe 23 ; $W = 25,500$; $p = 0,003$) et plus élevée pour le Corsi/GNG.

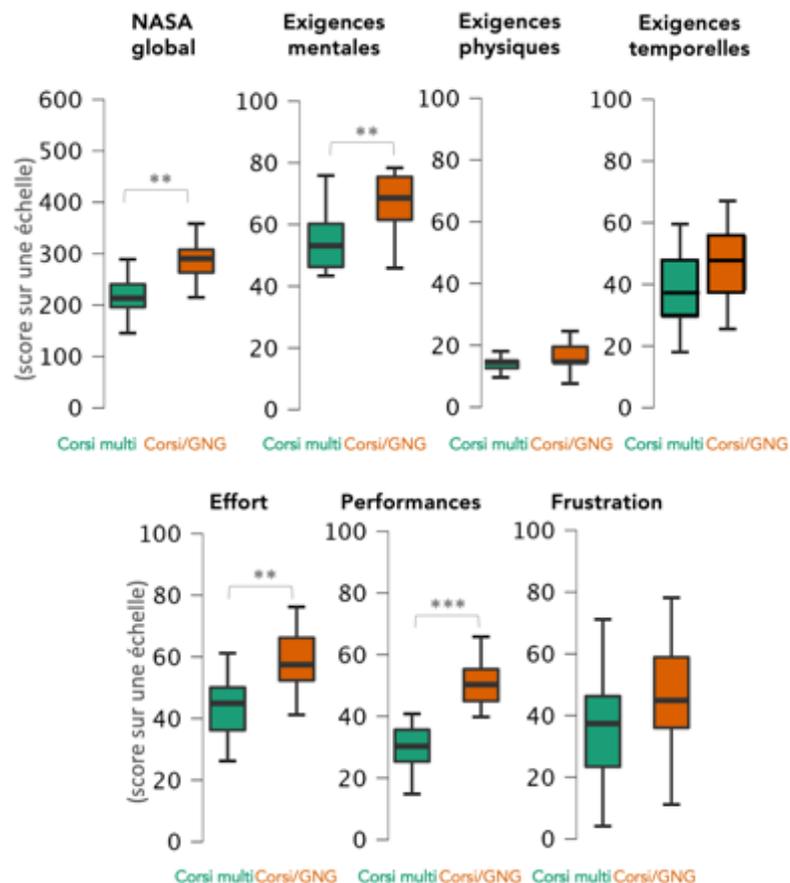


Figure 24 : La charge mentale basée sur le questionnaire NASA-TLX selon le Corsi ludique multicolore et le Corsi/GNG.

Analyse dimension par dimension

Concernant le score des dimensions du questionnaire NASA-TLX, nous considérerons d'abord les *Exigences mentales* (Cf. **Figure 24**). Entre les deux tâches, la dimension *Exigences mentales* était significativement différente et plus élevée pour le Corsi/GNG par rapport au Corsi multicolore (Annexe 23 ; $W = 35,500$; $p = 0,010$), de même, pour l'*Effort* (Annexe 23 ; $W = 29,000$; $p = 0,008$) et la dimension *Performances* (Annexe 23 ; $W = 1,000$; $p < 0,001$). Cependant, entre les deux tâches, les dimensions *Exigences physiques* (Annexe 23 ; $W = 33,000$; $p = 0,401$), *Exigences temporelles* (Annexe 23 ; $W = 56,000$; $p = 0,206$) et *Frustration* (Annexe 23 ; $W = 63,000$; $p = 0,205$) n'étaient pas significativement différentes.

Questionnaire du WP

Concernant le score des dimensions du questionnaire WP, nous considérerons d'abord le WP3 (Cf. **Figure 25**). Entre les deux tâches, la dimension *WP3-spatial* n'était pas significativement différente (Annexe 23 ; $W = 81,500$; $p = 0,390$).

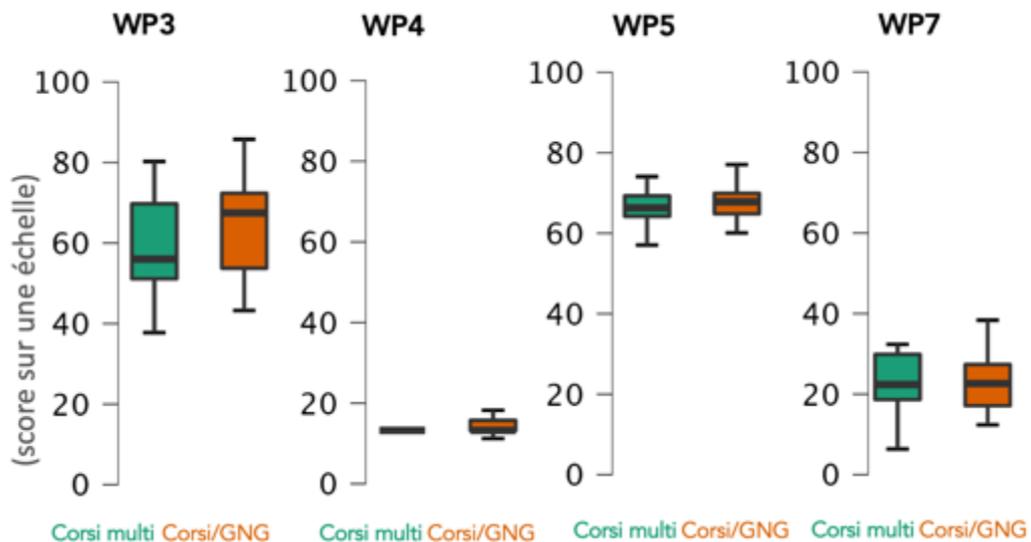


Figure 25 : La charge mentale à travers quatre dimensions du questionnaire WP selon le Corsi ludique multicolore et le Corsi/GNG.

De même, pour la dimension *WP4-verbal* (Annexe 23 ; $W = 14,000$; $p = 0,342$), pour la dimension *WP5-visuel* (Annexe 23 ; $W = 71,000$; $p = 0,541$) et pour le *WP7-manuel*

(Annexe 23 ; $W = 93,000$; $p = 0,760$) les comparaisons n'étaient pas significativement différentes entre les deux tâches.

Question Ludique et Motivation

Concernant le ressenti ludique (**Figure 26**), entre les deux tâches, il était significativement différent et plus élevé pour le Corsi seul, par rapport au Corsi/GNG (Annexe 23 ; $W = 161,000$; $p = 0,008$), contrairement à la dimension *Motivation* (**Figure 26**, Annexe 23 ; $W = 50,000$; $p = 0,409$) qui n'était pas significativement différente entre les deux comparaisons.

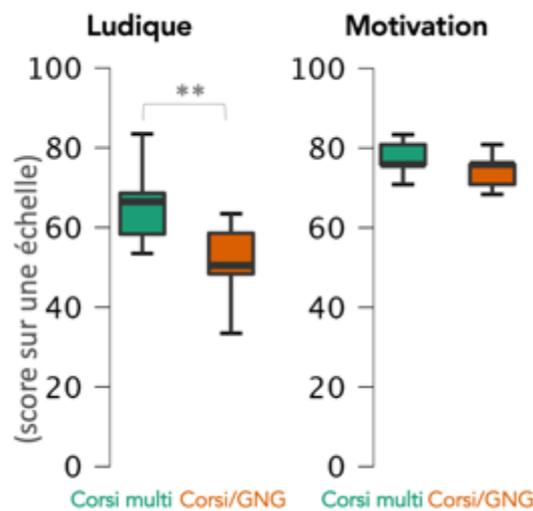


Figure 26 : Le ressenti Ludique (à gauche) et de Motivation (à droite) selon le Corsi ludique multicolore et le Corsi/GNG.

En conclusion, la tâche combinée obtenait de meilleures *Réponses attendues* par rapport au Corsi multicolore. Une mesure objective qui était appuyée par le ressenti subjectif de la dimension *Performance* du NASA-TLX qui était plus élevée pour le Corsi/GNG que pour le Corsi multicolore. Une piste d'explication est le *challenge* que représente une tâche avec deux fonctions cognitives combinées par rapport à une tâche simple.

Concernant la charge mentale globale, elle était plus élevée pour le Corsi/GNG que pour le Corsi multicolore. De même, pour la charge mentale liée aux exigences de l'activité (*Exigences mentales*) et aux stratégies du participant (*Effort mental*). Cependant, entre ces deux tâches, la charge mentale liée aux ressources attentionnelles mobilisées (mesurée via le WP) n'était pas significativement différente. Par conséquent, ce qui impactait la charge mentale correspondait davantage aux caractéristiques intrinsèques de la tâche combinée (ce que doivent faire les participants sur le plan cognitif) qui se

traduisait par une augmentation de la charge mentale liée aux *Exigences mentales* et à l'*Effort mental*, plutôt que les aspects liés à la mobilisation des ressources attentionnelles. Enfin, concernant le ressenti ludique, le Corsi multicolore était jugé comme plus ludique que le Corsi/GNG, mais la *Motivation* n'était pas différente entre ces deux tâches. Une piste d'explication est que le ressenti ludique est influencé par la complexité de la tâche en termes de multiplicité de fonctions cognitives sollicitées. Toutes ces pistes d'explication seront discutées davantage lors de la discussion de l'étude.

Go/No-Go ludique vs Corsi combiné au Go/No-Go

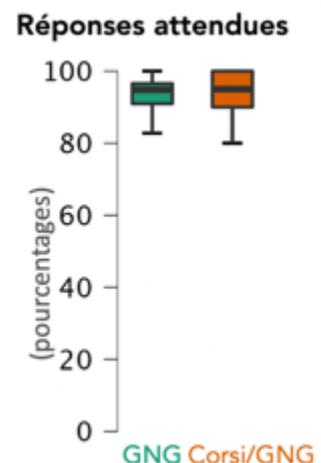
Dans cette section d'analyse, nos résultats permettront de vérifier, avec d'autres tâches cognitives, la troisième hypothèse de l'étude 2, la **H3**.

En effet, pour cette section, nous avons comparé le GNG ludique multicolore avec la Partie A (Corsi/GNG). Avant de comparer la charge mentale subjective entre ces deux tâches, nous voulions comparer les performances, en termes de *Réponses attendues*. Nous avons également identifié le *Temps de réaction pour tous les items confondus* et le *Temps de réaction pour les bonnes réponses*. Cependant, d'un point de vue technique, ces deux dernières dimensions de performance sont indisponibles, car nous n'avions pas les temps de mise en surbrillance des bouées. Puis, nous comparerons les données subjectives mesurées via le questionnaire NASA-TLX, le WP, la question *Ludique* et la question *Motivation*.

Mesures de performances

Dans un premier temps, nous avons donc comparé les performances en termes de *Réponses attendues* entre les deux tâches (Cf. **Figure 27**). Considérées deux à deux, elles avaient des *Réponses attendues* qui n'étaient pas significativement différentes (Annexe 23 ; $W = 94,000$; $p = 0,695$).

Figure 27 : Performances en termes de Réponses attendues selon le GNG et le Corsi/GNG.



Questionnaire du NASA-TLX

Score global

Désormais, nous nous intéressons aux mesures subjectives, à commencer par la charge mentale globale. Entre le GNG et le Corsi/GNG, la charge mentale globale du NASA-TLX (**Figure 28**) était significativement différente (Annexe 23 ; $W = 49,500$; $p = 0,040$) et plus élevée pour le Corsi/GNG par rapport au GNG seul.

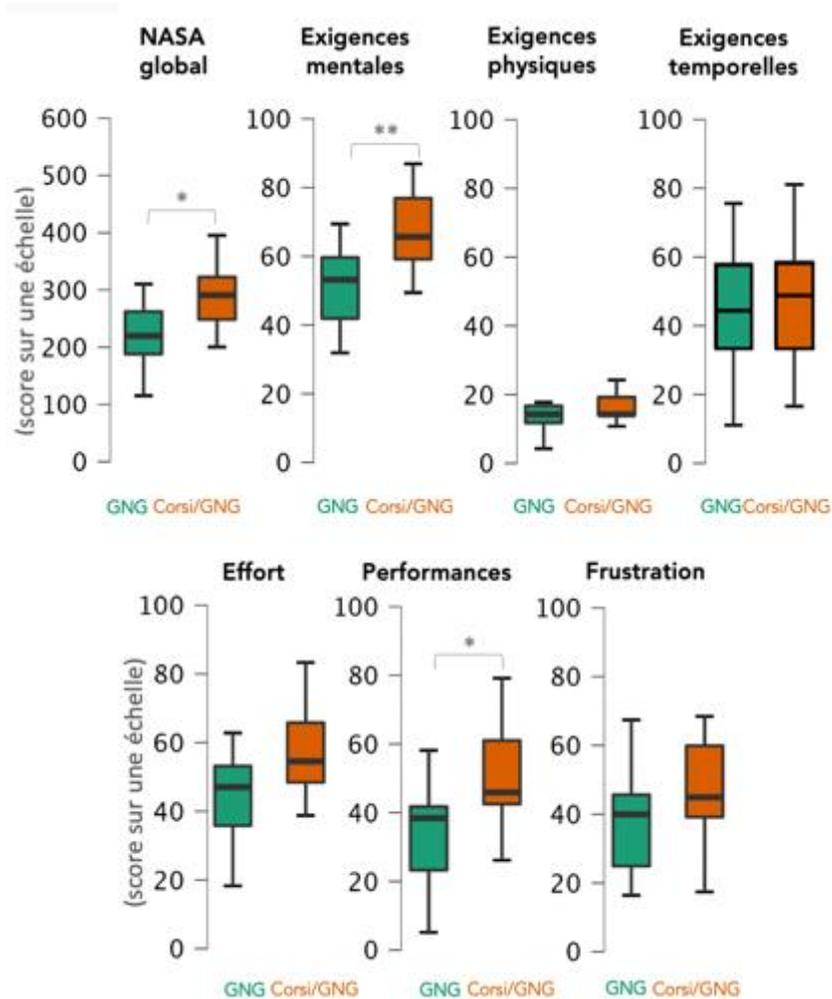


Figure 28 : La charge mentale basée sur le questionnaire NASA-TLX selon le GNG et le Corsi/GNG.

Analyse dimension par dimension

Concernant le score des dimensions du questionnaire NASA-TLX, nous considérerons d'abord les *Exigences mentales* (Cf. **Figure 28**). Entre les deux tâches, la dimension *Exigences mentales* était significativement différente et plus élevée pour le Corsi/GNG par rapport au GNG seul (Annexe 23 ; $W = 29,500$; $p = 0,009$), de même, pour la dimension *Performances* (Annexe 23 ; $W = 34,500$; $p = 0,016$). Cependant, entre les

deux tâches, les dimensions *Exigences physiques* (Annexe 23 ; $W = 42,000$; $p = 0,320$), *Exigences temporelles* (Annexe 23 ; $W = 96,500$; $p = 0,765$), *Effort mental* (Annexe 23 ; $W = 42,500$; $p = 0,064$) et *Frustration* (Annexe 23 ; $W = 48,000$; $p = 0,185$) n'étaient pas significativement différentes.

Questionnaire du WP

Concernant le score des dimensions du questionnaire WP, nous considérerons d'abord le WP3 (Cf. **Figure 29**). Entre les deux tâches, la dimension *WP3-spatial* n'était pas significativement différente (Annexe 23 ; $W = 74,500$; $p = 0,647$).

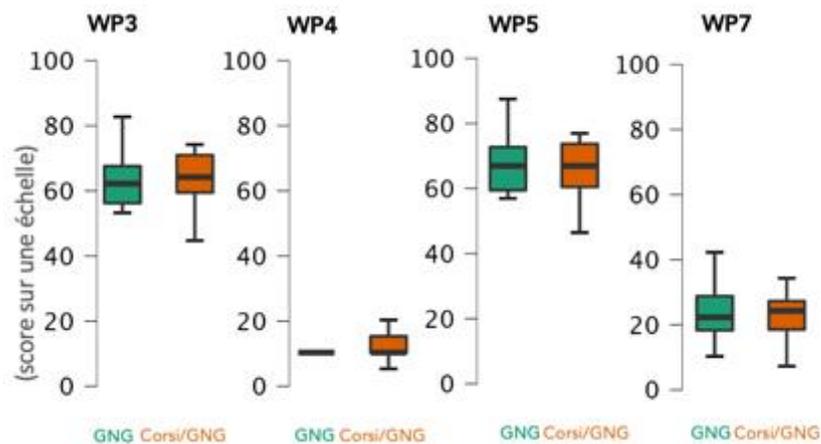


Figure 29 : La charge mentale à travers quatre dimensions du questionnaire WP selon le GNG et le Corsi/GNG.

De même, pour la dimension *WP4-verbal* (Annexe 23 ; $W = 13,000$; $p = 0,153$), pour la dimension *WP5-visuel* (Annexe 23 ; $W = 66,500$; $p = 0,959$) et pour le *WP7-manuel* (Annexe 23 ; $W = 87,000$; $p = 0,965$).

Question Ludique et Motivation

Concernant le ressenti ludique (**Figure 30**), entre les deux tâches, il était significativement différent et plus élevé pour le GNG ludique seul par rapport au Corsi/GNG (Annexe 23 ; $W = 148,500$; $p = 0,006$). Contrairement à la dimension *Motivation* (**Figure 30**), qui entre les deux tâches, n'était pas significativement différente (Annexe 23 ; $W = 43,000$; $p = 0,398$).

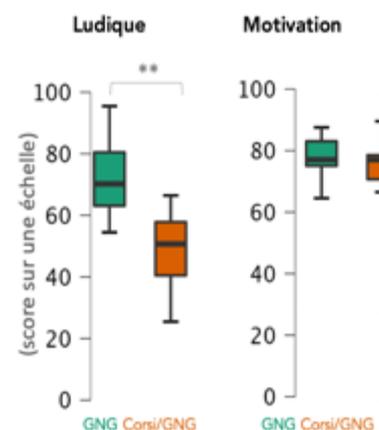


Figure 30 : Le ressenti Ludique (à gauche) et de Motivation (à droite) selon le GNG et le Corsi/GNG.

En conclusion, la tâche qui combinait le Corsi et le GNG lors d'un même essai obtenait des *Réponses attendues* qui n'étaient significativement pas différentes du GNG multicolore. Cependant, le ressenti de *Performance* était plus élevé pour le Corsi/GNG que pour le GNG seul. Concernant la charge mentale globale, elle était plus élevée pour le Corsi/GNG que pour le GNG, de même, pour la dimension *Exigences mentales*. Cependant, entre ces deux tâches, concernant les autres dimensions du NASA-TLX et l'ensemble des dimensions identifiées du WP, la charge mentale n'était pas significativement différente. Par conséquent, ce sont davantage les caractéristiques intrinsèques de la tâche combinée (ce que doivent faire les participants sur le plan cognitif) qui se traduisent par une augmentation de la charge mentale liée aux *Exigences mentales*, plutôt que les aspects liés à la mobilisation des ressources attentionnelles. Concernant le ressenti ludique, le GNG était jugé comme plus ludique que le Corsi/GNG, mais la *Motivation* n'était pas significativement différente entre ces deux tâches. Une piste d'explication est que le ressenti ludique est influencé par la complexité de la tâche en termes de multiplicité de fonctions cognitives sollicitées. Toutes ces pistes d'explication seront discutées davantage lors de la discussion de l'étude.

Comparaisons des moyennes d'une tâche à l'autre

Pour compléter l'analyse des résultats des deux premières comparaisons inter-tâches, qui ont permis d'étudier si une tâche sollicitant deux fonctions cognitives combinées impactait de manière différente la charge mentale et l'engagement par rapport à une tâche sollicitant une fonction cognitive. Nous avons également vérifié l'hypothèse **H3** (impact de la multiplicité des fonctions cognitives sur la charge mentale et sur l'engagement) en réalisant d'autres statistiques : des comparaisons de moyennes.

Premièrement, la moyenne de la charge mentale globale pour le Corsi multicolore (Cf. Annexe 23) était de 216,600 (écart-type ; $\sigma = 40,456$). La moyenne de la charge mentale globale pour le GNG ludique était de 226,250 ($\sigma = 53,490$). Enfin, la moyenne de la charge mentale globale pour le Corsi/GNG était de 284,450 ($\sigma = 40,456$ pour la comparaison avec le Corsi et $\sigma = 53,490$ pour la comparaison avec le GNG). Par conséquent, la charge mentale globale lorsqu'il y avait deux fonctions cognitives était

différente de la somme de chaque charge mentale pour les fonctions cognitives isolées (Somme ; $\Sigma = 442,85$). Ainsi, une tâche sollicitant deux fonctions cognitives (Corsi/GNG) augmentait la charge mentale globale ressentie de 31,32 % et de 25,72 % par rapport à une tâche sollicitant une fonction cognitive (pour respectivement le Corsi et le GNG). Le calcul utilisé pour le taux d'évolution était le suivant : Taux d'évolution = ((seconde valeur - première valeur) / première valeur) * 100.

Concernant les dimensions de la charge mentale subjective, nous avons comparé les dimensions *Exigences Mentales* et *Performances*, s'agissant des deux dimensions où il y avait des différences significatives entre les deux comparaisons inter-tâches (Corsi multicolore vs Corsi/GNG et GNG ludique vs Corsi/GNG). Premièrement, la moyenne de la dimension *Exigences mentales* pour le Corsi multicolore (Annexe 23) était de 54,200 ($\sigma = 9,415$). La moyenne de la dimension *Exigences mentales* pour le GNG ludique était de 51,300 ($\sigma = 10,837$). Enfin, la moyenne de la dimension *Exigences mentales* pour le Corsi/GNG était de 67,550 ($\sigma = 10,837$ pour la comparaison avec le Corsi et $\sigma = 9,415$ pour la comparaison avec le GNG). Par conséquent, la dimension *Exigences mentales* lorsqu'il y avait deux fonctions cognitives était différente de la somme de chaque charge mentale pour les fonctions cognitives isolées ($\Sigma = 105,5$). Ainsi, une tâche sollicitant deux fonctions cognitives (Corsi/GNG) augmentait les *Exigences mentales* ressenties de 24,63 % et de 31,68 % par rapport à une tâche sollicitant une fonction cognitive (pour respectivement le Corsi et le GNG).

Concernant la moyenne de la dimension *Performances* pour le Corsi multicolore (Cf. Annexe 23), elle était de 30,450 ($\sigma = 7,030$). La moyenne de la dimension *Performance* pour le GNG ludique était de 34,600 ($\sigma = 12,971$). Enfin, la moyenne de la dimension *Performances* pour le Corsi/GNG était de 50,200 ($\sigma = 7,030$ pour la comparaison avec le Corsi et $\sigma = 12,971$ pour la comparaison avec le GNG). Par conséquent, la dimension *Performance* lorsqu'il y avait deux fonctions cognitives était différente de la somme de chaque charge mentale pour les fonctions cognitives isolées ($\Sigma = 65,05$). Ainsi, une tâche sollicitant deux fonctions cognitives (Corsi/GNG) augmentait les *Performances* ressenties de 64,86 % et de 45,09 % par rapport à une tâche sollicitant une fonction cognitive (pour respectivement le Corsi et le GNG).

Enfin, nous avons comparé les différentes moyennes de la dimension du ressenti *Ludique*, car il y avait des différences significatives entre les deux comparaisons inter-tâches. Concernant la moyenne de la dimension *Ludique* pour le Corsi multicolore (Annexe 23), elle était de 66,000 ($\sigma = 11,893$). La moyenne de la dimension *Ludique* pour le GNG ludique était de 70,050 ($\sigma = 14,182$). Enfin, la moyenne de la dimension *Ludique* pour le Corsi/GNG était de 50,900 ($\sigma = 11,893$ pour la comparaison avec le Corsi et $\sigma = 14,182$ pour la comparaison avec le GNG). Par conséquent, la dimension *Ludique* lorsqu'il y avait deux fonctions cognitives était différente de la somme de chaque charge mentale pour les fonctions cognitives isolées ($\Sigma = 136,05$). Ainsi, une tâche sollicitant deux fonctions cognitives (Corsi/GNG) diminuait le ressenti *Ludique* de 22,88 % et de 27,33 % par rapport à une tâche sollicitant une fonction cognitive (pour respectivement le Corsi et le GNG).

En conclusion, une tâche sollicitant deux fonctions cognitives (Corsi/GNG) augmentait la charge mentale globale ressentie de manière proportionnelle (et non de manière cumulative) par rapport à une tâche sollicitant une fonction cognitive (le Corsi ou le GNG seuls). De même, pour les dimensions de charge mentale basée sur les *Exigences mentales* et les *Performances*. Une piste d'explication est que le Corsi et le GNG ont des aspects en commun (comme le type d'items), qui font que lorsque ces deux tâches se retrouvent combinées, l'augmentation de la charge mentale n'est pas cumulative, mais proportionnelle. Une piste d'explication que l'on discutera davantage lors de la discussion de l'étude 2.

Sollicitation d'une fonction cognitive en amont d'une tâche sollicitant plusieurs fonctions cognitives (i.e. tâche combinée) vs la tâche combinée seule.

Dans cette section d'analyse, nos résultats permettront de vérifier, avec d'autres tâches cognitives, la troisième hypothèse de l'étude 2, la **H3**. Effectivement, pour cette section, nous avons comparé la tâche combinée Corsi/GNG (également nommée partie A à la section 5.2.2.3 présentant les tâches cognitives combinées) avec la tâche combinée NBT + Corsi/GNG (nommée partie B). Cette comparaison permettait d'étudier si une tâche réalisée en amont influence la réalisation et le ressenti de charge mentale d'une tâche combinée, par rapport à une tâche combinée seule dans l'essai.

Avant de comparer la charge mentale subjective entre ces deux tâches, nous voulions comparer les performances en termes de *Réponses attendues* et du *Temps total*. Pour ces dimensions, seule la partie Corsi/GNG était comparable entre les deux tâches combinées. En effet, les dimensions de performances du NBT concernaient uniquement la Partie B. Par la suite, nous comparerons les données subjectives mesurées à travers le questionnaire NASA-TLX, le WP, la question *Ludique* et la question *Motivation*.

Mesures de performances

Dans un premier temps, nous avons donc comparé les performances en termes de *Réponses attendues* entre les deux parties (graphique de gauche **Figure 31**). Considérées deux à deux, elles avaient des *Réponses attendues* qui n'étaient pas significativement différentes (Annexe 23 ; $W = 40,500$; $p = 0,456$), de même pour le *Temps total pour la partie Corsi/GNG* qui n'était pas significativement différent (Annexe 23 ; $W = 77,000$; $p = 0,312$).

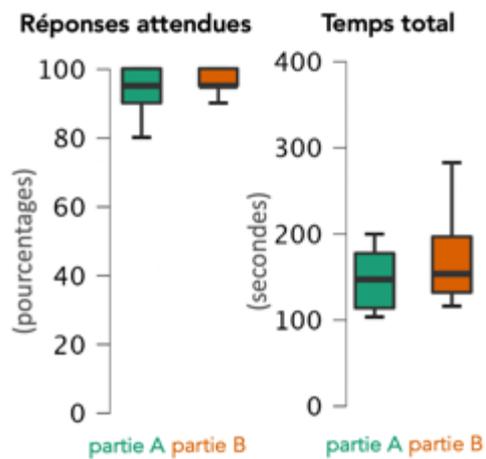


Figure 31 : Performances en termes de Réponses attendues (à gauche) et de Temps Total pour la partie Corsi/GNG selon la partie A et la partie B.

Questionnaire du NASA-TLX

Score global

Désormais, nous nous intéressons aux mesures subjectives, à commencer par la charge mentale globale. Entre les deux tâches, la charge mentale globale du NASA-TLX (Cf. **Figure 32**) n'était pas significativement différente (Annexe 23 ; $W = 85,500$; $p = 0,478$).

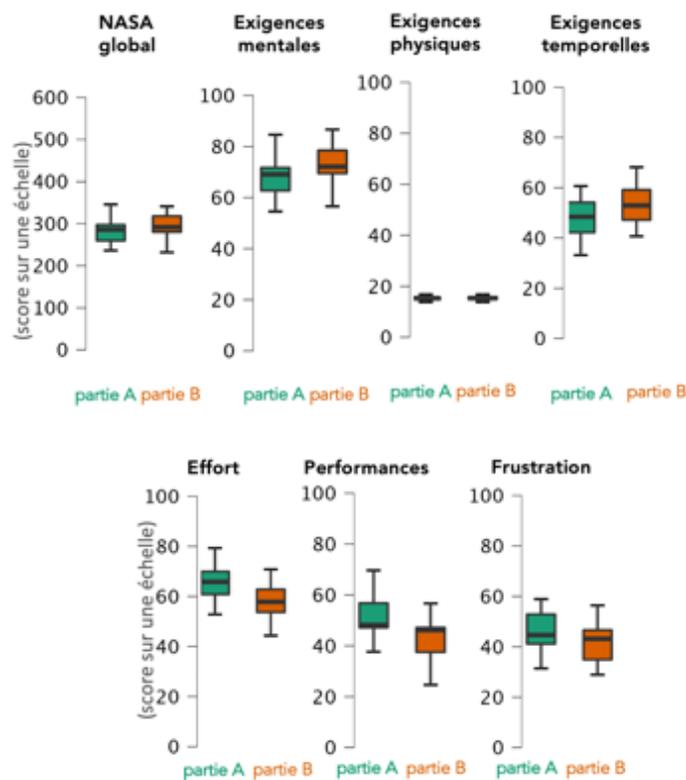


Figure 32 : La charge mentale basée sur le questionnaire NASA-TLX selon la partie A et la partie B.

Analyse dimension par dimension

Concernant le score des dimensions du questionnaire NASA-TLX, nous considérerons d'abord les *Exigences mentales* (Cf. **Figure 32**). Entre les deux parties, la dimension *Exigences mentales* n'était pas significativement différente (Annexe 23 ; $W = 41,500$; $p = 0,102$). De même, pour la dimension *Exigences physiques* (Annexe 23 ; $W = 37,00$; $p = 0,755$). Mêmes tendances pour toutes les autres dimensions du NASA-TLX, la dimension *Exigences temporelles* (Annexe 23 ; $W = 57,00$; $p = 0,131$), l'*Effort mental* (Annexe 23 ; $W = 51,500$; $p = 0,083$), la dimension *Performances* (Annexe 23 ; $W = 88,500$; $p = 0,111$) et enfin, la *Frustration* (Annexe 23 ; $W = 106,000$; $p = 0,383$), n'étaient pas significativement différentes entre les deux parties.

Questionnaire du WP

Concernant le score des dimensions du questionnaire WP, nous considérerons d'abord le WP3 (Cf. **Figure 33**). Entre les deux parties, la dimension *WP3-spatial* n'était pas significativement différente (Annexe 23 ; $W = 32,000$; $p = 0,116$).

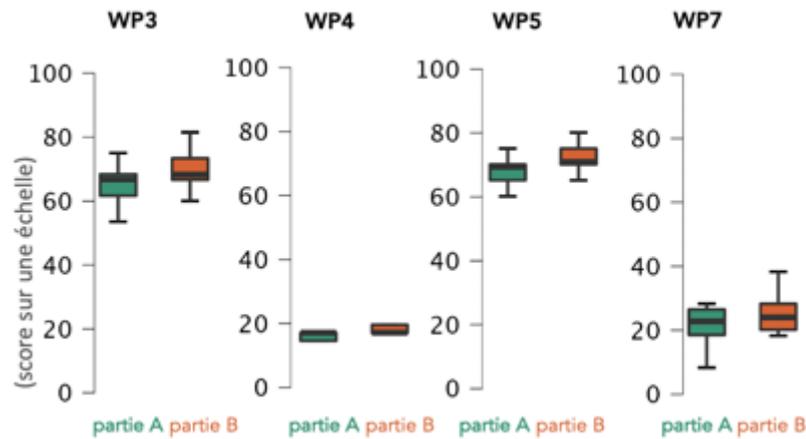


Figure 33 : La charge mentale à travers quatre dimensions du questionnaire WP selon la partie A et la partie B.

Concernant la dimension *WP4-verbale*, entre les deux parties, elle n'était pas significativement différente (Annexe 23 ; $W = 8,000$; $p = 0,181$), de même, pour le *WP5-visuel* (Annexe 23 ; $W = 21,000$; $p = 0,051$) et le *WP7-manuel* (Annexe 23 ; $W = 64,000$; $p = 0,359$).

Question Ludique et Motivation

Concernant la dimension *Ludique* (Figure 34), entre les deux parties, elle était significativement différente (Annexe 23 ; $W = 21,000$; $p = 0,029$) et la partie B (NBT + Corsi/GNG) était plus élevée que la partie A (Corsi/GNG seul). Inversement, la dimension *Motivation* (Figure 34), entre les deux parties, n'était pas significativement différente (Annexe 23 ; $W = 28,500$; $p = 0,722$).

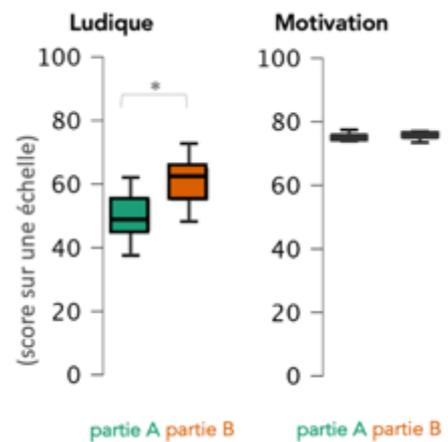


Figure 34 : Le ressenti Ludique (à gauche) et de Motivation (à droite) selon la partie A et la partie B.

En conclusion, la partie B (avec NBT) était jugée comme plus ludique que la partie A (sans NBT), mais la motivation restait la même, de même pour la charge mentale mesurée via le NASA-TLX, le WP et les performances (*Réponses attendues* et *Temps total pour Corsi/GNG*). Une explication serait que le NBT ludique seul est considéré

comme une tâche plus ludique que les autres. Une piste d'explication qui sera davantage discutée lors de la discussion de l'étude.

Une tâche simple lorsqu'elle précède une tâche combinée vs la tâche simple seule.

Dans cette section d'analyse, nos résultats permettront de vérifier, avec d'autres tâches cognitives, la troisième hypothèse de l'étude 2, la **H3**. Nous souhaitons comparer le NBT ludique avec le NBT ludique suivi du Corsi/GNG. Cette comparaison permettra d'étudier si la réalisation d'une tâche combinée qui suit une tâche simple (non combinée) au cours d'un même essai, influence la réalisation de la tâche simple.

Dans cette section, nous n'avions accès qu'aux performances pour les deux tâches. En effet, les données de charge mentale subjective de la partie NBT de la partie B, n'étaient pas disponibles, car elles concernaient l'ensemble de l'essai (NBT + Corsi/GNG). Nous ne pouvons donc pas dissocier les deux sous-parties de la partie B (NBT + Corsi/GNG) sans en changer la tâche. Au cours de l'essai de la Partie B, par exemple, nous ne pouvions pas interrompre les participants après la réalisation du NBT pour répondre aux questions du NASA-TLX, puis continuer l'essai avec le Corsi/GNG.

Parmi toutes les dimensions de performances du NBT, nous avons choisi d'en comparer trois, en se basant sur la logique de la tâche cognitive. Les dimensions de performance concernaient les *Réponses attendues*, le *Temps de réaction tous items confondus* et les *Fausses alarmes* (correspondant à l'appui sur la touche gauche (cible) lors de l'affichage d'un item non-cible où le participant doit appuyer sur la touche droite normalement).

