



Development and evaluation of AI-based personalization algorithms for attention training

Maxime Adolphe

► To cite this version:

Maxime Adolphe. Development and evaluation of AI-based personalization algorithms for attention training. Psychology. Université de Bordeaux, 2024. English. ⟨NNT : 2024BORD0176⟩. ⟨tel-04884647⟩

HAL Id: tel-04884647

<https://theses.hal.science/tel-04884647v1>

Submitted on 13 Jan 2025

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ECOLE DOCTORALE SOCIÉTÉS, POLITIQUE,
SANTÉ PUBLIQUE

Spécialité Sciences Cognitives

Par **Maxime ADOLPHE**

**Développement et évaluation d'algorithmes de
personnalisation basés sur l'IA pour l'entraînement de
l'attention.**

-
Development and evaluation of AI-based personalization
algorithms for attention training.

Sous la direction de : **Hélène Sauzéon**

Co-directeur : **Pierre-Yves Oudeyer**

Soutenue le 27 septembre 2024

Membres du jury et invités:

M. Julien DIARD	Chargé de Recherche	Université Grenoble Alpes	Rapporteur
M. Walter R BOOT	Professeur	Weill Cornell Medicine	Rapporteur
Mme. Vanda LUENGO	Professeure	Sorbonne Université	Examinatrice
Mme. Claudia VON BASTIAN	Associate professor	Université de Sheffield	Examinatrice
Mme. Hélène SAUZEON	Professeure des universités	Université de Bordeaux	Directrice
M. Pierre-Yves OUDEYER	Directeur de Recherche	INRIA	Directeur
Mme. Alexandra DELMAS	Chercheuse consultante	Onepoint	Invitée
M. Denis MAUREL	Chercheur consultant	Onepoint	Invité

Parce que l'époque que nous traversons nous oblige à réfléchir sur l'héritage de ceux qui nous ont précédés, ce manuscrit est dédié aux géants sur les épaules desquels je me tiens, mes grands-parents : Marie-Thérèse, Nadine et Dominique.

Because the times we live in compel us to reflect on the legacy of those who came before us, this manuscript is dedicated to the giants on whose shoulders I stand, my grandparents: Marie-Thérèse, Nadine, and Dominique.

Développement et évaluation d'algorithmes de personnalisation basés sur l'IA pour l'entraînement de l'attention.

Résumé : Les Systèmes Tutoriels Intelligents (STI) offrent des solutions éducatives innovantes proposant des expériences d'apprentissage personnalisées qui s'adaptent à la variabilité individuelle. L'équipe Flowers (centre inria de l'université de Bordeaux) développe des STI dont l'originalité est d'être basée sur l'Hypothèse de Progrès d'Apprentissage (HPA). L'HPA postule que les individus sont intrinsèquement motivés à s'engager dans des activités où ils perçoivent un progrès d'apprentissage élevé et que l'expérience de ce progrès renforce encore leur motivation. Le système, nommé Zone of Proximal Development and Empirical Success (ZPDES), utilise un algorithme d'apprentissage automatique qui personnalise les trajectoires d'apprentissage en identifiant dynamiquement les activités maximisant les progrès.

L'étude de l'Entraînement Cognitif (EC), une intervention visant à améliorer des fonctions cognitives telles que la mémoire, l'attention et la résolution de problèmes, révèle que les bénéfices d'un EC sont fortement soumis aux différences interindividuelles, soulignant la nécessité de la personnalisation. Dans ce contexte, notre recherche explore le potentiel de l'algorithme ZPDES utilisé dans le cadre de l'EC pour améliorer les performances cognitives, l'engagement et la motivation. Nous avons d'abord mené une revue systématique pour identifier les stratégies actuelles de personnalisation de l'EC. La revue a révélé l'imaturité relative du domaine, caractérisée par un nombre limité d'études ($n=19$), des problèmes méthodologiques et une grande variété d'approches au sein de cet échantillon restreint. Puis, à travers une revue subjective, nous nous sommes intéressés à la tâche Multi-Object Tracking (MOT), et nous avons documenté son efficacité comme EC en fonction des paramètres manipulés dans la tâche : sont observés chez des individus variés en âge et en neurodiversité, des effets de transfert proches, tels que l'amélioration sur des tâches nécessitant de l'attention (par exemple, la tâche Useful field of view), des transferts lointains (comme la mémoire de travail ou les fonctions exécutives), et des transferts écologiques, notamment une perception améliorée des mouvements biologiques et des performances en football. Sur cette base, nous avons conçu et testé un programme d'EC individualisé utilisant la tâche MOT. D'abord, nous avons développé une batterie cognitive complète comprenant sept tâches couvrant l'attention, la mémoire de travail et les fonctions exécutives. Ensuite, deux expérimentations ont été menées : l'une avec des jeunes adultes ($n=72$) et l'autre avec des adultes âgés ($n=50$), avec à chaque fois un groupe contrôle (EC non personnalisé) et un groupe expérimental (EC personnalisé). Les participants effectuaient trois heures d'évaluation pré-et post EC, et l'EC durait huit heures (réparties sur 2 semaines), en gérant leurs horaires de manière autonome via une plateforme en ligne.

Les résultats ont montré que ZPDES pouvait être plus efficace qu'une condition contrôle, avec des performances améliorées dans les tâches entraînées dans les deux études, soulignant les avantages des parcours individualisés. Cependant, la motivation et l'engagement étaient plus faibles dans les groupes utilisant ZPDES, probablement en raison de la charge cognitive et de facteurs métacognitifs.

En conclusion, la personnalisation de l'EC par des systèmes comme ZPDES représente une direction prometteuse pour les recherches futures. Elle propose des méthodes automatiques qui prennent en compte les différences individuelles tout en respectant les standards méthodologiques d'évaluation de l'efficacité d'un EC. Ce travail enrichit les connaissances dans les domaines des STI et de l'EC. Il démontre le potentiel des stratégies d'apprentissage individualisées pour améliorer les résultats d'un EC, tout en soulignant l'importance cruciale de la motivation et de l'engagement pour optimiser l'efficacité de ces approches en termes d'effets cognitifs et éducatifs.

Mots-clés : Entraînement Cognitif, Personnalisation, STI, Attention, MOT

Development and evaluation of AI-based personalization algorithms for attention training.

Abstract: Intelligent Tutoring Systems (ITS) offer innovative educational solutions by providing personalized learning experiences that adapt to individual variability. This adaptability is crucial for tailoring curricula to maximize student engagement and learning outcomes. The Flowers team (INRIA Bordeaux) has developed an ITS based on the Learning Progress Hypothesis (LP-H). The LP-H suggests that individuals are intrinsically motivated to engage in learning activities when they perceive their own progress. This perception of progress acts as an internal motivator, encouraging them to pursue more activities where they can continue to make significant progress. The system, named Zone of Proximal Development and Empirical Success (ZPDES), uses a machine learning algorithm to customize learning trajectories by dynamically identifying and exploiting activities that yield maximal learning progress, thereby enhancing student motivation.

The study of cognitive Training (CT), which involves structured tasks designed to improve specific cognitive functions such as memory, attention, and problem-solving, reveals that the benefits of CT are highly sensitive to inter-individual differences, highlighting the need for CT personalization. In this context, our research explores the potential of applying the ZPDES framework to CT to improve cognitive performance, engagement, and motivation.

We first conducted a systematic review to identify current strategies for individualizing cognitive training. The review revealed the field's relative immaturity, characterized by a limited number of studies ($n=19$), methodological issues, and a wide variety of approaches within this small sample. Then, through a subjective review, we examined the Multi-Object Tracking (MOT) task and documented its effectiveness as CT and its dependence on manipulated task parameters. In individuals varying in age and neurodiversity, we observed near transfer effects, such as improvements in attention-related tasks (e.g., the Useful Field of View task), far transfer effects (e.g., working memory and executive functions), and ecological transfers, including better perception of biological movements and soccer performance. Based on these insights, we designed and tested an individualized CT program using the MOT task. First, we developed a comprehensive cognitive battery comprising seven tasks covering attention, working memory and executive functions. Then, two experiments were carried out: one with young adults ($n=72$) and the other with older adults ($n=50$), each time with a control group (non-personalized CT) and an experimental group (personalized/ZPDES CT). Participants completed three hours of pre- and post-assessment, and the CT lasted eight hours (spread over 2 weeks), managing their schedules autonomously via an online platform.

The results showed that ZPDES could be more effective than a control condition, with improved performance on trained tasks in both studies, underlining the benefits of individualized training paths. However, motivation and engagement were lower in the groups using ZPDES, probably due to cognitive load and metacognitive factors.

Overall, individualizing cognitive training through systems like ZPDES provides a promising direction for future research by providing automatic methods for taking individual differences into account in CT programs while respecting methodological standards for evaluating the effectiveness of CT. As a result, our work contributes to the growing body of knowledge in both ITS and CT domains while stressing the crucial role of challenges related to motivation and engagement to optimize the effectiveness of these individualized approaches for cognitive and educational outcomes.

Keywords: Cognitive Training, Individualized, ITS, Attention, MOT

Unité de recherche

Flowers team, INRIA centre de l'université de Bordeaux.

Remerciements

Quelle incroyable aventure que le doctorat. Incroyable par son lot de découvertes, de problèmes, de curiosité, d'apprentissage, mais surtout pour sa remarquable capacité à vous faire rencontrer et travailler avec une multitude d'interlocuteurs. Merci à tous les participants, jeunes et moins jeunes, qui ont donné du temps pour que ce travail soit possible.

Je souhaite aussi remercier les membres du jury et particulièrement ceux qui auront le courage de lire ce manuscrit en entier. C'est un honneur de recevoir le retour de chercheuses et chercheurs dont les travaux ont largement nourri mon éveil scientifique ainsi que la mémoire de mon iPad.

Mes remerciements vont ensuite à ma directrice et à mon directeur de recherche, Hélène Sauzéron et Pierre-Yves Oudeyer. Je ne serai jamais assez reconnaissant pour le cadre que vous m'avez offert pendant ces 4 années. D'abord, sur la dimension scientifique, vous m'avez montré les voies du pragmatisme, de la rigueur et de l'excellence tout en me laissant une autonomie formatrice. Dans l'équipe, il n'est pas simplement question d'étudier la motivation intrinsèque; c'est aussi un moteur quotidien dont vous assurez le bon réapprovisionnement. Ensuite, sur le plan humain, vous m'avez montré les voies de l'ouverture et de l'écoute. Pour finir, vous avez mis en place quelque chose qui a beaucoup joué dans l'évolution de ce travail: la collaboration avec deux post-docs d'exception, Masataka Sawayama et Marion Pech. Je n'aurais jamais réalisé tout cela sans eux et vous. Merci.

Les conditions apportées par Hélène et Pierre-Yves ne pouvaient conduire qu'à une équipe haute en couleurs: Flowers. Je remercie grandement tous ses membres pour tous les moments d'échanges et les nombreux cafés. J'espère pouvoir vous laisser en héritage un ensemble de fléchettes, armes en plastique et autres projets de serveur Dofus. Un grand merci aux permanentes et permanents (Cécile et Clément), ainsi qu'aux doctorantes et doctorants que j'ai pu croiser, aux anciennes et anciens (Mayalen, Laetitia, Tristan, Remy, Cédric, Alex) tout comme à celles et ceux qui ont encore un peu de route (Thomas, Julien, Jérémy, Chloé, Corentin). À tous les autres stagiaires, étudiantes et étudiants qui m'ont aidé sur le projet (Thomas, Eléa, Anna, Agathe, Stéphanie) et celles et ceux que j'oublie, bon courage! Une pensée aussi à celles et ceux dont j'ai un peu plus de chances de recroiser le chemin: Matisse au pic à Lacanau ou à Bali, Grgur le fou du bus, Rania au détour d'une manif ou à un cours de charleston, Marion sur un marathon ou avec un verre de vin blanc, Isabeau en train de réparer les serveurs d'Inria tout en finissant leur bilan comptable, Gaia en road trip en Californie ou en Italie, Clément à droite sur un terrain de padel ou Gauthier en train de perdre de sombres paris tout en s'entraînant à faire la planche.

Cette thèse n'aurait pas été possible sans l'autre acteur important de ce travail: onepoint. Tout d'abord, je souhaite remercier Erwan Le Bronec sans qui cette collaboration n'existerait pas. Ensuite, merci à Alexandra Delmas et Denis Maurel pour la supervision, tous les conseils, le soutien et l'humanité qui vous caractérisent. Vous avez toujours eu pour objectif de ne pas reproduire ce qui vous avait déplu lors de vos doctorats respectifs. C'est mission réussie, vous m'avez offert le cadre le plus valorisant et accueillant qu'on puisse espérer. Une autre pensée à tous les collaborateurs avec qui j'ai pu interagir, et particulièrement les membres de la (meilleure) équipe de la boîte (Rémy, Jérôme, Juliette, Sébastien, Sylvia, Jean, Renaud, Mégane) et celles et ceux qui sont partis (Ikram, Coralie, Laura-Joy, Dan).

Ce travail n'aurait pas été possible sans l'incroyable soutien que j'ai reçu de mes amis, qu'ils soient normands (Bérénice, Vincent, Adèle, Hubert, Axelle, Mathis, Pierre), cogniticiens (Nathan, Marie, Nina, Emma, Marie-Camille, Cyril, Raphaël, Mathilda, Maria, Aziliz, Vincent, Doche, Lucas, Antoine, Colin, Nicolas, Eugénie, Jules, Mathilde), bordelais (Thibaut, Ludivine, Alice, Jean-Baptiste, Simon, Marion) et tous ceux que j'oublie (désolé), merci.

Je remercie aussi sincèrement ma famille (Marie-Thérèse, Nadine, Dominique, David, Isma, Léandre, Jean-Baptiste, Anne, Marc, Lennie, Érine, Sandrine, Manolo, Mathilde, Romain, Sylvia, Emma) pour la vitalité et la bienveillance qui vous caractérisent. Une pensée aussi à ma belle-famille, soutien incontesté pour passer de bons moments (Béatrice, Aymeric, Aloys, Matisse, Papou et les autres).

Particulièrement, je remercie Séverine, Yannis et Guillaume, mes parents et mon frère. Les premiers pour m'avoir donné tout ce qu'ils possédaient d'amour, d'intelligence, d'intégrité, de rire et de gentillesse. Le dernier en ce qu'il représente un modèle et une aide sur laquelle je pourrai toujours m'appuyer.

Enfin, ces remerciements ne suffiront jamais à traduire la gratitude que je ressens envers celle qui partage ma vie, Emma. Sans toi, ces derniers mois n'auraient pas été si simples. Merci pour ta présence, ta cuisine délicieuse, ton oreille attentive quand je décris mes nouvelles découvertes, ton militantisme et ta compassion, ton humour, ta douceur et ton incroyable intelligence. Bref, merci d'exister. C'est fait, nous avons traversé ensemble l'aventure du doctorat; la dernière page de ce manuscrit terminée, il est enfin temps de démarrer un nouveau chapitre.

Acknowledgments

What an incredible adventure a PhD is. Incredible for its discoveries, challenges, curiosity, learning, but most of all for its remarkable ability to connect you with a multitude of people. Thank you to all the participants, young and old, who took the time to make this work possible.

I also wish to thank the members of the jury, particularly those who will have the courage to read this manuscript in its entirety. It is an honor to receive feedback from researchers whose work has greatly enriched my scientific development as well as the memory of my iPad.

My thanks go to my supervisors, Hélène Sauzéron and Pierre-Yves Oudeyer. I will never be grateful enough for the environment you provided during these 4 years. First, on the scientific level, you showed me the paths of pragmatism, rigor, and excellence while giving me formative autonomy. In the team, it is not just a matter of studying intrinsic motivation; it is also a daily driver that you ensure is well replenished. On a personal level, you taught me the values of openness and listening. Finally, you facilitated something that greatly influenced the progress of this work: the collaboration with two exceptional post-docs, Masataka Sawayama and Marion Pech. I could never have accomplished all of this without them and you. Thank you.

The conditions provided by Hélène and Pierre-Yves led to a vibrant team: Flowers. I greatly thank all its members for the moments of exchange and the numerous coffees. I hope to leave you a legacy of a set of darts, plastic weapons, and other Dofus server projects. A big thank you to the permanent members (Cécile and Clément), as well as the PhD students I have met, the former members (Mayalen, Laetitia, Tristan, Remy, Cédric, Alex), and those who still have some journey ahead (Thomas, Julien, Jérémy, Chloé, Corentin). To all the other interns and students who helped me on the project (Thomas, Eléa, Anna, Agathe, Stéphanie), and those I forget, good luck! A thought also to those I am more likely to cross paths with again: Matisse at the peak in Lacanau or Bali, Grgur le fou du bus, Rania at a demonstration or a Charleston class, Marion in a marathon or with a glass of white wine, Isabeau repairing Inria's servers while finishing their accounting report, Gaia on a road trip in California or Italy, Clément on the right on a padel court, or Gauthier losing shady bets while practicing planching.

This thesis would not have been possible without another important player in this work: onepoint. First, I would like to thank Erwan Le Bronec, without whom this collaboration would not exist. Then, thank you to Alexandra Delmas and Denis Maurel for the supervision, all the advice, support, and humanity that characterize you. You always aimed not to replicate what displeased you during your respective PhDs. Mission accomplished, you provided me with the most rewarding and welcoming environment one could hope for. Another thought to all the collaborators I interacted with, particularly the members of the (best) team in the company (Rémy, Jérôme, Juliette, Sébastien, Sylvia, Jean, Renaud, Mégane) and those who left (Ikram, Coralie, Laura-Joy, Dan).

This work would not have been possible without the incredible support I received from my friends, whether they are Normans (Bérénice, Vincent, Adèle, Hubert, Axelle, Mathis, Pierre), cognitive scientists (Nathan, Marie, Nina, Emma, Marie-Camille, Cyril, Raphaël, Mathilda, Maria, Aziliz, Vincent, Doche, Lucas, Antoine, Colin, Nicolas, Eugénie, Jules, Mathilde), from Bordeaux (Thibaut, Ludivine, Alice, Jean-Baptiste, Simon, Marion, Antoine), and all those I forget (sorry), thank you.

I also sincerely thank my family (Marie-Thérèse, Nadine, Dominique, David, Isma,

Léandre, Jean-Baptiste, Anne, Marc, Lennie, Érine, Sandrine, Manolo, Mathilde, Romain, Sylvia, Emma) for the vitality and kindness that characterize you. A thought also to my in-laws, an undeniable support for having a good time (Béatrice, Aymeric, Aloys, Matisse, Papou, and others).

Particularly, I thank Séverine, Yannis, and Guillaume, my parents and my brother. My parents for giving me all their love, intelligence, integrity, laughter, and kindness. My brother for being a model and a support I can always rely on.

Finally, these acknowledgements will never be enough to express the gratitude I feel towards the one who shares my life, Emma. Without you, these past months would not have been as simple. Thank you for your presence, your delicious cooking, your attentive ear when I describe my new discoveries, your activism and compassion, your humor, your gentleness, and your incredible intelligence. In short, thank you for existing. We have made it through the adventure of the PhD together; with the last page of this manuscript completed, it is finally time to start a new chapter.

Contents

I	Introduction and background	27
1	Introduction	28
1.1	Non-Invasive Cognitive Enhancement: A Promising Future?	29
1.1.1	What is a Cognitive Training ?	29
1.1.2	Cognitive Training: a field in rapid progress	34
1.1.3	Cognitive Training: many challenges	37
1.1.4	Perceptual-Cognitive Training and the multi-object tracking task .	40
1.2	Educational Technologies: Where Algorithms Meet Human Learning . . .	42
1.2.1	Zone of Proximal Development (ZPD)	42
1.2.2	Motivation and Optimal difficulty	43
1.2.3	The Learning Progress (LP) hypothesis	44
1.2.4	Educational Technologies	45
1.3	Objectives, Scientific and Industrial Context of the Thesis	47
1.3.1	Collaboration between Bavelier Lab and the Flowers Team	47
1.3.2	Main Research Questions	49
1.3.3	Organization of the thesis	50
II	Litterature Review	51
2	Exploring the Potential of Artificial Intelligence in Individualized Cognitive Training: a Systematic Review	52
2.1	Introduction	53
2.1.1	Research question	55
2.1.2	Background	55
2.2	Material and method	61
2.2.1	Databases and Search query	62
2.2.2	Eligibility criteria	62
2.2.3	Screening and selection method	63
2.2.4	Data extraction	63
2.3	Results	79
2.3.1	Descriptive results	79
2.3.2	Q1 & Q2. What Type of AI Techniques have been used in the field of computerized CT? What are the Subject/Domains of CT for which adaptive techniques have been designed?	79
2.3.3	Q3. What populations are targeted and what are the characteristics of the CT design?	81

2.3.4	Q4. How effective are they in empirical CT studies? What effects are reported (Near and Far transfer (NFT) and everyday life transfer effects)? Are the effects dependent on the CT design (content, dose, location) and the targeted sample?	82
2.3.5	Q5. What Type of Validation have been conducted for these new generations of computerized CT?	84
2.4	Discussion	84
2.5	Conclusion	87
3	Broadening the Lens: A Narrative Review of Parameters Shaping Multi-Object Tracking Performance and Training Efficacy	91
3.1	Introduction	92
3.2	The MOT underlying theories	94
3.2.1	Q1. (a) How do changes in parameters affect performance in Multi-Object Tracking (MOT) tasks, and what theories explain these effects?	94
3.2.2	Q1. (b) Does the performance of neurodiverse populations offer clues about the fundamental processes of MOT task performance?	103
3.3	Focus on MOT training	108
3.3.1	Q2.(a) How does cognitive training using MOT tasks affect MOT performance (near transfer effect), and does changing MOT task parameters influence these effects?	108
3.3.2	Q2.(b) How does MOT Practice affect cognitive functioning (near to far transfer)?	109
3.3.3	What about a non neurotypical population ?	111
3.4	Outcomes from MOT practices in Real-Life	114
3.4.1	Q3.(a) What practices-related changes in MOT performance are linked to real-world transfer effects in neurotypical individuals?	114
3.4.2	Q3.(b) Do neurodiverse conditions alter these effects?	116
3.5	Conclusion	117
III	Experimental work	119
4	An Open-Source Cognitive Test Battery to Assess Human Attention and Memory	120
4.1	Introduction	121
4.2	Materials and Methods	124
4.2.1	Participants	124
4.2.2	Apparatus	124
4.2.3	General procedure	124
4.2.4	Stimuli and procedure of each task	124
4.3	Results	131
4.3.1	Basic performance	131
4.3.2	Reliability across days	132
4.3.3	Latent factor analysis	135
4.4	Discussion	139
4.5	Conclusions	142

5	AI-Enhanced Multi-Object Tracking: Toward Individualized Cognitive Training	143
5.1	Introduction	144
5.1.1	Background	145
5.2	Material and method	150
5.2.1	Schedule	150
5.2.2	Participants	150
5.2.3	The training task: Multi-Object Tracking	150
5.2.4	AI-based individualization procedure	152
5.2.5	Cognitive battery	154
5.2.6	Online experiment management tool and Gamification	157
5.2.7	MOT Intra-training evaluation	158
5.2.8	Subjective questionnaire	158
5.2.9	Data analysis procedure	158
5.3	Results	163
5.3.1	Participants	163
5.3.2	RQ1- How do learning trajectories differ when using our individualized AI-based adjustment procedure compared to a traditional staircase method?	163
5.3.3	RQ2- How does the individualized AI-based procedure affect the progression of task performance in the training activity and in overall cognitive performance as assessed by a broader set of cognitive tasks?	165
5.3.4	RQ3- How does our individualized AI-based procedure affect participants' self-reported experiences?	170
5.4	Discussion	171
5.5	Conclusion	177
5.5.1	Questionnaires	186
6	Benefits of Individualized Computerized Cognitive Training on older adults	194
6.1	Introduction	195
6.1.1	Cognitive Training (CT) for typical older adults : impact of interindividual differences	195
6.1.2	Computerized Cognitive Training for typical older adults : a new promise for healthy aging?	197
6.2	Design and Methods	199
6.2.1	Participants	199
6.2.2	Procedure and Materials	199
6.2.3	Schedule	199
6.2.4	Pseudo-randomization procedure	199
6.2.5	Training procedure	200
6.2.6	Assessments	201
6.2.7	Analyses	202
6.3	Results	204
6.3.1	Descriptive statistics	204
6.3.2	Q1. What is the impact of training conditions (Zone of Proximal Development and Empirical Success (ZPDES) group [individualized] vs. control group [Staircase]) on observed pre-post assessment?	204

6.3.3	Q2. What is the impact of training conditions (ZPDES group [individualized] vs. control group [Staircase]) on the MOT intra-training performance? How do individual differences impact MOT performance changes, and do training conditions influence them? .	205
6.3.4	Q3. What is the impact of training conditions on trainee experience in terms of engagement, reported cognitive load and motivation ? .	209
6.4	Discussion	212
6.4.1	Q1. Impact of training (ZPDES vs. Control) conditions on cognitive gains (pre- vs. post-measures)	212
6.4.2	Q2. Impact of Training Conditions (ZPDES vs. Control) on MOT training Performance	212
6.4.3	Q.3. The impact of cognitive load on performance, engagement, competence and motivation during trained tasks.	214
6.4.4	Lessons learned from the participants' feedback	215
6.4.5	How can we increase, boost objectify transfer effects?	215
6.4.6	Limitations and perspectives	216
6.5	Conclusion	216