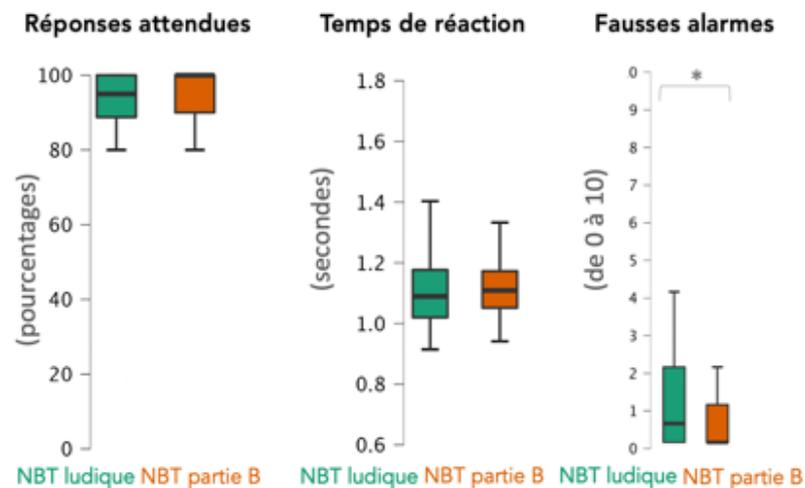


Figure 35 : Performances en termes de Réponses attendues (à gauche), de Temps de réaction (au milieu) et de Fausses alarmes (à droite) selon le NBT ludique unicolore et le NBT ludique de la partie B.

Ainsi, dans un premier temps, nous avons comparé les performances en termes de *Réponses attendues* entre les deux NBT (Cf. graphique de gauche **Figure 35**). Considérés deux à deux, ils avaient des *Réponses attendues* qui n'étaient pas significativement différentes (Annexe 23 ; $W = 17,500$; $p = 0,094$). De même, les deux NBT avaient un *Temps de réaction* qui n'était significativement pas différent (Annexe 23 ; $W = 88,000$; $p = 0,546$). Cependant, les deux NBT avaient un nombre de *Fausses alarmes* significativement différent et plus élevé pour le NBT ludique seul par rapport au NBT suivi du Corsi/GNG (Annexe 23 ; $W = 67,000$; $p = 0,028$).

En conclusion, que ce soit la sollicitation de la mémoire visuo-spatiale uniquement (à travers un NBT ludique) ou lorsqu'elle est suivie d'une tâche combinée dans un même essai (NBT + Corsi/GNG), l'impact sur les performances en termes de *Réponses attendues* et du *Temps de réaction* n'était pas significativement différent. Cependant, le NBT ludique seul générait plus de *Fausses alarmes* que le NBT suivi du Corsi/GNG. Une piste d'explication est qu'au moment de réaliser le NBT lors de la tâche combinée, il y a une anticipation de la partie suivante qui est plus compliquée. Une piste d'explication qui sera davantage discutée lors de la discussion de l'étude 2 ci-après.

5.4 Discussion de l'étude 2

Pour rappel, les hypothèses suivantes ont été posées :



H1. Un modèle de classification réalisé dans une situation expérimentale A (en ligne), peut prédire la charge mentale subjective d'une situation expérimentale B (en présentiel).

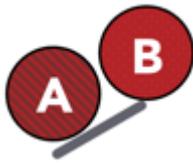
H2a. Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, alors le design ludique, par rapport à un design classique, augmente le ressenti ludique.

H2b. Si le ressenti ludique provient du design ludique et non de la multiplicité des fonctions cognitives, la charge mentale subjective liée aux caractéristiques de la tâche (comme les exigences mentales) ou aux stratégies de l'individu (comme l'effort mental) devrait diminuer. En parallèle le design ludique devrait augmenter la charge mentale subjective en lien avec la mobilisation des ressources attentionnelles.

H2c. Si le design ludique est accentué par la couleur et que le ressenti ludique est lié au design, alors il devrait augmenter et les effets consécutifs sur la charge mentale devraient être accentués.

H3a. Si le ressenti ludique provient de la multiplicité des fonctions cognitives, alors une tâche combinée modifie le ressenti ludique, par rapport à une tâche simple.

H3b. Une augmentation du nombre de fonctions cognitives entraîne une augmentation de la charge mentale. Cette augmentation ne correspond pas à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées), mais plutôt à une augmentation proportionnelle de la charge mentale.



Discussion sur l'hypothèse H1

L'hypothèse **H1** porte sur la transférabilité d'un modèle de classification réalisé dans une situation expérimentale A (en ligne), pouvant prédire la charge mentale subjective d'une situation expérimentale B (en présentiel). En effet, il y a des défis liés aux modèles de prédictions de la charge mentale, celui d'avoir des modèles plus précis, plus fiables, puis plus génériques et donc des modèles de classification inter-contextes. Or, après une revue de la littérature, nous n'avons pas identifié d'étude ayant réalisé de la classification de la charge mentale (à trois classes) inter-contextes (en ligne et en présentiel pour une même tâche donnée). Nous n'avons donc aucun moyen de comparer nos résultats de transférabilité inter-contextes.

Concernant nos résultats, ils indiquent que nous pouvons envisager la transférabilité d'un modèle de prédiction inter-contextes pour le test de Corsi et le GNG. En effet, nous n'observons pas de chute du score de précision à la limite du niveau de hasard (égal à 33 % pour trois classes de charge mentale) d'un contexte A à un contexte B pour le Corsi et le GNG. Ainsi, nous pouvons envisager d'entraîner un unique modèle de classification basé sur le test de Corsi en ligne, par exemple, pouvant prédire la charge mentale subjective du Corsi lorsqu'il est réalisé dans une situation expérimentale en laboratoire. Ce qui permettrait un gain de temps (exemples : pas de déplacement des participants en laboratoire, moins d'organisation pour l'équipe de recherche). Cependant, la transférabilité inter-contextes reste pour l'instant au stade d'une perspective envisageable et ceci, uniquement pour le test de Corsi et le GNG. L'échec du modèle de classification inter-contextes à classer correctement les données de charge mentale subjective de l'étude 2 pour la tâche de NBT peut provenir du type de tâche. En effet, la différence entre les trois tâches, le Corsi, le NBT et le GNG est que la première est une tâche non-séquentielle, contrairement aux deux autres. Ce qui laisse penser qu'un modèle de classification inter-contextes dépend du type de tâche. Mais, cette hypothèse est à tester dans une étude comparative avec plusieurs autres tâches (non) séquentielles.



Discussion sur l'hypothèse H2

L'hypothèse **H2** porte sur l'augmentation du ressenti ludique et de la charge mentale subjective avec le design ludique (par rapport à un design classique). Nous décomposons donc cette discussion en deux sous-

sections : l'impact du design ludique sur le ressenti ludique, puis l'impact du design ludique sur la charge mentale subjective.

Mais avant toute chose, intéressons-nous à l'indicateur de réussite à la tâche, les performances. Pour voir si les deux situations expérimentales sont équivalentes en termes de complexité. En nous basant sur la précision de la réponse (*Réponses attendues*), elle n'est pas différente entre un Corsi classique et un Corsi ludique. Ainsi, les deux situations expérimentales sont équivalentes en termes de complexité. Cependant, un design ludique diminue le temps de réponse (*Temps total*). Cette différence peut être due à la nature intrinsèquement *challengeante* de la plupart des jeux ayant par définition un design ludique. En effet, selon plusieurs auteurs (Aponte et al., 2011 ; Cole et Gillies, 2021), le *challenge*/défi est la clé de l'expérience de jeu et le défi a un impact sur la performance (Zook et Riedl, 2014).

L'impact d'un design ludique sur le ressenti ludique

Désormais, intéressons-nous à l'impact du design ludique sur le ressenti ludique. Tout d'abord, à la suite de nos résultats, cette partie de l'hypothèse **H2** (*H2a* - le design ludique augmente le ressenti ludique) est infirmée. En effet, entre un design classique et un design ludique, nous n'observons pas de différence au niveau du ressenti ludique (et de la motivation, faisant partie de la notion d'engagement avec le ressenti ludique). En parallèle, quelle que soit la tâche, ces deux dimensions sont jugées avec des valeurs au-dessus de la moyenne, plus précisément supérieure à 60 et à 70 (sur une échelle de 100) pour respectivement le ressenti Ludique et le ressenti Motivation. Ainsi, nos résultats laissent penser que ces deux ressentis ne dépendent pas du design de la tâche en lui-même.

Par ailleurs, si nous modifions une dimension, la couleur des bouées, de l'environnement de la tâche (design ludique multicolore), nous n'observons aucune différence significative de ressenti ludique avec le design ludique unicolore. Or, nous avons posé l'hypothèse (*H2c*) qu'il y aurait plus de ressenti ludique, car Goethe (2019) a indiqué que la couleur a un effet sur nos émotions. Ainsi, la couleur est une première étape vers un design ludique. Donc, en augmentant les couleurs, nous devons accentuer les émotions et donc le ressenti ludique. Une explication est que la couleur concerne une dimension restreinte du design ludique qui comprend d'autres éléments, comme ceux

du contexte (vagues, nénuphars) ou de la consigne qui ont plus d'impact sur le ressenti ludique.

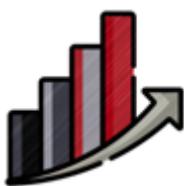
L'impact d'un design ludique sur la charge mentale subjective

À la suite de nos résultats, cette partie de l'hypothèse **H2** (*H2b* - augmentation de la charge mentale avec le design ludique) est confirmée en partie. En effet, un design ludique augmente seulement la charge mentale évaluée à travers les dimensions du WP entrant en jeu dans le test de Corsi (*Traitement spatial, Traitement manuel, Traitement visuel*). Cependant, le design ludique n'a pas d'impact sur la charge mentale liée aux exigences de la tâche qui est mesurée à travers les trois premières dimensions du questionnaire NASA-TLX (*Exigences mentales, physiques et temporelles*). Une explication probable concerne le nombre d'éléments visuels dans un design ludique qui est plus élevé que dans un design non ludique. En effet, il y a des éléments visuels supplémentaires, tels que des vagues d'eau ou des nénuphars, ce qui augmente la charge mentale basée sur le *Traitement visuel*. La même logique se retrouve pour la charge mentale basée sur le *Traitement spatial*, probablement parce qu'elle nécessite de se concentrer sur un seul aspect (la configuration visuelle des bouées) et d'ignorer les autres éléments visuels (comme les vagues ou les nénuphars). Par ailleurs, un design ludique augmente la charge mentale basée sur le *Traitement manuel*, probablement en raison de la propriété intrinsèque de l'élément qu'il faut pointer. En effet, le design ludique avec des bouées (dans lesquelles les participants entrent par le milieu) peut inciter les participants à cliquer au niveau du centre de la bouée, un espace plus petit que les cubes du test de Corsi classique (qui ont une surface plane) et donc plus complexe à atteindre avec la souris d'ordinateur, d'où la perception d'un plus grand effort manuel. Ce rapport taille-difficulté de la tâche est expliqué par la loi de Fitts (Fitts, 1954). La loi de Fitts permet de caractériser une tâche de pointage (avec souris d'ordinateur ou pas) par un indice de difficulté qui est fonction de la largeur et de la distance de la cible.

Si nous modifions une dimension (la couleur des bouées) du design ludique, nous n'observons aucune différence significative de charge mentale (NASA-TLX et WP) entre un Corsi avec des bouées unicolores (toutes jaunes) et un Corsi avec des bouées de différentes couleurs. Or, nous supposons une augmentation de la charge mentale subjective, car les différentes couleurs sont des stimuli visuels qui demandent plus de ressources attentionnelles visuelles qu'un design avec des items unicolores. Une

explication est que la couleur concerne une dimension restreinte du design ludique qui comprend d'autres éléments, comme ceux du contexte (vagues, nénuphars) ou de la consigne qui ont plus d'impact sur les ressources attentionnelles que la couleur. Lorsque l'on change l'ensemble (les couleurs, les formes, la narration, le contexte) à partir d'un contexte classique, nous observons un impact sur la charge mentale subjective. Cependant, si à partir d'un design ludique, nous modulons uniquement la couleur, nous n'observons pas d'impact sur la charge mentale subjective. Ainsi, dans une prochaine expérimentation, nous pourrions étudier l'impact d'une seule dimension du design ludique, en comparant un Corsi classique avec des cubes gris (comme celui de notre étude) avec un Corsi classique ayant des cubes colorés.

En conclusion, l'impact du design n'est pas perçu consciemment par les participants. Même s'ils déclarent plus de charge mentale liée aux ressources attentionnelles mobilisées, la tâche n'est pas jugée différemment sur la question portant sur le ressenti du ludique. Ainsi, le design ludique, qui est une dimension extrinsèque aux exigences de la tâche, n'impacte pas le ressenti ludique, mais impacte la charge mentale liée à la mobilisation des ressources attentionnelles. Par conséquent, en nous basant sur le modèle proposé à la section 1.2.3, nous pouvons conclure que le design ludique est bel et bien un facteur extrinsèque aux exigences de la tâche qui impacte la charge mentale.



Discussion sur l'hypothèse H3

L'hypothèse **H3** porte sur la modulation du ressenti ludique avec la multiplicité des fonctions cognitives (par rapport à une tâche sollicitant une fonction cognitive seule). Par ailleurs, nous posons l'hypothèse qu'une augmentation du nombre de fonctions cognitives ne correspond pas à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées), mais plutôt à une augmentation proportionnelle de la charge mentale. Nous décomposons donc cette discussion en deux sous-sections : l'impact de la multiplicité des fonctions cognitives sur le ressenti ludique (*H3a*), puis l'impact de la multiplicité des fonctions cognitives sur la charge mentale subjective (*H3b*).

Mais avant toute chose, intéressons-nous à l'indicateur de réussite à la tâche, les performances. Afin de savoir si la complexité était similaire entre les deux situations expérimentales. Les résultats indiquent que le Corsi/GNG obtient de meilleures

performances en termes de *Réponses attendues* par rapport au Corsi multicolore. Ce qui est corroboré avec le ressenti de *Performance* qui est plus élevé pour le Corsi/GNG que pour le Corsi simple. Ce résultat peut s'expliquer par le plan expérimental. En effet, le test de Corsi (classique) et ses variantes (ludiques) sont redondants dans le plan expérimental de l'étude 2. De manière aléatoire, les participants réalisent cinq variantes du Corsi tout au long de l'étude, contrairement au Corsi/GNG qui est réalisé qu'une seule fois. Ainsi, l'effet de nouveauté peut expliquer de meilleure performance du Corsi/GNG par rapport au Corsi simple qui peut avoir un effet de lassitude auprès des participants. Effectivement, ce résultat ne se retrouve pas pour la comparaison entre le GNG simple et le Corsi/GNG où l'on n'obtient pas de différence significative en termes de performance. C'est pour cela que nous avons privilégié l'explication liée à la lassitude par rapport à un éventuel effet d'entraînement du test de Corsi qui se retrouve à plusieurs reprises dans le plan expérimental de l'étude 2. De fait, contrairement au Corsi, le GNG n'a pas cet aspect de redondance au niveau du plan expérimental de l'étude 2.

L'impact de la multiplicité des fonctions cognitives sur le ressenti ludique

Désormais, intéressons-nous à l'impact de la multiplicité des fonctions cognitives sur le ressenti ludique. Tout d'abord, à la suite de nos résultats, cette partie de l'hypothèse **H3** (*H3a* - la multiplicité des fonctions cognitives modifie le ressenti ludique) est confirmée. Effectivement, le ressenti ludique est plus élevé pour une tâche simple, que pour une tâche combinée. Ainsi, la diminution du ressenti *ludique* quand les deux fonctions cognitives sont cumulées ne vient pas du fait que l'une ou l'autre des fonctions cognitives serait perçue comme moins ludique, mais bien de l'impact du cumul des deux fonctions cognitives au cours d'un même essai. Par conséquent, lorsque l'on manipule le design ludique (une variable extrinsèque aux exigences de la tâche, rattachée au visuel et à la narration de l'activité), cela n'impacte pas le ressenti *ludique* perçu. En revanche, lorsque l'on augmente la complexité en termes de fonctions cognitives sollicitées (une variable extrinsèque aux exigences de la tâche), à ce moment, le ressenti *Ludique* perçu diminue, alors que le design ludique, lui, ne change pas (on n'a pas augmenté ou diminué la *ludification* de la tâche combinée).

L'impact de la multiplicité des fonctions cognitives sur la charge mentale

Enfin, à la suite de nos résultats, l'hypothèse (*H3b*) selon laquelle la multiplicité des fonctions cognitives augmenterait la charge mentale de manière proportionnelle (et non de manière cumulative) est confirmée. Effectivement, les résultats pour le Corsi vs Corsi/GNG indiquent une augmentation proportionnelle de la charge mentale globale, mais aussi de la charge mentale basée sur les *Exigences mentales* et l'*Effort mental* (les deux dimensions entrant le plus en jeu dans ce type de tâche ; section 4.3.1 de l'étude 1). Ainsi, l'augmentation des exigences de la tâche (doublement du nombre de clics par exemple) induit une augmentation proportionnelle (et non cumulative) de la charge mentale mesurable à travers le NASA-TLX. Cependant, il n'y a aucune différence au niveau de la charge mentale liée à la mobilisation des ressources attentionnelles (mesurée via le WP). Nous retrouvons ce même résultat pour le GNG ludique seul comparé au Corsi/GNG (mise à part pour la dimension *Effort* qui n'est pas différente entre ces deux tâches). Tout d'abord, une piste expliquant que la charge mentale basée sur le NASA-TLX augmente de manière proportionnelle lors d'une tâche combinée par rapport à une tâche simple pourrait provenir du fait que le Corsi et le GNG ont des aspects en commun (comme le type d'items). Ainsi, lorsque ces deux tâches se retrouvent combinées, l'augmentation de la charge mentale n'est pas cumulative, mais proportionnelle. Effectivement, ce sont deux tâches cognitives ayant le même type d'items, des bouées dans un environnement aquatique. En outre, ces deux tâches cognitives se basent sur la mémoire. En effet, le Corsi sollicite la mémoire visuo-spatiale. Quant au GNG, il sollicite principalement l'inhibition motrice. Mais pour inhiber sa réponse motrice, il faut se remémorer à quel moment il faut le faire.

Concernant les résultats basés sur le WP, ils sont surprenants, car nous nous attendions à une augmentation du *Traitement manuel* (WP7) pour le Corsi/GNG par rapport au Corsi ou au GNG simples. En effet, par exemple, il faut cliquer deux fois sur les bouées lors de la restitution de la mémorisation visuo-spatiale, au lieu d'une pour le Corsi simple. De même, nous nous attendions à une augmentation du *Traitement visuel* (WP5) pour le Corsi/GNG, car il y a plus d'éléments visuels (fleurs) que le Corsi simple. Or, les résultats ne révèlent aucune différence significative en termes de charge mentale basée sur le WP. Par conséquent, ce qui impacte la charge mentale, ce sont davantage les caractéristiques intrinsèques de la tâche combinée (ce que doivent faire les participants sur le plan cognitif) qui se traduisent par une augmentation de la charge

mentale liée aux *Exigences mentales* et à l'*Effort mental*, plutôt que les aspects liés à la mobilisation des ressources attentionnelles.

Par ailleurs, la sollicitation d'une fonction cognitive seule dans un contexte ludique induit de prime abord un niveau élevé de charge mentale liée aux ressources attentionnelles mobilisées (liée au WP). L'ajout d'une fonction cognitive au cours de la tâche ne change pas ce ressenti de charge mentale (lié au WP), mais change le ressenti de charge mentale lié aux exigences de la tâche, aux stratégies et à l'état émotionnel de l'individu (évalué via le NASA-TLX). Autrement dit, le design ludique (un facteur extrinsèque rattaché à une dimension visuelle et narrative de la tâche) impacte la charge mentale liée aux ressources attentionnelles alors que l'ajout de fonctions cognitives (un facteur extrinsèque rattaché aux mécaniques intrinsèques de la tâche) impacte la charge mentale liée aux exigences de la tâche et aux stratégies de l'individu, ainsi que celle liée à son état émotionnel. Ce résultat supporte une fois de plus l'argument sur la complémentarité des deux questionnaires sur la charge mentale ressentie (Cf. section 2.1.2.4).

Toujours concernant l'hypothèse 3 sur la multiplicité des fonctions cognitives, nous nous étions également demandé si la réalisation d'une tâche combinée précédée d'une tâche simple au cours d'un même essai, influencerait la réalisation de la tâche combinée. En posant l'hypothèse que la charge mentale augmenterait lorsqu'on sollicitait une fonction cognitive en amont de la tâche combinée, par rapport à la tâche combinée seule. En effet, ajouter une fonction cognitive en amont peut générer l'impression que les ressources mentales sont déjà impactées par ce qu'on vient de faire et donc sont moins disponibles au moment de faire la tâche combinée. Cette hypothèse est infirmée. Que la tâche combinée (Corsi/GNG) soit précédée ou non d'une tâche simple (le NBT), cela ne change en rien le ressenti de charge mentale (NASA-TLX et WP), de même pour les performances et la motivation. Cependant, une dimension mesurée diffère significativement, celle du ressenti *Ludique* qui est plus élevée lorsque les participants réalisent le NBT suivi du Corsi/GNG par rapport au Corsi/GNG seul. Une explication serait que le NBT ludique seul est considéré comme une tâche plus ludique que les autres. Une hypothèse confirmée par le résultat du ressenti ludique du NBT qui est significativement plus élevé que celui du Corsi/GNG (Table 6 en Annexe 23 ; $W =$

125,500 ; $p = 0,022$). Ainsi, lorsqu'il y a un ajout du NBT ludique en amont du Corsi/GNG, il est logique d'avoir une augmentation du ressenti *Ludique*.

Enfin, concernant la dernière comparaison, les performances lors du NBT simple, par rapport aux performances du NBT qui est suivi d'une autre tâche, les résultats sont mitigés. Nous nous attendions à avoir une différence de performances entre les deux tâches, car les participants peuvent anticiper la seconde partie de la tâche lors du NBT, suivi du Corsi/GNG, contrairement au NBT seul. Or, cette hypothèse est infirmée, car cela dépend des dimensions de performances considérées. Les *Réponses attendues* et le *Temps de réaction* n'étaient pas significativement différents entre les deux tâches. Cependant, le NBT ludique seul génère plus de *Fausses alarmes* que le NBT suivi du Corsi/GNG. Une explication pourrait être qu'au moment de réaliser le NBT lors de la tâche combinée, il y a une anticipation de la partie suivante qui est plus compliquée, par conséquent, plus *challengeante*. Ainsi, dès le départ, l'exigence de réussite est plus élevée que pour la tâche simple.

En conclusion de l'hypothèse 3, le ressenti ludique est influencé par une caractéristique intrinsèque aux exigences de la tâche, la multiplicité de fonctions cognitives, et non par une caractéristique extrinsèque aux exigences de la tâche, le design ludique, comme nous aurions pu le penser. Lors d'une tâche combinée, le ressenti ludique est plus faible que lors d'une tâche simple. De plus, la charge mentale augmente que ce soit avec le design ludique ou la multiplicité des fonctions cognitives. Néanmoins, il est important de distinguer la charge mentale liée aux caractéristiques de la tâche qui augmente avec la variable intrinsèque aux exigences de la tâche (la multiplicité de fonctions cognitives) et la charge mentale liée à la mobilisation des ressources attentionnelles qui augmente avec la variable extrinsèque exigences de la tâche (le design ludique).

Par conséquent, l'augmentation de la charge mentale lors de la tâche combinée pourrait être expliquée par le ressenti ludique qui est plus faible dans cette situation expérimentale, par rapport à une tâche simple. En effet, comme vu à la section théorique 1.2.3, le ressenti ludique peut être un médiateur interne de la charge mentale ayant pour effet de la diminuer quand il est élevé.

Perspective de la triangulation des données

Pour finir, le plan expérimental de cette seconde étude a été guidé par une des mesures, celle de l'EEG. L'objectif est de pouvoir la trianguler avec les mesures de performances et les mesures subjectives, afin d'avoir une mesure plus juste de la charge mentale (Cf. section 2.3). Ainsi, l'objectif est de prendre en considération les trois types de mesure pour établir le niveau de charge mentale d'un individu. Un exemple est donné au niveau de la **Figure 36** lors de la réalisation de la tâche de NBT. Si nous observons une charge mentale élevée lors des mesures déclaratives (questionnaires issus de deux modèles conceptuels complémentaires), de l'EEG (illustration avec neuf électrodes sur seize dans notre protocole ; Cf. section 5.2.3) et des performances, nous pouvons conclure à une charge mentale élevée lors de cet essai du NBT.

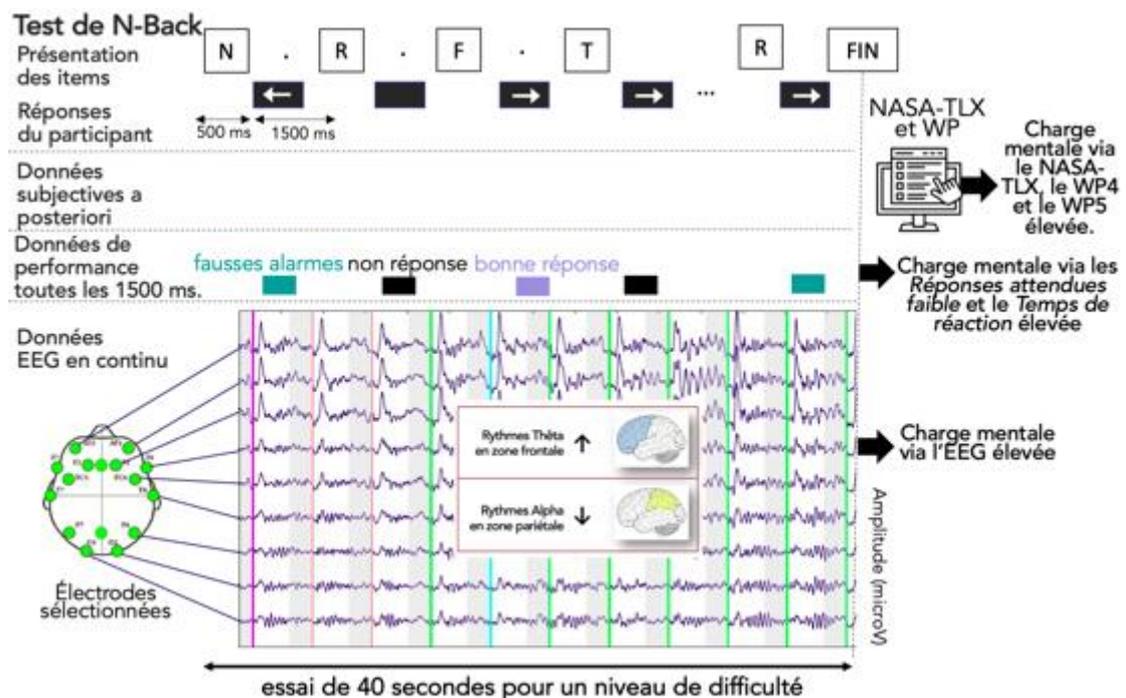


Figure 36 : Illustration à titre d'exemple de triangulation de la mesure de la charge mentale lors de la tâche de N-back en prenant en compte les données EEG, les performances et les questionnaires.

Enfin, dans cet objectif de triangulation des mesures de la charge mentale, nous présentons ci-après un tableau (Cf. **Table 2**) basé sur les résultats de nos deux études expérimentales. Il a pour vocation d'aider à la sélection des mesures de performances et des dimensions du WP les plus efficaces pour chaque tâche cognitive. Le questionnaire NASA-TLX étant de facto sélectionné, car il est complémentaire au WP et est l'outil

d'évaluation subjective de la charge mentale le plus couramment utilisé (Cf. section 2.1.2.4). Dans le cadre de ce travail de thèse, il n'a pas été possible d'approfondir davantage les analyses du signal EEG et de mettre effectivement en place la triangulation. Mais cela fera l'objet d'un travail futur réalisé par le consortium GHOST présenté en introduction de ce manuscrit (pp. 5).

Tâches	WP	Performances
NBT	WP4 et WP5	<i>Réponses attendues, Fausses alarmes, Omissions, Temps de réaction pour tous les items et pour les bonnes réponses et Non-réponses.</i>
Corsi	WP3 et WP5	<i>Faux cubes, Omissions, Séquences exactes et Temps total</i>
GNG	WP5 et WP7	<i>Réponses attendues, Temps de réaction pour tous les items et les bonnes réponses et Omissions</i>
WCST	WP3 et WP5	<i>Réponses attendues, Nombre d'erreurs, Erreurs persévératives et Temps de réaction pour tous les items</i>
DT	WP3 et WP5	<i>Réponses attendues, Temps du calcul mental et du Corsi, Cubes faux, Omissions et Séquences exactes.</i>

Table 2 : Sur la base des résultats des études expérimentales présentées dans le manuscrit de thèse, sélection, pour chaque tâche, des données de performance et des dimensions du WP pour la mesure triangulaire de la charge mentale.

Conclusion et limites

Pour cette seconde, nous nous sommes à nouveau basés sur le modèle conceptuel de la charge mentale proposé au Chapitre 1. Lors de l'étude 1, nous avons étudié l'impact sur la charge mentale et sur le ressenti ludique d'une variable intrinsèque aux exigences de l'activité, la complexité de la tâche. Cette fois-ci, lors de l'étude 2, nous avons étudié l'impact sur la charge mentale et sur le ressenti ludique d'une autre variable intrinsèque aux exigences de la tâche, la multiplicité des fonctions cognitives sollicitées. Par ailleurs, nous avons étudié une variable extrinsèque aux exigences de la tâche, le design ludique, qui avait, une influence sur la charge mentale liée aux ressources attentionnelles mobilisées.

De plus, à la suite de nos conclusions, il nous semble d'autant plus important de distinguer la notion de *variables* et de *facteurs extrinsèques aux exigences de la tâche* en fonction de leur influence sur la charge mentale, que nous avons initiée lors de la

partie théorique (Cf. section 1.2.3). Toutes les variables extrinsèques aux exigences de la tâche ne sont pas forcément des facteurs qui impactent la charge mentale. De plus, les variables internes à l'individu, pouvant être des médiateurs de la charge mentale, se comportent différemment en fonction de la manipulation de variables intrinsèques ou extrinsèques aux exigences de la tâche. Le ressenti ludique (variable interne à l'individu) n'est pas impacté par les changements de design ludique (facteur extrinsèque aux exigences de la tâche influençant la charge mentale), mais par la multiplication de fonctions cognitives (facteur intrinsèque aux exigences de la tâche influençant également la charge mentale).

Enfin, il est important de préciser que notre seconde étude a un certain nombre de limites. À commencer, par le nombre de participants égal à 20, qui a été plafonné par la technique de mesure de l'EEG. Une autre limite, que nous identifions, concerne la tâche combinée ayant trois fonctions cognitives lors d'un même essai. Nous avons fait précéder une tâche de NBT au Corsi/GNG. Nous aurions pu également faire le contraire, avec dans un premier temps, la réalisation du Corsi/GNG suivi du NBT. Dans une prochaine étude, il serait donc intéressant de modifier l'agencement des fonctions cognitives au cours d'un même essai.

Enfin, pour des travaux futurs, nous pourrions étudier d'autres éléments du design ludique (voir de la gamification) comme l'impact sur la charge mentale des retours/*feedbacks* visuels lors de bonnes réponses ou avoir un système de récompense après avoir accompli une tâche cognitive.

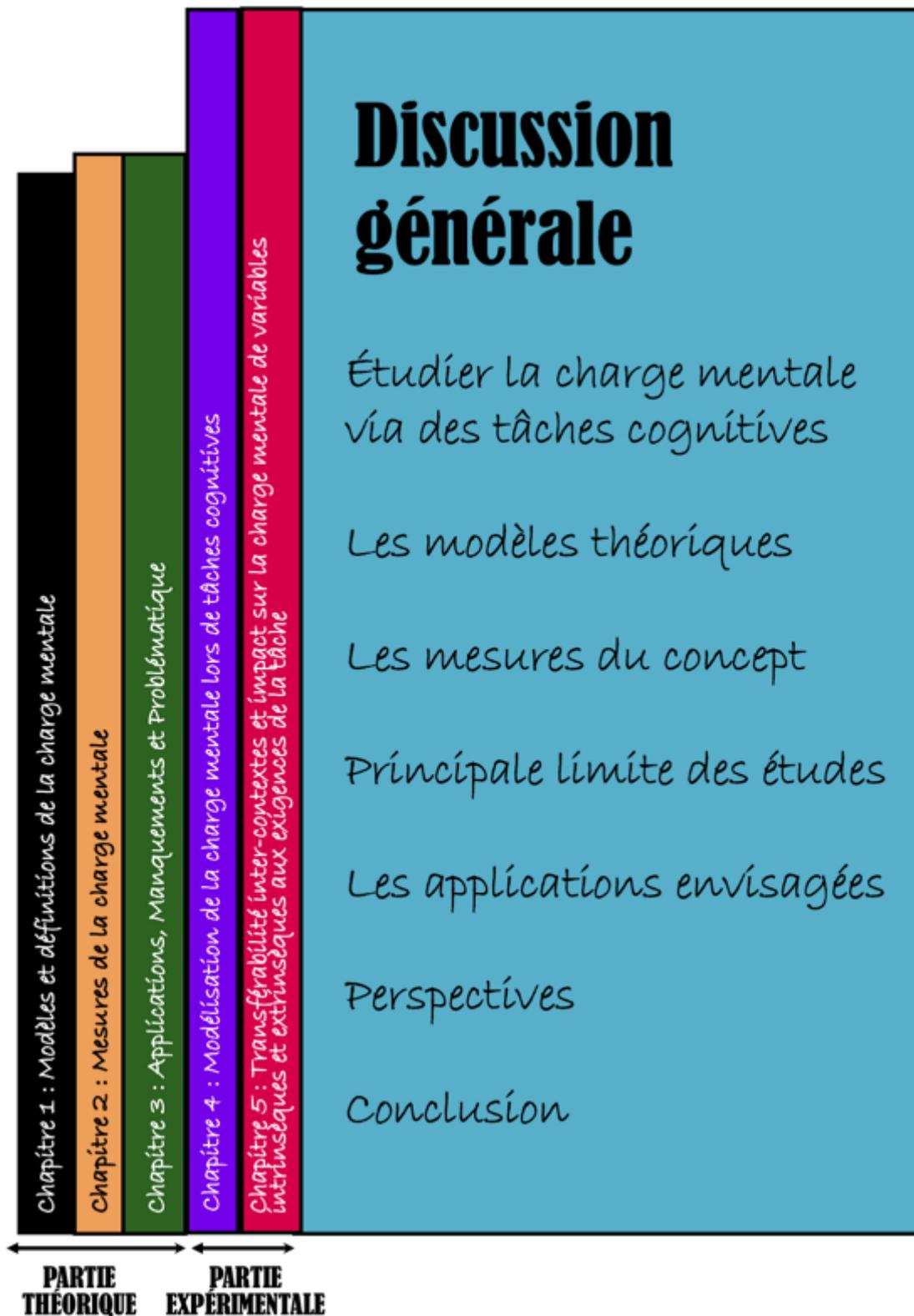


Figure 1 : Plan général du manuscrit de thèse et détail des sections de la Discussion générale.

La Discussion générale reprend en partie l'organisation du manuscrit de thèse (Cf. **Figure 1**). Dans un premier temps, nous allons discuter du titre de la thèse "*La charge mentale lors de tâches cognitives*". Quelles sont les visions qui s'opposent sur ce choix de contexte et peut-on généraliser nos résultats à des situations opérationnelles ? Puis, nous discuterons des modèles théoriques de la charge mentale et des techniques de mesure de ce concept. Par la suite, nous discuterons de la principale limite de nos études expérimentales et nous reprendrons les applications envisagées lors de la problématique en précisant si nos conclusions expérimentales permettent de les réaliser. Enfin, nous discuterons des perspectives de notre recherche avant de ponctuer le manuscrit par une conclusion générale.