IV Discussions and conclusion 220

7	Conclusion and discussion	221
7.1	Main Takeaway of the Results from Chapters 6 and 7	222
7.1.1	Training Performance	222
7.1.2	Intra-training evaluation	223
7.1.3	Pre-post cognitive evolution	224
7.1.4	Questionnaires	225
7.2	General summary of the work	226
7.2.1	Training trajectories	226
7.2.2	Cognitive Performance	227
7.2.3	Is ZPDES really providing an individualized path?	229
7.2.4	Questionnaires	230
7.3	Looking at the LP hypothesis	231
7.3.1	Cognitive Load and Metacognition	231
7.3.2	Inter-individual differences	232
7.4	Perspectives of the work	234
7.4.1	ZPDES algorithm	234
7.4.2	Perspectives on future individualized training	236
7.4.3	Enhancing the psychologist' toolbox with AI	239
7.5	Limits of the work	240
7.5.1	Ecological validity	240
7.5.2	Bayesian vs Frequentist view	241
7.5.3	Future directions	241
7.6	Conclusions	243

List of Figures

1	Illustrations of the examples taken in the general introduction	21
2	Illustration of Socrates and Meno - generated by gpt4o (05/16/2024) . . .	22
3	Portrait from René Descartes, after Frans Hals. (wikipedia)	23
4	Portrait from John Locke, after Godfrey Kneller. (wikipedia)	23
5	The cognitive hexagon	26
1.1	Illustration of activities proposed by Catherine Aiken, taken from the book <i>Methods of Mind-Training: Concentrated Attention and Memory</i> (Aiken, 1896)	29
1.2	Illustration of the integration of the CHC taxonomy into task performance, inspired by W. J. Schneider and McGrew, 2013	32
1.3	Overview of different theoretical models of working memory, executive functions, and attention, and how several tasks are used to measure some of their components. The right-hand side shows how all these tasks are integrated into the broad CHC-M taxonomy proposed in Webb et al., 2018c	33
1.4	Illustration of linear vs branched path designs.	42
1.5	The Region of Proximal Learning, <i>Figure taken from Metcalfe et al., 2020.</i>	44
1.6	The Learning Progress Hypothesis (<i>adapted from Oudeyer et al., 2016a</i>)	45
1.7	Core modules of an ITS.	46
1.8	A visual representation of the multi-armed bandit problem, illustrating its application in both a classical casino setting and an educational context. .	48
1.9	A visual representation of the evolution of the model of ZPD through time, implemented in ZPDES.	49
2.1	PRISMA Flow chart.	56
2.2	Illustration of macro and micro-adaptive strategies. (a): Macro-adaptive strategy exemplified by two trajectories within a CT program (unique trajectory or individualized design) spanning sessions 1 to 4, each session offering three potential cognitive tasks (A1 to A3). Arrows depict task order for each session. Non individualized trajectory always propose same order A1, A2 and A3 while individualized path adapts the trajectory according to training objectives. (b): Micro-adaptive strategy demonstrated by two trajectories of task difficulty adjustments specifically for task A3 during session 2. The non-individualized trajectory relies on a staircase approach that falls short of identifying the optimal zone of progress when contrasted with the individualized procedure, which proves to be a more suitable fit.	60
2.3	Distribution of AI techniques depending on type of CT studied (multi or single domain)	81

3.1	An example of a variant of the Multi-object tracking task (1) Three discs are briefly colored red to mark them as tracking targets. (2) All items look the same and move around randomly on the screen. (3) (a) At the end, one item is highlighted and the participant reports if it was a target or not. (3) (b) At the end of the motion phase, the participant must recognize the initial discs identified as targets. Animations of many different variants of this task can be viewed at or downloaded from https://perception.yale.edu/Brian/demos/MOT.html	93
3.2	A non exhaustive set of studies manipulating MOT parameters as a way to better understand tracking mechanisms	95
3.3	Neurotracker protocol, <i>image taken from Romeas et al., 2016b</i>	113
4.1	(a) interface example of the test battery. (b) Stimuli and tasks in the multiple-object tracking (MOT) task. After presenting the fixation point, five of the ten discs were briefly highlighted in red to show they are the targets to track. Then, ten objects started to move for eight seconds. Participants answered which are the target objects by clicking black boxes after the moving scene ended.	127
4.2	(a) Stimuli and tasks in the enumeration task. After presenting the fixation point, a brief flash of multiple white circles was presented. Participants answered how many circles were shown by using a slider. (b) Stimuli and tasks in the load-induced blindness task. Participants were asked to perform a dual-task, answering the length of the gazing point and the contrast of the images presented in the surroundings. After showing the fixation dot, a cross target with four Gabor stimuli was briefly presented. Participants first answered which of the lines was longer using mouse clicking. Then they answered which of the four Gabor stimuli had the enhanced contrast by clicking one of the four buttons.	128
4.3	(a) Stimuli and tasks in the go/no-go task. On each trial, a digit was briefly presented one by one. Participants were asked to focus on the number "7" and answer whether the number after the "7" was the "3" or not. If the number after "7" was "3," they had to press the key "J" as soon as possible. (b) Stimuli and tasks in the task-switching task. Participants' tasks changed with the task cue. When the task cue was the blue diamond-shaped background, participants had to answer whether the target digit was odd/even. In contrast, when the task cue was the red square background, they had to answer whether the target digit was higher/lower than five. . .	129
4.4	(a) Stimuli and tasks in the working memory task. On each trial, one of sixteen squares was sequentially flashed with a reddish color briefly. After the presentation, participants answered the sequence by clicking on the squares in the same order. (b) Stimuli and tasks in the memorability task. On each trial, a natural scene photograph was presented one by one. Participants were asked to remember each photograph and answer whether the photograph is presented twice or not.	130

4.5	Results of all tasks. The response probability or reaction time is shown for each task. The horizontal axis of each panel indicates the stimulus conditions. The small blue circle depicts the individual performance. The thin green line connects each individual performance across different stimulus parameters. The red square and error bars show the mean probability and 95 % credible intervals calculated from Bayesian statistical simulation. . . .	133
4.6	Bland-Altman plot for the accuracy data. Each participant's performance difference was plotted as a function of the mean performance of the two days. Different colors indicate different participant ages. The accuracy data is converted to the inverse normal cumulative distribution function, as in the latent factor analysis. The positive difference means that the second day performance is better than the first day. The solid red line indicates the mean difference across participants, while the dashed red line indicates the ± 1.96 SD of the differences.	134
4.7	Cross-day performance difference. Each panel shows the probability density of the posterior distribution difference between the first and second-day performance (Day 1 -Day 2 for the RT difference and Day 2 - Day 1 for the accuracy difference). The positive value means the better performance in Day2 for either RT or accuracy differences. The vertical red solid and dotted lines indicate the mean and 95% credible intervals, respectively. The condition names put on the left-top for each panel with blue color correspond to the ones shown as the abscissa in Figure 4.5. We show here cross-day performance for the parameters used in the latent factor analysis.	135
4.8	Scree plot of the latent factor analysis.	136
4.9	Individual data for PCA components 1 and 2. Different colors indicate different participant ages	137
4.10	The hierarchical clustering and the loading in (a) PCA and (b) ICA. The components in PCA are numbered in order of the magnitude of the explained variance, as in the legend. The component order in ICA is arbitrary because the analysis does not have the priority of the order.	138
5.1	(a) The MOT task. (b) Several visual snapshots of our intervention. (c) Schedule proposed to participants	151
5.2	Illustration of the beta-binomial model	162
5.3	Illustration of the methodology used to derive the posterior post-pre evolution difference and the computation of the SDDR	162

5.4	Two handpicked participant from ZPDES (a) and control (b). For each participant, the activity space is segmented into six discrete levels based on the number of targets, ranging from two to seven. For a given level of number of targets, each other parameter — speed, tracking duration, response time, and spacing — is symbolized by dots of varying sizes; the size correlates with the parameter’s relative frequency of proposition to participants. Larger dots indicate more frequent selection of that parameter value for practice. Transitions between different target difficulty levels are depicted by lines of varying thickness, where the thickness reflects the frequency with which participants experienced changes between levels. A thicker line denotes a more common transition. Color coding is utilized to indicate the sequence of the parameters’ first occurrence or the initial transition between levels; each color corresponds to a different index in the sequence of presentation or transition experienced by the participant. . . .	165
5.5	(a) Proportion of the population on each different number of targets (2 to 7 targets) through time. The total number of participants diminished over time as individuals experienced varying numbers of trials throughout their training.	166
5.6	Performance evolution across the four evaluation sessions for each of the 12 activities in the testing set is organized such that rows correspond to a fixed number of targets (easy, medium, and hard, with 2, 4, and 6 targets respectively), and columns correspond to fixed pairs of speed and spacing.	167
5.7	Pre-post evolution for accuracy (left) and RT (right) metrics. Probability in each bar represents the probability of effect. Corresponding BF is displayed right to the bars.	168
5.8	PCA results: scree plots (left) and Principal Components contribution (right)	169
5.9	Different trajectories from the control group.	183
5.10	Different trajectories from the zpdes group.	184
5.11	LP Questionnaires interaction and distribution.	187
5.12	NASA-tlx Questionnaires interaction and distribution.	188
5.13	TENS Questionnaires interaction and distribution.	189
5.14	UES Questionnaires interaction and distribution.	191
5.15	SIMS Questionnaires interaction and distribution.	193
6.1	Flowchart of the study	205
6.2	Pre-post evolution for MOT task accuracy (left) metric. Horizontal bars represent the 94% High Density Interval (HDI) of the posterior distribution for each parameter. Inside these bars, the probability of the parameter being either greater than zero ($P(\text{param} > 0)$) or less than zero ($P(\text{param} < 0)$) is shown, depending on which probability is higher.	206
6.3	Scree plot of the PCA (left) and Matrix loadings of the PCA (right). . . .	206
6.4	Proportion of the population on each different number of targets (2 to 7 targets) through time.	207
6.5	Number of transitions between activities in each group for the two studied populations.	208
6.6	Change of performance score across the sessions	210
6.7	on linear latent growth model of performance in intra-training evaluations for both groups. Estimated coefficients are provided for both control (c) and ZPDES (z). Non-null significant ($p < 0.05$) values are displayed in bold.	211

6.8	Subjective questionnaires with a) load Index related to the NASA-TLX; b) Self-Determination Index (SDI) related to the SIMS; d) feeling of competence related to the TENS.	211
6.9	Feeling of learning across the training.	219
7.1	Proportion of the population on each different number of targets (2 to 7 targets) through time. More details on this figures can be found in chapter 6.	222
7.2	Number of transitions between activities in each group for the two studied populations. More details on this figures can be found in chapter 6	222
7.3	Descriptive graph of the success rates for both groups in both studies across the four sessions of intra-training evaluation. Scatter plots represent individual performances.	223
7.4	Latent Growth Curve Model parameter estimates. Variables from the older adult studies are suffixed with “_o,” while variables from the young adult studies are suffixed with “_y.” Significant non-zero values are bolded. More details on this figure can be found in Chapter 6.	223
7.5	Forest plots of posterior distribution of the pre-test performance (white bars) and post-pre evolution (colored bars) for the young population. More details can be found in chapter 5.	224
7.6	Forest plots of the posterior distribution of pre-test performance (white bars) and post-pre evolution (colored bars) for the older population. More details can be found in Chapter 6.	224
7.7	Table summarizing the questionnaire results. Bayes Factors were used to assess the null hypothesis for each parameter estimate of the linear mixed model (see Chapter 5 for details). A credible non-null estimate is defined by a criterion of $BF > 1.5$	225
7.8	A new perspective on the LP hypothesis: Integrating cognitive load and metacognition	232
7.9	The CAMML model, taken from K. S. McGrew, 2022	234
7.10	An illustration of two possible hypotheses: a ZPD matching an optimal metacognitive development zone (solid white borders) and an overlapping but larger metacognitive development zone (dashed white borders). . . .	238

List of Tables

2.1	Inclusion and exclusion criteria	62
2.3	Overall descriptive results, AI techniques and Cognitive domains of CT for the selected studies.	64
2.4	CT characteristics	68
2.5	Risks of bias, proof level rating and authors conclusions	75
2.6	Grading scale results.	89
3.1	MOT Task	106
5.1	Summary of Questionnaires used. Details of all questionnaires can be retrieve in Annexe 5.5.1	159
5.3	Corrected summary of pre-test differences and post-pre evolution for each principal component and group, with significant p-values in bold for improved readability.	169
5.4	All hyperparameters used in our ZPDES version.	182
5.5	Null Hypothesis Models for Different Parameters. N stands for Normal distribution.	185
6.1	Models fit indices and comparisons. Significant χ^2 and $\Delta\chi^2$ are in bold (p<0.05).	209
6.2	Impact of individual differences variables on baseline and change performance. Notes: b are estimated parameters, and p are their respective p-values (assessing null). Predictors are regressed on both change and baseline, therefore model comparison between null (no predictor) and alternative (free estimation of association between individual differences variables and both slopes and initial performance) are provided in the same column. . .	209
6.3	Pre-test Difference and Post-Pre Evolution for Each Component	218

List of Acronyms

CT Cognitive Training

WM Working Memory

EF Executive Functions

AI Artificial Intelligence

ACT-R Adaptive Control of Thought—Rational

MOT Multi-Object Tracking

ITS Intelligent Tutoring System

ZPDES Zone of Proximal Development and Empirical Success

ZPD Zone of Proximal Development

CAI Computer-Assisted Instruction

RL Reinforcement Learning

AVG Action Video Games

LP Learning Progress

CHC Cattell-Horn-Carroll

RPL Region of Proximal Learning

UFOV Useful Field Of View

HDI High Density Interval

BKT Bayesian Knowledge Tracing

GAI Generative Artificial Intelligence

LLM Large Language Model

CAMML Cognitive-Affective-Motivation Model of Learning

LGCM Latent Growth Curve Models

NFT Near and Far transfer

OA Older Adults

SDI Self-Determination Index

BF Bayes Factor

SDDR Savage-Dickey density ratio

UES User Engagement Scale

BIC Bayesian Information Criterion

Preface

Historical Context

Cognitive Science

The human mind is an extraordinary entity, demonstrating remarkable capabilities in learning, adaptation, and problem-solving. An illustrative example can be drawn from the iconic film "Home Alone" (Figure 1a, "Maman, j'ai raté l'avion !" for the french readers), in which the protagonist, Kevin McCallister, devises a series of traps to defend his home against burglars. This seemingly playful scenario engages a diverse array of cognitive skills, warranting a closer examination. Firstly, Kevin's actions necessitate advanced planning and decision-making, as he must anticipate the burglars' movements and develop effective strategies to thwart them. The setup of the traps requires sophisticated spatial reasoning to ensure correct placement and functionality. Creativity and problem-solving are paramount in the design of these traps, while managing household items for this purpose showcases resourcefulness and ingenuity. Furthermore, Kevin must adhere to a strict timeline and precisely coordinate his actions, demonstrating significant control and attention to detail. Flexibility is also crucial, as he must adapt to the burglars' unexpected actions and modify his plans accordingly. While often regarded as mere entertainment, Kevin's elaborate preparations in "Home Alone" exemplify the impressive cognitive abilities that humans employ in routine activities. This underscores the sophistication inherent in human cognitive processes, revealing the complexity of tasks that are frequently underestimated.



(a) Illustrative poster of the movie Home Alone



(b) Portrait of Salvador Dali¹

Fig 1. Illustrations of the examples taken in the general introduction

Extending this perspective, individuals can acquire a wide array of skills within just a few years, such as playing musical instruments, mastering complex mathematical theories, and cooking gourmet meals. Humans navigate a world with intricate social norms, demonstrating an impressive ability to learn in diverse environments. Moreover, they are exceptional storytellers, capable of combining various skills in innovative ways that shape our world. For example, consider the life and work of Salvador Dalí (see Figure 1b). Known for his surrealist artwork, Dalí was a precursor of personal branding, a concept where individuals create and manage their public persona to stand out and be memorable. Dalí merged his artistic talent with a keen understanding of media and public perception to craft a unique identity that captured global attention (Archipub, 2021; Star Arts, 2016). This strategy helped him become not only a celebrated artist but also a cultural icon whose influence extended beyond the art world. Furthermore, Dalí’s distinctive persona and imaginative works continue to inspire and intrigue audiences, demonstrating humanity’s ability to craft narratives that resonate deeply with others. This example illustrates not only the human capability for innovation but also our remarkable ability to learn, adapt, and communicate the value of our ideas. The fusion of technical skills, social understanding, and storytelling exemplifies the multifaceted nature of human intelligence.

This thesis primarily examines a fundamental feature of intelligent systems: learning. Learning is at the core of all previously described achievements, enabling humans to integrate sensory input, produce fine motor control reactions, adapt to novel situations, and interact with others. As a key characteristic of intelligent systems, understanding the mechanisms driving intelligence has been a subject of great interest for many centuries. This manuscript aims to contribute to this long-standing inquiry by exploring the processes and principles underlying human learning and intelligence.

Philosophical foundations

Given the capacity for learning and adaptation in humans, it is unsurprising that the nature and origin of knowledge have been central concerns throughout history. Epistemology, the branch of philosophy focused on the theory of knowledge, has long sought to understand the foundations, scope, and validity of human understanding. Long before the advent of modern platforms like Facebook, this field was already a central concern of ancient philosophy, as exemplified by the quote from Plato’s *Meno* in Figure 2.

”And how will you enquire, Socrates, into that which you do not know? What will you put forth as the subject of enquiry? And if you find what you want, how will you ever know that this is the thing which you did not know?” *Socrates, Meno - 402 B.C.*



Fig 2. Illustration of Socrates and Meno - generated by gpt4o (05/16/2024)

In this widely studied dialogue, Plato, through the character of Socrates, introduces

¹Source: <https://www.artworkarchive.com/profile/thelilley/artist/salvador-dali>

the idea of anamnesis, which posits that knowledge is innate and resides within the soul. According to this theory, learning is essentially a process of recollection, where one accesses pre-existing knowledge within the soul by engaging in dialectical reasoning and introspection. While this theory may seem debatable today, it laid the groundwork for a long tradition of philosophers who have continued to explore this question for centuries. For instance, the 17th century saw a clear division among philosophers into two camps: rationalists, like Descartes (Figure 3), who proposed that truths are innate or can be discovered through logical reasoning, and empiricists, like Locke (Figure 4), who proposed that knowledge arises from sensory experience.

"I think, therefore I am."
*René Descartes - Dis-
course On The Method,*
1637



Fig 3. Portrait from René Descartes, after Frans Hals. (wikipedia)



Fig 4. Portrait from John Locke, after Godfrey Kneller. (wikipedia)

"Let us then suppose the mind to be, as we say, white paper [tabula rasa], void of all characters without any ideas; how comes it to be furnished? Whence comes it by that vast store, which the busy and boundless fancy of man has painted on it with an almost endless variety? Whence has it all the materials of reason and knowledge? To this I answer, in one word, From experience: in that all our knowledge is founded, and from that it ultimately derives itself."
*John Locke - Essay Concerning Humand
Understanding, 1689*

Building on these foundational debates, Immanuel Kant sought to reconcile the views of the rationalists and empiricists. In his seminal work, *Critique of Pure Reason*, Kant introduced the concept of transcendental idealism. He argued that while all knowledge begins with experience, not all knowledge arises from experience. According to Kant, the mind actively shapes experiences through a set of a priori concepts and categories that structure our understanding of the world. This synthesis implies that while sensory data provides the content of knowledge, the mind's inherent structures organize this data into coherent experiences.

The aim of this introduction is not to provide an exhaustive account of the history of human understanding of knowledge acquisition. Instead, this historical perspective is

intended to broadly highlight how this manuscript is grounded in a significant accumulation of theories, models, and experiments. Notably, the ideas presented thus far originate predominantly from the field of philosophy. This emphasis is intentional, as philosophy has opened up a large set of fundamental questions that are now closely investigated by the scientific community (for an insightful example of the links between modern physics and philosophy, see Etienne Klein's book "*Matière à contredire*", Klein, 2018).

The Cognitive Revolution

During the 20th century, a new community of researchers became particularly interested in the previously mentioned epistemological questions, developing innovative ideas and methodologies to understand the inner workings of the human mind. This endeavor, known as the "Cognitive Revolution," emerged in the first half of the last century, with numerous researchers engaging in interdisciplinary work. Once again, while it is not the purpose of this manuscript to exhaustively present all relevant contributions from this historical era, it is essential to understand the historical context that directly shaped subsequent research.

At the start of the 1900s, the field of psychology was dominated by the behaviorist paradigm. Behaviorists aimed to provide rigorous methods to understand behavior, in stark opposition to introspection, which was deemed subjective and unscientific. Researchers like John B. Watson and B.F. Skinner focused on observable behaviors and their responses to environmental stimuli, largely ignoring internal mental processes (J. Moore, 2011). However, this approach faced criticism for its limitations in explaining complex cognitive functions.

A significant example of the early interdisciplinary interest in studying the internal mechanisms of the mind is the Hixon Symposium of 1948 (Jeffress, 1951). Among the participants of this symposium, John von Neumann, a mathematician, presented a comparison between the brain and the computer, highlighting the potential for computational models of cognition. Warren McCulloch, a neurophysiologist, delivered a talk titled "*Why is the brain in the head?*" which explored the physiological basis of mental processes. Additionally, Karl Lashley, a psychologist, presented a critique of behaviorism, arguing that it was insufficient for studying certain types of behaviors, such as those "serially ordered" where steps in a sequence of actions are too rapid to be triggered by the preceding one. These interdisciplinary exchanges at the Hixon Symposium exemplify the early collaborative efforts that propelled the cognitive revolution (Gardner, 1985).

The efforts initiated during the first half of the 20th century eventually led to the establishment of the new field of "Cognitive Science." Historians often date the formal inception of this field to the 1960s, with the Symposium on Information Theory at MIT in 1956 frequently cited as a pivotal moment. George A. Miller, one of the organizers and a key figure in the development of cognitive science, expressed his vision in the following quote:

"I went away from the Symposium [on Information Theory at MIT, 1956,] with a strong conviction, more intuitive than rational, that human experimental psychology, theoretical linguistics and computer simulation of rational cognitive processes were all pieces of a larger whole, and that the future would see progressive elaboration and coordination of their shared concerns."

Miller, 2003

According to H. Gardner in "*The Mind's New Science: A History of the Cognitive Revolution*" (Gardner, 1985), cognitive science aims to address epistemological questions by studying the processes of the mind. The author identifies several properties shared by researchers in the field. First, cognitivists are interested in understanding mental processes at different levels of representation. This means that a certain cognitive ability can be described using symbols, schemas, or other abstract representations, as long as these descriptions provide useful insights into how the system functions. The choice of representation level is often guided by its utility in explaining and predicting cognitive phenomena. This principle is echoed in David Marr's tri-level hypothesis (Marr, 2010), which proposes that an Information Processing System can be understood at three levels: the computational level (what the system does and why), the algorithmic level (how the system performs computations and what representations are used), and the implementational level (how the system is physically realized). By selecting the most informative level of analysis for a given cognitive process, researchers can effectively bridge abstract representations with their underlying mechanisms and physical substrates.

Second, cognitive scientists emphasize the importance of computers in their work. Computers serve both as analogies and models for mental processes and as tools for designing and analyzing experimental data. Third, cognitive science is inherently interdisciplinary. Researchers recognize the benefits of integrating diverse fields and methodologies. For example, cognitive psychologist Zenon Pylyshyn described the connection between Artificial Intelligence (AI) and psychology (Friesen and Feenberg, 2007):

"Both fields are concerned with the same problems and thus must ultimately be judged by the same criteria of success. I believe that the field of AI is coextensive with that of cognitive psychology."

Zenon Pylyshyn, 1981

While subject to discussion, a report by the Sloan Foundation (Gardner, 1985) identified six core fields involved in cognitive science: psychology, philosophy, AI, neuroscience, linguistics, and anthropology. These fields are often illustrated by the cognitive hexagon (Figure 5, Gardner, 1985).

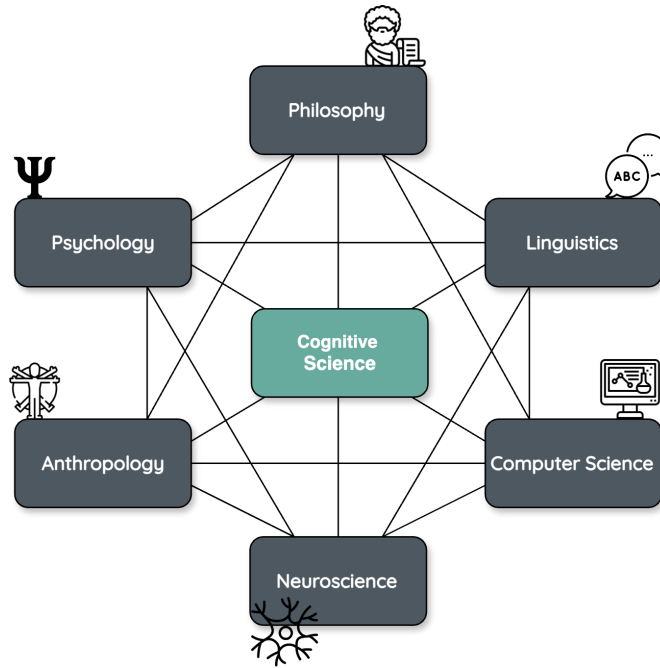


Fig 5. The cognitive hexagon

Positioning the Dissertation within this Historical Context

Building on the rich historical foundation of cognitive science, this research aims to advance our understanding of the mind's capacities and explore methods for their development or restoration. The proposed approach is deeply interdisciplinary, integrating principles from cognitive psychology, developmental psychology, Artificial Intelligence (AI), and educational sciences. As the rest of the manuscript will describe, our focus is on adapting training to participants' differences using individualized methodologies. Through this, we aim to address fundamental questions about human cognition. After this brief introduction and historical perspective, we can now turn to the details of our investigation and the core concepts explored throughout the dissertation.

Part I

Introduction and background

Chapter 1

Introduction

Aims

The aim of this general introduction is to provide an overview of the main concepts addressed in this manuscript. The scope will be intentionally broad, offering readers insight into the scientific and industrial context and highlighting the key questions that will be explored.