1. Étudier la charge mentale à travers des tâches cognitives

Le titre de la thèse nous informe, dès le départ, que nous étudions la charge mentale lors de tâches cognitives. Il nous semble donc pertinent de discuter en premier lieu de ce choix et de préciser les deux visions opposées concernant ce parti pris. Tout d'abord, la pensée de Miller (2001) défend l'idée selon laquelle la charge mentale ne peut être étudiée dans un environnement expérimental. Si tel est le cas, les conclusions de la recherche expérimentale basée sur des tâches cognitives ne peuvent être appliquées à des situations réelles. Une vision qui s'oppose à celle d'auteurs plus récents (Berka et al., 2007 ; Radüntz, 2017 ; Guan et al., 2021) qui recommandent d'étudier la charge mentale à travers des tâches cognitives. Ces derniers auteurs se basent sur le constat que les situations réelles peuvent être décomposées en plusieurs fonctions cognitives. À la suite de notre recherche, nous nous positionnons dans la lignée des derniers auteurs, tout en nuanciant le propos. En effet, étudier la charge mentale à travers des tâches cognitives a l'avantage de permettre d'avoir des situations expérimentales contrôlées, moins sujettes aux variabilités imprévues d'un terrain opérationnel. De plus, comme nous l'avons vu à la section *3.1 Applications*, la plupart des études sur la charge mentale sont contextualisées dans une situation opérationnelle spécifique (comme la conduite automobile). L'avantage d'utiliser des tâches cognitives est qu'on se décontextualise d'une situation opérationnelle. Nous rejoignons ainsi l'argument de Berka et al. (2007), Radüntz (2017) et Guan et al. (2021) sur les situations réelles qui peuvent se décomposer en plusieurs fonctions cognitives sollicitées. Cependant, l'hypothèse selon laquelle les

conclusions de la recherche concernant la charge mentale lors de tâches cognitives pourraient être transférées à des situations réelles est excessive. Nous pouvons l'affirmer à la suite de notre seconde étude où nous avons comparé la sollicitation d'une fonction cognitive avec la sollicitation de plusieurs fonctions cognitives au cours d'un même essai. La nature des tâches comparées étaient donc différentes. Nos résultats nous ont permis de conclure que l'étude de deux fonctions cognitives isolément ne nous guide pas totalement lorsque ces dernières sont combinées, comme c'est le cas dans la vie réelle. La charge mentale ressentie lors d'une tâche sollicitant deux fonctions cognitives ne correspond pas à une augmentation de la charge mentale pour chacune des tâches sollicitant une fonction cognitive seule, de même pour la motivation et le ressenti ludique. Pour ce dernier par exemple, il peut être élevé pour les deux tâches prises isolément et avoir un score plus faible lorsque ces deux tâches sont combinées. Ainsi, pour comprendre une activité réelle dans une étude expérimentale, il faut identifier clairement les différentes fonctions cognitives sollicitées, ainsi que leur imbrication. Afin, d'identifier, par la suite, de façon plus systémique les différents facteurs, ainsi que les médiateurs internes et externes de la charge mentale pour cette activité donnée.

Cependant, en dehors d'un cadre expérimental, les variables externes imprévues de situations réelles sont omniprésentes. Ainsi, nos conclusions telles qu'elles sont à nuancer, bien qu'utiles pour certains cas, et non de manière générale. Dans les situations réelles, les activités sollicitent plus ou moins fortement une fonction cognitive. Chercher une adresse sur un plan sollicite la mémoire visuo-spatiale par exemple. Nos conclusions sur le test de Corsi peuvent donc être intéressantes pour établir des hypothèses de recherche. À l'appui de nos résultats présentés dans ce manuscrit, nous pouvons poser l'hypothèse selon laquelle, plus le nombre de rues à retenir est élevé, plus la charge mentale visuelle et spatiale augmenteront ; de même, pour la charge mentale liée au ressenti des exigences mentales et temporelles, ou celle liée à l'émotion (avec une frustration plus grande). Si nous cherchons une adresse avec un plan au design ludique par rapport à un plan classique, nous pouvons supposer à l'appui de nos résultats que cela augmentera la charge mentale visuelle par exemple, mais pas l'effort mental.

Par conséquent, nous confirmons en partie la pensée de Miller (2001) sur l'impossibilité d'appliquer de but en blanc les conclusions d'étude sur la charge mentale lors de tâches cognitives à des situations réelles. Néanmoins, faire appel à des tâches cognitives permet des situations moins sujettes aux imprévus et une compréhension plus fine de l'évolution

de la charge mentale pour une fonction cognitive donnée. Enfin, les études sur la charge mentale lors de tâches cognitives peuvent être de précieux guides pour orienter les hypothèses de recherche de contextes opérationnels.

2. Discussions sur les modèles théoriques

Dans cette partie, au regard des résultats des études expérimentales présentées aux Chapitres 4 et 5, nous allons discuter des modèles théoriques de la charge mentale présentés au Chapitre 1. Par conséquent, dans un premier temps, nous allons discuter des modèles centrés sur les exigences de la tâche.

Modèles centrés sur les exigences de la tâche

Le modèle contraintes/astreintes (Karasek, 1979 ; Sperandio, 1988 ; Schlegel, 1993 ; Raufaste et al., 2004) ou le modèle temps/charge mentale (Hancock et Chignell, 1988 ; Hancock et Caird, 1993) accordent une importance capitale aux exigences de la tâche. Le modèle temps/charge mentale met en relation la charge mentale et les exigences temporelles. Par rapport à la notion temporelle, dans nos deux études expérimentales, nous demandions aux participants de réaliser les tâches cognitives le “*plus rapidement possible*” (Cf. consignes en Annexe 7). Cependant, au cours de la tâche, aucun indicateur de temps, tel qu’un minuteur, n’était présenté aux participants. Cette précision temporelle au niveau des consignes seulement, avait un impact sur les résultats de la dimension *Exigences temporelles* qui pour les deux études allaient de 20 (sauf pour le niveau 1 du test de Corsi) à 60 sur une échelle de 100. Ainsi, au cours de la tâche, les participants intériorisaient en partie cette exigence de la tâche, même si aucun rappel n’était donné au cours de l’activité. Néanmoins, ces modèles centrés sur les exigences de la tâche mesurent la charge mentale uniquement à travers les exigences de l’activité. Dans notre première étude, nous avons conclu, que les exigences de l’activité seules (correspondant dans notre cas à la complexité de la tâche) n’étaient pas suffisantes pour prédire la charge mentale subjective pour certaines tâches cognitives. Pour le WCST, par exemple, malgré l’augmentation des exigences de la tâche, la charge mentale subjective n’augmentait pas à chacun des niveaux de complexité. Une des pistes d’explication est que des médiateurs internes de la charge mentale, qui ne sont donc pas liés aux exigences de la tâche, ont modéré les effets des exigences sur la charge mentale

ressentie. Effectivement, lors du WCST, le ressenti ludique, un médiateur interne de la charge mentale, est élevé, ce qui peut expliquer que pour le niveau de complexité le plus haut, la charge mentale ressentie ne soit pas significativement plus élevée que le niveau de complexité intermédiaire. Néanmoins, les modèles de classification du test de Corsi et de la double tâche, se basaient principalement sur les exigences de la tâche pour déduire la charge mentale (résultat des *Permutation Feature Importances* - PFI), ce qui soutient les modèles théoriques se basant principalement sur les exigences de la tâche.

Approches centrées sur les performances

Dans ces approches, la charge mentale est directement liée à la performance. Par ailleurs, l'habituation à la tâche, l'entraînement ou encore la fatigue peuvent influencer sur la charge mentale en l'augmentant ou en la diminuant. Ainsi, la charge mentale fluctue au cours du temps. Par conséquent, la performance, indicateur de la charge mentale pour plusieurs modèles théoriques, était pertinente à prendre en compte dans nos études.

La relation entre la charge mentale et la performance peut être illustrée à travers une courbe en U inversé (Yerkes et Dodson, 1908 ; Boyer, 2016). Or, nos résultats expérimentaux montrent une relation non-linéaire et non-curvilinéaire entre la charge mentale et les performances, comme soutenu par plusieurs auteurs (Reid et Nygren, 1988 ; Raufaste et al., 2004 ; Cain, 2007 ; Radüntz, 2017). En effet, au cours d'une activité, un individu peut maintenir des performances élevées, mais ce maintien peut conduire également à une surcharge mentale. Par ailleurs, les individus peuvent varier l'effort mental et subir davantage de charge mentale pour maintenir un niveau de performance constant (Meijman et O'Hanlon, 1984). En accord avec ces modèles, nos résultats indiquent que la prise en compte des performances seules n'est pas suffisante pour déduire le niveau de charge mentale, malgré la multiplication des dimensions de performances qui augmente la précision des modèles de classification. Toujours dans les approches centrées sur les performances, Meister (1976) propose un modèle mettant en lien les performances et la charge mentale, mais en ajoutant la notion d'exigences de la tâche. Nos résultats ont montré que la prise en compte du niveau de complexité (exigences de la tâche) et des performances, permettaient d'augmenter le niveau de précision des modèles de classification, par rapport aux performances prises isolément.

Approches centrées sur les capacités limitées de l'individu

Ces approches considèrent la charge mentale comme le rapport entre les exigences de la tâche et la capacité disponible de traitement de l'information pour réaliser la tâche. Comme nous l'avons vu au Chapitre 1, les théories dites *des ressources* suggèrent qu'il existe un ou des réservoirs de ressources mentales dans lequel/lesquels il est possible de puiser pour répondre aux exigences de la tâche (Staal, 2004). Pour discuter de ce modèle théorique, nous nous sommes appuyés sur le questionnaire du WP qui se base sur le modèle à réservoirs multiples de Wickens (1984, 1987, 2002, 2008). Ce modèle défend l'idée qu'il existerait plusieurs réservoirs de ressources lors du traitement de l'information. La modélisation de ces différents réservoirs de ressources met donc en lumière l'existence de charges de natures différentes (Kostenko, 2017). En effet, par exemple, au niveau de la complexité la plus élevée du test de Go/No-Go de notre étude 1, les participants avaient une charge mentale visuelle élevée (supérieure à 60 sur une échelle sur 100) et une charge mentale verbale faible (inférieure à 25 sur une échelle sur 100). Ainsi, un individu peut bel et bien être en surcharge mentale sur certaines ressources, et être en sous-charge mentale sur d'autres (Kostenko, 2017). Néanmoins, l'idée selon laquelle deux tâches complexes peuvent être réalisées parallèlement sans interférence (sans augmentation significative de la charge mentale), si elles n'utilisent pas les mêmes réservoirs de ressources (Le Gonidec, 2022) est à nuancer. Effectivement, en comparant le test de Corsi et la double tâche (cette dernière correspondant à un test de Corsi avec un ajout de calcul mental), nous avons observé que pour les niveaux 1 (ayant 3 cubes pour les deux tâches) des traitements attentionnels liés au traitement spatial et visuel de la double tâche sont significativement plus élevés que ceux du test de Corsi (Partie E, Table 4, Annexe 13 ; pp. 276 et 278). Or, le calcul mental sollicitant la Mémoire de Travail verbale n'utilise pas les mêmes réservoirs de ressources que le test de Corsi sollicitant la mémoire visuo-spatiale. Ces résultats infirment le postulat des modèles à réservoirs multiples, mais confirment celui des modèles à réservoirs uniques comme ceux de Moray (1967), Kahneman (1973), Young et Stanton (2002), Wickens et al. (2015).

Approches centrées sur l'effort mental fourni par l'individu

Comme nous l'avons vu au Chapitre 1, l'opérateur a une marge de manœuvre pour réguler sa charge mentale. Les approches centrées sur l'effort mental fourni par l'individu considèrent la charge mentale comme la quantité de ressources mentales mobilisées volontairement pour répondre aux exigences d'une tâche (Leplat, 1977 ; Le Gonidec, 2022). Cette modulation volontaire de la quantité de ressources mentales investies lors de la réalisation d'une tâche correspond à la notion d'effort mental (Kahneman, 1973 ; Delignières et Deschamps, 2000). Pour nos deux études, la dimension liée à l'individu *Effort mental* est une des plus élevées avec la dimension liée aux exigences de l'activité, les *Exigences mentales* (mesurées via le questionnaire du NASA-TLX). Le modèle de De Waard (1996) centré sur l'effort mental fourni par l'individu, est donc pertinent lorsqu'il précise que l'augmentation des exigences de la tâche, à partir de l'état optimal de l'opérateur, induit une augmentation de l'effort mental et plus globalement de la charge mentale et une baisse des performances.

Dans le modèle de De Waard (1996), il est intéressant de noter la temporalité selon laquelle la mobilisation volontaire des ressources mentales se réalise pour réguler la charge mentale. En effet, l'effort mental se fait uniquement lorsque les exigences de la tâche sont moyennes (ni faibles, ni élevées). De plus, De Waard (1996) distingue l'effort mental lié à la tâche (nécessitant un traitement contrôlé de l'information) pour des exigences moyennement élevées, et l'effort mental lié à l'état de l'individu (dans le cas où il doit modifier l'état actuel vers un état requis) pour des exigences moyennement faibles. Dans nos études expérimentales, les tâches cognitives avaient des niveaux de complexité induisant des niveaux de charge mentale dans la zone basse et intermédiaire de la charge mentale. Ainsi, la dimension *Effort mental* lors de la réalisation des niveaux de complexité 1 et 2 (les moins exigeants) toutes tâches confondues, correspond à l'effort lié à l'état actuel de l'individu. La dimension *Effort mental* lors de la réalisation des niveaux de complexité 3 correspond quant à elle à l'effort mental lié à la tâche. Ces deux types d'effort mental, qui selon De Waard (1996), permettent d'éviter le basculement au-delà des lignes rouges (sous-charge ou surcharge mentale).

Approches incluant les stratégies de l'individu

La dernière catégorie de modèles théoriques que nous avons vu au Chapitre 1 correspond aux approches où l'individu est de nouveau acteur de sa charge mentale, mais cette fois-ci, par la mobilisation de stratégies. Nos résultats ne permettent pas de discuter les modèles théoriques incluant les stratégies de l'opérateur, car nous n'avons pas eu d'objectifs et donc pas d'hypothèses sur ce sujet précis, parmi la multitude de notions abordées. Ainsi, nous n'avons pas pu le tester. Cependant, nous discuterons des perspectives de la prise en compte des stratégies de l'opérateur à la section 6. *Perspectives de la Discussion générale.*

Discussion du modèle et de la définition proposés

À la suite des résultats expérimentaux, notre modèle théorique présenté à la section 1.2.3 a évolué. Nous le représentons donc, de nouveau, à la **Figure 2** avec des ajouts (par rapport au modèle présenté à la section théorique 1.2.3) correspondant aux nouvelles flèches jaunes et vertes. En effet, comme nous l'avons vu à la partie expérimentale, lorsque la variable extrinsèque aux exigences de la tâche, correspondant au design ludique, est manipulée, la charge mentale liée à la mobilisation des ressources attentionnelles augmente. Ainsi, l'individu a dû engager des ressources pour faire face à la situation. C'est donc la *conduite de l'opérateur* qui est davantage impactée dans cette situation est qui est symbolisée par le cheminement des **flèches orangées** sur la **Figure 2**. Par ailleurs, lorsque la variable intrinsèque aux exigences de la tâche correspondant à la multiplicité des fonctions cognitives, est manipulée, la charge mentale liée aux caractéristiques de la tâche augmente. Le cheminement est représenté par les **flèches vertes** au niveau de la **Figure 2**. La perception qu'à l'individu sur les caractéristiques intrinsèques à la tâche est modulée par des médiateurs internes à l'individu comme le ressenti ludique. À la suite de nos résultats expérimentaux, nous pensons que le ressenti ludique (potentiel médiateur de la charge mentale), pourrait impacter en retour, la perception qu'à l'individu sur les caractéristiques de la tâche (d'où les flèches vertes en pointillées). Une perception qui ensuite impactera la conduite de l'opérateur comme l'effort mental.

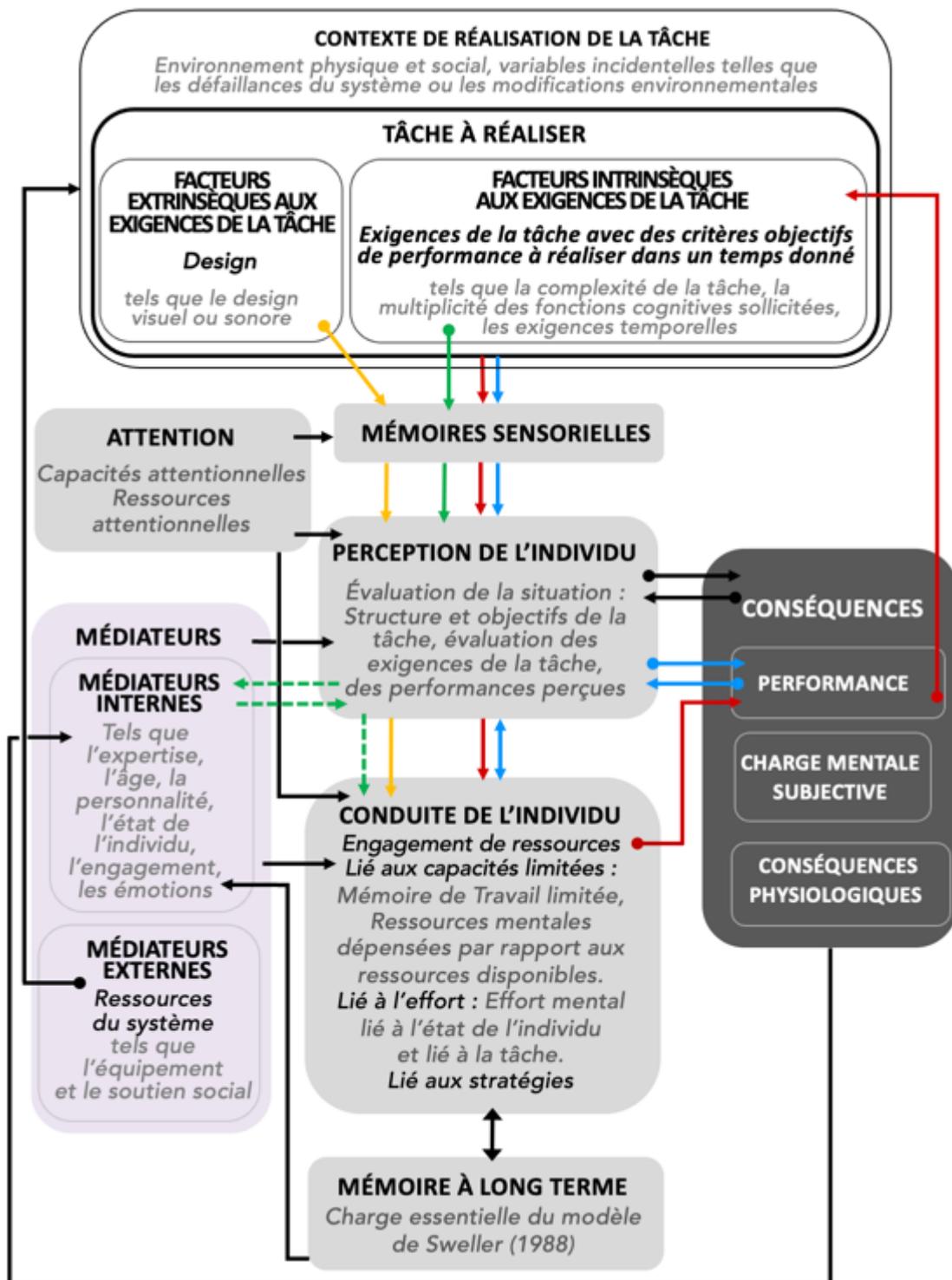


Figure 2 : Notre modèle conceptuel de la charge mentale, combinant différents modèles théoriques, présenté à la section 1.2.3, avec l'ajout des flèches jaunes et vertes à la suite de nos résultats expérimentaux.

En outre, la notion des stratégies fait partie intégrante de la définition de la charge mentale que nous avons proposé à la suite de la revue de la littérature. Ci-après, nous

proposons un rappel de la définition avec un code couleur pour faciliter la discussion qui suivra. Nous proposons que la charge mentale soit *“le niveau d'activation des ressources mentales limitées pour répondre aux critères de performance (objectifs et/ou subjectifs) d'une activité accomplie en un temps donné par un individu qui y consacre de l'attention, de l'effort mental et déploie des stratégies de régulation pour maintenir une charge mentale dans des limites acceptables. La mobilisation délibérée de ressources mentales (effort mental) et des stratégies résulte de la mise en relation d'informations pertinentes issues de la mémoire à long terme ou du contexte avec les nouveaux éléments d'information issus de l'activité à réaliser. Par ailleurs, cet individu est influencé par des variables internes (expertise, âge, engagement ...) et externes (soutien social, équipement technologique ...) qui peuvent être des médiateurs de la perception et de la conduite de l'opérateur. En conséquence, l'individu a une expérience subjective de la charge mentale, ainsi que des répercussions comportementales et physiologiques qui, à leur tour, ont un impact sur la perception et les médiateurs internes de l'individu”*.

La partie violette représente le niveau d'activation des ressources mentales limitées qui fait référence aux approches centrées sur les capacités limitées de l'individu. Les capacités limitées peuvent s'appréhender à travers les tâches d'empan mnésiques, telles que le test de Corsi. Dans notre première étude, plus l'empan mnésique était élevé, plus les performances diminuaient, car la Mémoire de Travail a des capacités de traitement de l'information limitées. **La partie rouge** représente les exigences de la tâche. Comme nous l'avons à la section théorique 1.2.3, il existe des variables intrinsèques et extrinsèques aux exigences de la tâche. Pour nos deux études, nous nous sommes focalisés sur les variables extrinsèques aux exigences de la tâche, tels que le design ludique et les variables intrinsèques aux exigences de la tâche (complexité de la tâche pour une fonction cognitive donnée et multiplicité de fonctions cognitives). Ainsi, à l'appui de nos résultats, nous savons qu'un design ludique impacte uniquement la mobilisation des ressources attentionnelles liées au traitement visuel et spatial, alors que l'ajout de fonctions cognitives impacte uniquement la charge mentale liée aux exigences de la tâche et liée au ressenti sur ses propres performances (de manière proportionnelle et non cumulative). Par ailleurs, à la suite de nos expérimentations, il nous semble d'autant plus important de distinguer la notion de *variables* et de *facteurs extrinsèques*

aux exigences de la tâche en fonction de leur influence sur la charge mentale (distinction initiée lors de la partie théorique ; Cf. section 1.2.3). Toutes les variables extrinsèques aux exigences de la tâche ne sont pas forcément des facteurs qui impactent la charge mentale. De plus, les variables internes à l'individu, pouvant être des médiateurs de la charge mentale, se comportent différemment en fonction de la manipulation de variables intrinsèques ou extrinsèques aux exigences de la tâche. Le ressenti ludique (variable interne à l'individu) n'est pas impacté par les changements de design ludique (facteur extrinsèque aux exigences de la tâche influençant la charge mentale), mais par la multiplication de fonctions cognitives (facteur intrinsèque aux exigences de la tâche influençant également la charge mentale). **La partie noire** correspond aux critères de performance subjectifs, que nous pouvions appréhender à travers la dimension *Exigences temporelles*. Comme nous l'avons vu, toutes les consignes des tâches cognitives demandaient d'aller le *plus rapidement possible*, mais au cours de la tâche, il n'y avait aucun indicateur de temps. Si un participant souhaitait aller à son rythme, il l'aurait pu, rien ne l'en aurait empêché. Surtout pour les tâches qui ne présentaient pas un défilement continu d'items, comme la double tâche. Or, la dimension *Exigences temporelles* n'était pas significativement différente entre le WCST et la double tâche par exemple. À travers ce résultat, la notion de critères de performance subjectifs prend tout son sens, car ce ne sont pas seulement les exigences de l'activité qui guident les critères de performance d'un individu. Mais, également ses propres critères de performances qui sont donc liés notamment à la compréhension de la situation et à sa motivation. **La partie verte** conçoit l'opérateur comme un acteur de sa charge mentale, dans la sélection des processus attentionnels et la mobilisation des ressources mentales (l'effort mental). Les différents processus attentionnels qui sont au cœur du questionnaire du WP et la dimension *Effort mental* qui est au cœur du questionnaire du NASA-TLX. Ainsi, l'opérateur est bel et bien acteur de sa charge mentale. **La partie orange** fait référence aux différents niveaux de la charge mentale, car malgré la terminologie de ce concept, la charge mentale est délétère uniquement aux niveaux extrêmes, de sous-charge ou surcharge mentale. L'enjeu de l'expérience utilisateur est donc de maintenir la charge mentale dans les limites acceptables appelées *zone de confort*. Vu cet enjeu, lors de nos études expérimentales, nous nous sommes focalisés sur ses limites acceptables. **La partie rose** précise les variables internes et externes à l'individu. Selon le modèle théorique à la section 1.2.3, si ces variables influencent la

charge mentale, ces variables peuvent être des médiateurs internes ou externes à la charge mentale. Étant donné que nous n'avons aucun objectif et hypothèse sur l'impact sur la charge mentale d'une variable externe à l'individu, nous ne pouvons avancer des conclusions sur ce point précis. Concernant les variables internes à l'individu, comme le ressenti ludique, nous supposons un lien entre le ressenti ludique et la charge mentale subjective. Effectivement, nous avons observé un effet plafond de la charge mentale du WCST de l'étude 1 qui était une tâche jugée comme ludique (par rapport aux autres tâches cognitives). Par ailleurs, nous avons observé une plus grande charge mentale et un ressenti ludique plus faible pour les tâches combinées, par rapport aux tâches simples. Par conséquent, nous supposons un lien entre le ressenti ludique et la charge mentale subjective, mais pour le tester, il faudrait distinguer deux groupes de participants dans notre seconde étude. Un groupe avec un niveau élevé de ressenti ludique (par exemple supérieur à 70 sur une échelle de 100) et un autre groupe avec un niveau faible de ressenti ludique (inférieur à 40) et écarter le reste des participants (se situant entre 40 et 70). Cependant, en reprenant les résultats du ressenti ludique (Cf. section 5.3.2), nous observons qu'il varie entre 55 et moins de 80 sur une échelle de 100, pour l'ensemble des participants. Ainsi, concernant nos résultats, il serait difficile d'établir deux groupes de participants en se basant sur l'intensité de leur ressenti ludique. Une prochaine étude, avec un plus grand nombre de participants pourrait permettre d'atteindre cet objectif. Par ailleurs, à la section 6 de la Discussion générale, nous discuterons de perspectives différentes permettant d'étudier si des variables internes à l'individu influencent la charge mentale, et donc sont des médiateurs internes à la charge mentale. **La partie bleue** représente, quant à elle, les conséquences de la charge mentale se traduisant au niveau du ressenti de la charge mentale, des performances et des aspects physiologiques que nous détaillerons plus en détail dans la section suivante de la Discussion générale.

3. Discussions sur les mesures de la charge mentale

Que ce soit à la suite de la revue de la littérature ou à la suite des études expérimentales, nos conclusions pour une mesure juste de la charge mentale étaient les mêmes. La mesure de la charge mentale issue des questionnaires et des performances, bien que pertinente, n'est pas suffisante pour qu'elle soit juste. La triangulation des trois

catégories de mesures (subjectives, performances et physiologiques) pourrait dépasser les limites de chaque mesure. Ci-après, la discussion des trois catégories de mesures.

Mesures subjectives

Nos deux études confirment la littérature sur la complémentarité des questionnaires du WP et du NASA-TLX (Rubio et al., 2004 ; Paxion, 2014). L'étude 1 a confirmé que le NASA-TLX permettait la mesure de la charge mentale en relation avec les exigences de l'activité (exigences mentales, temporelles et physiques), les exigences de l'opérateur (performance, effort mental) et l'état émotionnel de l'individu (frustration). Alors que le WP permet la mesure de la charge mentale liée aux ressources attentionnelles mobilisées dans chaque réservoir de ressources du modèle de Wickens (1984, 1987, 2002, 2008). L'étude 2, quant à elle, nous a indiqué que le NASA-TLX et le WP pouvaient discriminer différemment la charge mentale pour deux facteurs extrinsèques. Le design de l'interface impactait uniquement la saturation des ressources attentionnelles liées au traitement visuel et spatial, alors que la multiplication de fonctions cognitives impactait uniquement la charge mentale liée aux exigences de la tâche et aux performances ressenties. Le WP permet donc de fournir des informations sur les variables extrinsèques des exigences de la tâche (l'impact du design ludique sur la charge mentale par exemple). Chose que ne fait pas le NASA-TLX. Mais, le NASA-TLX est plus étroitement lié à la façon dont une tâche est mise en œuvre dans un contexte donné, via les trois dimensions associées à l'activité (les exigences mentales, les exigences physiques, la pression temporelle), contrairement au WP.

Par conséquent, le type de mesure doit être choisi avec soin en fonction des caractéristiques de la tâche et de l'impact sur la charge mentale que l'on souhaite évaluer. Il serait donc judicieux de s'interroger au préalable sur les sources de charge mentale pour une tâche donnée, au risque d'établir des conclusions hâtives, voire faussées par la sélection d'un questionnaire inadapté.

Performances

Les deux études expérimentales ont conclu sur la nécessité de prendre en compte une diversité de dimensions de performances. En effet, il se peut qu'une dimension soit sensible à une variable externe et d'autres nettement moins, voire pas du tout (résultats des *Permutation Feature Importance* - PFI - de l'étude 2) ou que des modèles de

classifications se basent sur certaines dimensions et pas d'autres pour déduire la charge mentale (résultats de l'étude 1). Néanmoins, Hancock et al. (2021) expriment une limite dans la prise en compte des performances dans les conditions expérimentales en laboratoire. En effet, dans ces situations, la majorité du temps, nous demandons aux participants de répondre le plus rapidement et le plus précisément possible. Cependant, les situations de la vie réelle sont plus diversifiées. Nous pouvons être confrontés à des exigences élevées, mais avoir le contrôle sur notre emploi du temps. De plus, Hancock et al. (2021) se réfèrent aux modèles motivationnels de Simon (1996) pour appuyer leur thèse sur la limite de la mesure des performances dans un contexte expérimental. En effet, dans la vie réelle, la satisfaction (faire suffisamment bien pour accomplir la tâche) est recherchée, alors que dans un contexte expérimental, l'optimisation de la réponse, la recherche du *meilleur absolu* sont attendues. Selon les auteurs, une grande partie de nos connaissances expérimentales découlent d'instructions visant à optimiser la réponse, bon nombre des mêmes résultats sont ensuite utilisés pour s'appliquer à des environnements opérationnels où la satisfaction prédomine. Il y a donc une déconnexion entre la performance des tâches en tant que mesure du rendement (optimisation de la réponse, *meilleur absolu*) et la performance des tâches en tant que reflet de la charge mentale. Une conception qui encourage à réfléchir à l'avenir à d'autres critères de performances se rapprochant de situations écologiques où la satisfaction prend le dessus sur l'optimisation de la réponse.

Physiologiques

Parmi toutes les mesures physiologiques, l'EEG est une des plus efficaces pour mesurer la charge mentale (Zhang et al., 2018), avec des perspectives encourageantes si l'on l'associe à la SPIRf (Aghajani et al., 2017). Cependant, la mesure de la charge mentale à travers l'EEG a plusieurs contraintes (comme le temps de mise place et de traitement des données, l'expertise, l'intrusivité du dispositif, ou le rapport signal/bruit élevé pouvant être dû aux signaux musculaires). En effet, au cours de la thèse, nous nous sommes rendus compte des contraintes liées à l'expertise et au temps pour traiter les données EEG récoltées lors de l'étude 2. Ainsi, nous rejoignons Miller (2001) sur l'idée qu'une mesure qui nécessite un équipement coûteux ou des experts pour superviser une partie du test n'est pas complètement rentable.

Or, comme nous l'avons vu au niveau des résultats des deux études expérimentales, quelle que soit la tâche cognitive, la dimension du *Traitement visuel* était une des plus élevée par rapport aux autres dimensions de ce questionnaire. Ce qui était cohérent avec la logique des tâches qui étaient réalisées sur écran avec un traitement visuel de l'information. Ainsi, comme précisé au niveau du Chapitre 2 - *Mesures de la charge mentale*, si la tâche comporte principalement des stimuli visuels, les mesures oculaires sont conseillées, car la tâche examinée sollicite d'une manière accrue l'attention visuelle (Veltman et Gaillard, 1996 ; Miller, 2001). Ce qui est notre cas avec des tâches cognitives numérisées. De plus, ces dernières années, les mesures oculaires ont vu leur utilisation augmenter grâce à l'accessibilité des appareils d'oculométrie (couramment nommés par l'anglicisme *eye-tracking*) selon Charles et Nixon (2019). En outre, l'oculométrie nécessite un coût moins élevé que celui des mesures centrales comme l'EEG, que ce soit le coût financier, le coût temporel dans la passation expérimentale, le coût de l'expertise (comme le traitement des données) et l'intrusivité pour le participant, car un oculomètre peut être uniquement intégré à l'ordinateur, sans dispositif physique pour l'individu. Par ailleurs, l'intégration croissante de la technologie de suivi oculaire dans les nouvelles technologies, telles que la Réalité Virtuelle ou Augmentée, permet une mesure de haute qualité dans une variété d'applications à l'avenir (Appel et al., 2021).

Ainsi, pour de futures recherches sur la mesure de la charge mentale lors de tâches cognitives numérisées, nous préconisons une triangulation entre des données subjectives issues des questionnaires du NASA-TLX et du WP, des données de performances avec une diversité au niveau des dimensions choisies et des données physiologiques oculaires (fixations, saccades, dilatation de la pupille et clignements) qu'on pourrait comparer avec des données cérébrales issues de l'EEG, qui sont pour l'heure, toujours considérées comme les mesures physiologiques de la charge mentale les plus efficaces.

4. Principale limite de nos études expérimentales

Bien qu'apportant de nombreuses réponses à nos questionnements de départ, nos deux études expérimentales ont une limite commune, celle du nombre de données. Dès le départ, dans notre problématique, nous posions des hypothèses opérationnelles en lien

avec des méthodes de classifications (non) supervisées des données. Or, l'utilisation de ces techniques dites de *Machine Learning* nécessitent un corpus de données important de l'ordre de 600 données pour permettre entre autres aux phases de validation croisée d'avoir assez de données pour s'entraîner. Dans notre cas, avec 159 données pour l'étude 1, nous étions bien en-deçà des préconisations de l'utilisation de ce type de méthode. Nous étions pourtant pour la première expérimentation sur une étude en ligne qui a l'avantage de pouvoir atteindre un large corpus de participants. Or, contraint par le temps, notre équipe avait décidé de s'arrêter à 53 participants, l'idéal étant d'en avoir 200 (pour atteindre 600 données avec trois classes de charge mentale). Concernant l'étude 2, nous étions contraints à un nombre de participants limité, étant donné le dispositif de mesure physiologique de l'EEG qui ralentit la procédure expérimentale (expérimentation en présentiel avec un expert BCI, temps d'installation et de désinstallation du matériel).

Par ailleurs, en plus du nombre faible d'essais, il est important de noter des biais potentiels dus à la séparation en *training* et en *testing* après les pré-traitements. Ces biais potentiels provenant du fait que l'ensemble des données aient été utilisées lors de l'étape de normalisation. Au préalable, nous avons étudié trois possibilités :

- Normalisation de toutes les données avant de scinder en deux populations (entraînement et test). C'est une logique d'analyse globale, même si elle peut contenir un léger biais.
- Normalisation de chaque sous-ensemble séparément. C'est une logique sous-optimale et peut engendrer une forte dispersion des résultats.
- Normalisation de l'ensemble de test pour l'amener aux mêmes statistiques que l'ensemble d'entraînement. C'est une logique plus proche de la réalité, mais plus complexe à mettre en place.

Nous avons retenu la première option, car c'était la plus simple à mettre en place et la réduction, voire la suppression des biais liés à la séparation *en training* et en *testing*, n'était pas un aspect au cœur du sujet et de l'expertise de cette thèse.

5. Applications envisagées

À l'appui de nos résultats, seul le Corsi pourrait être une tâche cognitive pertinente pour réaliser les applications envisagées au *Chapitre 3*, comme le maintien d'une classe de

charge mentale donnée durant une tâche pour mesurer son impact sur différentes variables. Effectivement, en nous basant uniquement sur le niveau de complexité (à trois niveaux) et les performances, nous étions capables de déduire la classe de charge mentale subjective avec une précision de plus de 80 % pour le test de Corsi. Une triangulation de la mesure, avec l'ajout de données oculométriques ou EEG, pourrait augmenter la précision du modèle de classification pour déduire la charge mentale subjective à plus de 90 % pour cette même tâche cognitive.

De plus, pour le choix du design du test de Corsi numérisé, à la suite de nos résultats, nous préconisons une version classique (cubes gris) et non une version avec un design ludique (bouées unicolores ou multicolores). En effet, cette dernière augmentait significativement la charge mentale basée sur les traitements attentionnels de stimuli visuels et spatiaux. Une charge mentale liée aux traitements visuels, par exemple, qui pour le Corsi classique était égale à 50 en moyenne (sur une échelle de 100 ; pouvant être considérée dans la *zone de confort* de la charge mentale) pour les deux études. Par ailleurs, le design ludique n'impacte pas significativement la motivation ou le ressenti ludique entre une version classique et ludique du Corsi. Ainsi, l'ajout du design semble avoir une balance *bénéfice/risque* délétère du fait de l'augmentation de la charge mentale ressentie sans avoir une augmentation significative de la motivation. Cependant, avant de mettre de côté cet apport du design, il semble intéressant d'interroger l'impact sur l'engagement des participants à travers des échelles dédiées, telles que la *User Engagement Scale (UES) short form* (O'Brien et al., 2018). Une prise en compte plus approfondie de l'engagement qui peut être vu comme une des perspectives envisagées à la suite de ce travail de recherche.

6. Perspectives

Prise en compte des données cérébrales

À la suite de ce travail de thèse, la première perspective est l'analyse des résultats de la charge mentale issue des données EEG que nous avons prise en compte au cours de l'étude 2. Le test de Corsi à trois niveaux de complexité serait la tâche de référence, car il permet d'avoir trois niveaux de charge mentale subjective distincts, de même pour les conséquences comportementales. Pour le futur algorithme de classification de la charge

mentale mesurée à travers l'EEG (en vue en autres d'adapter la rééducation à la charge mentale du patient) qui sera en partie alimenté par les observations issues des études 1 et 2. En effet, nous savons que plus le niveau de complexité augmente, plus la charge mentale subjective en lien avec les exigences de la tâche augmente et plus les performances diminuent. Ainsi, pour l'étiquetage EEG, nous partirons des données de subjectives et de performances pour pouvoir étiqueter les données EEG liées à la charge mentale telles que les ondes Thêta en zone frontale ou les ondes Alpha en zone pariétale. L'analyse des données EEG est en cours par notre équipe de recherche. Mais cette étape nécessitant des expertises humaines diverses et étant chronophage, nous n'avons pu présenter les résultats au cours de ce manuscrit. Dans la continuité de notre argumentaire, il serait pertinent dans une prochaine étude, de comparer la mesure de la charge mentale issue de l'EEG avec celle issue de données oculaires lors de la réalisation du test de Corsi à plusieurs niveaux de complexité. Étude, qui a notre connaissance (à la suite d'une revue de la littérature), n'a pas encore été faite.

Perspectives du Machine Learning

De plus, il y a un défi lié aux modèles de prédictions de la charge mentale, celui d'avoir des modèles plus précis, plus fiables, puis plus génériques et donc des modèles inter-tâches et inter-contextes. Cependant, la transférabilité reste pour l'instant au stade d'une perspective envisageable. Il faut encore améliorer la fiabilité des modèles de classification inter-tâches et inter-contextes, avant d'envisager des modèles génériques. Par ailleurs, pour améliorer le résultat de précision de la prédiction de la charge mentale subjective, nous pourrions réaliser une adaptation de domaine, afin de faire correspondre dans un même domaine d'analyse, les données d'entraînement et de tests issus de distributions différentes, comme ce fut le cas pour plusieurs études (Magliacane et al., 2018 ; Zhou et al., 2022). Étant donné que la thèse n'était pas centrée sur le développement algorithmique, nous n'avons pas lancé une étude sur ce point précis, nécessitant une expertise supplémentaire. Mais, il nous semble important de souligner que l'adaptation de domaine est une perspective intéressante qui vaut aussi bien pour la transférabilité inter-tâches (étude 1), que la transférabilité inter-contextes (étude 2).

Autre perspective concernant les perspectives sur le *Machine Learning*, l'étude comparative de plusieurs algorithmes de classification de données. Dans le cadre de la

thèse, nous avons utilisé la LDA, qui par la suite, pourrait être comparée au Support Vector Machine (SVM), dont l'hyperplan utilisé pour séparer les différentes classes est différent de la LDA. Pour ce dernier, l'hyperplan de séparation permet d'identifier les caractéristiques qui maximisent les variances entre les classes, tout en minimisant la variance à l'intérieur de la classe. Concernant l'hyperplan du SVM, il est positionné de façon à maximiser la distance avec les données les plus proches (Mohanavelu et al., 2022). La comparaison algorithmique n'était pas au cœur du sujet de thèse, mais l'on peut l'envisager à la suite de ce travail de recherche. Une perspective est donc de comparer ces deux méthodes de classification, que ce soit dans la prédiction de la classe de charge mentale subjective, de la transférabilité d'un modèle de classification inter-tâches ou inter-contextes.

Influence des médiateurs internes

Comme nous l'avons vu au Chapitre 1, les variables internes et externes à l'individu peuvent être des médiateurs de la perception et de la conduite de l'opérateur lors de la gestion de sa charge mentale. Or, dans notre recherche, nous avons manipulé uniquement des variables intrinsèques aux exigences de la tâche (niveaux de complexité pour une fonction cognitive donnée et complexité liée à la multiplicité de fonctions cognitives) ou des variables extrinsèques aux exigences de la tâche (designs ludiques de l'interface). Mais qu'en est-il de l'influence des variables internes à l'individu comme l'influence de son expertise, de sa personnalité sur la gestion de sa charge mentale ? Ainsi, en plus de la perspective concernant l'étude de l'impact de l'engagement sur la charge mentale (détaillé à la section 2 de la Discussion générale), des études sur l'impact sur la charge mentale de variables internes (telles que l'expertise ou la personnalité) lors de tâches cognitives peut être une perspective envisagée.

De plus, comme nous l'avons vu, la charge mentale n'est pas un concept que l'on subit. Il serait donc intéressant d'étudier les différents types de stratégies de régulation de la charge mentale. Pour ce faire, des tâches cognitives simples ne sont pas les plus adéquates. En effet, selon Loft et al. (2007), la prise en compte des stratégies est d'autant plus importante lors de l'évaluation de la charge mentale associée à l'exploitation de systèmes complexes. Ainsi, l'idéal est d'à minima étudier les différents types de stratégies lors de tâches sollicitant plusieurs fonctions cognitives au cours d'un même essai, comme ce que nous avons proposé lors de l'étude 2. Pour étudier les stratégies de

régulation de la charge mentale, les questionnaires du NASA-TLX et du WP ne sont pas suffisants, car aucune dimension ne se rapporte à cette notion. La dimension *Effort mental* peut s'y rapprocher, mais elle ne peut distinguer les différents types d'efforts (lié à l'état de l'individu, à la tâche ou l'effort mental lié aux stratégies déployées ; Cf. Mulder, 1986 à la section 1.2.1). Ainsi, il est nécessaire d'ajouter des dimensions ou des questionnaires complémentaires, comme ceux issus de la théorie de la charge cognitive (*Cognitive Load Theory* - CLT) qui centre son modèle sur les stratégies permettant l'apprentissage. Dans ce cas, les modèles de la charge mentale et de la CLT sont complémentaires et s'apportent mutuellement. Concernant un questionnaire de la CLT qui serait pertinent d'ajouter, nous pouvons mentionner celui de Klepsch et al. (2017), qui permet de mesurer entre autres la charge essentielle, qui selon la CLT est imposée par l'utilisation délibérée de stratégies cognitives par l'individu.

Lien charge mentale et charge émotionnelle

Une autre perspective est issue des propos de Van Acker (2018) précisant que les ressources émotionnelles et cognitives puisent dans le même bassin de ressources énergétiques physiologiques. Ainsi, si la charge émotionnelle prend le dessus, elle peut empêcher les ressources cognitives d'exploiter pleinement leur potentiel. Dans notre étude, la dimension subjective *Frustration* n'est pas prégnante (inférieure à 50 sur une échelle de 100 pour l'ensemble des niveaux de complexité, toutes tâches confondues). Pour tester cette hypothèse d'un réservoir commun entre les ressources cognitives et la charge émotionnelle, une perspective possible est d'utiliser une tâche génératrice de frustration, telle que la tâche de *Frustration Go/No-Go* utilisée par Seymour et al. (2020) ou encore la *Mirror Tracing Frustration Task* utilisée par Meindl et al. (2019).

Niveaux de complexité plus nombreux

Enfin, une prochaine étude peut être envisagée avec des niveaux de complexité plus élevés. Un des objectifs serait de vérifier le lien entre la motivation, l'effort mental et la difficulté perçue. Effectivement, dans la partie théorique, nous avons exposé la théorie de l'intensité de la motivation proposée par Brehm et Self (1989 ; Cf. section 1.2.1 *Approches centrées sur l'effort mental fourni par l'individu*). Selon cette théorie, si la motivation potentielle est élevée, alors l'individu mobilise de l'effort mental selon la difficulté perçue. Enfin, lorsque l'opérateur perçoit la difficulté de la tâche comme

impossible à surmonter, il se désengage de l'activité et ne mobilise plus d'effort mental. Contrairement à l'étude 1, la seconde étude prend en compte la motivation perçue (à travers cette question : « *Quel était votre niveau de motivation pour réaliser la tâche ?* »). De plus, la notion d'effort mental est la quatrième dimension du NASA-TLX. Afin de définir si nos données pourraient correspondre au modèle, nous pouvons considérer le Corsi classique, car cette tâche a trois niveaux de complexité générant trois niveaux distincts d'exigences mentales et temporelles (résultats issus de l'étude 1 ; Cf. section 4.3.1 *Correspondance entre les niveaux de charge mentale subjective et les niveaux de complexité*), qu'on peut définir comme de la difficulté perçue. Lors de la réalisation du test de Corsi, les participants ont une motivation élevée (supérieure à 68 sur une échelle de 100) quel que soit le niveau de complexité. De plus, l'effort mental augmente avec la difficulté perçue (exigences mentales et temporelles ; Cf. Figure 15 du Chapitre 4). Nous obtenons les mêmes tendances pour le test de Corsi de l'étude 2. Ce résultat confirme la théorie de Brehm et Self (1989). Mais nous n'atteignons pas le *point de rupture* où l'individu perçoit la difficulté comme insurmontable et se désengage de l'activité en ne mobilisant plus d'effort mental. Pour cela, il faut davantage de niveaux de complexité. Par ailleurs, l'avantage d'avoir des niveaux de complexité plus nombreux est qu'on peut déduire les *lignes rouges*, les basculements dans la zone de la sous-charge ou la surcharge mentale. Des basculements au-delà des *lignes rouges* que nous n'avons pas pu observer dans nos études qui avaient pour objectifs d'avoir une charge mentale dans les limites dites *acceptables*, d'un niveau intermédiaire.

7. Conclusion générale

Au cours de cette recherche, nous avons pu proposer une nouvelle définition de la charge mentale basée sur un cadre conceptuel issu d'une revue de la littérature articulée en fonction de la passivité ou non de l'opérateur vis-à-vis de sa charge mentale. Puis, nous avons pu étudier la charge mentale à travers plusieurs tâches cognitives. Pour ce faire, nous nous sommes basés sur le modèle théorique que nous avons proposé au cours du Chapitre 1. Ainsi, nous avons étudié l'impact sur la charge mentale de deux variables intrinsèques aux exigences de l'activité, le niveau de complexité pour une fonction cognitive donnée et la complexité liée à la multiplicité de fonctions cognitives. Par ailleurs, nous avons étudié une variable extrinsèque aux exigences de la tâche, le design

ludique, pouvant exercer une influence sur la charge mentale en fonction du type de design ludique. Enfin, en se basant sur les défis des modèles de classification de la charge mentale, d'être plus fiables et plus génériques, nous avons proposé des modèles de classification, pour déduire les classes de charge mentale subjective, lors de la réalisation de tâches cognitives et nous avons étudié la transférabilité de modèles de classification inter-tâches et inter-contextes.

Les principales conclusions de cette thèse sont que les exigences de la tâche ne suffisent pas à elles seules pour déduire la charge mentale subjective, même lorsqu'elles sont prises en compte avec un indicateur disponible en continu, la performance. Cependant, plus les dimensions de performances sont diverses et variées, plus la précision de la prédiction augmente. Par ailleurs, les variables intrinsèques et extrinsèques aux exigences de la tâche sont importantes à prendre en compte, car elles peuvent réguler le ressenti de la charge mentale, donc permettre qu'elle reste dans les limites acceptables. Enfin, nous avons identifié une nouvelle tâche cognitive, le test de Corsi, pour laquelle nous avons pu établir un modèle prédictif de la charge mentale. Un modèle dont nous avons testé la validité inter-tâches (pour un même contexte de passation) et la validité inter-contextes (pour une même tâche donnée). Ce modèle prédictif pourrait, par la suite, être amélioré en comparant différents types d'algorithmes de classification de données, qui est une des perspectives à la suite de ce travail de recherche. Autre perspective, tester d'autres niveaux de complexité du test de Corsi, par exemple, qui soient davantage distincts. Afin, d'avoir une prédiction des classes de charge mentale plus élevées, car elles seront davantage distinctes avec moins de zones de chevauchements.

Cette thèse a donc proposé un regard enrichi et novateur sur la charge mentale, alliant travail théorique et expérimental.

Contributions

Article dans des revues internationales

Louis, L-E., Moussaoui, S., Van Langenhove, A., Ravoux, S., Le Jan, T., Roualdes, V., Milleville-Pennel, I. (2023). Cognitive task and combined statistical methods to evaluate, model and predict mental workload. *Frontiers in Psychology-Cognitive Science (In press)*.

Communications dans des conférences internationales

Louis, L-E., Moussaoui, S., Van Langenhove, A., Ravoux, S., Le Jan, T., Roualdes, V., Milleville-Pennel, I. (2022). Design impact on Mental Workload. *63rd annual meeting of Psychonomic Society*, Boston, MA, 2022.

Louis, L-E., Moussaoui, S., Van Langenhove, A., Ravoux, S., Le Jan, T., Roualdes, V., Milleville-Pennel, I. (2022). Subjective Mental Workload modeling during cognitive tasks, *22nd conference of the European Society for Cognitive Psychology, ESCOP*, Lille, France.

Communications dans des conférences nationales

Louis, L-E., Moussaoui, S., Van Langenhove, A., Ravoux, S., Le Jan, T., Roualdes, V., Milleville-Pennel, I. (2023). Monitoring Mental Workload by EEG during a game in Virtual Reality. *IHM 2023*. Troyes, France.

Bibliographie

- Abibullaev, B., & An, J. (2012). Classification of frontal cortex haemodynamic responses during cognitive tasks using wavelet transforms and machine learning algorithms. *Medical engineering & physics*, 34(10), 1394-1410.
- Al-Mohair, H. K., Saleh, J. M., & Suandi, S. A. (2015). Hybrid human skin detection using neural network and K-means clustering technique. *Applied Soft Computing*, 33, 337-347.
- Andre A.D. (2001). The value of workload in the design and analysis of consumer products. In P.A. Hancock & P.A. Desmond PA (Eds.), *Stress, workload, and fatigue* (pp. 373–382). Mahwah, NJ: L. Erlbaum.
- Andreessen, L. M., Gerjets, P., Meurers, D., & Zander, T. O. (2021). Toward neuroadaptive support technologies for improving digital reading: a passive BCI-based assessment of mental workload imposed by text difficulty and presentation speed during reading. *User Modeling and User-Adapted Interaction*, 31(1), 75-104.
- Appel, T., Gerjets, P., Hoffman, S., Moeller, K., Ninaus, M., Scharinger, C., ... & Kasneci, E. (2021). Cross-task and cross-participant classification of cognitive load in an emergency simulation game. *IEEE Transactions on Affective Computing*, 1-1
- Aponte, M. V., Levieux, G., & Natkin, S. (2011). Measuring the level of difficulty in single player video games. *Entertainment Computing*, 2(4), 205-213.
- Arvaneh, M., Umilta, A., & Robertson, I. H. (2015). Filter bank common spatial patterns in mental workload estimation. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4749-4752). Milan, IEEE.
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Progress in brain research*, 228, 295-328.
- Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., & Collet, C. (2002). Complexity and controller workload: Trying to bridge the gap. In *Proceedings of the 2002 International Conference on Human-Computer Interaction in Aeronautics* (pp. 56-60). Cambridge, MA, USA: Massachusetts Institute of Technology.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory, Vol. II*. (pp. 89-195). New York: Academic press.
- Auray, N. (2003). L'engagement des joueurs en ligne : ethnographie d'une sociabilité distanciée et restreinte. *Les Cahiers du numérique*, 4, 83-100.
- Averty, P., Collet, C., Dittmar, A., Athènes, S., & Vernet-Maury, E. (2004). Mental workload in air traffic control: an index constructed from field tests. *Aviation, space, and environmental medicine*, 75(4), 333-341.
- Babiloni, F. (2019) Mental workload monitoring: new perspectives from neuroscience. In L. Longo and M. Leva (Eds.), *Human Mental Workload: Models and Applications, H-WORKLOAD 2019. Communications in computer and information science. Vol. 1109*. (pp. 3-19). Springer, Cham.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. OPH. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 8*. (pp. 47-89). New York: Academic press.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Barachant, A. (2013). *Commande robuste d'un effecteur par une interface cerveau machine EEG asynchrone*. Thèse de doctorat en Signal, image, parole, télécoms, Université de Grenoble, France.
- Bayssié, L. & Chaudron, L. (2002). Évaluation de la performance d'un opérateur en fonction de sa tâche. Application aux IHM. In P., Girard, T., Baudel, M., Beaudouin- Lafon, E., Lecolinet & D. L. Scapin (Eds.), *IHM 2002* (pp. 219-222). Poitiers, France.

- Beh, W. K., & Wu, Y. H. (2021). MAUS: A Dataset for Mental Workload Assessment on N-back Task Using Wearable Sensor. In: arXiv preprint arXiv:2111.02561.
- Bel, O. M. S. (2020). *Dynamic performance enhancement of scientific networks and systems*. Thèse de doctorat en philosophie en science informatique, Université de Californie, Santa Cruz, États-Unis.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5), B231-B244.
- Bertet, K., Visani, M., & Girard, N. (2009). Treillis dichotomiques et arbres de décision. *Traitement du Signal*, 26(5), 409-418.
- Besançon, L., Issartel, P., Ammi, M., & Isenberg, T. (2017). Mouse, tactile, and tangible input for 3D manipulation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 4727-4740). Denver, CO: ACM.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44, 58-75.
- Boyer, S. (2016). *Contribution de l'analyse du signal vocal à la détection de l'état de somnolence et du niveau de charge mentale*. Thèse de doctorat en Neurosciences, Université Paul Sabatier-Toulouse III, France.
- Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, 40(109), 23.
- Brigham, T. J. (2015). An introduction to gamification: adding game elements for engagement. *Medical reference services quarterly*, 34(4), 471-480.
- Brookhuis, K. A., & De Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, 42(3), 898-903.
- Brooks, L. R. (1967). The suppression of visualization by reading. *The Quarterly journal of experimental psychology*, 19(4), 289-299.
- Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, 9(4), 045008.
- Bucy, E. P., & Gregson, K. S. (2001). Media participation: A legitimizing mechanism of mass democracy. *New media & society*, 3(3), 357-380.
- Burgess, P.W. & Shallice, T. (1996). Bizarre responses, rule detection and frontal lesions. *Cortex*, 32, 241-259.
- Butmee, T., Lansdown, T. C., & Walker, G. H. (2019). Mental workload and performance measurements in driving task: A review literature. In *Proceedings of the 20th Congress of the International Ergonomics Association Volume VI: Transport Ergonomics and Human Factors, Aerospace Human Factors and Ergonomics* (pp. 286-294). Florence, Italy: IEA.
- Byers, J. C., Bittner, A. C., Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? In Mital, A. (Ed.), *Advances in industrial ergonomics and safety* (pp. 481-485). London: Taylor and Francis.
- Byrne, A., (2011). Measurement of mental workload in clinical medicine: a review study. *Anesthesiol. Pain Med.* 1(2), 90.

- Cain, B. (2007). *A review of the mental workload literature*. Technical report, Defense Research and Development: Toronto, Canada.
- Carelli, L., Solca, F., Faini, A., Meriggi, P., Sangalli, D., Cipresso, P., et al. (2017). Brain-computer interface for clinical purposes: cognitive assessment and rehabilitation. *BioMed Research International*, 2017, 1695290.
- Cegarra, J., & Morgado, N. (2009). Étude des propriétés de la version francophone du NASA-TLX. In *2009 5ème Colloque de Psychologie Ergonomique* (pp. 233-239). Nice, EPIQUE.
- Chanquoy, L., Tricot, A., & Sweller, J. (2007). *La charge cognitive : théories et applications*. Paris : Armand Collin.
- Charles, R. L., and Nixon, J. (2019). Measuring mental workload using physiological measures: a systematic review. *Applied Ergonomics*, 74, 221–232.
- Chen, S., Epps, J., Ruiz, N., & Chen, F. (2011). Eye Activity As a Measure of Human Mental Effort in HCI. In *Proceedings of the 16th International Conference on Intelligent User Interfaces* (pp. 315–318). New York, NY: ACM.
- Cinaz, B., Arnrich, B., La Marca, R., & Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal and ubiquitous computing*, 17(2), 229-239.
- Cole, T., & Gillies, M. (2021). Thinking and doing: Challenge, agency, and the eudaimonic experience in video games. *Games and Culture*, 16(2), 187-207.
- Collette F. (2004). Exploration des fonctions exécutives par imagerie cérébrale. In T. Meulemans, F. Collette, and M. Van der Linden (Eds.), *Neuropsychologie des fonctions exécutives*. Marseille : Solal.
- Colombi, J. M., Miller, M. E., Schneider, M., McGrogan, M. J., Long, C. D. S., & Plaga, J. (2012). Predictive mental workload modeling for semi-autonomous system design: Implications for systems of systems. *Systems Engineering*, 15(4), 448-460.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1), 87-114.
- Cooper, G. E., & Harper, R. P. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities*. (NASA Ames Technical Report NASA TN-D-5153). Moffett Field, CA: NASA Ames Research Center.
- Corsi, P.M. (1972). *Human memory and the medial temporal region of the brain*. Thèse de doctorat en Psychologie, Université de McGill, Montréal, Canada.
- Curran, E. A., & Stokes, M. J. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain and cognition*, 51(3), 326-336.
- Cuvelier, L. (2012). Mesures quantitatives de la charge mentale : avancées, limites et usages pour la prévention des risques professionnels. *Archives des Maladies Professionnelles et de l'Environnement*, 73(2), 120-126.
- Dehais, F., Lafont, A., Roy, R., and Fairclough, S. (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Front. Neurosci.* 14, 268.
- Delignières, D. & Deschamps, T. (2000). L'effort mental. In D. Delignières (Ed.), *L'effort* (pp. 25-40). Paris : Editions Revue EPS.
- Delignières, D. (2004). Difficulté de la tâche et performance. *Manuel de Psychologie du Sport*, 1, 85- 112.

- Della Sala S., Gray, C., Baddeley, AD., & Wilson, L. (1997). *The visual pattern Test: A new Test of short-term visual Recall*. Feltham, Suffolk: Thames Valley Test Company.
- Delliaux, S., Delaforge, A., Deharo, J. C., & Chaumet, G. (2019). Mental workload alters heart rate variability, lowering non-linear dynamics. *Frontiers in physiology*, *10*, 565.
- De Waard, D. (1996). *The measurement of drivers' mental workload*. Thèse de doctorat en Sciences psychologiques, pédagogiques et sociologiques, Université de Groningen, Haren, Pays-Bas.
- Dhamala, M., Pagnoni, G., Wiesenfeld, K., & Berns, G. S. (2002). Measurements of brain activity complexity for varying mental loads. *Physical Review E*, *65*(4), 041917.
- Dimitrakopoulos, G. N., Kakkos, I., Dai, Z., Lim, J., deSouza, J. J., Bezerianos, A. et Sun, Y. (2017). Task-independent mental workload classification based upon common multiband eeg cortical connectivity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(11), 1940–1949.
- Di Stasi, L. L., Antolí, A., & Cañas, J. J. (2013). Evaluating mental workload while interacting with computer-generated artificial environments. *Entertainment Computing*, *4*(1), 63-69.
- Djokic, J., Lorenz, B., & Fricke, H. (2010). Air traffic control complexity as workload driver. *Transportation research part C: emerging technologies*, *18*(6), 930-936.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and social psychology bulletin*, *21*(3), 215-225.
- Eslinger, P. J., & Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia*, *31*, 17-28.
- Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a biocybernetic loop. *Frontiers in human neuroscience*, *10*, 223.
- Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/No-Go tasks and their relation to inhibition. *Acta psychologica*, *101*(2-3), 267-291.
- Fallahi, M., Motamedzade, M., Heidari Moghadam, R., Soltanian, A. R., & Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied Ergonomics*, *52*, 95–103.
- Falzon, P., & Sauvagnac, C. (2004). Charge de travail et stress. In Falzon, P. (Ed.) *Ergonomie* (pp. 175-190). Paris : Presses Universitaires de France.
- Felton, E. A., Williams, J. C., Vanderheiden, G. C., & Radwin, R. G. (2012). Mental workload during brain-computer interface training. *Ergonomics*, *55*(5), 526-537.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381-391.
- Fournier, M., & Albaret, J. M. (2013). Étalonnage des blocs de Corsi sur une population d'enfants scolarisés du CP à la 6e. *Développements*, (3), 76-82.
- Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied ergonomics*, *73*, 90-99.
- Gabaude, C., Baracat, B., Jallais, C., Bonniaud, M., & Fort, A. (2012). Cognitive load measurement while driving. In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, & S. Röttger, et al. (Eds.), *Human factors: A view from an integrative perspective* (pp. 67–80). Toulouse, France: Human Factors and Ergonomics Society.

- Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, 83(3), 269-275.
- Godefroy, O., & les membres du GREFEX. (2008). *Fonctions exécutives et pathologies neurologiques et psychiatriques : Évaluation en pratique clinique*. Marseille : Solal.
- Goethe, O. (2019). Visual aesthetics in games and gamification. In *Gamification Mindset* (pp. 85-92). Springer, Cham.
- Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 41.1-41.49). New York: Wiley.
- Guan, K., Chai, X., Zhang, Z., Li, Q., & Niu, H. (2021). Evaluation of Mental Workload in Working Memory Tasks with Different Information Types Based on EEG. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society* (pp. 5682-5685). Mexico, IEEE.
- Hancock, P. A., & Chignell, M. (1988). Mental Workload Dynamics in Adaptive Interface Design. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(4), 647-658.
- Hancock, P. A., & Caird, J. K. (1993). Experimental evaluation of a model of mental workload. *Human Factors*, 35(3), 413-429.
- Hancock, P. A. & Williams, G. (1993). Effect of task load and task load increment on performance and workload. In: R. S. Jensen (Ed.), *Seventh International Symposium on Aviation Psychology* (pp. 328-334). Columbus: Ohio.
- Hancock, G. M., Longo, L., Young, M. S., & Hancock, P. A. (2021). Mental workload. *Handbook of human factors and ergonomics*, 203-226.
- Harmony, T., Alba, A., Marroquín, J. L., & González-Frankenberger, B. (2009). Time-frequency-topographic analysis of induced power and synchrony of EEG signals during a Go/No-Go task. *International Journal of Psychophysiology*, 71(1), 9-16.
- Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland Press.
- Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 904-908). Los Angeles, CA: Sage publications.
- Heaton, R. K (1981). *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources.
- Hebb, D. O. (1955). Drives and the CNS (Conceptual nervous system). *Psychological Review*, 62(4), 243–254.
- Heuer, R. J. (1999). *Psychology of intelligence analysis*. Technical report, Center for the Study of Intelligence: CIA.
- Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., ... & Gabaude, C. (2019). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Frontiers in Human Neuroscience*, 12, 525.
- Hockey, G. R. J. (1986). A state control theory of adaptation and individual differences in stress management. In *Energetics and Human Information Processing* (pp. 285-298). Springer, Dordrecht.

- Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task-evoked pupillary response to mental workload in human-computer interaction. In *CHI'04 extended abstracts on Human factors in computing systems* (pp. 1477-1480). Vienna Austria: ACM.
- ISO (2017). ISO 1075-1:2017 - *Ergonomic Principles Related to Mental Workload - Part 1: General Issues and Concepts, terms and definitions*. Geneva: ISO.
- Jersild, A. T. (1927). Mental set and shift. Archives of Psychology, Whole No. 89. *J. Exp. Psychol.* 49, 29–50.
- Kahneman, D., (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Karasek Jr, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative science quarterly*, 285-308.
- Katmah, R., Al-Shargie, F., Tariq, U., Babiloni, F., Al-Mughairbi, F., & Al-Nashash, H. (2021). A review on mental stress assessment methods using EEG signals. *Sensors*, 21(15), 5043.
- Ke, Y., Qi, H., Zhang, L., Chen, S., Jiao, X., Zhou, P., ... & Ming, D. (2015). Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression. *International Journal of Psychophysiology*, 98(2), 157-166.
- Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: standardization and normative data. *Applied neuropsychology*, 7(4), 252-258.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*, 55(4), 352.
- Klein, S. B. (1982). *Motivation: biosocial approaches*. New-York: McGraw-Hill College.
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, 8.
- Ko, S. M., & Ji, Y. G. (2018). How we can measure the non-driving-task engagement in automated driving: comparing flow experience and workload. *Applied ergonomics*, 67, 237-245.
- Kostenko, A. S. (2017). *Évaluation multidimensionnelle et dynamique de la maîtrise de la situation par l'opérateur : création d'un indicateur temps réel de charge mentale pour l'activité de supervision de drones*. Thèse de doctorat en Facteurs Humains, Université de Bretagne Sud, France.
- Kuribayashi, A., Takeuchi, E., Carballo, A., Ishiguro, Y., & Takeda, K. (2021). A Comparison of Methods for Sharing Recognition Information and Interventions to Assist Recognition in Autonomous Driving System. In *2021 IEEE Intelligent Vehicles Symposium (IV)* (pp. 622-629). Nagoya, Japan, IEEE.
- Laouar-Zouyed, A. (2021). *Étude de la prise de décision chez les pilotes d'aviation commerciale : relation entre le contrôle cognitif et la charge mentale*. Thèse de doctorat en Sciences cognitives, Université de Bretagne Sud, France.
- Laurent, L. (2013). *Électromyographie laryngée du langage intérieur dans la flexibilité mentale et la mémoire*. Thèse de doctorat en Psychologie, Université de Franche-Comté, France.
- Lean, Y., & Shan, F. (2012). Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(3), 177-187.
- Leduc, S. (2019). Charge de travail. In G. Valléry (Ed.), *Psychologie du Travail et des Organisations : 110 notions clés* (pp. 83-88). Paris : Dunod.

- Le Gonidec, N. (2022). *Conceptualiser et évaluer la charge mentale de salariés dans un contexte d'usage d'outils numériques : Le cas d'une entreprise de télécommunications*. Thèse de doctorat en Psychologie, Université Côte d'Azur, France.
- Leplat, J. (1977). Les facteurs déterminant la charge de travail. *Le Travail humain*, 40, 195- 202.
- Leplat, J. (2006). La notion de régulation dans l'analyse de l'activité. *Perspectives interdisciplinaires sur le travail et la santé*, 8(1).
- Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merriënboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and instruction*, 30, 32-42.
- Li, W. C., Chiu, F. C., Kuo, Y. S., & Wu, K. J. (2013). The investigation of visual attention and workload by experts and novices in the cockpit. In *Engineering Psychology and Cognitive Ergonomics. Applications and Services: 10th International Conference*. Las Vegas, NV, Springer Berlin Heidelberg.
- Li, W., Li, R., Xie, X., & Chang, Y. (2021). Evaluating mental workload during multitasking in simulated flight. *Brain and Behavior*, 12(4), e2489.
- Lim, W. L., Sourina, O., & Wang, L. P. (2018). STEW: simultaneous task EEG workload data set. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(11), 2106-2114.
- Loft, S., Sanderson, P., Neal, A., & Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. *Human factors*, 49(3), 376-399.
- Logan, G. D. (1994). On the ability to inhibit thought and action: A user's guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego, CA: Academic Press.
- Longo, L. (2017). Subjective usability, mental workload assessments and their impact on objective human performance. In *IFIP Conference on Human- Computer Interaction* (pp. 202–223). Mumbai, Springer International Publishing.
- Longo, L., Wickens, C. D., Hancock, P. A., & Hancock, G. M. (2022). Human Mental Workload: A Survey and a Novel Inclusive Definition. *Frontiers in psychology*, 13, 883321.
- Louis, L-E., Moussaoui, S., Van Langenhove, A., Ravoux, S., Le Jan, T., Roualdes, V., Milleville-Pennel, I. (2023). Cognitive task and combined statistical methods to evaluate, model and predict mental workload. *Frontiers in Psychology-Cognitive Science (In press)*.
- Luximon, A., & Goonetilleke, R. S. (2001). Simplified subjective workload assessment technique. *Ergonomics*, 44(3), 229–243.
- Magliacane, S., Van Ommen, T., Claassen, T., Bongers, S., Versteeg, P., & Mooij, J. M. (2018). Domain adaptation by using causal inference to predict invariant conditional distributions. *Advances in neural information processing systems*, 31.
- Malakoutikhah, M., Kazemi, R., Rabiei, H., Alimohammadlou, M., Zare, A., & Hassanipour, S. (2021). Comparison of mental workload with N-Back test: A new design for NASA-task load index questionnaire. *International Archives of Health Sciences*, 8(1), 7-13.
- Mallat, C. (2019). *Évaluation et mise en évidence des effets de la charge mentale de travail sur l'adaptabilité comportementale et physiologique*. Thèse de doctorat en Ergonomie, Université de Toulouse, France.
- Mandrick, K. (2013). *Application de la spectroscopie proche infrarouge dans la discrimination de la charge de travail*. Thèse de doctorat en Sciences du mouvement humain, Université Montpellier I, France.