Contents

1.1 Non-Invasive Cognitive Enhancement: A Promising Future? . . .	29
1.2 Educational Technologies: Where Algorithms Meet Human Learning .	42
1.3 Objectives, Scientific and Industrial Context of the Thesis	47

Abstract

Following a brief presentation on the connection of this work to the field of Cognitive Science and its history, this chapter will introduce the core concepts of Cognitive Training, a field focused on enhancing or restoring cognitive function, and more broadly on cognitive skills development at all ages. Initially, an overview of the main concepts, methodologies, and challenges will be presented. This will be followed by a specific example: an attention task known as MOT. Introducing this task and discussing the challenges faced by current research will provide context for the scientific and industrial background of this thesis, particularly how individual differences can be addressed using intelligent, machine learning-based procedures to tailor interventions. After presenting the main research question, an outline of the dissertation will be provided.

Collaborators

Adolphe Maxime, Pech Marion, Oudeyer Pierre-Yves, Delmas Alexandra, Maurel Denis, Sauzéon Hélène ¹

¹Following Contributor Roles Taxonomy, all collaborators contributed for "Writing - Review & Editing".

1.1 Non-Invasive Cognitive Enhancement: A Promising Future?

1.1.1 What is a Cognitive Training ?

In “*Methods of Mind-Training: Concentrated Attention and Memory*”, Catharine Aiken (Aiken, 1896), an American teacher, describes how she trained her pupils’ concentration through various exercises. Her teaching philosophy is well encapsulated in this quote from the book:

“Botany, Natural History, Mineralogy, Geology, and other studies drawn from the realm of nature, may be successfully taught in a class well trained to listen, with little need of textbooks except for reference or for reading.” *Catharine Aiken, 1896*

As illustrated in Figure 1.1, Aiken asked students to look at a column of numbers and, after a few seconds of presentation, she reversed the switching blackboard. She began with a small set of numbers, gradually increasing the set size and assigning more challenging tasks, such as recalling the numbers in reverse order or performing arithmetic operations on each number. Through these methods, she aimed to enhance students’ perception, attention, and memory. In her book, in addition to describing the method, she finally discusses the positive impact of these exercises on academic performance and includes a letter from a music instructor who was delighted by the improvement in students’ musical abilities.

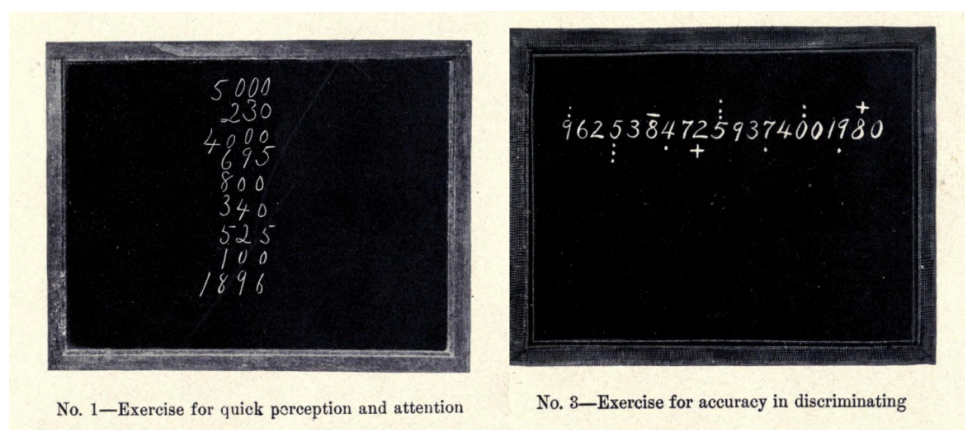


Fig 1.1. Illustration of activities proposed by Catherine Aiken, taken from the book *Methods of Mind-Training: Concentrated Attention and Memory* (Aiken, 1896)

Although Aiken’s description lacks scientific evidence and has been widely discussed (Whipple, 1910), her approach contains several features of what we now refer to as Cognitive Training (CT). CT is formally defined as the sustained and repetitive engagement in activities designed to improve or restore cognitive abilities through training or the acquisition of new cognitive skills. The activities studied in the field of CT are diverse and can include different cognitive exercises specifically designed to target certain abilities, referred to as the “direct route” (F. Joessel, 2022a), such as Working Memory (WM) (Jaeggi et al., 2008b; Klingberg et al., 2002), Executive Functions (EF) (Karch and Kray, 2021), or Perceptual (Romeas et al., 2016a) training. Alternatively, they can involve

everyday activities, termed the "indirect route" (F. Joessel, 2022a), such as playing chess, music training (Sala and Gobet, 2017b), meditation (Taylor et al., 2013), or playing video games (Boot, 2015, Bediou et al., 2023a). In practice, complex skill training usually offers a significant advantage over simple behavioral interventions, as research demonstrates that the former more effectively generalizes to real-life tasks outside the laboratory (Kovbasiuk et al., 2022). However, when focusing on complex skill training, it becomes more challenging to discern the specific properties that contribute to this enhancement. Despite the variety of these activities, all forms of CT share the common goal of understanding how the mind's abilities can be acquired, enhanced or restored. To assess the effects of CT, researchers use a variety of tasks that measure outcomes ranging from very similar to highly different from the trained task. This range is referred to as near to far transfer, indicating the distance from the trained activities (Barnett and Ceci, 2002a). By employing this approach, researchers can determine the extent to which skills or knowledge from the trained task transfer to other tasks.

Before proceeding further, it is important to clarify what is meant by "cognitive abilities" or "cognitive functions" terms we have used interchangeably and referred to vaguely since the beginning of this manuscript. As introduced in the historical perspective (), for some readers, the level of representation we will use might seem incomplete or open to interpretation. However, this description will generally be sufficient to discuss the impact and benefits of CT interventions. Cognitive functions can be defined as a wide range of mental processes that allow us to carry out tasks from the simplest to the most complex.

Attentional Control Attentional control is described as *"the ability to focus on the task at hand and to ignore sources of distraction or noise while at the same time constantly monitoring one's environment for new sources of information"* (Bavelier and Green, 2019). Essentially, it is the ability to direct cognitive processing towards relevant information. McDowd, 2007 distinguishes different modes of attention based on task context: selective attention, divided attention, sustained attention, and switching attention. Selective attention is the ability to focus on a specific object or task while ignoring irrelevant stimuli, allowing individuals to prioritize essential information. Divided attention refers to the capacity to attend to and perform multiple tasks or process multiple information sources simultaneously, crucial for multitasking. Sustained attention, also known as vigilance, is the ability to maintain focus over prolonged periods, especially during monotonous tasks. Switching attention involves shifting focus between different tasks or mental processes, allowing flexibility in responding to changing demands. Common tasks to assess these modes of attention include visual search tasks for selective attention (Müller and Krummenacher, 2006), dual-task paradigms such as the Useful Field Of View (UFOV) for divided attention (Yung et al., 2015a), the Continuous Performance Test for sustained attention (Shaked et al., 2020), and the Trail Making Test for switching attention (Bowie and Harvey, 2006).

Working Memory WM, introduced by A. D. Baddeley and Hitch, 1974, is defined as *"a system of components that holds a limited amount of information temporarily in a heightened state of availability for use in ongoing processing"* (Adams et al., 2018). This system comprises several sub-systems: the visuospatial sketchpad, responsible for storing visuospatial information; the phonological loop, which manages verbal information; and the episodic buffer, introduced later (A. Baddeley, 2000), which integrates information

from long-term memory. These sub-systems are coordinated by the central executive, which oversees processing and manipulation of information across the components. The capacity or span limit of WM has been a major research focus. Miller proposed the well-known “magic number 7 plus or minus 2” (Miller, 1956), indicating the number of items an individual can hold in WM. However, the embedded-processes model by Cowan, 2001 suggests a more limited span of approximately four slots. In Cowan’s model, WM is not a separate system but an activated subset of long-term memory, with the central executive controlling attentional resources. The span limit of four items is attributed to the restricted capacity of attention. More recently, research by W. J. Ma et al., 2014 has further refined this understanding by proposing that the span limit could be a function of the precision needed to encode information. In addition to capacity limits, the temporal dynamics of WM have also been explored. The Time-Based Resource-Sharing (TBRs) model by Barrouillet et al., 2007; Logie et al., 2021 posits that cognitive resources are shared over time between processing and storage. This model predicts that increased processing demands reduce the time available for maintaining information, leading to memory decay. Common tasks used to measure WM include the N-back task (Jaeggi et al., 2010), which assesses the ability to monitor and update information, and complex span tasks such as the Operation Span (OSPAN) task (Unsworth et al., 2005), which evaluate the ability to store and manipulate information simultaneously.

Executive Functions Closely related to the construct of WM, EF have been described as “*general-purpose control mechanisms that modulate the operation of various cognitive subprocesses and thereby regulate the dynamics of human cognition*” (Miyake et al., 2000b). Seminal works by Miyake and colleagues (Miyake et al., 2000b) propose that EF can be separated into three correlated constructs: updating, which refers to the ability to refresh the contents of WM; shifting, which involves the ability to switch between different mental representations; and inhibition, which is the ability to suppress dominant or automatic responses. Later, Diamond, 2013 expanded this model by including additional components such as reasoning, problem-solving, and planning, thus providing a more comprehensive framework for understanding the broad scope of EF. Common tasks used to measure EF include the Stroop test (MacLeod, 1991; Stroop, 1935), which assesses inhibitory control, or the Wisconsin Card Sorting Test (WCST), which evaluates cognitive flexibility and set-shifting abilities (Dehaene and Changeux, 1991; MILNER, 1963).

A broad set of other cognitive functions Learning, long-term memory, processing speed, language and verbal skills, perceptual speed and accuracy, spatial abilities are other critical cognitive functions. We will not provide an exhaustive overview of all cognitive functions but a useful taxonomy can be found in Webb et al., 2018d. Notably, this taxonomy has been developed using previous works relying on confirmatory factor analysis from intelligence researchers. For instance, the Cattell-Horn-Carroll (CHC) model (W. Schneider and McGrew, 2012, Flanagan and McDonough, 2018), one of the most influential frameworks, categorizes cognitive abilities into broad and narrow strata. The broad abilities include fluid intelligence (Gf), crystallized intelligence (Gc), visual processing (Gv), auditory processing (Ga), and processing speed (Gs), among others. Within these broad categories, narrow abilities represent more specific skills, such as quantitative reasoning or reading comprehension. The CHC model also illustrates how general intelligence (g) is an overarching factor that influences performance across various cognitive domains, highlighting both the interconnectedness and distinctiveness of

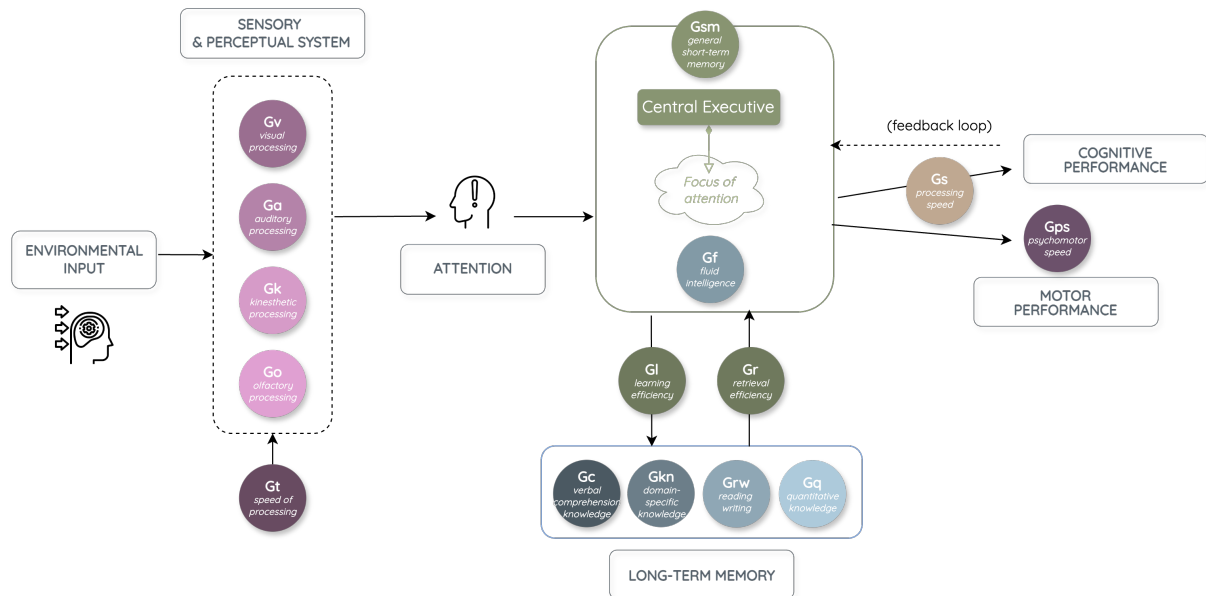


Fig 1.2. Illustration of the integration of the CHC taxonomy into task performance, inspired by W. J. Schneider and McGrew, 2013

individual cognitive abilities. Inspired by W. J. Schneider and McGrew, 2013, Figure 1.2 aims to demonstrate how this taxonomy can be integrated and used as parameters in an information processing model. Additionally, it underscores the importance of attention as a gateway to the cognitive system. Complementary to this, Figure 1.3 seeks to provide an overarching view of how different models and tasks relate to this taxonomy. It is important to note that all of these are constructs, and the terminology might sometimes vary across different studies and theoretical frameworks.

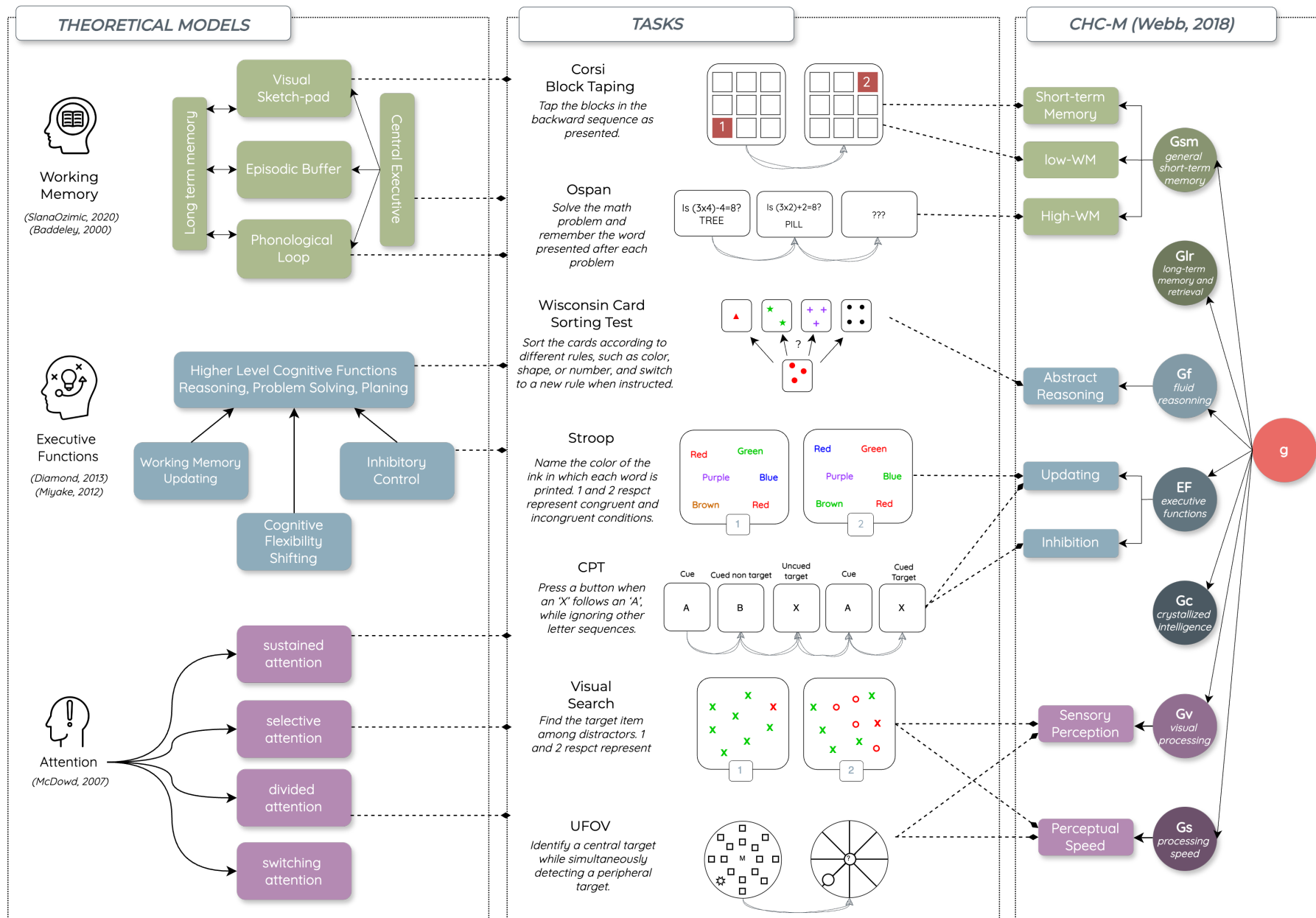


Fig 1.3. Overview of different theoretical models of working memory, executive functions, and attention, and how several tasks are used to measure some of their components. The right-hand side shows how all these tasks are integrated into the broad CHC-M taxonomy proposed in Webb et al., 2018c

1.1.2 Cognitive Training: a field in rapid progress

As highlighted in Strobach and Karbach, 2016, the CT research field has gained increasing importance over the last decades. A search on PubMed for the term “CT” from 1973 to 2016 revealed 2,471 peer-reviewed contributions. Extending this search to the period from 1973 to May 2024 results in 4,587 peer-reviewed contributions, underscoring the field’s rapid growth. This growth can be attributed to several factors, including technical progress (e.g., computerized setups, more sophisticated analysis tools) and theoretical advances (e.g., findings in cognitive and neuronal plasticity). Moreover, another significant factor contributing to the increasing number of studies and papers is the ongoing debate about the impact of CT. While some observations suggest that CT can lead to near to far transfer, many researchers argue that far transfer does not exist (Melby-Lervåg and Hulme, 2013a; Melby-Lervåg et al., 2016; Sala and Gobet, 2017b).

This debate is particularly intense due to the influence of commercial programs also known as “Brain training”, which often introduce noise into the global assessment of CT’s effectiveness (Simons et al., 2016a). The promises hold by CT such as improving cognitive performance, enhance memory or even delay effect of aging on the brain set the stage for an important market (Simons et al., 2016a). Therefore, in 2014, 70 researchers wrote an open letter, pointing out the challenge of insufficient compelling evidence in this complex and multifaceted field². In response, 111 researchers in 2016 acknowledged the need for improvement while highlighting the continued promise of various research directions³.

This debate is also notably important because it holds promise in several domains. First, in the educational context, some cognitive functions have been observed to correlate with academic performance. For instance, WM has been linked to reading comprehension, reading efficiency, and mathematical abilities (Titz and Karbach, 2014a). Additionally, EFs, particularly cognitive flexibility, have been associated with reading and mathematical abilities. There is promising evidence of the effectiveness of WM training and EF training in terms of both near and far transfer (i.e., improvement in untrained tasks or academic performance) for children with cognitive or learning difficulties, such as those using the Cogmed Training program (Klingberg et al., 2005) or the Jungle Memory program (Alloway et al., 2013). These interventions have also shown benefits for typically developing children, as seen with the Braintwister WM training (Karbach et al., 2015). However, like many areas in CT research, the results are mixed. A meta-analysis (Sala and Gobet, 2017a) found a significant overall effect size for mathematics and a marginally significant effect size for literacy/word decoding, yet concluded that WM training is not effective at improving children’s cognitive or academic abilities. These mixed results highlight the complexity of the field. Moreover, there are still open questions regarding meta-analysis outcomes — even when considering several moderators, we might still ask whether the absence of far-transfer effects is genuinely evidence of no far-transfer effect or a lack of modeling complex interactions between several moderators (Johann and Karbach, 2020).

Beyond educational contexts, CT could be highly valuable for activities requiring high-level expertise. This is particularly relevant when training specific skills related to a specific task at hand, such as in military training (Blacker et al., 2019) or athletic training, where context-specific CT, especially those relying on perceptual-cognitive elements, might provide a competitive edge (Harris et al., 2018b).

CT offers non-invasive, non-drug methods to restore cognitive functions. Systematic reviews and meta-analyses have found evidence of CT effects, showing overall improvement

²<https://www.cognitivetrainingdata.org/the-controversy-does-brain-training-work/stanford-letter/>

³<https://www.cognitivetrainingdata.org/wp-content/uploads/2014/open-letter.pdf>

in cognition (with small to moderate effect sizes) for post-stroke patients (van de Ven et al., 2016), post-acute traumatic brain injury (Hallock et al., 2016b), multiple sclerosis (Lampit et al., 2019), and Parkinson’s disease (I. H. Leung et al., 2015b). However, when looking at specific cognitive domains to assess far transfer, results were mixed, indicating that further research is needed to fully understand the effects of training.

Finally, in a world where aging is increasingly prevalent, cognitive decline has become a significant concern. For instance, in France, 20.5% of the population was above 65 in 2020 compared to 15.8% in 2000. Studies have shown that older adults with decreased cognitive abilities, such as processing speed, response inhibition, or attention, face a higher risk of falls (Muir et al., 2012) and increased healthcare costs (Pech, 2023). Numerous systematic reviews have identified small to moderate effect sizes on global cognition and transfer in older adults, with significant variability in observed outcomes across different reviews, depending on the specific domains trained and assessed. These observations hold true for both healthy older adults (Kelly et al., 2014; Kueider et al., 2012b; Lampit et al., 2014a; Shah et al., 2017b; Shao et al., 2015; Webb et al., 2018c) and older adults with mild cognitive impairment (MCI) (Butler et al., 2018; da Silva et al., 2022), dementia, or Alzheimer’s disease (Hill et al., 2017a). However, it remains challenging to gain a clear understanding of these outcomes, as reviews differ in their inclusion criteria, such as the type of training (e.g., CT, video games, the addition of physical activities) and differences in age range.

In sum, while more research is needed, CT offers a promising avenue for populations with cognitive decline, impairment, or neurodevelopmental disorders. This approach might also be beneficial for the general neurotypical population, as more and more press articles highlight a general shrinking attention span⁴, ⁵. For instance, in 2021, a survey of 2093 UK adults aged 18+ examined their perception of their own attention spans⁶. Results indicated that half (49%) felt their attention span was shorter than it used to be, and 66% believed that young people have shorter attention spans than those in the past.

Is There a Widespread Attention Decline in Modern Society?

If the public self-perceived a shrinking of their attentional span, it needs to be discussed in light of modern habits. Indeed, in today’s digitally pervasive environment, the prevalence of screens and the practice of using multiple screens simultaneously—a phenomenon known as “multiscreening”—raise significant questions about the impacts on human attention. In his book “Apocalypse cognitive”, Gerald Bronner describes a deregulated “cognitive market” where an enormous quantity of information is competing to get our attention (Bronner, 2021). As an example, a 2016 survey showed that people checked on their smartphones more than 221 times a day (once every 6 minutes Bronner, 2021, p71). Moreover, according to a Nielsen survey, a quarter of tablet or smartphone owners use their devices daily while watching TV, embedding multiscreening into daily routines (C. M. Segijn et al., 2017). This increasing integration of digital devices in our lives prompts a critical examination of how such habits may influence our attentional capacities, both positively and negatively. This question draws a lot of attention - notably for children development: for instance, in

⁴<https://edition.cnn.com/2023/01/11/health/short-attention-span-wellness/index.html>

⁵<https://time.com/6302294/why-you-cant-focus-anymore-and-what-to-do-about-it/>

⁶<https://www.kcl.ac.uk/policy-institute/assets/how-people-focus-and-live-in-the-modern-information-environment.pdf>

april 2024, a commissioned report ("*Enfants et écrans: A la recherche du temps perdu*" ⁷) was submitted to the french government - showing the interest to this potential public health issue.

Empirical studies present a complex picture of the cognitive consequences of multi-screening. C. M. Segijn et al., 2017 illustrates that individuals can effectively manage multiple screens and maintain their ability to recall information under conditions of adequate visual attention. This finding has already been demonstrated in earlier work by Jeong and Hwang, 2012, yet it contrasts with other studies suggesting potential cognitive costs associated with frequent screen switching (C. Segijn, 2016). Such contradictory findings underscore the need for a nuanced understanding of how divided attention, potentially trained through regular multiscreen usage, affects our ability to sustain focus over longer periods.

More generally, several systematic reviews (Domingues-Montanari, 2017, Radesky and Christakis, 2016) have discussed the impact of screens on health and development. First, studies have linked excessive screen exposure to poorer sleep quality, which in turn affects cognitive performance. For instance, poor sleep quality is associated with a negative cognitive bias toward negatively valenced stimuli and diminished performance in sustained attention tasks (Gobin et al., 2015, Marinelli et al., 2014). Second, prolonged screen time has shown to induce an impact by substituting with other activities. For example, screen time has been associated with higher obesity rates among children who watch more than two hours of TV daily (Twarog et al., 2015) or induce less exposure to natural environment as highlighted in this scoping review (Oswald et al., 2020).

In terms of cognitive development, results are mixed. It appears in the meta-analysis from Madigan et al., 2020 that the more the screen time is limited, better are the language skills. For instance, in this line, Pagani et al., 2013 show that the risk of language delay increased proportionately with time spent in front of the TV. Other works showed more nuanced results, for instance Nichols and Walker, 2005 showed that the viewing of "Dora the Explorer" resulted in 13.30 more vocabulary words acquired at the age of 30 months compared with non-viewers, in comparison with the viewing of "Teletubbies", which was negatively related to vocabulary acquisition. Finally, other studies emphasize the necessity of parental co-viewing to realize these benefits in younger children (Richert et al., 2010).