- Marinescu, A. C., Sharples, S., Ritchie, A. C., Sanchez Lopez, T., McDowell, M., & Morvan, H. P. (2018). Physiological parameter response to variation of mental workload. *Human factors*, 60(1), 31-56.
- Martin, C. (2013). *La gestion de la charge mentale des contrôleurs aériens en-route: apports de l'eye-tracking dans le cadre du projet européen SESAR*. Thèse de doctorat en Ergonomie, Université de Toulouse, France.
- McEvoy, L. K., Pellouchoud, E., Smith, M. E., & Gevins, A. (2001). Neurophysiological signals of working memory in normal aging. *Cognitive Brain Research*, 11(3), 363-376.
- Meijman, T.F. & O'Hanlon, J.F. (1984). Workload. An introduction to psychological theories and measurement methods. In P.J.D. Drenth, H. Thierry, P.J. Willems & C.J. de Wolff (Eds.), *Handbook of Work and Organizational Psychology* (pp. 257-288). New York: Wiley.
- Meindl, P., Yu, A., Galla, B. M., Quirk, A., Haeck, C., Goyer, J. P., ... & Duckworth, A. L. (2019). A brief behavioral measure of frustration tolerance predicts academic achievement immediately and two years later. *Emotion*, 19(6), 1081.
- Meister, D. (1976). *Behavioral foundations of system development*. New York: Wiley.
- Mélan, C., & Cascino, N. (2014). A multidisciplinary approach of workload assessment in real-job situations: Investigation in the field of aerospace activities. *Frontiers in psychology*, 5, 964.
- Meshkati, N. (1988). Toward development of a cohesive model of workload. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 305–314). Amsterdam: Elsevier.
- Milleville-Pennel, I., & Charron, C. (2015). Do mental workload and presence experienced when driving a real car predispose drivers to simulator sickness? An exploratory study. *Accident Analysis & Prevention*, 74, 192-202.
- Milleville-Pennel, I., & Marquez, S. (2020). Comparison between elderly and young drivers' performances on a driving simulator and self-assessment of their driving attitudes and mastery. *Accident Analysis & Prevention*, 135, 105317.
- 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.
- Miller, S. (2001). *Workload measures*. Iowa City, IA: National Advanced Driving Simulator.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology*, 41(1), 49-100.
- Mohanavelu, K., Poonguzhali, S., Janani, A., & Vinutha, S. (2022). Machine learning-based approach for identifying mental workload of pilots. *Biomedical Signal Processing and Control*, 75, 103623.
- Moray, N. (1967) Where is capacity limited? A survey and a model. *Acta Psychologica*, 27, 84-92.
- Moray, N. (1979). *Mental workload: Its theory and measurement*. New York: Plenum.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2), 61-64.
- Mulder, G. (1979). Mental load, mental effort and attention. In N. Moray (Ed.), *Mental Workload: Its Theory and Measurement*. New York: Plenum Press.
- Mulder, G. (1986). The concept and measurement of mental effort. In G.R.J. Hockey, A.W.K. Gaillard and M.G.H. Coles (Eds.), *Energetics and human information processing* (pp. 175-198). Dordrecht, The Netherlands: Martinus Nijhoff Publishers.

- Muñoz-de-Escalona, E., & Cañas, J. J. (2019). Latency differences between mental workload measures in detecting workload changes. In Longo, L., Leva, M.C. (Eds.) *H-WORKLOAD 2018* (pp. 131-146). Berlin: Springer
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383.
- Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex*, 12(4), 313-324.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R.J. Davidson, G.E. Schartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research* (pp. 1-18). New York: Plenum Press.
- Nguma, H. (2020). *Étude de l'effet de synchronie pour la conception d'IHM adaptatives: Effet sur la vigilance, la charge mentale, la conscience de la situation et les performances*. Thèse de doctorat en Psychologie, Université Côte d'Azur, France.
- O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112, 28-39.
- O'Donnell, R.D. and Eggemeier, F.T. (1986). Workload assessment methodology. In Kaufman, B.K., Wiley, T.J. (Eds.), *Handbook of Perception and Human Performance, Volume II, Cognitive Processes and Performance* (pp. 41–42). New York: John Wiley.
- Onken, R., & Schulte, A. (2010). *System-ergonomic design of cognitive automation: dual-mode cognitive design of vehicle guidance and control work systems, Vol. 235*. Berlin: Springer.
- Orru G., & Longo, L. (2019). The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: a review. In Longo, L., Leva, MC. (Eds.) *Human mental workload: models and applications* (pp. 23–48). Berlin: Springer.
- Oyshi, M. T., Saifuzzaman, M., & Tumpa, Z. N. (2018). Gamification in children education: Balloon shooter. In *2018 4th International Conference on Computing Communication and Automation* (pp. 1-5). Greater Noida, India, IEEE.
- Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), 351 6 371.
- Paxion, J. (2014). *Complexité des situations, expérience, tension et vigilance : quels impacts sur la charge de travail et les performances de jeunes conducteurs ?* Thèse de doctorat en Ergonomie, Université de Grenoble, France.
- Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016). A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5776-5788). San Jose, California, CHI.
- Piéron, H. (1959). Les échelles subjectives peuvent-elles fournir la base d'une nouvelle loi psychophysique ? *L'Année psychologique*, 59(1), 1-34.
- Poirier, F., Condat, A., Laufer, L., Rosenblum, O., & Cohen, D. (2019). Non-binarité et transidentités à l'adolescence : une revue de la littérature. *Neuropsychiatrie de l'Enfance et de l'Adolescence*, 67(5-6), 268-285.
- Radüntz, T. (2017). Dual frequency head maps: A new method for indexing mental workload continuously during execution of cognitive tasks. *Frontiers in physiology*, 8, 1019.
- Rainieri, G., Fraboni, F., Russo, G., Tul, M., Pingitore, A., Tessari, A., & Pietrantonio, L. (2021). Visual scanning techniques and mental workload of helicopter pilots during simulated flight. *Aerospace Medicine and Human Performance*, 92(1), 11-19.

- Raufaste, É., Daurat, A., Mélan, C., & Ribert-Van De Weerd, C. (2004). Aspects intensifs de la cognition en situation de travail. In Darses, F., & Hoc, J. M. (Eds.), *Psychologie ergonomique : tendances actuelles* (pp. 175-199). Paris: Presses Universitaires de France.
- Raufi, B., & Longo, L. (2022). An Evaluation of the EEG alpha-to-theta and theta-to-alpha band Ratios as Indexes of Mental Workload. *Frontiers in Neuroinformatics*, 16, 44.
- Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of experimental psychology: Applied*, 9(2), 119.
- Reid, G.B., & Nygren, T.E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185-218). Amsterdam: Elsevier.
- Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and motor skills*, 8(3), 271-276.
- Ries, A. J., Slayback, D., & Touryan, J. (2018). The Effect of Visual Task Difficulty on the Fixation-Related Lambda Response. In *2016 49th Hawaii International Conference on System Sciences* (pp. 1226-1235). Hawaii, HICSS.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207-231.
- Roscoe, a H. (1992) Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, 34(2-3), 259-287.
- Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI. In *35th annual international conference of the IEEE Engineering in Medicine and Biology Society* (pp. 6607-6610). Osaka, Japan: IEEE.
- Roy, R. N. (2015). *Étude de corrélats électrophysiologiques pour la discrimination d'états de fatigue et de charge mentale : apports pour les interfaces cerveau-machine passives*. Thèse de doctorat en Signal, image, parole, télécoms, Université de Grenoble, France.
- Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied psychology*, 53(1), 61-86.
- Rueb, J., Vidulich, M., & Hassoun, J. (1992). Establishing workload acceptability: an evaluation of a proposed KC-135 cockpit redesign. In *Proceedings of the Human Factors Society 36th annual meeting* (pp. 17-21). Santa Monica, CA: Human Factors Society.
- Ruscio, D., Bos, A. J., & Ciceri, M. R. (2017). Distraction or cognitive overload? Using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system. *Accident Analysis & Prevention*, 103, 105-111.
- Salicé, G., & Juhel, J. (2005). Effets du vieillissement normal sur les composantes « réactive » et « auto-générée » de la flexibilité cognitive. In L. Taconnat, D. Clarys, S. Vanneste & M. Isingrini (Eds.), *Manifestations cognitives du vieillissement psychologique* (pp. 399-418). Paris : Editions Publibook Université.
- Sanders, A. (1983). Towards a model of stress and human performance. *Acta psychologica*, 53(1), 61-97.
- Schulte, A., Donath, D., & Honecker, F. (2015). Human-system interaction analysis for military pilot activity and mental workload determination. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1375-1380). Hong Kong, China: IEEE.

- Shaheen, N., Raza, B., Shahid, A. R., & Alquhayz, H. (2020). A novel optimized case-based reasoning approach with K-means clustering and genetic algorithm for predicting multi-class workload characterization in autonomic database and data warehouse system. *IEEE Access*, 8, 105713-105727.
- Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2017). Comparison of the working memory load in n-back and working memory span tasks by means of EEG frequency band power and P300 amplitude. *Frontiers in human neuroscience*, 11, 6.
- Schlegel, R. E. (1993). Driver mental workload. In B. Peacock & W. Karwowski (Eds.), *Automotive ergonomics* (pp. 359–382). London: Taylor & Francis.
- Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2014). Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. *Instructional Science*, 43, 93-114.
- Seymour, K. E., Rosch, K. S., Tiedemann, A., & Mostofsky, S. H. (2020). The validity of a frustration paradigm to assess the effect of frustration on cognitive control in school-age children. *Behavior therapy*, 51(2), 268-282.
- Shin, J., Von Lühmann, A., Kim, D. W., Mehnert, J., Hwang, H. J., & Müller, K. R. (2018). Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Scientific data*, 5(1), 1-16.
- Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell, J. F. (1993). Assessment of Pilot Performance and Mental Workload in Rotary Wing Aircraft. *Ergonomics*, 36(9), 1121-1140.
- So, W. K., Wong, S. W., Mak, J. N., & Chan, R. H. (2017). An evaluation of mental workload with frontal EEG. *PloS one*, 12(4), e0174949.
- Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., ... & Jacob, R. J. (2009). Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (pp. 157-166). Victoria BC Canada: ACM.
- Sperandio, J. C. (1971). Variation of operator's strategies and regulating effects on workload. *Ergonomics*, 14(5), 571-577.
- Sperandio, J.-C. (1988). *Ergonomie du travail mental*. Paris : Éd. Masson.
- Sperandio, J.-C. (2019). La charge mentale... Au travers de quelques recherches anciennes d'ergonomie. *Ergonomie et nouvelles Technologies*, 10, 57-61.
- Staal, M. A. (2004). *Stress, cognition, and human performance: A literature review and conceptual framework*. Moffett Field, CA: National Aeronautics and Space Administration.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, 22(2), 123–138.
- Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5), 740-748.
- Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The quarterly journal of experimental psychology*, 37(4), 571-590.

- Törnros, J., & Bolling, A. (2006). Mobile phone use—effects of conversation on mental workload and driving speed in rural and urban environments. *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(4), 298-306.
- Tran, Y., Craig, A., Craig, R., Chai, R., & Nguyen, H. (2020). The influence of mental fatigue on brain activity: Evidence from a systematic review with meta-analyses. *Psychophysiology*, 57(5), 1-17.
- Trejo, L. J., Knuth, K., Prado, R., Rosipal, R., Kubitz, K., Kochavi, R., ... & Zhang, Y. (2007). EEG-based estimation of mental fatigue: convergent evidence for a three-state model. In D. D. Schmorow & L. M. Reeves (Eds.), *Augmented Cognition, HCII, LNAI 4565* (pp. 201–211). Berlin: Springer.
- Tremmel, C., Herff, C., Sato, T., Rechowicz, K., Yamani, Y., & Krusienski, D. J. (2019). Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG. *Frontiers in Human Neuroscience*, 13.
- Tsang, P.S., Velazquez, V.L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics* 39(3), 358–381.
- Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, technology & work*, 20, 351-365.
- Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education*, 44(1), 85-93.
- Veltman, J.A. and Gaillard, W.K. (1996) 'Physiological indices of workload in a simulated flight task', *Biological Psychology*, 42(3), 323–342.
- Vidal, J., Mills, T., Pang, E. W., & Taylor, M. J. (2012). Respiration inhibition in adults and teenagers: spatiotemporal differences in the prefrontal cortex. *Brain and cognition*, 79(1), 49-59.
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walter, C., Schmidt, S., Rosenstiel, W., Gerjets, P., & Bogdan, M. (2013). Using cross-task classification for classifying workload levels in complex learning tasks. In *Proceedings of the 5th IEEE Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013 (Geneva)*, 876–881.
- Wanyan, X., Zhuang, D., & Zhang, H. (2014). Improving pilot mental workload evaluation with combined measures. *Bio-medical materials and engineering*, 24(6), 2283-2290.
- Weinger, M. B., Reddy, S. B., & Slagle, J. M. (2004). Multiple measures of anesthesia workload during teaching and nonteaching cases. *Anesthesia & Analgesia*, 98(5), 1419-1425.
- Welford, A. T. (1978). Mental Workload as a Function of Demand, Capacity, Strategy and Skill. *Ergonomics*, 21(3), 151–167.
- Wickens, C.D. (1984). Processing resources in attention. In R. Parasuraman & D.R. Davis (Eds.), *Varieties of Attention* (pp. 63-101). New-York: Academic Press.
- Wickens, C. D. (1987). Attention. In P. A. Hancock (Ed.), *Human Factors Psychology* (pp. 29-80). New York: North- Holland Press.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2), 159–177.
- Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 449-455.

- Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). Mental workload, stress, and individual differences: cognitive and neuroergonomic perspectives. In *Engineering Psychology and human performance* (pp. 346-376). New Jersey: Pearson Education.
- Wierwille, W. W., & Gutmann, J. C. (1978). Comparison of primary and secondary task measures as a function of simulated vehicle dynamics and driving conditions. *Human Factors*, *20*, 233–244.
- Wierwille, W. W., & Casali, J. G. (1983). A validated rating scale for global mental workload measurement applications. *Proceedings of the Human Factors Society*, *27*, 129-133.
- Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human factors*, *35*(2), 263-281.
- Wilson, J. L., Scott, J. H., & Power, K. G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology*, *5*(3), 249-255.
- Wilson, G.F. (2002). “An analysis of mental workload in pilots during flight using multiple psychophysiological measures.” *International Journal of Aviation Psychology* *12*(1), 3-18.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.
- Yoo, S. H., Woo, S. W., Shin, M. J., Yoon, J. A., Shin, Y. I., & Hong, K. S. (2020). Diagnosis of mild cognitive impairment using cognitive tasks: A functional near-infrared spectroscopy study. *Current Alzheimer Research*, *17*(13), 1145-1160.
- Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human factors*, *44*(3), 365-375.
- Young, M. S., & Stanton, N. A. (2004). Taking the load off: investigations of how adaptive cruise control affects mental workload. *Ergonomics*, *47*(9), 1014-1035.
- Young, M. S., & Stanton, N. A. (2007). What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics*, *50*(8), 1324-1339.
- Young, G., Zavelina, L., & Hooper, V. (2008). Assessment of workload using NASA Task Load Index in perianesthesia nursing. *Journal of PeriAnesthesia Nursing*, *23*(2), 102-110.
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1-17.
- Young, J. Q., & Sewell, J. L. (2015). Applying cognitive load theory to medical education: construct and measurement challenges. *Perspectives on medical education*, *4*, 107-109.
- Young, J. Q., John, M., Thakker, K., Friedman, K., Sugarman, R., Sewell, J. L., & O'sullivan, P. S. (2021). Evidence for validity for the cognitive load inventory for handoffs. *Medical Education*, *55*(2), 222-232.
- Zagermann, J., Pfeil, U., & Reiterer, H. (2016). Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization* (pp. 78-85). Baltimore MD USA: BELIV.
- Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*(1), 37-63.
- Zhang, P., Wang, X., Zhang, W. et Chen, J. (2018). Learning spatial–spectral–temporal eeg features with recurrent 3d convolutional neural networks for cross-task mental workload assessment. *IEEE Transactions on neural systems and rehabilitation engineering*, *27*(1), 31–42.
- Zhou, Y., Xu, Z., Niu, Y., Wang, P., Wen, X., Wu, X., & Zhang, D. (2022). Cross-Task cognitive workload recognition based on EEG and domain adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 50-60.

Zijlstra, F. R. H. (1993). Efficiency in work behavior. A design approach for modern tools. Thèse de doctorat en Design, Delft University of Technology, Delft, The Netherlands.

Zook, A., & Riedl, M. O. (2014). Temporal game challenge tailoring. *IEEE Transactions on Computational Intelligence and AI in Games*, 7(4), 336-346.

Annexes

Annexe 1 : Version française du questionnaire du NASA-TLX

L'échelle des réponses est sur 100. Vous pouvez glisser le curseur de gauche à droite.
Ou alors, vous pouvez mettre un chiffre directement dans la case vide à côté du curseur.

0 : EXIGENCE MENTALE : Dans quelle mesure des opérations mentales et perceptives ont-elles été requises (par ex.: penser, décider, calculer, se rappeler, regarder, chercher, etc.) ? Ont-elles conduit à une tâche plutôt facile ou difficile, simple ou complexe ? abordable ou exigeante ?		<input type="text" value=""/>
	Faible	Forte
1 : EXIGENCE PHYSIQUE : Dans quelle mesure des opérations physiques ont-elles été requises (par ex.: pousser, tirer, tourner, superviser, activer, etc.) ? Ont-elles conduit à une tâche plutôt facile ou difficile, lente ou rapide, lâche ou vigoureuse, reposante ou ardue ?		<input type="text" value=""/>
	Faible	Forte
2 : EXIGENCE TEMPORELLE : Quelle pression temporelle avez-vous ressentie (que ce soit à cause de la cadence du rythme, de l'allure des tâches ou de l'apparition des éléments de la tâche) ? L'allure était-elle lente et tranquille ou rapide et frénétique ?		<input type="text" value=""/>
	Faible	Forte
3 : EFFORT : Quelle a été la difficulté d'accomplir (mentalement et physiquement) la tâche avec un niveau de performance tel que le vôtre ?		<input type="text" value=""/>
	Faible	Forte
4 : PERFORMANCE : Quelle réussite vous attribuez-vous en ce qui concerne l'atteinte des buts de la tâche, fixés par l'experimentateur (ou par vous même) ? Dans quelle mesure êtes-vous satisfait de votre performance dans l'accomplissement de ces buts ?		<input type="text" value=""/>
	Faible	Bonne
5 : FRUSTRATION : Au cours de la tâche, quelle proportion de sentiment d'être peu sûr de vous, découragé, irrité, stressé et agacé avez-vous ressenti par rapport au fait d'être sûr de vous, satisfait, content, détendu, et complaisant ?		<input type="text" value=""/>
	Faible	Forte

Annexe 2 : Questionnaire du *Workload Profile* issu de Raufi (2019)

WP	Version originale du WP	Traduction française du WP
WP1	How much attention was required for activities like remembering, problem-solving, decision-making, perceiving (detecting, recognizing, identifying objects)?	Quel niveau d'attention vous a été nécessaire pour des activités telles que la mémorisation, la résolution de problèmes, la prise de décision, la perception (détecter, reconnaître, identifier des objets) ?
WP2	How much attention was required for selecting the proper response channel (manual - keyboard/mouse, or speech - voice) and its execution?	Quel niveau d'attention vous a été nécessaire pour sélectionner le canal de réponse approprié (manuel-clavier/souris, ou parole-voix) et pour exécuter cette réponse ?
WP3	How much attention was required for spatial processing (spatially pay attention around you)?	Quel niveau d'attention vous a été nécessaire pour le traitement spatial de l'information (pour porter votre attention sur l'espace autour de vous) ?
WP4	How much attention was required for verbal material (eg. reading, processing linguistic material, listening to verbal conversations)?	Quel niveau d'attention vous a été nécessaire pour le traitement du matériel verbal (par exemple, la lecture, le traitement de matériel linguistique, l'écoute des conversations verbales) ?
WP5	How much attention was required for executing the task based on the information visually received (eyes)?	Quel niveau d'attention vous a été nécessaire pour exécuter la tâche sur la base des informations perçues visuellement (yeux) ?
WP6	How much attention was required for executing the task based on the information auditorily received (ears)?	Quel niveau d'attention vous a été nécessaire pour exécuter la tâche sur la base des informations perçues auditivement (oreilles) ?
WP7	How much attention was required for manually respond to the task (eg. keyboard/mouse)?	Quel niveau d'attention vous a été nécessaire pour répondre manuellement à la tâche (par exemple, clavier/souris) ?
WP8	How much attention was required for producing the speech response (eg. engaging in a conversation, talking, answering questions)?	Quel niveau d'attention vous a été nécessaire pour produire la réponse verbale (par exemple, engager une conversation, parler, répondre à des questions) ?

Annexe 3 : Tableau comparatif des différentes variantes du NBT

Auteurs	Kirchner (1958)	Dimitrakopoulos et al. (2017)	Shin et al. (2018)	Zhang et al. (2018)	Tremmel et al. (2019)	Étude 1 de la thèse
Objectifs de l'étude	Mesurer la performance des plus âgés et des plus jeunes sur une tâche impliquant des informations en évolution rapide.	Permettre une discrimination de la charge de travail inter-tâches en utilisant l'EEG.	Faciliter l'évaluation des performances et la comparaison de nombreuses techniques d'analyse de neuroimagerie.	Evaluation inter-tâches afin d'étendre un modèle pré-entraîné à de nouvelles conditions.	Evaluer la faisabilité de surveiller la charge de travail via l'EEG dans un environnement VR interactif.	Cf. section 3.3. <i>Problématique</i>
Niveaux de complexité	0-back ; 1-back ; 2-back ; 3-back	0-back et 2-back	0-back ; 2-back ; 3-back	1-back et 3-back	0-back ; 1-back ; 2-back	0-back ; 1-back ; 2-back
Nb d'essais par niveau	36 essais par niveau	150 (30 % de cibles)	180 (30 % de cibles)	200 (30 % de cibles)	80 (>10 % de cibles)	30 (30 % de cibles)
Durée des stimuli + durée inter-essais	1,5 s. + 0	0,5 + 1,5 (aucun repos)	0,5 + 1,5 (+ 20 s de repos)	0,5 + 2,5	4 s + 0	0,5 + 1,5 (+10s de repos)
Durée	Non précisée	10 mins	27 mins	70 mins	15 mins	17 mins
N	N = 40 participants	N = 28	N = 26	N = 20	N = 15	N = 53
Feedbacks lors de réponses correctes	Si le participant a une bonne réponse, un buzzer retentit ; sinon, silence	Aucune rétroaction n'est donnée quelle que soit la précision de la réponse.	Aucune rétroaction n'est donnée quelle que soit la précision de la réponse.	Aucune rétroaction n'est donnée quelle que soit la précision de la réponse.	La performance (pourcentage correct) a été affichée après chaque essai.	Aucune rétroaction n'est donnée quelle que soit la précision de la réponse.
Double action cible/non-cible	Non (aucun argument) le participant appuie sur une touche de clavier en cas de cibles uniquement	Oui (aucun argument) aucune précision sur les actions à réaliser	Oui (argument : plus grand engagement dans l'expérience) : Cible : touche 7 Non-cible : touche 8	Non (aucun argument) Cible : touche "A" du clavier	Oui (aucun argument) Cible : balle dans un réceptacle Non-cible : balle dans réceptacle de droite	Oui
Type d'environnement	Espace physique réel	2D non gamifié	2D non gamifié	2D non gamifié	Réalité Virtuelle gamifiée	2D non gamifié
Mesures de la charge mentale	Aucun. Ce n'est pas l'objectif de l'étude.	EEG	EEG	NASA-TLX + EEG	EEG	Mesures subjectives et de performances
Mesures de la performance	- Pourcentage de réponses correctes Nombre et le type d'erreurs (comme l'erreur d'omission)	-Précision - Temps de réaction (TR)	-Réponses correctes (%) ; Mauvaises réponses (%) ; Temps de réaction (ms)	- Temps de réaction - Précision	-Pourcentage de réponses correctes	Cf. section 4.2.2 <i>Tâches et procédures</i>

Annexe 4 : Tableau comparatif des différents tests de Corsi

Auteurs	Corsi (1972)	Kessels et al. (2000)	Fournier et Albaret (2013)	Brunetti et al. (2014)	Björngrim, (2019)	Étude 1 de la thèse
Objectifs de l'étude	Mettre en évidence le rôle des lésions cérébrales dans les troubles de la mémoire.	Comparer les résultats du test de Corsi entre des participants « sains » et des patients.	Étalonnage du test de Corsi sur une population d'enfants scolarisés du CP à la 6e	Comparer la version traditionnelle et numérisée du test de Corsi	Comparer la version traditionnelle et numérisée du test de Corsi	Cf. section 3.3 <i>Problématique</i>
Niveaux de complexité	Ensemble des niveaux dans l'ordre croissant	Ensemble des niveaux dans l'ordre croissant	Ensemble des niveaux dans l'ordre croissant	Ensemble des niveaux dans l'ordre croissant	Ensemble des niveaux dans l'ordre croissant	3 – 5 – 7 cubes à retenir
N	N = 98	N = 70	N = 456	N = 107	N = 91	N = 53
Type d'environnement	Première version physique du matériel	Version traditionnelle du Corsi	Version traditionnelle du Corsi	Planche traditionnelle du Corsi + version numérisée	Planche traditionnelle du Corsi + version numérisée	Version numérisée du Corsi
Mesures de la charge mentale	Non	Non	Non	Non	Non	Mesures subjectives et de performances
Mesures de la performance	Le nombre de séquences de blocs récurrentes qui ont été tapées dans le bon ordre Le nombre de séquences non récurrentes correctement tapées	L'empan mnésique, égal à la longueur de la dernière séquence correctement répétée. Le score total, produit de l'empan mnésique et du nombre de séquences correctement répétées jusqu'à ce que le test soit interrompu.	La note attribuée correspond au niveau de la dernière séquence correctement reproduite.	Précision et temps de réaction	Précision et temps de réaction	Cf. section 4.2.2 <i>Tâches et procédures</i>

Annexe 5 : Tableau comparatif des différents tests de Go/No-Go

Auteurs	Vidal et al. (2012)	Étude 1 de la thèse
Objectifs de l'étude	Étude des caractéristiques spatio-temporelles de l'inhibition de la réponse lors d'une tâche visuelle de Go/No-Go à l'adolescence et à l'âge adulte.	Cf. section 3.3 <i>Problématique</i>
Stimuli GO	9 formes géométriques noires solides différentes	2 formes géométriques possibles (carré ou cercle noirs)
Stimuli No-Go	Mêmes images avec un « X » gris supplémentaire superposé au centre.	Mêmes images avec un « X » gris supplémentaire superposé au centre.
Matériel	La forme est affichée dans un carré blanc de 10 cm au milieu d'un fond d'écran noir.	La forme est affichée dans un carré blanc de 10 cm au milieu d'un fond d'écran noir.
Complexité	Un Go/No-Go où la majorité des essais (66 %) nécessitent une réponse active (c.-à-d. Stimuli « Go »)	3 niveaux : Facile : 90% de No-Go Moyen : 50% de No-Go Très difficile : 20% de No-Go
Entraînement	Non	Oui.
Procédure	2 essais = contrôle et tâche de Go/No-Go contrebalancés	Entraînement + 3 niveaux dans un ordre aléatoire.
Mesures de la charge mentale	Non	Mesures subjectives et de performances
Mesures de la performance	Temps de réaction pour les réponses correctes uniquement. La précision. Les fausses alarmes (pressions sur des boutons pour des stimuli interdits).	Cf. section 4.2.2 <i>Tâches et procédures</i>

Annexe 6 : Tableau comparatif des différents WCST

Auteurs	Nelson (1976) et base de Godefroy et GREFEX (2008)	Laurent (2013)	Étude 1 de la thèse
Objectifs de l'étude	Développer un outil utile pour le travail clinique diagnostique	Confirmer de précédents résultats.	Cf. section 3.3 <i>Problématique</i>
Interface	Interface physique avec des cartes de jeux	Version numérisée	Version numérisée
Dimensions utilisées	Tri en fonction de la couleur, de la forme ou du nombre.	Tri en fonction de la forme, du nombre, de la taille et de la couleur	Tri en fonction de la Forme, du Nombre et de la Couleur.
N	53 patients répartis en sous-groupes selon le site de la lésion.	31 jeunes adultes	N = 53
Complexité	La complexité est fonction des réponses du participant en cours d'essai.	4 niveaux du WCST (simple répété, simple alterné, complexe répété et complexe alterné)	Trois niveaux de complexité basés sur de l'alternance de dimensions.
Matériel	2 jeux de cartes de 24 cartes. Tri en fonction de la Forme (étoile, croix, cercle, triangle), du Nombre (1, 2, 3 et 4) et de la Couleur (jaune, rouge, bleu, vert).	24 cartes tests par paquets. 3 séries de 4 cartes cibles. Les formes pouvaient être des triangles, des étoiles, des croix ou des ronds, être au nombre de 1, 2, 3 ou 4, mesurer 4.8, 3.3, 2.3, ou 1.5 cm de hauteur et avoir pour remplissage un fond gris, un pointillage important, un pointillage léger ou un fond blanc.	Version numérisée du MCST décrit dans Godefroy et GREFEX (2008) et Laurent (2013) qu'on a adapté pour les besoins de cette étude.
Procédure	Initialement, le sujet doit correspondre en fonction de la couleur, mais après 6 réponses correctes consécutives, ces exigences sont modifiées sans avertissement en forme, puis en nombre. Le test est interrompu après l'achèvement des six catégories, ou lorsque les 48 cartes sont épuisées.	2 tâches de complexité simple et complexe avec entraînement + consignes au début de chaque bloc.	Tâche répétée + tâche alternée + tâche alternée complexe.
Entraînement	Non	Oui	Oui
Mesures de la charge mentale	Non	Non	Mesures subjectives et de performances
Mesures de la performance	Nombre d'erreurs ; Type d'erreurs (erreurs persévératives et non persévératives) ; Score d'erreurs total ; Nombre de catégories achevées	Temps de réponse ; Taux d'erreurs et le coût du temps d'alternance	Cf. section 4.2.2 <i>Tâches et procédures</i>

Annexe 7 : Les consignes des tâches cognitives de l'étude 1

Consigne de la tâche de N-back

Vous êtes dans un jeu basé sur la mémoire.

Toutes les 2 secondes, des lettres défilent devant vous, l'une après l'autre.

Votre objectif : **repérer si la lettre qui s'affiche a déjà été présentée auparavant.**

Si c'est le cas, la lettre présentée est une cible et vous appuyez sur la touche gauche du pavé directionnel de votre clavier. Si ce n'est pas le cas, la lettre présentée est une non-cible et vous appuyez sur la touche droite du pavé directionnel de votre clavier.

Ce jeu de mémoire à 3 niveaux de difficulté.

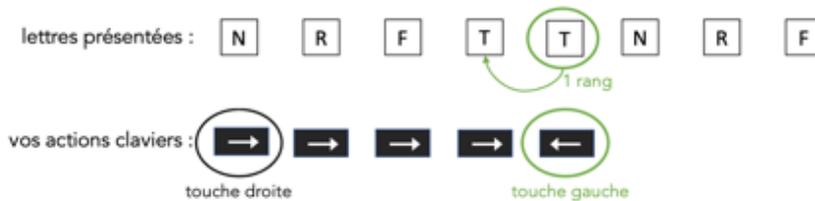
Dans le niveau « 0 rang » la cible est la lettre « X » alors que toutes les autres lettres sont des non-cibles. Ce niveau de jeu est un cas particulier. Et pour les autres niveaux de difficulté de jeu, vous devez comparer les lettres entre elles et indiquer si la lettre présentée à l'instant T est apparue 1 ou 2 lettres (rangs) avant elle.



Exemple 1 :

Pour la consigne « 1 rang », vous devez comparer chaque lettre présentée à la lettre apparue 1 rang avant elle. Si vous voyez défiler ceci : N - R - F - T - T.

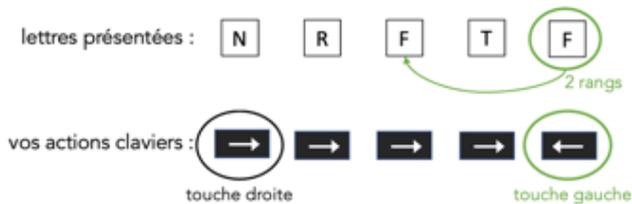
La cible est le deuxième T puisqu'il est identique à la lettre apparue 1 rang auparavant (T). Les autres lettres sont des non-cibles.



Exemple 2 :

Pour la consigne "2 rangs", vous devez comparer chaque lettre à la lettre apparue 2 rangs avant elle. Si vous voyez défiler ceci : N - R - F - T - F.

La cible est le deuxième F puisqu'il est identique à la lettre apparue 2 rangs auparavant (F). Les autres lettres sont des non-cibles.



Indications :

Le son *beep* et l'affichage de « N-rang » indique le début d'un niveau de jeu. Le son *beep* et l'affichage de « FIN » indique la fin d'un niveau. Entre chaque niveau de tâche, vous avez un temps pour vous reposer qui sera indiqué par l'affichage « REPOS ».

Afin de faciliter vos réponses, il est préconisé de mettre l'index et le majeur de la main droite sur respectivement les touches gauche et droite des touches directionnelles de votre clavier. Durant la tâche, il est normal de ne pas avoir de retours/feedbacks de ses actions.

Récapitulatif pour les niveaux 1, 2, 3 rangs :

Les lettres sont présentées les unes après les autres. Vous devez les mémoriser pour les comparer entre elles et indiquer si la lettre présentée est apparue 1, 2 ou 3 rangs avant elle. Ne vous inquiétez pas, avant chaque niveau, vous aurez un rappel de la consigne spécifique au niveau et également une phase d'entraînement avec la possibilité de réaliser un essai.

[Le participant visionne la consigne vidéo basée sur le niveau du N-back N=1]

Ne vous inquiétez pas, avant chaque niveau, vous avez un rappel de la consigne spécifique au niveau.

[Question avant chaque début de jeu]

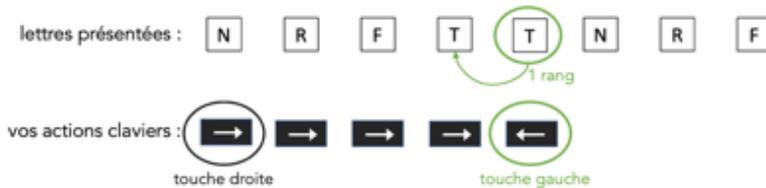
Pour cette tâche, utilisez-vous un pavé tactile ou une souris d'ordinateur ? [Souris/pavé tactile]

Entraînement :

On commence par un petit entraînement.

Vous devez comparer chaque lettre à la lettre apparue 1 rang avant elle.

Si la lettre est identique à celle apparue 1 rang avant, c'est une cible et vous devez appuyer sur la touche directionnelle de gauche. Pour tous les autres cas, lorsque vous rencontrez une non-cible, vous devez appuyer sur la touche directionnelle de droite.

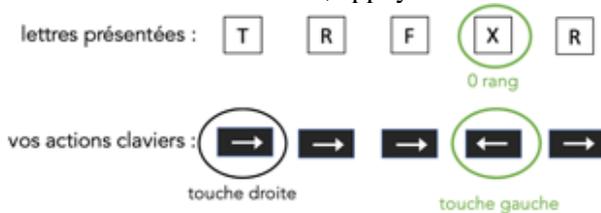


Une fois que vous avez bien compris la consigne d'entraînement, appuyez sur la touche directionnelle droite de votre clavier pour démarrer.

Consignes par niveau :

Consigne pour N=0.

Appuyez sur la touche cible (touche directionnelle gauche) lorsque vous voyez la lettre X. Pour toutes les autres lettres, appuyez sur la touche non-cible (touche directionnelle droite).

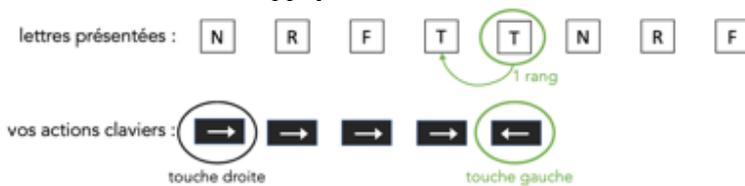


Pour démarrer ce niveau, appuyez sur la touche directionnelle droite de votre clavier.

Consigne pour N=1.

Vous devez comparer chaque lettre à la lettre apparue 1 rang avant elle.

Si la lettre est identique à celle apparue 1 rang avant, c'est une cible et vous devez appuyer sur la touche directionnelle de gauche. Pour tous les autres cas, lorsque vous rencontrez une non-cible, vous devez appuyer sur la touche directionnelle de droite.

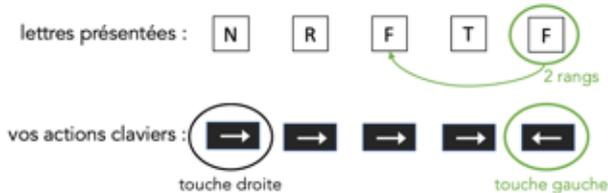


Pour démarrer ce niveau, appuyer sur la touche directionnelle droite de votre clavier.