In terms of the impact of screen specifically on attention, a review by R. M. S. Santos et al., 2022, showed that screen times - higher than the guidelines - for children under 12 might be associated with a diminished attention. In adult population, some experimental data tends to show that there is a decreased in sustained attention: a 2004 study found that in average, workers (analyst, software developers, and managers) switch task every 2 minutes (González and Mark, 2004) while in 2016 a study by Mark et al., 2016 found that the median time was only 40 second. This findings have to be taken with care, as methods to calculate these durations were different and are noisy (as taken directly in the workplace, with small sample sizes). In sum, even though more scientific evidence is needed to confirm all these findings, it appears that this new way of living and working is a real concern to consider.

⁷<https://www.elysee.fr/admin/upload/default/0001/16/fbec6abe9d9cc1bff3043d87b9f7951e62779b09.pdf>

1.1.3 Cognitive Training: many challenges

After reviewing the potential importance of developing effective CT interventions for various populations in an ultra-connected society, this section aims to identify and understand the current challenges that need to be addressed.

Methodological considerations

As will be presented later in the manuscript, our work involves developing an intervention to study a specific CT. Like other researchers in the field, we have strived to ensure high standards in the methodologies employed. Despite these efforts, there remains a lack of published studies adhering to these gold standards. For example, as Chapter 2 of this manuscript will demonstrate, and as many systematic reviews and meta-analyses (Schmiedek, 2016a) in the field show, the quality assessment of experimental designs and methodologies is still, on average, quite poor. Randomized controlled trials remain the gold standard for assessing interventions. Unfortunately, many studies still lack randomization, participant and/or researcher blinding, and even control groups. The choice of a proper control group, particularly an active one, can be challenging (Au et al., 2020). For instance, previous works have shown that trained intervention groups showed effects when the control group practiced crosswords (e.g., J. L. Hardy et al., 2015) or read books (e.g., N. T. Y. Leung et al., 2015), but there was no effect if the control group also practiced video games (e.g., R. E. Mayer et al., 2019).

Additionally, the broader issue of the “replication crisis” is a concern, with a lack of standardization in training and assessment methods (Schmiedek, 2016a), as well as a lack of data sharing and other materials useful for replicating studies (such as scripts used for data pre-processing, cleaning, and analysis). More work needs to be conducted to measure the impact of interventions accurately. While well-conducted studies use sets of cognitive tasks—also known as cognitive batteries—different from the trained task to assess the impact of training, there is still a lack of standardization in how to properly assess interventions. An insightful example of this issue is the re-analysis of a systematic review conducted in 2014 (Lampit et al., 2014a), which showed different conclusions when using a different taxonomy to classify the cognitive tasks used for assessment (Webb et al., 2018c). More generally, assessing the impact of training is somewhat limited to computerized laboratory tasks. More work is needed to understand the potential transfer of training to real-world activities.

The quest for transfer

Transfer is of principal importance in the CT field. Historically, this idea was already developed in Plato’s *Republic* or John Locke’s *Some Thoughts Concerning Education* (Graves, 1915). Plato emphasized rigorous training in disciplines like mathematics and philosophy to cultivate critical and abstract thinking, aiming for well-rounded individuals capable of societal contribution. Locke, on the other hand, viewed the mind as a blank slate and stressed developing good habits and reasoning through practice and repetition, believing that training in one area could enhance overall cognitive abilities. In practice, contemporary schools sometimes promote this idea to students, suggesting that seemingly useless knowledge today will build and prepare their minds to handle future problems “Learning “Useless” Things in School Is (Usually) NOT Useless”, 2020.

However, there is currently a significant debate about the possible existence of transfer from trained to far, untrained tasks. Known as “the curse of specificity,” (Fulvio et al.,

2013) this phenomenon has been observed in various tasks, including perceptual tasks (Sowden et al., 2002) or chess (Gobet and Simon, 1996). In a famous experiments by Ericsson et al., 1980, authors trained a participant to hear random digits at a rate of one digit per second and recall the sequence. The participant trained on this memory span task for about an hour a day, 3 to 5 days a week, for over 1.5 years (230 hours of practice). The result showed an increase from 7 to almost 80 digits recalled, but the participant did not transfer this ability to a very similar task with letters instead of digits. While several studies do not show far transfer, others have demonstrated at least near to far transfer (e.g Jaeggi et al., 2008b). Therefore, the question remains open, as the lack of results could be due to the type of CT intervention or the way it is assessed (e.g., is the cognitive battery well-designed to detect a relevant effect?).

Existing theoretical models of transfer In line with the idea of formal education, CT transfer is often explained through the “muscle analogy” (N. A. Taatgen, 2021). Cognitive functions are likened to the “muscles of the mind,” suggesting that training them can enhance performance in other tasks. For instance, a tennis player who trains their biceps might see improved tennis performance because these muscles are used in the game. However, this muscle analogy doesn’t hold up due to the mixed results from CT studies. For instance, if we apply some of the mixed results to the tennis player example, it would imply that while training their biceps improves their tennis ability, the stronger biceps do not enhance performance in other tasks, such as climbing or swimming. This discrepancy highlights the limitations of the muscle analogy for CT transfer. It does not make sense that increased strength in one area would not benefit other activities requiring those same muscles, indicating a potential flaw in the aforementioned analogy.

This question relates to the mechanisms involved when inducing training in a participant. According to Lövdén et al., 2010a, brain plasticity is induced by the mismatch between the functional capacity of the system and the environmental demands that operate through experience. While this view is important for accounting for structural changes observed in imaging studies, it is limited in its ability to fully understand or predict the outcomes of CT. Although it provides guidelines for challenging participants beyond their current abilities, it does not describe how to achieve this. To this end, von Bastian et al., 2022a proposes that transfer could occur through two potential pathways: capacity enhancement or efficiency enhancement.

To illustrate the capacity-efficiency model, we can use an analogy with physical activity. An athlete can increase power in a movement by improving both speed and strength, as power is the product of speed and strength. By increasing muscle mass, a powerlifter enhances strength and, consequently, performance, representing capacity enhancement. Alternatively, the powerlifter could refine their technique to increase speed, representing efficiency enhancement.

Applying this to cognitive training, the first pathway expands cognitive capacity, allowing improvements to generalize to untrained tasks that rely on the same capacity limits. Neurobiologically, this increase in capacity is reflected by an increase in grey matter. The second pathway enhances efficiency in using the available capacity, evidenced by changes in functional brain connectivity and/or a reduction in the overall energy required to complete tasks. These two mechanisms are not mutually exclusive; training can provide broad benefits by enhancing both efficiency and capacity (von Bastian et al., 2022a).

On the capacity-enhancement dimension, a recent study (J. Ericson and Klingberg, 2023) proposed that there is task-specific capacity improvement and that improvements

manifest with different temporal dynamics. They showed that a two-factor model (task-specific and general capacity factors) explained the evolution of performance in both trained and untrained tasks. Their computational model analysis revealed that task-specific improvements plateaued after the first three days and that the transition time was participant-dependent. This finding suggests that averaging data can obscure individual trajectories.

However, in their review, von Bastian et al., 2022a identified that, given the limited transfer observed, it is possible that transfer mainly occurs through efficiency enhancement. Several frameworks have proposed that efficiency enhancement is related to the acquisition of cognitive skills that can be reused in untrained contexts. For instance, Woodworth and Thorndike, 1901 proposed the "identical elements theory of transfer," suggesting that shared knowledge components between two skills can lead to transfer. This was further developed by Singley and Anderson, 1989, who specified that these knowledge components are production rules (i.e. an action triggered by a condition as described in the Adaptive Control of Thought—Rational (ACT-R) architecture J. R. Anderson et al., 1997; Ritter et al., 2019). However, these production rules are quite task-specific and explain limited transfer between tasks. Therefore, N. Taatgen, 2013 introduced an extension of the ACT-R model, named the primitive information processing element (PRIMS), to explain far transfer. In this model, learning specific cognitive tasks involves acquiring general cognitive skills, or patterns of routing information through the cognitive system. This model proposes categorizing production rules into task-specific and task-general steps, allowing for zero-shot learning (i.e., the ability to apply learned knowledge to new, unseen tasks) by combining already mastered skills. This computational model has shown promise, with N. A. Taatgen, 2021 demonstrating how such models can predict transfer, such as enhanced proactive cognitive control strategies during multitasking training that translate into far transfer.

The idea of creating or enhancing reusable cognitive skills has also been proposed by Gathercole et al., 2019, who suggested that cognitive routines might transfer to other tasks if the needed cognitive routine is novel and has been created during training. Additionally, this idea of cognitive routine acquisition can be linked to the triarchic theory of learning (Chein and Schneider, 2012), which proposes that novice learners of a new cognitive task initially rely on their metacognitive system to generate and establish new behavioral routines. These routines may involve strategies such as grouping information or mental imagery. Once they are formed, the metacognitive system's role diminishes, and learners engage their cognitive control network to execute them. With sufficient practice, learners progress from controlled to automatic task execution (Feltovich et al., 2018).

Another account attempting to explain transfer, not mutually exclusive from efficiency enhancement through cognitive skill acquisition, is the learning-to-learn hypothesis. Previous works on Action Video Games (AVG) have shown that these games positively impact vision and cognitive function (Bavelier, Green, et al., 2012) and lead to faster learning on various untrained tasks (R.-Y. Zhang et al., 2021b). Rather than positing that trained tasks enhance task-specific strategies shared with the assessment task, Bavelier, Green, et al., 2012 suggested an other hypothesis: training leads to learning how to learn. In practice, quickly learning the assessment task can slightly impact the overall measured performance, as participants who adapt faster to the evaluation guidelines may perform better, despite having similar cognitive skills to those who adapt more slowly (Kattner et al., 2017). Indeed, in most CT intervention designs, performance is assessed through a cognitive battery where performance is averaged across a block of several trials. If a group is faster than an other group to learn or to adapt to the task, the average performance

will be higher. As a result, the final between group difference is not related to the actual performance but a learning component within the evaluation session. Specifically, Bavelier, Green, et al., 2012 posits that three computational properties of the learning-to-learn theory can be enhanced via AVG (or CT): resource allocation (i.e better distribution of attention for efficient decision making), knowledge acquisition, and improved learning algorithms. While knowledge acquisition and learning algorithms are properties closer to cognitive skills acquisition, resource allocation can be improved by enhancing attentional control or probabilistic inference (C. S. Green et al., 2010, Kemp et al., 2010, Greenwood and Parasuraman, 2016). The next section will focus on tasks that foster the learning-to-learn hypothesis, specifically perceptual-cognitive training and video games.

1.1.4 Perceptual-Cognitive Training and the multi-object tracking task

Properties of Action Video Games

Previous sections emphasized that transfer might occur via different mechanisms (e.g acquisition of new cognitive routines or learning-to-learn hypothesis). Particularly, a domain where this transfer appears to be particularly present is AVG training. A recent metanalysis (Bediou et al., 2023a) show that AVG players included in 105 cross-sectional studies outperformed non gamers with large effect ($g = 0.64$, 95% CI [0.53, 0.74]) and that the causal relationship was confirmed within the 28 intervention studies included in the study with small effect size ($g = 0.30$, 95% CI [0.11, 0.50]). More specifically, the cognitive constructs that seemed to be enhanced by AVG training are perceptual processing, spatial cognition (Bediou et al., 2018b; Blumberg, 2014; J. Hilgard et al., 2019), top down attention (C. S. Green and Bavelier, 2003b), verbal cognition or components of executive functioning such as multitasking or inhibition (Bavelier and Green, 2019). This was also confirmed by Mark et al., 2016, who systematically reviewed the neuroanatomical correlates of complex skill acquisition after video game training and found evidence of increases in both cortical and subcortical areas. In young adults, these included the frontal cortex (prefrontal cortex, dorsolateral prefrontal cortex, precentral areas), parietal cortex (precuneus), temporal cortex (middle temporal gyrus), and cerebellum. For older adults, increases were observed in the frontal cortex (precentral gyrus, frontal gyrus, frontal eye field), superior parietal gyrus, cerebellum, and hippocampus.

However - as observed in the literature - not all video games are equal (Blumberg, 2014). As such, previous works have started to extract intrinsic properties of AVG that might be beneficial (C. S. Green et al., 2010). Particularly, F. Joessel, 2022a identified six features leading to the attentional control enhancements previously observed after AVG training. These features are:

- Difficulty scaffolding and activities in the Zone of Proximal Development (ZPD), i.e nor to difficult, nor to easy)
- Variable feedback and a valid reward structure (immediate and clear feedback and a composition of short term reward and long term rewards)
- High variability
- High pacing (feeling of time pressure)

- Encourage model-based learning (i.e in addition to enhancing attentional control, participant must built useful internal representation that would guide on what ressources should be allocated to)
- Need for switching between divided attention and focused attention

Against this backdrop, a natural hypothesis is that cognitive training based on these properties should lead to similar progress as those observed in video game training. Moreover, bridging the gap between video games (i.e., complex and uncontrolled tasks) and cognitive training tasks (i.e., simpler but controllable exercises) by adding gamified components to cognitive training interventions would likely enhance adherence. Indeed, as previous works have demonstrated (Belchior et al., 2019), video games are more engaging and can be used to leverage intrinsic motivation (**<empty citation>**). As such, the next section will present the Multi-Object Tracking (MOT), a perceptual and cognitive task that might move us in this direction.