Consigne pour N=2.

Vous devez comparer chaque lettre à la lettre apparue 2 rangs avant elle.

Si la lettre est identique à celle apparue 2 rangs avant, c'est une cible et vous devez appuyer sur la touche directionnelle de gauche. Pour tous les autres cas, lorsque vous rencontrez une non-cible, vous devez appuyer sur la touche directionnelle de droite.



Pour démarrer ce niveau, appuyer sur la touche directionnelle droite de votre clavier.

Consigne du test de Corsi

En tant qu'expérimentatrice virtuelle, je vais tapoter sur une séquence de cubes plus ou moins longue. Lorsque j'aurai fini, votre objectif est de reproduire la séquence en cliquant ces cubes dans le même ordre que moi. Lorsque vous entendez un son "beep", c'est à votre tour de jouer. Dès que vous avez terminé votre tour, cliquez sur "terminer".

Regardez cet exemple vidéo.

[Le participant visionne la consigne vidéo basée sur le niveau à 2 cubes.]

[Puis, question avant chaque début de jeu]

Pour cette tâche, utilisez-vous un pavé tactile ou une souris d'ordinateur ? [Souris/pavé tactile]

Entraînement :

Vous êtes dans une phase d'entraînement où vous allez réaliser uniquement un seul niveau. On commence par un petit entraînement. Répétez cette séquence immédiatement après moi et dans le même ordre que moi.

Une fois que vous avez bien compris la consigne, cliquez sur valider pour vous entraîner.

Consigne avant chaque essai :

Répétez cet enchaînement dans le même ordre que moi.

Consigne du test de Go/No-Go

Vous allez effectuer une tâche d'inhibition motrice durant laquelle des formes noires vont vous être présentées les unes après les autres.

On peut voir défiler soit un carré vide noir, soit un cercle vide noir. Et ces formes noires peuvent aussi apparaître l'une après l'autre avec un X au centre.

Votre objectif est de répondre aussi rapidement que possible sur la touche « ENTRÉE » du clavier avec votre main droite, lorsque la forme est vide.

Et vous ne devez **pas** appuyer, lorsque la lettre X apparaît sur la forme.



Durant la tâche, c'est normal de ne pas avoir de retours/feedbacks de ses actions.

Cas particulier : La règle générale du jeu pourra subir quelques variations qui vous seront expliquées dans le détail durant le jeu.

Indications :

Le son *beep* indique le début d'un niveau. Le son *beep* et l'affichage de « FIN » indiquent la fin d'un niveau. Entre chaque niveau de tâche, vous avez un temps pour vous reposer qui sera indiqué par l'affichage « PAUSE ».

Regardez l'exemple d'un niveau de la tâche [Le participant visionne la vidéo avec le niveau No-Go = X]

[Puis, question avant chaque début de jeu]

Pour cette tâche, utilisez-vous un pavé tactile ou une souris d'ordinateur ? [Souris/pavé tactile]

Entraînement

Appuyez sur la touche ENTRÉE quand la forme est vide.

Si vous voyez un « X », n'appuyez pas.



Une fois que vous avez bien compris la consigne d'entraînement, appuyez sur la touche ENTRÉE de votre clavier pour démarrer.

Consigne pour No-Go = X

Appuyez sur la touche ENTRÉE quand la forme est vide.

Si vous voyez un « X », n'appuyez pas.



Pour démarrer ce niveau, appuyez sur la touche ENTRÉE de votre clavier.

Consigne pour No-Go = X dans cercle

Appuyez sur la touche ENTRÉE :

- Quand la forme est vide
- Quand le X est dans un carré.

Et n'appuyez **pas**, si vous voyez un « X » dans un cercle.



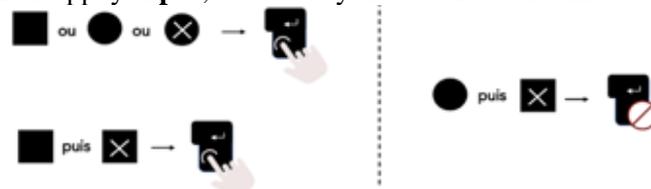
Pour démarrer ce niveau, appuyez sur la touche ENTRÉE de votre clavier.

Consigne pour No-Go = X dans carrés précédés de cercles vides

Appuyez sur la touche ENTRÉE :

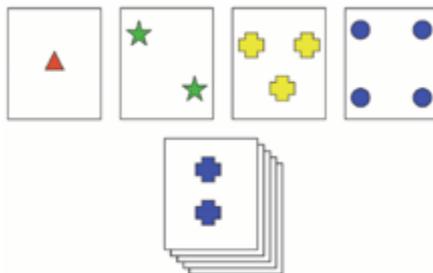
- quand la forme est vide
- quand le X est dans un cercle
- quand le X est dans un carré précédé d'un carré vide

Et n'appuyez **pas**, si vous voyez un « X » dans un carré précédé d'un cercle vide



Pour démarrer ce niveau, appuyez sur la touche ENTRÉE de votre clavier.

Consigne du WCST



Vous avez en haut de l'écran quatre cartes différentes : un triangle rouge, deux étoiles vertes, trois croix jaunes et quatre ronds bleus. En bas de l'écran, je vais faire **défiler** une à une, chacune des cartes du paquet situé en bas de l'écran.

Je vous demande d'**associer** chacune des cartes que je vous montre en bas de l'écran, avec l'une des quatre cartes du haut en fonction d'une règle que je vous annoncerai par écrit. Autrement dit, à chaque nouvelle carte visible sur le paquet de cartes du bas, cliquez sur la carte du haut lui correspondant en fonction de la règle que je vous annoncerai par écrit. Durant la tâche, c'est normal de ne pas avoir de retours/feedbacks de ses actions.

Particularité : au cours d'un même niveau, la consigne peut changer plusieurs fois.

Indications :

Les cartes en bas de l'écran défilent automatiquement toutes les 2 secondes.

Un son beep et l'affichage de la règle de tri indique le début d'un niveau de jeu. Le son beep et l'affichage de « FIN » indiquent la fin d'un niveau. Entre chaque niveau de jeu, vous avez un temps pour vous reposer. Regardez l'exemple vidéo ci-dessous. [Le participant visionne la consigne vidéo basée sur le niveau 1].

[Puis, question avant chaque début de jeu]

Pour cette tâche, utilisez-vous un pavé tactile ou une souris d'ordinateur ? [Souris ou pavé tactile]

Entraînement

Dans ce niveau, je vous demande d'**associer** chacune des cartes que je vous montre en bas avec l'une des quatre cartes qu'il y a devant vous en fonction d'une règle que je vous annonce par écrit.

Par exemple, pour la consigne « associez en fonction de la COULEUR »

Vous allez associer toutes les cartes du bas aux cartes du haut en fonction du nombre d'éléments sur la carte. Et cela, jusqu'au changement de consigne.

Ainsi, au cours d'un même niveau, la consigne peut changer plusieurs fois.

Une fois que vous avez bien compris la consigne, cliquez sur valider pour vous entraîner.

Début du jeu

La phase d'entraînement est terminée. Maintenant, le jeu commence réellement. En tout, vous allez réaliser 4 niveaux de jeu. Je vous rappelle la consigne.

[Étant donné que l'ordre de passation des niveaux est aléatorisé pour éviter l'effet d'apprentissage, les 4 consignes ci-dessous sont présentées dans des ordres aléatoires].

Consigne pour le niveau 0

Dans ce niveau, je vous demande d'**associer** chacune des cartes que je vous montre en bas avec l'une des quatre cartes qu'il y a devant vous en fonction d'une règle que je vous annonce par écrit.

Par exemple, pour la consigne « associez en fonction de la COULEUR BLEUE »
Vous allez associer toutes les cartes du bas aux cartes du haut en fonction de la couleur bleue. Autrement dit, vous devez associer (cliquer) la carte BLEUE en haut si, et uniquement si, la carte du bas est BLEUE. Et cela, jusqu'au changement de consigne.
Ainsi, au cours de ce même niveau, la consigne peut changer plusieurs fois.
Regardez l'exemple vidéo.
[Le participant regarde la vidéo avec le niveau 0]

Consigne pour le niveau 1

Dans ce niveau, je vous demande d'**associer** chacune des cartes que je vous montre en bas avec l'une des quatre cartes qu'il y a devant vous en fonction d'une règle que je vous annonce par écrit.
Par exemple, pour une consigne « associez en fonction du NOMBRE »
Vous allez associer toutes les cartes du bas aux cartes du haut en fonction du nombre d'éléments sur la carte. Et cela, jusqu'au changement de consigne.
Ainsi, au cours de ce même niveau, la consigne peut changer plusieurs fois.
Regardez l'exemple vidéo.
[Le participant regarde la vidéo avec le niveau 1]

Consigne pour le niveau 2

Dans ce niveau, je vous demande d'**associer** chacune des cartes que je vous montre en bas de l'écran avec l'une des quatre cartes qu'il y a devant vous en fonction d'une règle que je vous annonce par écrit.
Particularité, nous sommes dans un niveau d'**alternance**. Autrement dit, il y a une alternance entre deux critères/règles toutes les deux cartes.
Par exemple, pour une consigne « associez en alternant entre le NOMBRE et la COULEUR », vous allez pour la première carte présentée en bas, l'associer à une carte du haut en fonction du nombre d'éléments sur la carte, et pour la deuxième carte présentée, l'associer en fonction de la couleur sur la carte. Et ainsi de suite jusqu'au changement de règles que je vous annonce.
Ainsi, au cours d'un même niveau, je vous dirai les changements de règles d'alternance.
Regardez l'exemple vidéo.
[Le participant regarde la vidéo avec le niveau 2]

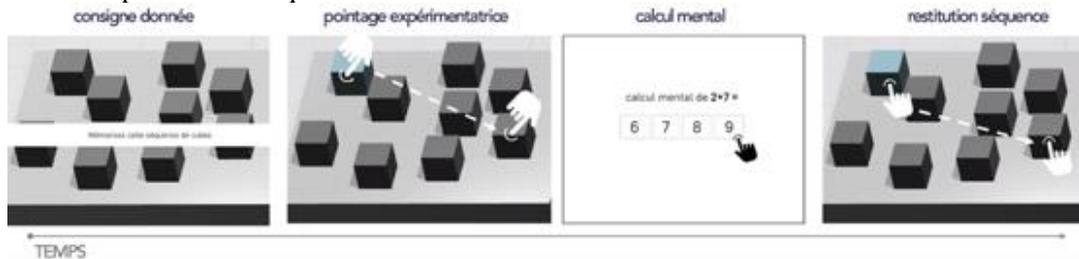
Consigne pour le niveau 3

Dans ce niveau, je vous demande d'**associer** chacune des cartes que je vous montre en bas de l'écran avec l'une des quatre cartes qu'il y a devant vous en fonction d'une règle que je vous annonce par écrit.
Particularité, nous sommes dans un niveau d'**alternance**. Autrement dit, il y a une alternance entre trois critères/règles toutes les trois cartes.
Par exemple, pour une consigne « associez en alternant entre le NOMBRE, la FORME et la COULEUR ». Vous allez pour la première carte présentée en bas, l'associer à une carte du haut en fonction du nombre d'éléments sur la carte, pour la deuxième carte présentée, l'associer en fonction de la forme et pour la troisième carte présentée, l'associer en fonction de la couleur sur la carte. Et ainsi de suite jusqu'au changement de règles que je vous annonce.
Ainsi, au cours d'un même niveau, je vous dirai les changements de règles d'alternance.
Regardez l'exemple vidéo.
[Le participant regarde la vidéo avec le niveau 3].

Consigne de la double tâche

En tant qu'expérimentatrice virtuelle, je vais cliquer sur une séquence de cubes plus ou moins longue. Votre objectif est de la mémoriser, puis répondre le **plus vite possible** à un calcul mental sans calculatrice. Vous indiquerez le résultat de ce calcul en cliquant sur la proposition de réponse qui vous semble juste parmi les 4 qui vous seront proposées à l'écran.

Une fois que vous aurez donné la réponse au calcul mental, vous devrez reproduire la séquence de cubes mémorisée en cliquant sur les cubes dans le même ordre que celui que je vous avais montré au début de l'exercice. Un son « *beep* » vous indiquera que c'est à votre tour de reproduire la séquence de cubes.



Regardez cet exemple vidéo

[Le participant visionne la consigne vidéo basée sur le niveau à 2 cubes et ce calcul mental (2+7)]

[Puis, question avant chaque début de jeu]

Pour cette tâche, utilisez-vous un pavé tactile ou une souris d'ordinateur ? [Souris ou pavé tactile]

Entraînement

Vous êtes dans une phase d'entraînement où vous allez réaliser deux essais avec deux cubes, contrairement au jeu où ce seront des séquences de pointage de 3 cubes. Je vous rappelle la consigne. Mémorisez la séquence de cubes pointés. Puis, répondez au calcul mental le plus vite possible. Pour finir, réalisez la séquence de cubes mémorisée. Une fois que vous avez bien compris la consigne, cliquez sur « valider » pour vous entraîner.

Début du jeu

La phase d'entraînement est terminée. Maintenant, le jeu commence réellement. En tout, vous allez réaliser 3 niveaux de jeu. Je vous rappelle la consigne. Mémorisez la séquence de cubes pointés. Puis, répondez au calcul mental le plus vite possible. Pour finir, réalisez la séquence de cubes mémorisée. Pour démarrer ce niveau, cliquez sur « valider » en bas à droite.

Consigne pour les deux autres niveaux

Mémorisez la séquence de cubes pointés. Puis, répondez au calcul mental le plus vite possible. Pour finir, reproduisez la séquence de cubes mémorisée. Pour démarrer ce niveau, cliquez sur « valider » en bas à droite

Annexe 8 : Avis favorable du CERNI



UNIVERSITÉ DE NANTES

Nantes, le 6 mai 2021

Dossier suivi par : Catherine BONTE
Direction de la recherche, des partenariats et de
l'innovation
catherine.bonte@univ-nantes.fr
+33 (0) 2 28 08 14 28
07 87 20 45 26 (en télétravail lundi et vendredi)

Madame Isabelle MILLEVILLE
LS2N - Equipe PACCE
Campus Ecole Centrale de Nantes
1 rue de la Noë
BP 92101
44321 NANTES cedex 03

N/Réf: SY/GD/CB 2021 DRPI n°290

Objet : Avis du CERNI sur le projet - Modélisation de la variation de la charge mentale auto-déclarée pour des tâches sollicitant différentes fonctions cognitives -.

Madame, Chère collègue,

Vous avez soumis à l'examen du comité d'éthique de la recherche non interventionnelle (CERNI) de l'Université de Nantes (n°IRB : IORG0011023), une étude intitulée - Modélisation de la variation de la charge mentale auto-déclarée pour des tâches sollicitant différentes fonctions cognitives -, dont vous assurez la responsabilité scientifique.

Ce projet a pour objectif l'étude, chez des volontaires sains, de l'évolution de la charge mentale auto-déclarée pour différents niveaux de difficultés de cinq tâches faisant appel à des fonctions cognitives bien distinctes. Cette démarche vous permettra d'identifier au minimum trois niveaux de charge mentale subjective et de modéliser la façon dont varient les différentes dimensions de la charge mentale.

Après un examen attentif, j'ai le plaisir de vous informer que la qualification réglementaire en recherche non interventionnelle a été soumise et validée par le comité éthique de la recherche non-interventionnelle de l'Université de Nantes avec le numéro de référence n°06052021.

Restant à votre disposition, je vous prie de croire, Madame, Chère collègue, en l'assurance de mes sentiments dévoués.

Guillaume DURAND
Président du CERNI

Annexe 9 : Présentation d'un résumé de l'étude 1 pour le recrutement

Bonjour à vous, et merci de prendre ce temps pour lire cette annonce.

Dans le cadre d'un projet collaboratif entre l'entreprise Onepoint, le CHU de Nantes, le CNRS et Central Nantes, nous avons conçu une étude sur la modélisation de la variation de la charge mentale auto-déclarée pour des tâches sollicitant différentes fonctions cognitives.

Dans ce contexte, nous aurions besoin de personnes motivées pour être volontaires pour nos passations. Ces passations se feront totalement en ligne en deux sessions d'une durée d'une heure grand maximum chacune et espacées de quelques jours à quelques semaines. Pour vous remercier de votre contribution à la recherche scientifique, un bon cadeau Fnac d'un montant de 15 euros vous sera offert. Ce bon vous sera remis à la toute fin de l'étude, il pourra donc s'écouler quelques semaines avant que vous ne le receviez.

Les données recueillies dans le cadre de l'étude seront totalement anonymisées et vous pourrez décider d'arrêter l'étude à tout moment. Votre participation ne pourra cependant être considérée comme effective que si vous réalisez l'intégralité des deux sessions de passation.

Pour pouvoir être candidate, candidat pour passer cette étude, il est nécessaire que vous soyez majeure, francophone, droitier, que vous ayez une vue normale ou corrigée à la normale et une audition normale ou corrigée à la normale, une perception correcte des couleurs (ne pas avoir une forme de daltonisme) et ne pas être épileptique. Pour participer rien de plus simple, cliquez (avec tous types de navigateurs Web, mis à part Safari) sur ce lien qui vous emmène directement dans l'expérience : [XXX](#)

Nous avons besoin de vous ! Merci beaucoup pour votre contribution à la recherche.

Annexe 10 : Consentement libre et éclairé de l'étude 1

Je suis Lina-Estelle LOUIS, doctorante en psychologie cognitive et je vous remercie de votre intérêt pour notre étude scientifique 100 % en ligne ! Elle concerne la modélisation de la variation de la charge mentale pour des tâches sollicitant différentes fonctions cognitives.

Informations sur l'étude

La charge mentale (fatigue/difficulté ressentie lorsque que l'on effectue une activité) a un impact sur les performances dans des exercices sollicitant des fonctions mentales comme la mémoire ou l'attention. Lorsque ces fonctions mentales sont détériorées, il est important de les ré-entraîner pour permettre aux personnes de retrouver un bon niveau de performance. Mais pour éviter que la charge mentale ressentie pendant le réentraînement ne soit trop forte, il est important de pouvoir la mesurer pour adapter le niveau de difficulté de l'entraînement. Ceci est l'objectif de notre étude.

5 tâches portant sur 5 fonctions mentales différentes ont été conçues. Chacune de ces tâches possède plusieurs niveaux de difficulté. À la fin de la réalisation de chaque niveau de difficulté d'une tâche, deux questionnaires sur la charge mentale ressentie vous seront proposés.

Nous souhaitons ainsi déterminer au minimum trois niveaux de charge mentale dans cinq tâches faisant appel à des fonctions mentales bien distinctes. Ainsi, dans cette étude, nous n'évaluons pas, à proprement parler, la performance, mais bien la charge mentale ressentie en effectuant les différentes tâches. Il sera néanmoins demandé de faire de votre mieux pour réaliser les tâches correctement de façon à ce que la mesure de la charge mentale soit réellement représentative.

Votre participation au projet

Si vous acceptez de participer à cette étude, vous serez dans une expérience pendant laquelle vous réaliserez cinq tâches. À la fin de chaque tâche, vous remplirez deux questionnaires sur votre charge mentale ressentie. Une fois l'étude terminée et si vous avez réalisé intégralement toutes les tâches, vous recevrez un bon cadeau d'une valeur de 15 euros.

Nous vous précisons que votre contribution à cette recherche est volontaire. À tout moment, vous pourrez vous retirer ou cesser votre participation. De plus, votre décision de participer, de refuser de participer, ou de cesser votre participation n'aura aucun effet sur vos notes, votre statut, vos relations futures avec le laboratoire, l'Université et l'entreprise qui mène cette étude expérimentale.

Enfin, en cas de refus de participation ou de cessation de participation, aucune justification n'est requise.

Concernant les données obtenues, elles seront traitées avec la confidentialité la plus stricte. Nous masquerons votre identité à l'aide d'un numéro aléatoire. De plus, aucun autre renseignement ne pourra révéler votre identité. Toutes les données seront gardées dans un endroit sécurisé et seuls le responsable scientifique, et les chercheurs y auront accès. Enfin, étant donné qu'il y a une impossibilité de faire correspondre l'identité des personnes aux données, il n'est pas possible de détruire les données pour un individu désigné.

La suite du projet

Les avantages attendus de cette étude sont d'obtenir une modélisation de la charge mentale pour différentes tâches sollicitant des fonctions cognitives spécifiques. Ainsi, d'obtenir une meilleure compréhension des facteurs qui vont influencer la charge mentale. Il sera ainsi possible d'adapter la difficulté d'un test sollicitant des fonctions cognitives de façon à maintenir un niveau optimum de charge mentale.

Cette recherche sera diffusée dans des colloques et elle sera publiée dans des actes de colloque et des articles de revues académiques. Les principaux résultats globaux de l'étude seront transmis aux participant.es sous la forme d'un petit compte rendu, mais aucun résultat individuel ne pourra être communiqué.

Vous pouvez poser des questions au sujet de la recherche en tout temps en communiquant avec le responsable scientifique du projet par courrier électronique via Lina-Estelle LOUIS (le.louis@groupeonepoint.com).

Votre adresse électronique :

Nom de famille :

Prénom :

[Case à cocher] J'ai lu et compris les renseignements ci-dessus et j'accepte de plein gré de participer à cette recherche.

Annexe 11 : Questionnaire pré-test de l'étude 1

Bonjour à vous, et merci de participer à notre première session d'étude en ligne. Je suis Lina-Estelle Louis et je vous accompagnerai pas à pas durant plusieurs tâches que vous allez réaliser aujourd'hui. Avant toute chose, je vous laisse répondre aux questions ci-après.

- La passation en ligne doit être faite sur n'importe quel navigateur SAUF Safari. Ainsi, êtes-vous sur un navigateur autre que Safari ? [Cocher pour oui]
 - Durant la passation, vous devez être dans une pièce calme où vous ne serez pas dérangé. Est-ce le cas ? [Cocher pour oui]
 - Combien d'heures avez-vous dormi la nuit passée ? [Réponse numérique]
 - Durant cette passation, vos haut-parleurs ou votre retour audio sur casque ou écouteur doivent-ils être actifs. Le sont-ils ? [Cocher pour oui]
 - Votre ordinateur doit être branché à une prise électrique. Est-ce bien le cas ?
 - Votre téléphone portable doit être en silencieux. Est-ce bien le cas ? [Cocher pour oui]
 - Cette page Web doit-elle être en plein écran. Est-ce le cas ? [Cocher pour oui]
 - Vous devez être assis, assise sur un bureau ou une table. Est-ce le cas ? [Cocher pour oui]
 - En cliquant sur ce bouton, m'entendez-vous ? [Mettre un bouton ayant le mot "Écouter" dessus et l'enregistrement de ma voix "m'entendez-vous ?"] [Cocher pour oui]
 - Le test en cours doit être le seul onglet ouvert sur votre navigateur Web. Est-ce le cas ? [Cocher pour oui]
 - À l'instant T, quel est votre niveau de fatigue ? [Échelle : grande fatigue, fatigue moyenne, fatigue basse, aucune fatigue].
 - Sexe [femme, homme, autre (à préciser)]
 - Votre âge [réponse alphanumérique]
 - Votre niveau d'étude maximal [collège, lycée, Bac + 2, Bac + 3, Bac + 5, au-delà].
 - Langue maternelle [réponse libre]
 - Votre niveau de Français ? [Bas, moyen, haut, langue courante]
 - A quelle fréquence jouez-vous aux jeux-vidéo ? [Jamais, très peu, un peu, souvent]
 - A quelle fréquence jouez-vous aux jeux en ligne ? [Jamais, très peu, un peu, souvent]
- Merci pour vos retours. Maintenant, nous allons passer aux tâches expérimentales. C'est parti ! Appuyez sur le bouton Valider pour commencer.

Consigne générale session 2 :

Bonjour à vous, et merci de participer à notre deuxième session d'étude. Vous serez toujours guidé par moi, Lina-Estelle Louis où je vous accompagnerai pas à pas durant les derniers jeux d'aujourd'hui. Mais avant, je vous laisse répondre tranquillement à ces questions ci-après.

- La passation en ligne doit être faite sur n'importe quel navigateur SAUF Safari. Ainsi, êtes-vous sur un navigateur autre que Safari ? [Cocher pour oui]
 - Durant la passation, vous devez être dans une pièce calme où vous ne serez pas dérangé. Est-ce le cas ? [Cocher pour oui]
 - Combien d'heures avez-vous dormi la nuit passée ? [Réponse numérique]
 - Durant cette passation, vos haut-parleurs ou votre retour audio sur casque ou écouteur doivent-ils être actifs. Le sont-ils ? [Cocher pour oui]
 - Votre ordinateur doit être branché à une prise électrique. Est-ce bien le cas ?
 - Votre portable doit être en silencieux. Est-ce bien le cas ? [Cocher pour oui]
 - Cette page Web doit-elle être en plein écran. Est-ce le cas ? [Cocher pour oui]
 - Vous devez être assis sur un bureau ou une table. Est-ce le cas ? [Cocher pour oui]
 - En cliquant sur ce bouton, m'entendez-vous ? [Cocher pour oui]
 - Le test en cours doit être le seul onglet ouvert sur votre navigateur Web. Est-ce le cas ? [Cocher pour oui]
 - À l'instant T, quel est votre niveau de fatigue ? [Échelle : grande fatigue, fatigue moyenne, fatigue basse, aucune fatigue].
- Merci pour vos retours. Maintenant, nous allons passer aux dernières tâches expérimentales à distance. C'est parti ! Appuyez sur le bouton Valider pour commencer.**

Annexe 12 : Questionnaire post-test

Questionnaire post-test après chaque niveau d'une tâche :

- Durant ce niveau, avez-vous été dérangé(e) ou distrait(e) ? [Oui/non]
- Suivi de la passation du NASA-TLX et des quatre items du WP
- Dans quelle mesure la tâche que vous venez de faire a été ludique, amusante ?

Situation post-test final :

- Dans le passé, avez-vous déjà réalisé cette tâche expérimentale ? [Oui/non]
- Si oui, depuis combien de temps ? [Moins d'1 an/plus d'un an]

Bravo ! Vous avez terminé l'ensemble des tâches de la session d'aujourd'hui. Avez-vous des retours, des remarques sur la session d'aujourd'hui ? [Champ libre de texte].

Je me joins à toute l'équipe pour vous remercier de votre participation.

Au plaisir d'échanger, Lina-Estelle LOUIS

Annexe 13 : Tableaux de résultats supplémentaires de l'étude 1

Réponses attendues

Friedman Test

A

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	16.735	4	0.002	0.030
Complexity levels	145.255	2	< .001	0.359

B

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	2.887	740	2829.500	2095.000	0.004
Level 1 Level 3	5.479	740	2829.500	1435.500	< .001
Level 2 Level 3	2.592	740	2095.000	1435.500	0.010

Note. Grouped by subject.

C

Measure 1	Measure 2	W	df	p
NBT1	- NBT2	575.000		< .001
NBT2	- NBT3	765.000		0.002
NBT1	- NBT3	932.500		< .001
Corsi1	- Corsi2	123.500		0.027
Corsi2	- Corsi3	955.000		< .001
Corsi1	- Corsi3	988.000		< .001
GNG1	- GNG2	581.000		< .001
GNG2	- GNG3	561.000		0.086
GNG1	- GNG3	984.000		< .001
WCST1	- WCST2	871.500		< .001
WCST2	- WCST3	967.000		0.004
WCST1	- WCST3	1205.000		< .001
DT1	- DT2	285.000		< .001
DT2	- DT3	460.000		0.102
DT1	- DT3	541.500		< .001

Note. Wilcoxon signed-rank test.

Paired Samples T-Test

D

Measure 1		Measure 2	W	df	p
NBT1	-	Corsi1	70.000		0.004
NBT1	-	GNG1	122.500		0.023
NBT1	-	WCST1	464.500		0.468
NBT1	-	DT1	80.000		0.005
Corsi1	-	GNG1	66.000		0.147
Corsi1	-	WCST1	384.000		0.002
Corsi1	-	DT1	19.000		0.944
GNG1	-	WCST1	477.000		0.002
GNG1	-	DT1	42.000		0.312
WCST1	-	DT1	107.000		0.003
NBT2	-	Corsi2	389.000		0.216
NBT2	-	GNG2	760.000		0.143
NBT2	-	WCST2	927.000		0.062
NBT2	-	DT2	576.000		0.346
Corsi2	-	GNG2	577.500		0.055
Corsi2	-	GNG2	577.500		0.055
Corsi2	-	WCST2	883.500		0.018
Corsi2	-	DT2	331.500		0.042
GNG2	-	WCST2	645.500		0.538
GNG2	-	DT2	530.500		0.682
WCST2	-	DT2	631.000		0.858
NBT3	-	Corsi3	1072.000		0.002
NBT3	-	GNG3	519.000		0.991
NBT3	-	WCST3	835.500		0.290
NBT3	-	DT3	730.000		0.533
Corsi3	-	GNG3	324.500		< .001
Corsi3	-	WCST3	468.500		0.045
Corsi3	-	DT3	365.000		0.036
GNG3	-	WCST3	777.500		0.423
GNG3	-	DT3	710.500		0.332
WCST3	-	DT3	656.500		0.771

Note. Wilcoxon signed-rank test.

Table 1 : Performance en termes de *Réponses attendues*. A) ANOVAs non paramétriques à mesures répétées. B) Indépendamment de la tâche, comparaisons post hoc de Conover pour les trois niveaux de complexité. C) Tâche par tâche, comparaisons post-hoc de Wilcoxon entre les niveaux de complexité. D) Niveau par niveau, comparaison post-hoc de Wilcoxon entre les tâches.

NASA-TLX global

Friedman Test

A

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	33.605	4	< .001	0.099
Complexity levels	179.626	2	< .001	0.572

B

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	3.554	740	1272.500	2187.000	< .001
Level 3	6.327	740	1272.500	2900.500	< .001
Level 2 Level 3	2.773	740	2187.000	2900.500	0.006

Note. Grouped by subject.

C

Measure 1		Measure 2	W	df	p
NBT1	-	NBT2	300.000		< .001
NBT2	-	NBT3	248.500		< .001
NBT1	-	NBT3	40.000		< .001
Corsi1	-	Corsi2	39.000		< .001
Corsi2	-	Corsi3	100.000		< .001
Corsi1	-	Corsi3	0.000		< .001
GNG1	-	GNG2	151.000		< .001
GNG2	-	GNG3	320.500		0.001
GNG1	-	GNG3	22.500		< .001
WCST1	-	WCST2	196.000		< .001
WCST2	-	WCST3	461.000		0.025
WCST1	-	WCST3	174.000		< .001
DT1	-	DT2	46.000		< .001
DT2	-	DT3	172.500		< .001
DT1	-	DT3	3.000		< .001

Note. Wilcoxon signed-rank test.

Measure 1		Measure 2	W	df	p D
NBT1	-	Corsi1	1203.000		< .001
NBT1	-	GNG1	766.000		0.658
NBT1	-	WCST1	820.000		0.357
NBT1	-	DT1	793.000		0.495
Corsi1	-	GNG1	270.000		< .001
Corsi1	-	WCST1	283.500		< .001
Corsi1	-	DT1	169.000		< .001
GNG1	-	WCST1	795.000		0.484
GNG1	-	DT1	638.000		0.495
WCST1	-	DT1	592.000		0.276
NBT2	-	Corsi2	1106.500		< .001
NBT2	-	GNG2	769.500		0.636
NBT2	-	WCST2	889.000		0.126
NBT2	-	DT2	635.000		0.479
Corsi2	-	GNG2	401.500		0.006
Corsi2	-	WCST2	500.000		0.057
Corsi2	-	DT2	206.500		< .001
GNG2	-	WCST2	832.000		0.304
GNG2	-	DT2	624.500		0.423
WCST2	-	DT2	446.000		0.017
NBT3	-	Corsi3	1012.000		0.009
NBT3	-	GNG3	848.000		0.149
NBT3	-	WCST3	1114.500		< .001
NBT3	-	DT3	828.500		0.319
Corsi3	-	GNG3	611.000		0.480
Corsi3	-	WCST3	1024.500		0.006
Corsi3	-	DT3	525.500		0.093
GNG3	-	WCST3	1058.500		0.002
GNG3	-	DT3	617.500		0.388
WCST3	-	DT3	250.000		< .001

Note. Wilcoxon signed-rank test.

Table 2 : NASA-TLX global. A) ANOVAs non paramétriques à mesures répétées. B) Indépendamment de la tâche, comparaisons post hoc de Conover pour les trois niveaux de complexité. C) Tâche par tâche, comparaisons post-hoc de Wilcoxon entre les niveaux de complexité. D) Niveau par niveau, comparaison post-hoc de Wilcoxon entre les tâches.

Exigences mentales

A

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	30.873	4	< .001	0.080
Complexity levels	164.416	2	< .001	0.477

B

Conover's Post Hoc Comparisons – Complexity levels

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	3.433	740	1309.000	2191.500	< .001
Level 3	6.031	740	1309.000	2859.500	< .001
Level 2 Level 3	2.598	740	2191.500	2859.500	0.010

Note. Grouped by subject.

C

Paired Samples T-Test

Measure 1		Measure 2	W	df	p
NBT1	-	NBT2	306.000		< .001
NBT2	-	NBT3	99.500		< .001
NBT1	-	NBT3	50.000		< .001
Corsi1	-	Corsi2	107.500		< .001
Corsi2	-	Corsi3	140.500		< .001
Corsi1	-	Corsi3	33.000		< .001
GNG1	-	GNG2	185.000		< .001
GNG2	-	GNG3	191.500		< .001
GNG1	-	GNG3	30.000		< .001
WCST1	-	WCST2	153.500		< .001
WCST2	-	WCST3	546.000		0.135
WCST1	-	WCST3	178.000		< .001
DT1	-	DT2	61.000		< .001
DT2	-	DT3	166.500		< .001
DT1	-	DT3	10.500		< .001

Note. Wilcoxon signed-rank test.

Paired Samples T-Test ▼

D

Measure 1		Measure 2	W	df	p
NBT1	-	Corsi1	947.000		0.003
NBT1	-	GNG1	671.000		0.697
NBT1	-	WCST1	465.000		0.097
NBT1	-	DT1	489.000		0.045
Corsi1	-	GNG1	362.000		0.002
Corsi1	-	WCST1	225.500		< .001
Corsi1	-	DT1	175.000		< .001
GNG1	-	WCST1	563.000		0.178
GNG1	-	DT1	434.000		0.020
WCST1	-	DT1	603.000		0.321
NBT2	-	Corsi2	667.000		0.591
NBT2	-	GNG2	608.000		0.841
NBT2	-	WCST2	492.000		0.048
NBT2	-	DT2	301.500		< .001
Corsi2	-	GNG2	503.000		0.278
Corsi2	-	WCST2	424.000		0.010
Corsi2	-	DT2	177.000		< .001
GNG2	-	WCST2	541.000		0.123
GNG2	-	DT2	383.500		0.003
WCST2	-	DT2	464.000		0.026
NBT3	-	Corsi3	764.500		0.343
NBT3	-	GNG3	748.000		0.288
NBT3	-	WCST3	824.000		0.016
NBT3	-	DT3	431.500		0.072
Corsi3	-	GNG3	657.000		0.662
Corsi3	-	WCST3	951.000		0.037
Corsi3	-	DT3	321.000		0.006
GNG3	-	WCST3	762.500		0.137
GNG3	-	DT3	389.500		0.027
WCST3	-	DT3	223.000		< .001

Note. Wilcoxon signed-rank test.

Exigences physiques

A

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	8.776	4	0.067	0.016
Complexity levels	14.769	2	< .001	0.061

B

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	0.884	740	1891.500	2118.000	0.377
Level 1 Level 3	1.792	740	1891.500	2350.500	0.074
Level 2 Level 3	0.908	740	2118.000	2350.500	0.364

Note. Grouped by subject.

C

Measure 1	Measure 2	W	df	p
NBT1	- NBT2	698.000		0.880
NBT2	- NBT3	226.000		0.022
NBT1	- NBT3	561.000		0.173
Corsi1	- Corsi2	411.000		0.007
Corsi2	- Corsi3	352.000		1.000
Corsi1	- Corsi3	370.000		0.002
GNG1	- GNG2	197.000		0.020
GNG2	- GNG3	271.500		0.153
GNG1	- GNG3	125.000		0.002
WCST1	- WCST2	361.000		0.260
WCST2	- WCST3	455.000		0.021
WCST1	- WCST3	402.000		0.006
DT1	- DT2	352.000		0.216
DT2	- DT3	346.000		0.126
DT1	- DT3	189.000		0.001

Note. Wilcoxon signed-rank test.

Exigences temporelles

Friedman Test

A

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	51.179	4	< .001	0.093
Complexity levels	49.422	2	< .001	0.159

B

		T-Stat	df	W _i	W _j	p
Level 1	Level 2	2.008	740	1666.500	2182.500	0.045
	Level 3	3.286	740	1666.500	2511.000	0.001
Level 2	Level 3	1.278	740	2182.500	2511.000	0.202

Note. Grouped by subject.

C

Measure 1		Measure 2	W	df	p
NBT1	-	NBT2	430.000		0.012
NBT2	-	NBT3	303.000		0.004
NBT1	-	NBT3	186.000		< .001
Corsi1	-	Corsi2	232.000		< .001
Corsi2	-	Corsi3	322.000		0.004
Corsi1	-	Corsi3	181.000		< .001
GNG1	-	GNG2	191.500		< .001
GNG2	-	GNG3	419.000		0.689
GNG1	-	GNG3	178.000		< .001
WCST1	-	WCST2	322.000		0.004
WCST2	-	WCST3	457.000		0.035
WCST1	-	WCST3	324.000		< .001
DT1	-	DT2	251.500		0.003
DT2	-	DT3	383.500		0.023
DT1	-	DT3	212.500		< .001

Note. Wilcoxon signed-rank test.

Paired Samples T-Test ▼

D

Measure 1		Measure 2	W	df	p
NBT1	-	Corsi1	907.500		< .001
NBT1	-	GNG1	634.000		0.473
NBT1	-	WCST1	838.000		0.102
NBT1	-	DT1	904.000		0.096
Corsi1	-	GNG1	258.000		< .001
Corsi1	-	WCST1	397.000		0.005
Corsi1	-	DT1	318.000		< .001
GNG1	-	WCST1	968.000		0.026
GNG1	-	DT1	758.500		0.081
WCST1	-	DT1	722.000		0.958
NBT2	-	Corsi2	1046.000		0.001
NBT2	-	GNG2	513.500		0.233
NBT2	-	WCST2	873.000		0.165
NBT2	-	DT2	957.500		0.032
Corsi2	-	GNG2	214.500		< .001
Corsi2	-	WCST2	547.000		0.137
Corsi2	-	DT2	395.500		0.012
GNG2	-	WCST2	942.000		0.045
GNG2	-	DT2	954.000		0.006
WCST2	-	DT2	746.000		0.791
NBT3	-	Corsi3	956.500		< .001
NBT3	-	GNG3	725.000		0.401
NBT3	-	WCST3	918.000		0.002
NBT3	-	DT3	1043.000		0.001
Corsi3	-	GNG3	386.500		0.010
Corsi3	-	WCST3	719.500		0.975
Corsi3	-	DT3	582.000		0.595
GNG3	-	WCST3	853.500		0.017
GNG3	-	DT3	965.500		0.005
WCST3	-	DT3	603.000		0.577

Note. Wilcoxon signed-rank test.

Effort

Friedman Test

A

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	37.273	4	< .001	0.097
Complexity levels	143.925	2	< .001	0.463

B

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	3.268	740	1358.500	2198.000	0.001
Level 3	5.624	740	1358.500	2803.500	< .001
Level 2 Level 3	2.357	740	2198.000	2803.500	0.019

Note. Grouped by subject.

C

Measure 1	Measure 2	W	df	p
NBT1	- NBT2	404.000		0.006
NBT2	- NBT3	161.000		< .001
NBT1	- NBT3	67.000		< .001
Corsi1	- Corsi2	56.000		< .001
Corsi2	- Corsi3	196.500		< .001
Corsi1	- Corsi3	34.000		< .001
GNG1	- GNG2	121.500		< .001
GNG2	- GNG3	307.000		0.001
GNG1	- GNG3	41.500		< .001
WCST1	- WCST2	239.000		< .001
WCST2	- WCST3	518.000		0.081
WCST1	- WCST3	212.000		< .001
DT1	- DT2	132.500		< .001
DT2	- DT3	147.500		< .001
DT1	- DT3	52.000		< .001

Note. Wilcoxon signed-rank test.

D

Paired Samples T-Test

Measure 1		Measure 2	W	df	p
NBT1	-	Corsi1	918.500		< .001
NBT1	-	GNG1	792.000		0.351
NBT1	-	WCST1	652.500		0.223
NBT1	-	DT1	561.500		0.174
Corsi1	-	GNG1	390.000		0.004
Corsi1	-	WCST1	287.000		0.001
Corsi1	-	DT1	147.000		< .001
GNG1	-	WCST1	665.500		0.661
GNG1	-	DT1	314.000		0.001
WCST1	-	DT1	389.000		0.004
NBT2	-	Corsi2	712.500		0.472
NBT2	-	GNG2	478.000		0.660
NBT2	-	WCST2	713.000		0.986
NBT2	-	DT2	364.000		0.014
Corsi2	-	GNG2	493.500		0.238
Corsi2	-	WCST2	654.000		0.589
Corsi2	-	DT2	287.500		< .001
GNG2	-	WCST2	771.000		0.626
GNG2	-	DT2	468.500		0.069
WCST2	-	DT2	352.000		0.001
NBT3	-	Corsi3	789.000		0.145
NBT3	-	GNG3	789.500		0.144
NBT3	-	WCST3	970.000		0.001
NBT3	-	DT3	453.500		0.169
Corsi3	-	GNG3	681.000		0.678
Corsi3	-	WCST3	785.000		0.008
Corsi3	-	DT3	388.000		0.006
GNG3	-	WCST3	781.000		0.022
GNG3	-	DT3	400.000		0.014
WCST3	-	DT3	166.000		< .001

Note. Wilcoxon signed-rank test.

Performance

A

Friedman Test ▼

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	31.676	4	< .001	0.064
Complexity levels	137.616	2	< .001	0.356

B

		T-Stat	df	W _i	W _j	p
Level 1	Level 2	2.821	740	1409.500	2133.000	0.005
	Level 3	5.490	740	1409.500	2817.500	< .001
Level 2	Level 3	2.669	740	2133.000	2817.500	0.008

Note. Grouped by subject.

C

Measure 1		Measure 2	W	df	p
NBT1	-	NBT2	207.000		< .001
NBT2	-	NBT3	363.500		0.008
NBT1	-	NBT3	78.500		< .001
Corsi1	-	Corsi2	31.500		< .001
Corsi2	-	Corsi3	58.000		< .001
Corsi1	-	Corsi3	0.000		< .001
GNG1	-	GNG2	187.000		0.003
GNG2	-	GNG3	357.000		0.011
GNG1	-	GNG3	131.000		< .001
WCST1	-	WCST2	272.500		0.010
WCST2	-	WCST3	409.500		0.028
WCST1	-	WCST3	228.500		< .001
DT1	-	DT2	77.500		< .001
DT2	-	DT3	299.000		0.001
DT1	-	DT3	48.000		< .001

Note. Wilcoxon signed-rank test.

Paired Samples T-Test ▼

D

Measure 1		Measure 2	W	df	p
NBT1	-	Corsi1	721.000		< .001
NBT1	-	GNG1	665.000		0.097
NBT1	-	WCST1	594.500		0.144
NBT1	-	DT1	683.000		0.121
Corsi1	-	GNG1	163.000		0.003
Corsi1	-	WCST1	154.500		< .001
Corsi1	-	DT1	100.000		< .001
GNG2	-	WCST2	763.000		0.228
GNG2	-	DT2	502.500		0.276
WCST2	-	DT2	353.000		0.002
NBT3	-	Corsi3	724.000		0.164
NBT3	-	GNG3	806.500		0.104
NBT3	-	WCST3	1096.500		< .001
NBT3	-	DT3	774.000		0.109
Corsi3	-	GNG3	678.500		0.515
Corsi3	-	WCST3	1053.500		< .001
Corsi3	-	DT3	673.000		0.929
GNG3	-	WCST3	986.500		0.007
GNG3	-	DT3	616.000		0.663
WCST3	-	DT3	237.500		< .001
GNG1	-	WCST1	534.500		0.759
GNG1	-	DT1	549.500		0.534
WCST1	-	DT1	625.000		0.905
NBT2	-	Corsi2	1015.500		< .001
NBT2	-	GNG2	741.500		0.201
NBT2	-	WCST2	1003.500		< .001
NBT2	-	DT2	776.000		0.595
Corsi2	-	GNG2	332.000		0.009
Corsi2	-	WCST2	394.000		0.111
Corsi2	-	DT2	208.000		< .001

Note. Wilcoxon signed-rank test.

Frustration

Friedman Test

A

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	19.209	4	< .001	0.036
Complexity levels	81.364	2	< .001	0.247

B

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	2.472	740	1557.500	2189.000	0.014
Level 3	4.134	740	1557.500	2613.500	< .001
Level 2 Level 3	1.662	740	2189.000	2613.500	0.097

Note. Grouped by subject.

C

Measure 1	Measure 2	W	df	p
NBT1	- NBT2	290.500		0.011
NBT2	- NBT3	426.000		0.098
NBT1	- NBT3	130.500		< .001
Corsi1	- Corsi2	159.000		0.004
Corsi2	- Corsi3	105.500		< .001
Corsi1	- Corsi3	44.000		< .001
GNG1	- GNG2	160.000		< .001
GNG2	- GNG3	413.500		0.112
GNG1	- GNG3	91.500		< .001
WCST1	- WCST2	253.000		0.008
WCST2	- WCST3	488.000		0.424
WCST1	- WCST3	322.500		0.011
DT1	- DT2	97.500		< .001
DT2	- DT3	262.500		0.030
DT1	- DT3	51.000		< .001

Note. Wilcoxon signed-rank test.

D

Paired Samples T-Test

Measure 1		Measure 2	W	df	p
NBT1	-	Corsi1	675.500		0.005
NBT1	-	GNG1	494.000		0.414
NBT1	-	WCST1	608.000		0.310
NBT1	-	DT1	631.000		0.202
Corsi1	-	GNG1	305.000		0.105
Corsi1	-	WCST1	317.500		0.145
Corsi1	-	DT1	263.000		0.049
GNG1	-	WCST1	537.500		0.624
WCST1	-	DT1	458.000		0.505
NBT2	-	Corsi2	812.500		0.003
NBT2	-	GNG2	581.000		0.662
NBT2	-	WCST2	741.500		0.061
NBT2	-	DT2	664.500		0.798
Corsi2	-	GNG2	289.000		0.006
Corsi2	-	WCST2	459.500		0.189
Corsi2	-	DT2	266.000		0.003
GNG2	-	WCST2	778.500		0.100
GNG2	-	DT2	565.000		0.793
WCST2	-	DT2	474.000		0.344
NBT3	-	Corsi3	701.000		0.381
NBT3	-	GNG3	637.500		0.440
NBT3	-	WCST3	930.500		0.002
NBT3	-	DT3	731.500		0.238
Corsi3	-	GNG3	606.500		0.956
Corsi3	-	WCST3	864.000		0.005
Corsi3	-	DT3	610.000		0.630
GNG3	-	WCST3	905.500		0.010
GNG3	-	DT3	666.500		0.595
WCST3	-	DT3	275.500		0.011

Note. Wilcoxon signed-rank test.

Table 3 : Dimensions du NASA-TLX. A) ANOVAs non paramétriques à mesures répétées. B) Indépendamment de la tâche, comparaisons post hoc de Conover pour les trois niveaux de complexité. C) Tâche par tâche, comparaisons post-hoc de Wilcoxon entre les niveaux de complexité. D) Niveau par niveau, comparaison post-hoc de Wilcoxon entre les tâches.

WP3

A

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	86.422	4	< .001	0.150
Complexity levels	27.797	2	< .001	0.109

B

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	1.446	740	1787.500	2158.000	0.149
Level 2 Level 3	2.447	740	1787.500	2414.500	0.015
Level 1 Level 3	1.001	740	2158.000	2414.500	0.317

Note. Grouped by subject.

C

	T-Stat	df	W _i	W _j	p
NBT Corsi	3.035	738	914.000	1456.500	0.002
NBT GNG	1.055	738	914.000	1102.500	0.292
NBT WCST	2.000	738	914.000	1271.500	0.046
NBT DT	3.925	738	914.000	1615.500	< .001
Corsi GNG	1.981	738	1456.500	1102.500	0.048
Corsi WCST	1.035	738	1456.500	1271.500	0.301
Corsi DT	0.890	738	1456.500	1615.500	0.374
GNG WCST	0.946	738	1102.500	1271.500	0.345
GNG DT	2.870	738	1102.500	1615.500	0.004
WCST DT	1.925	738	1271.500	1615.500	0.055

Note. Grouped by subject.

D

Measure 1	Measure 2	W	df	p
NBT1	- NBT2	233.000		0.047
NBT2	- NBT3	298.000		0.296
NBT1	- NBT3	223.500		0.033
Corsi1	- Corsi2	91.000		< .001
Corsi2	- Corsi3	227.000		0.003
Corsi1	- Corsi3	49.500		< .001
GNG1	- GNG2	341.000		0.169
GNG2	- GNG3	274.000		0.010
WCST1	- WCST2	414.500		0.350
WCST2	- WCST3	366.000		0.134
WCST1	- WCST3	330.500		0.035
DT1	- DT2	387.500		0.212
DT2	- DT3	338.500		0.340
DT1	- DT3	392.500		0.045

Note. Wilcoxon signed-rank test.

E

Measure 1	Measure 2	W	df	p
Corsi1	- DT1	312.500		< .001

WP4

A

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	55.564	4	< .001	0.092
Complexity levels	13.809	2	0.001	0.059

B

		T-Stat	df	W _i	W _j	p
Level 1	Level 2	0.814	740	1906.500	2114.000	0.416
	Level 3	1.699	740	1906.500	2339.500	0.090
Level 2	Level 3	0.885	740	2114.000	2339.500	0.376

Note. Grouped by subject.

C

		T-Stat	df	W _i	W _j	p
NBT	Corsi	2.164	738	1360.500	976.000	0.031
	GNG	1.083	738	1360.500	1168.000	0.279
	WCST	1.111	738	1360.500	1558.000	0.267
	DT	0.355	738	1360.500	1297.500	0.723
Corsi	GNG	1.080	738	976.000	1168.000	0.280
	WCST	3.275	738	976.000	1558.000	0.001
	DT	1.809	738	976.000	1297.500	0.071
GNG	WCST	2.195	738	1168.000	1558.000	0.029
	DT	0.729	738	1168.000	1297.500	0.466
WCST	DT	1.466	738	1558.000	1297.500	0.143

Note. Grouped by subject.

D

Measure 1		Measure 2	W	df	p
NBT1	-	NBT2	325.500		0.372
NBT2	-	NBT3	99.500		< .001
NBT1	-	NBT3	227.500		0.003
Corsi1	-	Corsi2	197.000		0.471
Corsi2	-	Corsi3	245.500		0.378
Corsi1	-	Corsi3	200.000		0.352
GNG1	-	GNG2	147.500		0.082
GNG2	-	GNG3	251.500		0.203
GNG1	-	GNG3	78.000		< .001
WCST1	-	WCST2	256.500		0.015
WCST2	-	WCST3	198.500		< .001
WCST1	-	WCST3	165.500		< .001
DT1	-	DT2	226.000		0.147
DT2	-	DT3	377.500		0.867
DT1	-	DT3	292.000		0.174

Note. Wilcoxon signed-rank test.

WP5

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	13.893	4	0.008	0.026
Complexity levels	43.682	2	< .001	0.165

A

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	1.818	740	1704.000	2169.500	0.069
Level 2 Level 3	3.056	740	1704.000	2486.500	0.002
Level 1 Level 3	1.238	740	2169.500	2486.500	0.216

B

Note. Grouped by subject.

	T-Stat	df	W _i	W _j	p
NBT Corsi	0.199	738	1146.500	1182.000	0.842
NBT GNG	1.501	738	1146.500	1414.500	0.134
NBT WCST	1.103	738	1146.500	1343.500	0.270
NBT DT	0.711	738	1146.500	1273.500	0.477
Corsi GNG	1.302	738	1182.000	1414.500	0.193
Corsi WCST	0.904	738	1182.000	1343.500	0.366
Corsi DT	0.512	738	1182.000	1273.500	0.609
GNG WCST	0.398	738	1414.500	1343.500	0.691
GNG DT	0.790	738	1414.500	1273.500	0.430
WCST DT	0.392	738	1343.500	1273.500	0.695

C

Note. Grouped by subject.

Measure 1	Measure 2	W	df	p
NBT1	- NBT2	343.500		0.050
NBT2	- NBT3	219.500		0.004
NBT1	- NBT3	206.000		< .001
Corsi1	- Corsi2	99.000		< .001
Corsi2	- Corsi3	185.500		0.002
Corsi1	- Corsi3	44.000		< .001
GNG1	- GNG2	350.500		0.060
GNG2	- GNG3	345.500		0.053
GNG1	- GNG3	233.500		< .001
WCST1	- WCST2	404.000		0.060
WCST2	- WCST3	371.000		0.220
WCST1	- WCST3	270.000		0.005
DT1	- DT2	235.500		0.001
DT2	- DT3	418.500		0.266
DT1	- DT3	224.000		< .001

D

Note. Wilcoxon signed-rank test.

E

Measure 1	Measure 2	W	df	p
Corsi1	- DT1	393.500		0.012

WP7 **A**

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
Cognitive tasks	5.944	4	0.203	0.009
Complexity levels	15.296	2	< .001	0.055

B

	T-Stat	df	W _i	W _j	p
Level 1 Level 2	1.101	740	1871.500	2153.500	0.271
Level 3	1.809	740	1871.500	2335.000	0.071
Level 2 Level 3	0.708	740	2153.500	2335.000	0.479

Note. Grouped by subject.

D

Measure 1	Measure 2	W	df	p
NBT1	- NBT2	329.500		0.129
NBT2	- NBT3	467.000		0.748
NBT1	- NBT3	315.000		0.014
Corsi1	- Corsi2	243.000		0.009
Corsi2	- Corsi3	427.000		0.217
Corsi1	- Corsi3	94.000		< .001
GNG1	- GNG2	278.000		0.077
GNG2	- GNG3	363.500		0.274
GNG1	- GNG3	173.000		0.001
WCST1	- WCST2	463.500		0.203
WCST2	- WCST3	323.500		0.111
WCST1	- WCST3	258.500		0.004
DT1	- DT2	354.500		0.154
DT2	- DT3	348.000		0.198
DT1	- DT3	402.000		0.057

Note. Wilcoxon signed-rank test.

Table 4 : Dimensions du WP. A) ANOVAs non paramétriques à mesures répétées. B) Indépendamment de la tâche, comparaisons post hoc de Conover pour les trois niveaux de complexité. C) Indépendamment du niveau de complexité, comparaisons post hoc de Conover pour les cinq tâches. D) Niveau par niveau, comparaison post-hoc de Wilcoxon entre les tâches.

Ludique

Nonparametrics ▼

Friedman Test

Factor	Chi-Squared	df	p	Kendall's W
CTs	161.566	4	< .001	0.307
Levels	0.523	2	0.770	6.995e-4

Conover Test ▼

Conover's Post Hoc Comparisons - CTs

		T-Stat	df	W _i	W _j	p	P _{bonf}	P _{holm}
NBT	Corsi	3.189	738	775.000	1347.000	0.001	0.015	0.010
	GNG	1.327	738	775.000	1013.000	0.185	1.000	0.370
	WCST	5.492	738	775.000	1760.000	< .001	< .001	< .001
	DT	3.847	738	775.000	1465.000	< .001	0.001	0.001
Corsi	GNG	1.862	738	1347.000	1013.000	0.063	0.630	0.252
	WCST	2.303	738	1347.000	1760.000	0.022	0.216	0.108
	DT	0.658	738	1347.000	1465.000	0.511	1.000	0.511
GNG	WCST	4.165	738	1013.000	1760.000	< .001	< .001	< .001
	DT	2.520	738	1013.000	1465.000	0.012	0.119	0.072
WCST	DT	1.645	738	1760.000	1465.000	0.100	1.000	0.301

Note. Grouped by subject.

Conover's Post Hoc Comparisons - Levels

		T-Stat	df	W _i	W _j	p	P _{bonf}	P _{holm}
Level 1	Level 2	0.309	740	2091.000	2170.500	0.757	1.000	1.000
	Level 3	0.029	740	2091.000	2098.500	0.977	1.000	1.000
Level 2	Level 3	0.280	740	2170.500	2098.500	0.780	1.000	1.000

Note. Grouped by subject.

Table 5 : Dimension Ludique. A) ANOVAs non paramétriques à mesures répétées. B) Indépendamment de la tâche, comparaisons post hoc de Conover pour les trois niveaux de complexité. C) Indépendamment du niveau de complexité, comparaisons post hoc de Conover pour les cinq tâches. D) Niveau par niveau, comparaison post-hoc de Wilcoxon entre les tâches.

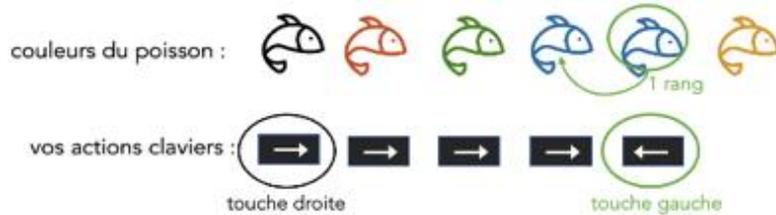
Annexe 14 : Les cinq consignes des tâches cognitives de l'étude 2

CONSIGNE DU NBT LUDIQUE

Consigne générale :

Dans ce jeu, vous allez voir un poisson sauter de bouée en bouée. À chaque saut, vous devez comparer la couleur du poisson à sa couleur précédente.

Si la couleur du poisson est identique à la précédente, vous devez appuyer sur la touche directionnelle gauche du clavier. Pour tous les autres cas, vous devez appuyer sur la touche directionnelle de droite.



Durant le jeu, c'est normal de ne pas avoir de retours/feedbacks de ses actions. Regardez cet exemple vidéo. [Vidéo N-back ludique N=1]

Entraînement :

On commence par un petit entraînement.

Vous devez comparer la couleur du poisson à sa couleur précédente.

Si la couleur du poisson est identique à la précédente, vous devez appuyer sur la touche directionnelle gauche du clavier. Pour tous les autres cas, vous devez appuyer sur la touche directionnelle de droite.

Une fois que vous avez bien compris la consigne d'entraînement, appuyez sur la touche directionnelle droite de votre clavier pour démarrer.

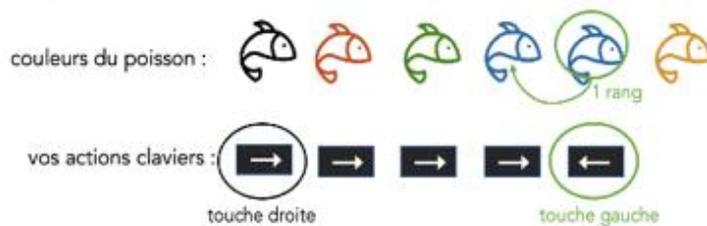
Après la phase d'entraînement

La phase d'entraînement est terminée. Maintenant, le jeu commence réellement. Je vous rappelle la consigne.

Vous devez comparer la couleur du poisson à sa couleur précédente.

Si la couleur du poisson est identique à la précédente, vous devez appuyer sur la touche directionnelle gauche du clavier. Pour tous les autres cas, vous devez appuyer sur la touche directionnelle de droite.

Pour rappel :



Une fois que vous avez bien compris la consigne, appuyez sur la touche directionnelle droite de votre clavier pour démarrer.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir, bouger avant d'être fixe au lancement du jeu.

Après la fin de la passation du N-back ludique

Bravo ! Vous avez terminé ce jeu.

CONSIGNE DU CORSI UNICOLORE

Dans ce jeu, je vais lancer le ballon dans une séquence de cinq bouées. Lorsque j'aurai fini, votre objectif sera de reproduire la séquence en lançant le ballon dans le même ordre que moi. Une fois que vous avez fini de répondre, cliquez sur le buzzer rouge TERMINER.

Regardez cet exemple vidéo.

[Vidéo d'exemple Corsi ludique unicolore avec bouées 2-6]

Entraînement :

On commence par un petit entraînement.

Lancez le ballon dans le même ordre que moi. Une fois que vous avez bien compris la consigne d'entraînement, cliquez sur la touche valider.

Après la phase d'entraînement

La phase d'entraînement est terminée. Maintenant, le jeu commence réellement. Je vous rappelle la consigne. Lancez le ballon dans le même ordre que moi.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir et bouger avant d'être fixe au lancement du jeu.

Avant chaque niveau de jeu

Répétez cet enchaînement dans le même ordre que moi.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir et bouger avant d'être fixe au lancement du jeu.

Après la fin de la passation du Corsi ludique unicolore

Bravo ! Vous avez terminé ce jeu.

CONSIGNE DU CORSI MULTICOLORE

Dans ce jeu, je vais lancer le ballon dans une séquence de cinq bouées. Lorsque j'aurai fini, votre objectif sera de reproduire la séquence en lançant le ballon dans le même ordre que moi. Une fois que vous avez fini de répondre, cliquez sur le buzzer rouge TERMINER.

Sachez qu'entre mon lancer de ballon et votre tour, les bouées vont changer de couleur. Du coup, mémorisez les positions des bouées et non leurs couleurs.

Regardez cet exemple vidéo.

[Vidéo d'exemple Corsi ludique multicolore avec bouées 2-6]

Entraînement :

On commence par un petit entraînement.

Lancez le ballon dans le même ordre que moi. Mémorisez les positions des bouées et non leurs couleurs. Une fois que vous avez bien compris la consigne d'entraînement, cliquez sur la touche valider.

Après la phase d'entraînement

La phase d'entraînement est terminée. Maintenant, le jeu commence réellement. Je vous rappelle la consigne. Lancez le ballon dans le même ordre que moi.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir et bouger avant d'être fixe au lancement du jeu.

Avant chaque niveau de jeu

Répétez cet enchaînement dans le même ordre que moi.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir et bouger avant d'être fixe au lancement du jeu.

Après la fin de la passation du Corsi ludique multicolore

Bravo ! Vous avez terminé ce jeu.

CONSIGNE DU GNG LUDIQUÉ

Consigne générale

De manière aléatoire, les bouées vont s'allumer les unes après les autres. Votre objectif sera d'appuyer sur la touche Entrée aussi rapidement que possible pour toutes les bouées qui s'allument. Sauf qu'il y a une exception : **n'appuyez pas sur la touche Entrée pour les bouées grises qui contiennent une fleur**. Pour les bouées avec d'autres couleurs contenant une fleur, appuyez sur la touche Entrée. Sachez qu'au cours du jeu, les couleurs des bouées changent. Regardez cet exemple vidéo.

Consigne d'entraînement

On commence par un petit entraînement.

Appuyez sur la touche Entrée pour toutes les bouées qui s'allument sauf pour celles qui sont grises avec une fleur. Une fois que vous avez bien compris la consigne d'entraînement, appuyez sur la touche Entrée.

Consigne générale après entraînement

La phase d'entraînement est terminée. Maintenant, le jeu commence réellement. Je vous rappelle la consigne. Appuyez sur la touche Entrée aussi rapidement que possible pour toutes les bouées qui s'allument. Sauf pour les bouées grises qui contiennent une fleur.

Appuyez sur la touche Entrée pour lancer le jeu.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir, bouger avant d'être fixe au lancement du jeu.

Après la fin de la passation du GNG ludique multicolore

Bravo ! Vous avez terminé ce jeu.

CONSIGNE DES TACHES COMBINÉES

Consigne générale

Bonjour à vous ! Nous sommes tous deux partis en vacances au bord de mer et je viens d'avoir une idée de nouveau jeu avec un ballon de plage. D'abord, je lancerai le ballon dans cinq bouées différentes. Ensuite, vous devrez reproduire la séquence en cliquant deux fois sur les mêmes bouées que moi.

Nous jouerons deux fois à ce jeu. Lors de l'une des deux fois, il sera précédé d'un autre jeu dont on verra par la suite les détails.

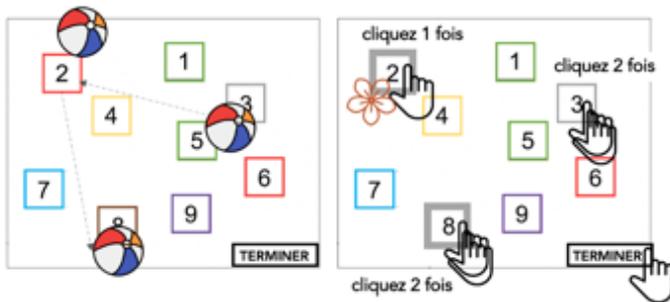
Partie A sans le poisson (GNG/Corsi ludique)

Consigne d'entraînement

Dans cette partie, il n'y a qu'une seule étape de jeu, à savoir : le lancer de ballon.

Concernant le jeu de ballon, je vais lancer le ballon sur une séquence de cinq bouées. Lorsque j'aurai fini, votre objectif sera de reproduire la séquence en cliquant deux fois sur les mêmes bouées que moi et dans le même ordre que moi. Mémorisez les positions des bouées et non leurs couleurs (car elles peuvent changer d'un essai à l'autre et ne sont donc pas pertinentes).

Après le premier clic sur une bouée, celle-ci s'allumera et vous devrez effectuer le second clic pour l'éteindre. Une fleur pourra parfois apparaître sur la bouée après le premier clic, vous devrez tout de même effectuer le second clic, sauf si la fleur apparaît sur une bouée grise. Dans ce dernier cas, il faudra passer à la bouée suivante de la séquence sans effectuer le second clic. Une fois que vous avez fini de répondre, cliquez sur le buzzer rouge TERMINER.



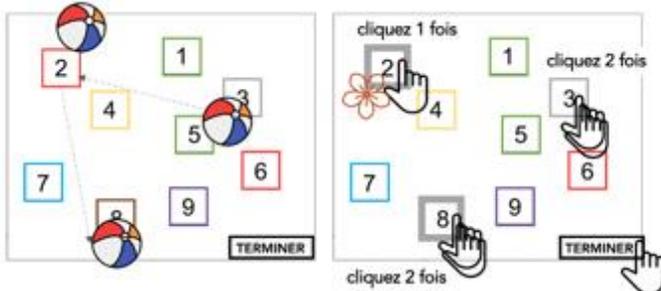
Regardez cette vidéo d'explication.

Une fois que vous avez bien compris la consigne d'entraînement, cliquez sur le bouton valider pour démarrer. Une fois que vous avez bien compris la consigne d'entraînement, cliquez sur le bouton valider pour démarrer.

Consigne générale après entraînement

La phase d'entraînement est terminée. Maintenant, le jeu commence réellement.

Pour rappel :



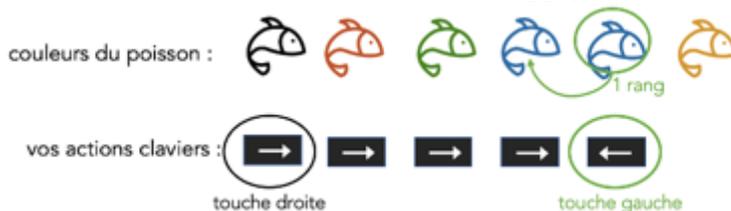
Pour commencer le jeu, cliquez sur le bouton valider.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir, bouger avant d'être fixe au lancement du jeu.

Partie B avec le poisson (N-back ludique + GNG/Corsi ludique)

Dans cette partie, il y aura deux étapes de jeu qui vont se succéder, à savoir : le lancer de ballon précédé d'un jeu de mémorisation des couleurs d'un poisson. Ce dernier va sauter de bouée en bouée. Mais ne prêtez pas attention à ses positions mais bien à sa couleur.

En effet, dans la première étape du jeu, vous allez voir un poisson sauter de bouée en bouée. À chaque saut, vous devez comparer la couleur du poisson à sa couleur précédente. Si la couleur du poisson est identique à la précédente, vous devrez appuyer sur la touche directionnelle gauche du clavier. Pour tous les autres cas, vous devrez appuyer sur la touche directionnelle de droite.

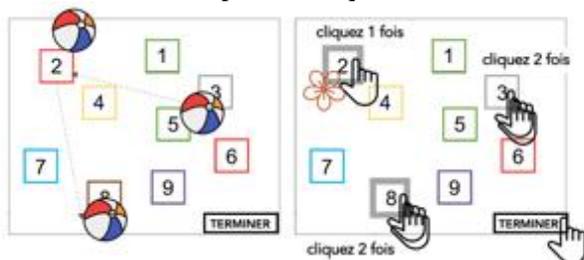


Regardez cet exemple vidéo.

Maintenant nous allons passer à la deuxième étape du jeu. Après que le poisson ait terminé de faire l'ensemble de ses sauts, on jouera ensemble. En effet, je vais lancer le ballon sur une séquence de cinq bouées. Lorsque j'aurai fini, votre objectif sera de reproduire la séquence en cliquant deux fois sur les mêmes bouées que moi et dans le même ordre.

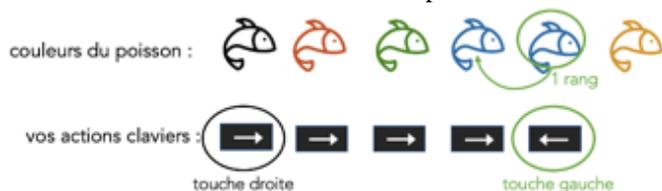
Après le premier clic sur une bouée, celle-ci s'allumera et vous devrez effectuer le second clic pour l'éteindre. Une fleur pourra parfois apparaître sur la bouée après le premier clic, vous devrez

tout de même effectuer le second clic, sauf si la fleur apparaît sur une bouée grise. Dans ce dernier cas, il faudra passer à la bouée suivante de la séquence sans effectuer le second clic. Une fois que vous avez fini de répondre, cliquez sur le buzzer rouge TERMINER.



Regardez cette vidéo d'explication.

Une fois que vous avez bien compris la consigne de cette étape, appuyez sur la touche directionnelle droite de votre clavier pour vous entraîner. Pour rappel :

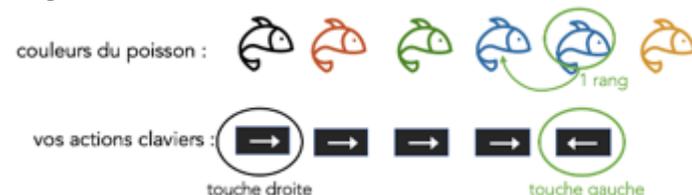


Consigne générale après entraînement

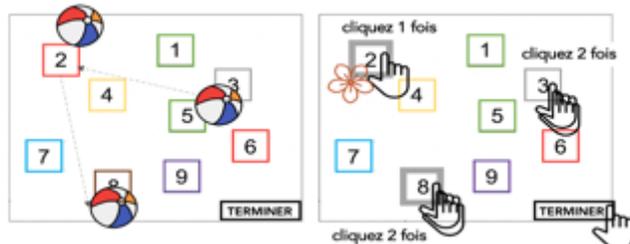
Maintenant nous allons pouvoir faire une partie ensemble. Les deux étapes de jeu vont donc se succéder.

Pour rappel les actions à faire pour chaque étape du jeu :

Étape 1



Étape 2

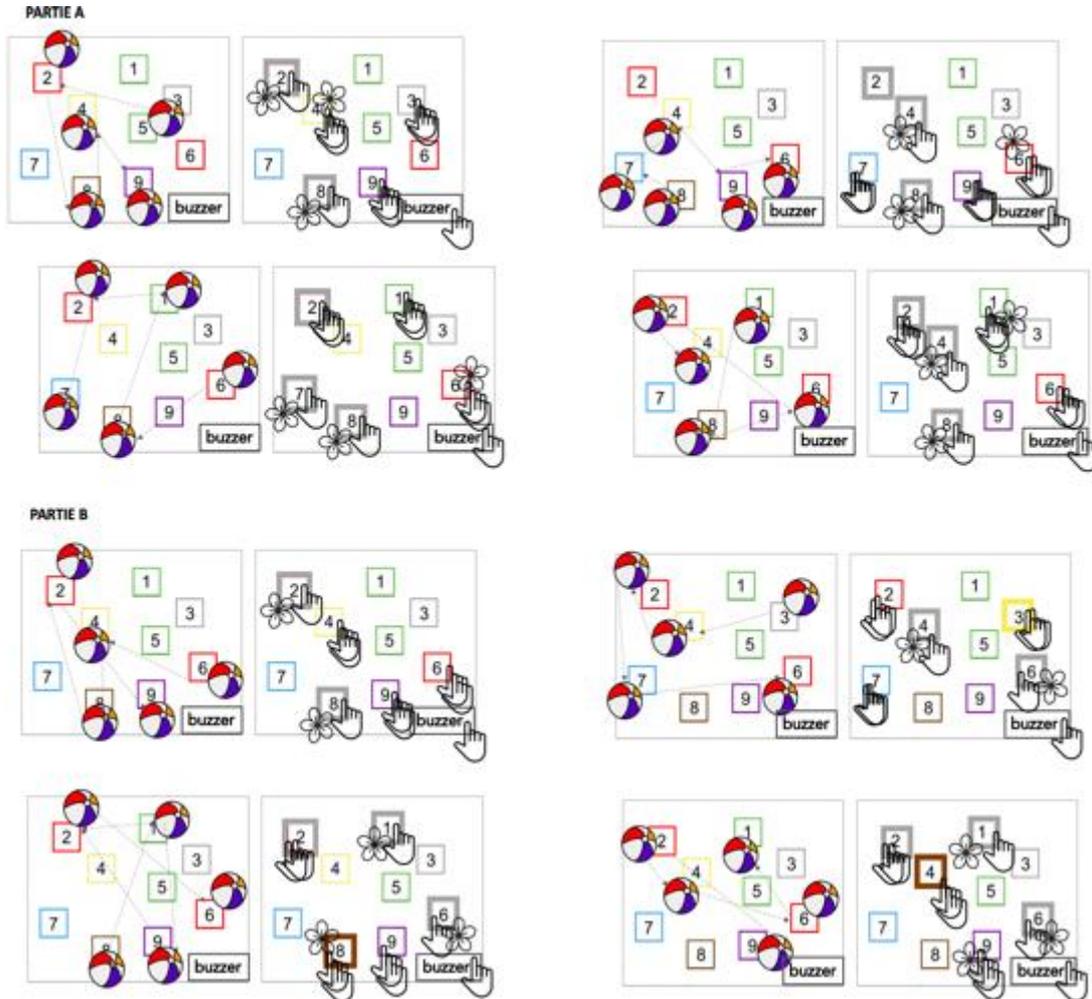


Une fois que vous êtes prêt, appuyez sur la touche directionnelle droite de votre clavier pour démarrer la partie réelle.

Pour garantir un bon enregistrement des ondes cérébrales, prenez quelques secondes pour déglutir, bouger avant d'être fixe au lancement du jeu.

Bravo ! Vous avez terminé ce jeu.

Annexe 15 : Représentation graphique des cubes pointés pour les deux parties de la tâche combinée



Annexe 16 : Avis favorable du comité du CPP pour l'étude 2

COMITE DE PROTECTION DES PERSONNES NORD OUEST III CHU – niveau 03 – porte 03-363 14033 Caen cedex 9

Tel : 09.64.08.19.44.
Fax : 02.31.47.57.81.
E-mail : cppnordouest3@orange.fr

Caen, le 10 mars 2022

CHU (DRCI)
A l'attention de Mme Monique MARGUERITE
5 allée de l'île Gloriette
44093 Nantes cedex 1

Réf. CPP : 2022-17
N° ID RCB : 2021-A03026-35
Titre abrégé : CHARGE MENTALE
Réf. SI RIPH 2G : 21.04612.000059

Tirage au sort : 17/12/2021
Demande de complément : 27/12/2021
Retour du promoteur : 30/12/2021
Recevabilité : 04/01/2022
Séance : 05/03/2022
Demande de modifications mineures : 07/03/2022
Retour du promoteur : 10/03/2022
Avis favorable : 10/03/2022
RIPH 3 HPS

Madame,

Le CPP NORD OUEST III, par délibération en date du 05 mars 2022, a opté pour une demande de modifications mineures concernant le projet de recherche « Étude CHARGE MENTALE : Modélisation de la variation de la charge mentale objective pour des tâches sollicitant différentes fonctions cognitives » et conduit par M. le Docteur Vincent ROUALDES.

Siégeaient ensemble ce 05 mars 2022, sous la présidence de Mme Charlotte GOURIO,

MEMBRES DU COLLEGE 1

Médecins ayant une qualification et une expérience approfondies en matière de RIPH

- Mme le Docteur Bérengère BEAUPLET
- Mme le Docteur Kelly MONTHE-SAGAN
- M. le Professeur Jean-François HERON
- M. le Docteur Rémy MORELLO
- M. le Docteur Damiano CERASUOLO

Pharmaciens hospitaliers

- Mme le Docteur Charlotte GOURIO

MEMBRES DU COLLEGE 2

Compétence à l'égard des questions éthiques

- M. Hervé PLATEL

Compétence en sciences humaines et sociales ou expérience dans le domaine de l'action sociale

- Mme Sylvie BALP

Compétence juridique

- Mme Fanny ROGUE
- Mme Véronique MIKALEF-TOUDIC
- Mme Victoire-Pauline LHERMITTE

Représentants des associations agréées de malades ou d'usagers du système de santé

- Mme le Docteur Marie-Thérèse BARRELLIER (Centre François Baclesse)
- Mme Nicole DELPERIE (Alliance Maladies Rares)
- M. Jean-Marc DUJARDIN (Association Française des Diabétiques)

Délibération

Considérant les documents suivants :

Premier examen [05 mars 2022]

- Courrier de soumission daté du 15/12/2022 et signé,
- Attestation selon laquelle l'étude est conçue et réalisée conformément aux dispositions du titre II de la première partie du Code de la Santé Publique datée 15/12/2021 et signée,
- MR03 n° 20001209 du 21/10/2016,
- Liste des investigateurs, version 01 du 15/12/2021,
- CV de Vincent ROUALDES daté et signé,
- Certificat des BP de Vincent ROUALDES,
- Protocole, version 01 du 15/12/2021,
- Résumé, version 01 du 15/12/2021,
- Document d'information et de non-opposition, version 01 du 15/12/2021,
- Mail de recrutement, version 01 du 15/12/2021,
- Questionnaire NASA-TLX (annexe 7 du 01 du 15/12/2021),
- Questionnaire WORKLOAD PROFILE (annexe 8 du 01 du 15/12/2021),

Documents transmis en réponse [10 mars 2022]

- Courrier de réponse daté du 08/03/2022 et signé,
- Attestation selon laquelle l'étude est conçue et réalisée conformément aux dispositions du titre II de la première partie du Code de la Santé Publique datée 08/03/2022 et signée,
- Protocole, version 02 du 08/03/2022,
- Document d'information et de non-opposition, version 02 du 08/03/2022,

Vu l'article R.1123-25 du CSP, le comité émet un :

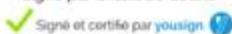
AVIS FAVORABLE

Je vous prie d'agréer, Madame, mes salutations distinguées.

Fait le 10/03/2022

Charlotte Gourio

Signé par Charlotte Gourio



Présidente

Annexe 17 : Présentation d'un résumé de l'étude 2 pour le recrutement

Mail de recrutement envoyé au sein de l'entreprise Onepoint Nantes, du laboratoire LS2N, du CHU de Nantes et de l'École Centrale de Nantes :

Intitulé du mail « Recherche volontaires pour une étude sur la charge mentale »

Contenu du mail « Bonjour à vous, et merci de prendre ce temps pour lire cette annonce. Dans le cadre d'un projet collaboratif entre le CHU de Nantes, l'entreprise Onepoint, le CNRS et l'École Centrale de Nantes, nous avons conçu une étude intitulée « Modélisation de la variation de la charge mentale objective pour des tâches sollicitant différentes fonctions cognitives ».

Cette étude est sous la responsabilité du CHU de Nantes. Elle porte sur l'identification de la charge mentale lors de différents jeux de réflexion faisant appel à des capacités mentales comme la mémoire ou l'attention (appelées fonctions cognitives). Ce type de jeux est souvent utilisé dans un contexte de rééducation et/ou d'entraînement médical. Dans notre étude, il s'agit de jeux gamifiés proposés sur ordinateur à une population non-médicale.

Le niveau de charge mentale sera mesuré de deux façons différentes :

- grâce à l'enregistrement de vos signaux d'Électroencéphalographie (EEG) à l'aide d'un casque spécifique que vous porterez durant l'étude. Ce casque permet d'enregistrer l'activité cérébrale de façon tout à fait inoffensive et indolore.
- grâce aux réponses que vous apporterez à deux questionnaires de mesure de la charge mentale et qui vous seront remis durant l'étude.

Dans ce contexte, nous aurions besoin de personnes motivées pour être volontaires pour nos passations qui se dérouleront au CHU de Nantes (Hôpital Guillaume et René Laënnec, Boulevard Jacques Monod - St Herblain, 44093 NANTES au laboratoire ICONIC au niveau - 2). Pour l'expérience, votre présence en ce lieu se fera en deux fois. La première visite (d'information) durera 1 heure où l'on vous expliquera plus en détail l'étude. Puis, minimum sept jours plus tard, la seconde visite (d'inclusion) durera 2 h 00 et consistera en une passation de tâches expérimentales durant lesquelles des questionnaires mesurant la charge mentale vous seront proposés, en parallèle d'un enregistrement EEG.

Pour vous remercier de votre contribution à cette recherche scientifique, un bon cadeau d'un montant de 50 € vous sera offert. Ce bon vous sera remis à la toute fin de l'étude, il pourra donc s'écouler quelques semaines avant que vous ne le receviez.

Les données recueillies dans le cadre de l'étude seront totalement anonymisées et vous pourrez décider d'arrêter à tout moment. Votre participation ne pourra cependant être considérée comme effective que si vous réalisez l'intégralité des deux sessions de visite. Pour pouvoir être candidat à cette étude, il est nécessaire que vous soyez majeure, francophone, droitier.es, que vous ayez une vue normale ou corrigée à la normale et une audition normale ou corrigée à la normale, une perception correcte des couleurs. Pour participer rien de plus simple, répondez à ce mail et l'expérimentatrice répondra à vos éventuelles questions et vous indiquera les étapes à suivre.

Nous avons besoin de vous, merci d'avance pour votre contribution. ».

Annexe 18 : Consentement libre et éclairé de l'étude 2



Note d'information pour la participation à la recherche

« Modélisation de la variation de la charge mentale objective pour des tâches sollicitant différentes fonctions cognitives. »

Titre abrégé : « Modélisation de la charge mentale objective »

Promoteur : CHU de Nantes

N° Enregistrement : 2021-A03026-35

Madame, Monsieur,

Le Centre Hospitalier et Universitaire (CHU) de Nantes souhaite mener une recherche dont il est le promoteur (c'est-à-dire qu'il en est responsable et qu'il l'organise). Cette recherche est réalisée en collaboration avec l'entreprise onepoint, le laboratoire LS2N et l'École Centrale de Nantes. Ces différents partenaires forment le groupe d'intérêt scientifique (GIS) nommé « Iconic ». Votre participation à la recherche, au cas où vous donneriez votre accord, ne pourra vous être confirmée qu'à la condition que vous remplissiez tous les critères d'inclusion pour participer à cette recherche. Cette dernière ne présente pas de risque pour votre santé. Les résultats qui en seront issus serviront à la recherche et n'apporteront pas d'informations concernant votre état de santé. Ils favoriseront le développement des connaissances dans le domaine de la psychologie cognitive, la neuroergonomie et plus largement sur les sciences humaines et sociales.

Cette recherche porte sur l'identification de la charge mentale lors de différents exercices de réflexion faisant appel à des capacités mentales comme la mémoire ou l'attention (appelées fonctions cognitives). Ce type d'exercices est souvent utilisé dans un contexte de rééducation et/ou d'entraînement médical. Dans ce genre de rééducation, il est essentiel de pouvoir moduler le niveau de difficulté de la tâche afin de maintenir un niveau de charge mentale constant. Ceci dans le but de ne pas fatiguer inutilement le patient et préserver l'aspect ludique de l'activité. Ainsi, il faut définir le niveau de sollicitation mentale de façon à ce qu'il soit suffisant pour maintenir l'implication de la personne dans la tâche sans pour autant être trop élevé.

Or, actuellement, il n'existe aucun modèle de variation de la charge mentale pour des tâches sollicitant diverses fonctions cognitives et sur lequel nous pourrions nous baser pour en moduler la difficulté. Notre objectif est donc d'étudier, chez des participant.es volontaires (dont vous faites partie), l'évolution de la charge mentale dans différentes activités mentales sollicitant les fonctions cognitives. Le niveau de charge mentale sera mesuré de deux façons différentes :

- grâce à l'enregistrement de vos signaux d'ÉlectroEncéphaloGraphie (EEG) à l'aide d'un casque spécifique que vous porterez durant l'étude. Ce casque permet d'enregistrer l'activité cérébrale de façon tout à fait inoffensive et indolore.
- grâce aux réponses que vous apporterez à deux questionnaires de mesure de la charge mentale et qui vous seront remis durant l'étude, à savoir le questionnaire du *National Aeronautics and Space Administration Task Load Index* (NASA-TLX) et du *Workload Profile* (WP).

Ces données seront récoltées uniquement lors de la deuxième session expérimentale (visite d'inclusion) de 2 heures au CHU de Nantes (Hôpital Guillaume et René Laënnec, Boulevard Jacques Monod - St Herblain, 44093 NANTES au laboratoire ICONIC au niveau -2). Cinq tâches différentes seront réalisées en tout. Chaque tâche comportera au minimum 3 niveaux de difficulté distincts pour lesquels seront recueillies les mesures EEG et les réponses aux questionnaires de charge mentale.

Concernant les tâches expérimentales, trois tâches feront appel à des fonctions cognitives bien distinctes qui sont la mémoire verbale, la mémoire visuo-spatiale et l'inhibition mentale motrice. La quatrième tâche expérimentale dite « tâche classique gamifiée » reprend les niveaux intermédiaires des tâches classiques dans un contexte ludique et la cinquième tâche dite « jeu vidéo composite » reprend les trois fonctions cognitives (mémoire visuo-spatiale, mise à jour de la mémoire verbale et inhibition) conjointement dans un même jeu vidéo.

Les données recueillies comprendront également des informations relatives à vos habitudes de vie comme la fréquence d'utilisation de jeux en ligne.

Grâce à cette étude, nous souhaitons identifier au minimum trois niveaux de charge mentale mesurés de façon objective (à travers des indicateurs EEG) tels que la charge mentale faible, la charge mentale intermédiaire et la charge mentale élevée. Enfin, à la fin de la visite d'inclusion, quand les tests auront été réalisés, les participant.es recevront une indemnisation sous forme d'un chèque cadeau de 50 euros.

Sachez que les indemnisations au titre de votre participation à une ou plusieurs recherches ne peuvent pas dépasser, sur une période déterminée, une somme fixée par arrêté.

TRAITEMENT DES DONNEES

Le traitement de vos données personnelles a pour fondement juridique l'intérêt public que revêt cette recherche (article 6.1.e du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (RGPD)). Il s'agit d'une des exceptions prévues par l'article 9 du même texte, permettant le traitement des données de santé.

Le traitement est conforme à la méthodologie de référence MR001/003/004 publiée par la CNIL.

Dans le cadre de cette recherche, un traitement informatique de vos données personnelles va être mis en œuvre : cela permettra d'analyser les résultats de la recherche et de remplir l'objectif de la recherche.

Pour cela, les données vous concernant seront recueillies dans un fichier informatique auquel aura accès le Promoteur de la recherche (c'est-à-dire Dr Vincent ROUALDES du CHU de Nantes).

Afin d'assurer leur confidentialité, vos données seront identifiées par un numéro de code et vos initiales. Les personnes intervenant dans la recherche dans le cadre de leurs missions auront accès à vos données codées.

Les seules personnes qui auront accès à vos données directement sont les personnes qui assurent votre prise en charge dans le cadre de la recherche et les personnes responsables du contrôle et de l'assurance qualité de la recherche.

Vos données pourront, dans des conditions assurant leur confidentialité, via des plateformes ou serveurs sécurisés, faire l'objet de transferts nationaux ou internationaux (comme, par exemple, aux personnes ou sociétés agissant pour le compte du CHU, aux autorités sanitaires habilitées) pour cette étude, ou pour d'autres recherches ultérieures, exclusivement à des fins scientifiques. Au sein de l'Europe la protection de vos données est garantie (Règlement européen UE 2016/679). Hors Europe

vos données pourront être transmises dans des états n'ayant pas le même niveau d'exigence en termes de protection des données. Le cas échéant, le Promoteur prendra toutes les mesures nécessaires pour protéger les données recueillies et devra s'engager à assurer un niveau de sécurité équivalent à celui couvert par les lois françaises et européennes pour les données envoyées à l'étranger. Si vous le souhaitez, vous pouvez obtenir la copie de l'ensemble des mesures prises pour assurer l'ensemble de la protection de vos données auprès du Délégué à la Protection des Données (DPD) voir en fin de document, paragraphe « vos contacts ».

Vos données pseudonymisées (identifiées par un numéro de code et vos initiales) seront susceptibles d'être exploitées dans le cadre de publications ou de communications ; dans ce cas votre anonymat sera préservé.

Conformément aux dispositions de la loi relative à l'informatique aux fichiers et aux libertés (loi modifiée du 6 janvier 1978), et du Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (RGPD), vous disposez d'un droit d'accès, de rectification, de limitation et d'opposition du traitement de vos données personnelles. Si vous décidez de retirer votre non-opposition pour participer à la recherche, les données obtenues avant que celui-ci n'ait été retiré seront utilisées conformément à l'article L.1122-1-1 du CSP. Les données recueillies après le retrait de votre non-opposition ne seront pas utilisées pour cette recherche et resteront destinées à l'usage strict du soin. Vous pouvez également porter une réclamation auprès d'une autorité de contrôle (CNIL pour la France : <https://www.cnil.fr/fr/webform/adresser-une-plainte/>).

Ces données pourront être utilisées lors de recherches ultérieures exclusivement à des fins scientifiques. Vous pouvez retirer votre non-opposition à cette utilisation ultérieure ou exercer votre faculté d'opposition à tout moment.

Vous pouvez également accéder directement ou par l'intermédiaire d'un médecin de votre choix à l'ensemble de vos données médicales en application des dispositions de l'article L 1111-7 du Code de la Santé Publique.

Vos données seront conservées tout au long de la recherche. Après la fin de la recherche, les données seront archivées pour une durée de 15 ans, puis détruites.

L'investigateur pourra vous informer, sur votre demande, des résultats globaux de cette recherche.

Pour en savoir plus ou exercer vos droits concernant vos données, voir en fin de document, paragraphe « vos contacts ».

Cette recherche est conforme :

- aux articles L. 1121-1 à L. 1126-12 du Code de la Santé Publique relatifs aux recherches impliquant la personne humaine
- à la loi « Informatique et Libertés » du 6 janvier 1978 modifiée et la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles
- au Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (RGPD)

Elle a reçu l'avis favorable du Comité de Protection des Personnes (CPP) Nord Ouest III le 10/03/2022.

Votre participation à cette recherche est libre. Vous pouvez refuser de participer à cette recherche, et vous pouvez à tout moment vous retirer de cette recherche, sans préjudice. Cela n'aura aucune

conséquence sur la qualité des soins qui vous seront donnés; vous devez simplement en informer l'investigateur.

VOS CONTACTS :

Pour toute question concernant l'étude, retrait de non-opposition, ou pour exercer vos droits concernant vos données (accès, rectification, etc...), votre contact privilégié est :

L'investigateur coordonnateur de la recherche :

Dr ROUALDES Vincent

✉ vincent.roualdes@chu-nantes.fr

☎ 02 401 65 173 (secrétariat) et 02 401 65 069 (DECT)

Pour toute question générale sur le traitement de vos données :

Le promoteur de la recherche, responsable du traitement :

CHU de Nantes, direction de la recherche et de l'innovation
5 allée de l'Île Gloriette, 44093 NANTES Cedex 1

Le Délégué à la Protection des Données (DPD) / Data Protection Officer (DPO) :

vosdonneespersonnelles@chu-nantes.fr

A compléter par la personne participant à la recherche

Prénom/Nom :

J'accepte que mes données soient utilisées pour cette recherche : oui non

Date :/...../.....

Signature :

Annexe 19 : Questionnaire pré-test de l'étude 2

Consigne générale session 1 :

Bonjour à vous, et merci de participer à notre session d'étude au CHU de Nantes. Je suis Lina-Estelle Louis et je vous accompagnerai pas à pas durant l'ensemble des jeux que vous allez réaliser aujourd'hui. Avant toute chose, je vous laisse répondre aux questions ci-après.

- Combien d'heures avez-vous dormi la nuit passée ? [Réponse numérique]
- Votre téléphone portable doit être en silencieux. Est-ce bien le cas ? [Cocher pour oui]
- À l'instant T, quel est votre niveau de fatigue ? [Échelle : grande fatigue, fatigue moyenne, fatigue basse, aucune fatigue].
- Sexe biologique [femme, homme]
- Votre âge [réponse alphanumérique]
- Votre niveau d'étude maximal [collège, lycée, Bac + 2, Bac + 3, Bac + 5, au-delà].
- Langue maternelle [réponse libre]
- Votre niveau de Français ? [Bas, moyen, haut, langue courante]
- A quelle fréquence jouez-vous aux jeux-vidéo ? [Jamais, très peu, un peu, souvent]
- A quelle fréquence jouez-vous aux jeux en ligne ? [Jamais, très peu, un peu, souvent]

Merci pour vos retours. Maintenant, nous allons passer aux jeux. C'est parti ! Appuyez sur le bouton Valider pour commencer.

Annexe 20 : Questionnaire post-test de l'étude 2

Questionnaires post-test après chaque niveau d'une tâche :

- Durant ce niveau, avez-vous été dérangé(e) ou distrait(e) ? [Oui/non]
- Suivi de la passation du NASA-TLX (Annexe 4)
- Et des quatre items du WP (Annexe 5)
- Dans quelle mesure la tâche que vous venez de faire a été ludique, amusante ?

Situation post-test final :

- Dans le passé, avez-vous déjà réalisé cette tâche expérimentale ? [Oui/non]
- Si oui, depuis combien de temps ? [Moins d'1 an/plus d'un an]

Bravo ! Vous avez terminé l'ensemble des tâches de la session au CHU de Nantes. Avez-vous des retours, des remarques sur la session d'aujourd'hui ? [Mettre un champ libre de texte]. [Puis, saut de ligne]

Je me joins à toute l'équipe pour vous remercier de votre participation.

Au plaisir d'échanger, Lina-Estelle LOUIS

Annexe 21 : Tableaux de résultats concernant l'H2 de l'étude 2

Réponses attendues

Paired Samples T-Test				
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	86.000		0.145

Note. Wilcoxon signed-rank test.

Temps total

Paired Samples T-Test				
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	208.000		< .001

Note. Wilcoxon signed-rank test.

Table 1 : Entre le Corsi classique et le Corsi ludique unicolore, considérés deux à deux, les comparaisons post hoc de Conover pour les Réponses attendues et le Temps total.

NASA-TLX global

Paired Samples T-Test				
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	101.000		0.896

Note. Wilcoxon signed-rank test.

Exigences mentales

Paired Samples T-Test				
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	81.000		0.380

Note. Wilcoxon signed-rank test.

Exigences physiques

Paired Samples T-Test				
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	63.000		0.233

Note. Wilcoxon signed-rank test.

Paired Samples T-Test		Exigences temporelles		
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	86.000		1.000

Note. Wilcoxon signed-rank test.

Paired Samples T-Test		Effort		
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	101.000		0.513

Note. Wilcoxon signed-rank test.

Paired Samples T-Test		Performances		
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	62.500		0.522

Note. Wilcoxon signed-rank test.

Paired Samples T-Test		Frustration		
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	71.000		0.813

Note. Wilcoxon signed-rank test.

Table 2 : Entre le Corsi classique et le Corsi ludique unicolore, considérés deux à deux, les comparaisons post hoc de Conover pour le NASA-TLX global et les différentes dimensions du NASA-TLX.

Paired Samples T-Test ▼		WP3		
Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiClassic	121.500		0.035

Note. Wilcoxon signed-rank test.

Paired Samples T-Test ▼				WP5
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	33.500		0.025

Note. Wilcoxon signed-rank test.

Paired Samples T-Test				WP7
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	26.000		0.010

Note. Wilcoxon signed-rank test.

Table 3 : Entre le Corsi classique et le Corsi ludique unicolore, considérés deux à deux, les comparaisons post hoc de Conover pour les différentes dimensions du WP.

Paired Samples T-Test				Ludique
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	78.500		0.519

Note. Wilcoxon signed-rank test.

Table 4 : Entre le Corsi classique et le Corsi ludique unicolore, considérés deux à deux, les comparaisons post hoc de Conover de la question Ludique.

Paired Samples T-Test				Motivation
Measure 1	Measure 2	W	df	p
CorsiClassic	- CorsiUni	51.000		0.133

Note. Wilcoxon signed-rank test.

Table 5 : Entre le Corsi classique et le Corsi ludique unicolore, considérés deux à deux, les comparaisons post hoc de Conover de la question Motivation.

Annexe 22 : Tableaux de résultats concernant la suite de l'H2 (multiplicité de couleurs) de l'étude 2

Paired Samples T-Test			Réponses attendues	
Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	97.000		0.342

Note. Wilcoxon signed-rank test.

Paired Samples T-Test			Temps total	
Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	101.000		0.898

Note. Wilcoxon signed-rank test.

Table 1 : Entre le Corsi ludique unicolore et le Corsi ludique multicolore, considérés deux à deux, les comparaisons post hoc de Conover pour les Réponses attendues et le Temps total,

Paired Samples T-Test			NASA-TLX global	
Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	55.000		0.065

Note. Wilcoxon signed-rank test.

Paired Samples T-Test			Exigences mentales	
Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	61.000		0.296

Note. Wilcoxon signed-rank test.

Paired Samples T-Test			Exigences physiques	
Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	28.000		0.408

Note. Wilcoxon signed-rank test.

Paired Samples T-Test **Exigences temporelles**

Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	62.000		0.316

Note. Wilcoxon signed-rank test.

Paired Samples T-Test **Effort**

Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	53.500		0.098

Note. Wilcoxon signed-rank test.

Paired Samples T-Test ▼ **Performances**

Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	47.000		0.288

Note. Wilcoxon signed-rank test.

Paired Samples T-Test **Frustration**

Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	58.000		0.239

Note. Wilcoxon signed-rank test.

Table 2 : Entre le Corsi ludique unicolore et le Corsi ludique multicolore, considérés deux à deux, les comparaisons post hoc de Conover pour le NASA-TLX global et les différentes dimensions du NASA-TLX.

Paired Samples T-Test **WP3**

Measure 1	Measure 2	W	df	p
CorsiUni	- CorsiMulti	106.000		0.383

Note. Wilcoxon signed-rank test.

Paired Samples T-Test					WP5
Measure 1		Measure 2	W	df	p
CorsiUni	-	CorsiMulti	60.000		0.697

Note. Wilcoxon signed-rank test.

Paired Samples T-Test					WP7
Measure 1		Measure 2	W	df	p
CorsiUni	-	CorsiMulti	105.000		0.185

Note. Wilcoxon signed-rank test.

Table 3 : Entre le Corsi ludique unicolore et le Corsi ludique multicolore, considérés deux à deux, les comparaisons post hoc de Conover pour les différentes dimensions du WP.

Paired Samples T-Test					Ludique
Measure 1		Measure 2	W	df	p
CorsiUni	-	CorsiMulti	71.500		0.831

Note. Wilcoxon signed-rank test.

Table 4 : Entre le Corsi ludique unicolore et le Corsi ludique multicolore, considérés deux à deux, les comparaisons post hoc de Conover de la question Ludique.

Paired Samples T-Test					Motivation
Measure 1		Measure 2	W	df	p
CorsiUni	-	CorsiMulti	29.000		0.919

Note. Wilcoxon signed-rank test.

Table 5 : Entre le Corsi ludique unicolore et le Corsi ludique multicolore, considérés deux à deux, les comparaisons post hoc de Conover de la question Motivation.

Annexe 23 : Tableau de résultats concernant l'H3 de l'étude 2

Paired Samples T-Test

Measure 1		Measure 2	W	df	p
RépAtt Corsi multi	-	RépAtt Corsi/GNG	27.000		0.007
NASA global Corsi multi	-	NASA global Corsi/GNG	25.500		0.003
EM Corsi multi	-	EM Corsi/GNG	35.500		0.010
EP Corsi multi	-	EP Corsi/GNG	33.000		0.401
ET Corsi multi	-	ET Corsi/GNG	56.000		0.206
Effort Corsi multi	-	Effort Corsi/GNG	29.000		0.008
Performance Corsi multi	-	Performance Corsi/GNG	1.000		< .001
Frustration Corsi multi	-	Frustration Corsi/GNG	63.000		0.205
WP3 Corsi multi	-	WP3 Corsi/GNG	81.500		0.390
WP4 Corsi multi	-	WP4 Corsi/GNG	14.000		0.342
WP5 Corsi multi	-	WP5 Corsi/GNG	71.000		0.541
WP7 Corsi multi	-	WP7 Corsi/GNG	93.000		0.760
Ludique Corsi multi	-	Ludique Corsi/GNG	161.000		0.008
Motivation Corsi multi	-	Motivation Corsi/GNG	50.000		0.409

Note. Wilcoxon signed-rank test.

Table 1 : Entre le Corsi ludique multicolore et la partie A (Corsi/GNG), considérés deux à deux, les comparaisons par paire de Wilcoxon pour les Réponses attendues, la charge mentale subjective (NASA-TLX et WP) et les dimensions Ludique et Motivation.

Paired Samples T-Test

Measure 1		Measure 2	W	df	p
NASA global GNG	-	NASA global Corsi/GNG	49.500		0.040
EM GNG	-	EM Corsi/GNG	29.500		0.009
EP GNG	-	EP Corsi/GNG	42.000		0.320
ET GNG	-	ET Corsi/GNG	96.500		0.765
Effort GNG	-	Effort Corsi/GNG	42.500		0.064
Performance GNG	-	Performance Corsi/GNG	34.500		0.016
Frustration GNG	-	Frustration Corsi/GNG	48.000		0.185
WP3 GNG	-	WP3 Corsi/GNG	74.500		0.647
WP4 GNG	-	WP4 Corsi/GNG	13.000		0.153
WP5 GNG	-	WP5 Corsi/GNG	66.500		0.959
WP7 GNG	-	WP7 Corsi/GNG	87.000		0.965
Ludique GNG	-	Ludique Corsi/GNG	148.500		0.006
Motivation GNG	-	Motivation Corsi/GNG	43.000		0.398
RépAtt GNG	-	RépAtt Corsi/GNG	94.000		0.695

Note. Wilcoxon signed-rank test.

Table 2 : Entre le GNG et la partie A (Corsi/GNG), considérés deux à deux, les comparaisons par paire pour la charge mentale subjective (NASA-TLX et WP), les dimensions Ludique, Motivation et les Réponses attendues.

Paired Samples T-Test

Measure 1		Measure 2	W	df	p
Réponses Attendues partie A	-	Réponses Attendues partie B	40.500		0.456
Temps total partie A	-	Temps total partie B	77.000		0.312
NASA global Partie A	-	NASA global Partie B	85.500		0.478
Exigences mentales Partie A	-	Exigences mentales Partie B	41.500		0.102
Exigences physiques Partie A	-	Exigences physiques Partie B	37.000		0.755
Exigences temporelles Partie A	-	Exigences temporelles Partie B	57.000		0.131
Effort Partie A	-	Effort Partie B	51.500		0.083
Performances Partie A	-	Performances Partie B	88.500		0.111
Frustration Partie A	-	Frustration Partie B	106.000		0.383
WP4 Partie A	-	WP4 Partie B	8.000		0.181
WP5 Partie A	-	WP5 Partie B	21.000		0.051
WP7 Partie A	-	WP7 Partie B	64.000		0.359
WP3 Partie A	-	WP3 Partie B	32.000		0.116
Ludique Partie A	-	Ludique Partie B	21.000		0.029
Motivation Partie A	-	Motivation Partie B	28.500		0.722

Note. Wilcoxon signed-rank test.

Table 3 : Entre la partie A et la partie B, considérés deux à deux, les comparaisons par paire pour les Réponses attendues et le Temps total pour la partie Corsi/GNG, pour la charge mentale ressentie (NASA-TLX et WP) et les dimensions Motivation et Ludique.

Paired Samples T-Test

Measure 1		Measure 2	W	df	p
Temps de réaction NBT ludique	-	Temps de réaction NBT partie B	88.000		0.546
Réponses Attendues NBT ludique	-	Réponses Attendues NBT partie B	17.500		0.094
Fausses Alarmes NBT ludique	-	Fausses Alarmes NBT partie B	67.000		0.028

Note. Wilcoxon signed-rank test.

Table 4 : Entre le NBT ludique et le NBT ludique de la partie B, considérés deux à deux, les comparaisons par paire pour les Réponses attendues, le Temps de réaction tous items confondus et les Fausses alarmes.

	N	Mean	SD	SE
NASA global GNG	20	226.250	53.490	11.961
NASA global Corsi/GNG	20	284.450	53.490	11.961
NASA global Corsi multi	20	216.600	40.456	9.046
NASA global Corsi/GNG	20	284.450	40.456	9.046
EM GNG	20	51.300	10.837	2.423
EM Corsi/GNG	20	67.550	10.837	2.423
EM Corsi multi	20	54.200	9.415	2.105
EM Corsi/GNG	20	67.550	9.415	2.105

Performance GNG	20	34.600	12.971	2.900
Performance Corsi/GNG	20	50.200	12.971	2.900
Performance Corsi multi	20	30.450	7.030	1.572
Performance Corsi/GNG	20	50.200	7.030	1.572
Ludique GNG	20	70.050	14.182	3.171
Ludique Corsi/GNG	20	50.900	14.182	3.171
Ludique Corsi multi	20	66.000	11.893	2.659
Ludique Corsi/GNG	20	50.900	11.893	2.659

Table 5 : Statistiques descriptives entre les tâches simples et combinées pour répondre à l'hypothèse qu'une augmentation du nombre de fonctions cognitives ne correspond pas à une augmentation de la charge mentale de manière cumulative (somme des fonctions cognitives sollicitées), mais plutôt à une augmentation proportionnelle de la charge mentale.

Paired Samples T-Test

Measure 1	Measure 2	W	df	p
Ludique NBT ludique	- Ludique Corsi/GNG	125.500		0.022

Note. Wilcoxon signed-rank test.

Table 6 : Entre le NBT ludique et le Corsi/GNG, considérés deux à deux, la comparaison par paire pour le ressenti Ludique.