The multi-object tracking task

In a subsequent experiment described in F. Joessel, 2022a, several properties were embedded in a training based on the Multi-Object Tracking (MOT) task. The MOT task is a dynamic task where participants track several targets on a screen while inhibiting several distractors. This task (the core subject of Chapter 3 of this dissertation) has been extensively studied to assess different models of visual attention and tracking (Cavanagh and Alvarez, 2005b; Oksama and Hyönä, 2004b; Z. W. Pylyshyn and Storm, 1988a) and as a training task (Romeas et al., 2016a) with observed near to far transfer effects (Corbin-Berrigan et al., 2018; Legault and Faubert, 2012b). As previous studies have investigated (J. Li et al., 2019a; Nyquist et al., 2016a), F. Joessel, 2022a proposed updating the MOT task by integrating either a concurrent detection or discrimination task in addition to tracking. Four conditions of the dual MOT was studied:

- RQ1- Pacing: constant low demand MOT + adaptive⁸ detection task
- RQ2- Divided attention: adaptive MOT + constant low demand detection task
- RQ3- Focused attention: constant low demand MOT + adaptive discrimination task
- RQ4- Combined: adaptive MOT + adaptive discrimination task

The authors demonstrated that all training groups improved attentional control compared to a no-contact control group. However, an analysis of the training trajectories revealed that the combined condition, where the difficulty was managed both in the MOT task (e.g., staircase adjustments on speed and number of targets) and the discrimination task (e.g., staircase adjustments on secondary stimulus presentation), showed less progress in training performance enhancement compared to the other conditions. Notably, the combined group was the only one without a participant reaching an excellent level of performance. This observation is intriguing, suggesting that while individual components work well independently, their combination may introduce complexities that hinder overall effectiveness.

To address the issue of providing a tailored path for each participant within their ZPD, F. Joessel, 2022a leveraged a staircase strategy. Staircase methods are common

⁸*Adaptive*: difficulty is managed through a staircase algorithm

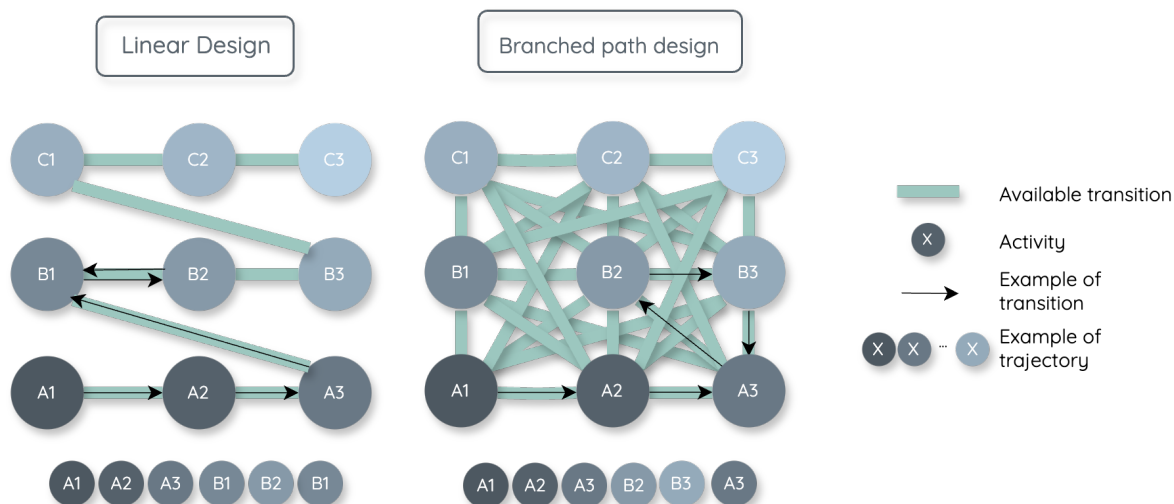


Fig 1.4. Illustration of linear vs branched path designs.

adaptive difficulty procedures that increase difficulty after three consecutive successes and decrease difficulty after one failure. While this strategy is a simple algorithm for scaffolding learning—a feature identified as crucial for fostering engagement and learning—the results obtained with the “combined” version revealed several issues. In F. Joessel, 2022a’s experiments, the difficulty of the combined condition was managed by simultaneously increasing all difficulty parameters when participants succeeded three times in a row. Although this strategy allows for evolving activity difficulty, the simultaneous increment might be too significant when participants approach their maximum skill levels, potentially hindering learning.

As discussed in Chapter 2 of this manuscript, the staircase procedure likely presents other drawbacks that can negatively impact learning, such as creating a unique trajectory and a lack of variability. Although the staircase method has shown promise (Pedullà et al., 2016b), in practice, the observed issues may relate to a fundamental property of the procedure: while these algorithms can meet participants’ limits in terms of ability and skills, they do so along a pre-specified route. This strategy is often described as following a “linear design”. Fortunately, educational scientists and technologists have long considered other ways to better handle participant heterogeneity.

1.2 Educational Technologies: Where Algorithms Meet Human Learning

If linear designs such as staircase methodologies scaffold learning in a straightforward manner, “branched-path” designs offer a way for participants to learn on a non-predefined route. As exemplified in Figure (1.4), educational content based on “branched-path” designs allows for learning trajectories tailored to individual learners’ needs. But what exactly are learners’ needs?

1.2.1 Zone of Proximal Development (ZPD)

We introduced this manuscript with a quote from Plato’s dialogue, *Meno*, and highlighted the idea that epistemological questions of interest in cognitive sciences take inspiration

from long-standing, ancestral interrogations. Reading this dialogue can be seen as an early example of a concept introduced by Vygotsky in 1978 (Vygotsky, 1978) and already discussed in previous sections: the ZPD. In *Meno*, Plato continuously questions a slave (the learner) with geometric problems, guiding him to realize what he knows and understand why his initial reasoning led to an incorrect outcome and how to eventually reach the correct solution. This social interaction exemplifies the idea that learners can exceed their current knowledge when given appropriate assistance. In chapter 6 of his book *Mind in Society* (Vygotsky, 1978), Vygotsky discusses the links between learning and development, proposing that effective pedagogical content lies in the zone between the learner’s current developmental stage (i.e what the learner can do independently) and the potential developmental stage (i.e what the learner can achieve with guidance or collaboration). This concept has led to the derivation of various theories and continues to guide the development of educational content (Gredler, 2012, Obukhova and Korepanova, 2009).

1.2.2 Motivation and Optimal difficulty

While the concept of the zone of proximal development (ZPD) underscores the importance of guided learning, it does not fully explain how individuals learn independently. Many people acquire knowledge and skills on their own, driven by motivations that have been extensively discussed. Historically, White, 1959 challenged the behaviorist perspective, exemplified by Hull, 1943, which posited that motivation stemmed from primary drives such as hunger, thirst, or sex. For example, Hull’s theory could not adequately explain a baby’s behavior when it explores a room and examines various objects without any apparent biological drive behind it. This innate curiosity and drive to understand the environment suggest a different source of motivation. White argued that such behaviors are better explained by a drive for competence, where engaging with the environment and mastering new skills are intrinsically rewarding experiences.

This idea opens the door for several accounts on what kind of activities or stimuli triggers motivation. For instance, Berlyne, 1960 introduced the concept of intermediate novelty, suggesting that individuals seek experiences that are optimally novel — not too familiar, yet not too unfamiliar. Other accounts proposed were optimal incongruity (Deci, 1975, discrepancy between perceived and standard level of a stimulus), knowledge gaps (Loewenstein, 1994) or the notion of optimal challenges (tasks that are neither too easy nor too difficult). On this later idea, Csikszentmihalyi, 2000 expanded and developed the concept of flow, a state where an optimal challenge, clear goals, and immediate feedback result in intense concentration, loss of self-awareness, a sense of mastery, and a distorted perception of time. Specifically, Ryan and Deci, 2000 proposed a conceptual framework in which motivation is categorized into intrinsic and extrinsic types. Intrinsic motivation involves engaging in activities for their inherent satisfaction, while extrinsic motivation involves performing tasks for external rewards or to avoid negative consequences. Curiosity, as more casually used, can be seen as particular form of intrinsic motivation (Oudeyer et al., 2016a).

As such, these concepts provide valuable insights for educational practices. For instance, Freeman et al., 2014 demonstrated through a meta-analysis that active learning and intrinsically motivated students achieve better outcomes in science, technology, engineering, and mathematics courses. In intervention settings, it was also observed that curiosity foster learning and memory retention. For instance, Kang et al., 2009 showed that human adults show greater long-term memory retention for verbal material for which they had expressed

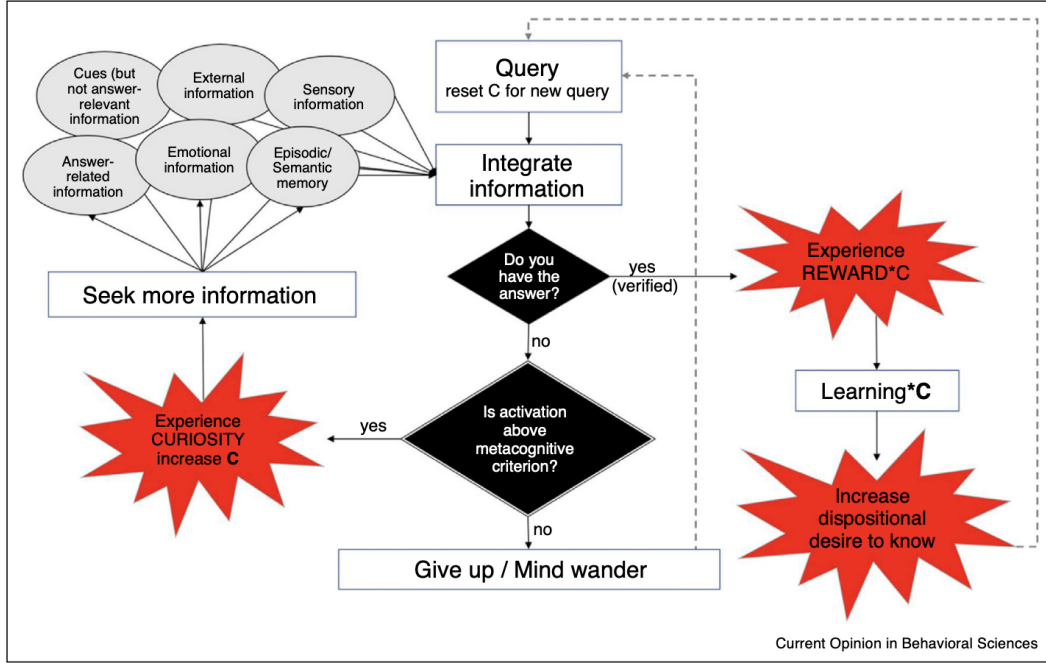


Fig 1.5. The Region of Proximal Learning, *Figure taken from Metcalfe et al., 2020.*

high curiosity than for low-curiosity questions. In line with these observations, Metcalfe et al., 2020 examines how epistemic curiosity (i.e., the desire to seek information) impacts self-regulated learning, proposing the framework of the Region of Proximal Learning (RPL). Within this framework, epistemic curiosity is characterized as a metacognitive feeling state that arises when an individual is on the verge of knowing or understanding (tip-of-the-tongue feeling). Metcalfe et al., 2020 argues that curiosity is influenced by a metacognitive appraisal: when too little information is available, individuals are unlikely to invest time in it, leading to a lack of curiosity (i.e., giving up or mind-wandering). Conversely, if the information is almost known, placing the individual in the RPL, they enter the left-hand loop of Figure 1.5, which enhances curiosity.

1.2.3 The Learning Progress (LP) hypothesis

If previously presented concepts and theories are intuitive and useful for explaining certain behaviors, recent works have questioned their applicability as mechanisms to actually generate curiosity (Oudeyer et al., 2016a). The concept of intermediate novelty or challenge can be hard to properly define, as it requires a reference frame for estimation. Additionally, heuristics based on searching for novelty or complexity can be inefficient in large environments (Schmidhuber, 1991). This is illustrated by the noisy TV problem (Oudeyer and Kaplan, 2007), where an intrinsically motivated agent exploring an open-ended environment, seeking high prediction error or novelty, will be attracted to sources of noise, such as a TV displaying static.

Therefore, the LP hypothesis has been proposed as an alternative mechanism. This hypothesis suggests that intrinsically motivated agents are not solely attracted to activities with high prediction error or uncertain outcomes. Instead, they are drawn to stimuli where their predictions improve over time. This idea relates to previous work, as activities that are too easy or too hard yield constant performance levels—either 100% or 0%, resulting in no LP. Consequently, exploration naturally gravitates toward activities of intermediate

difficulty.

Moreover, while several works have suggested that the learning process itself does not impact curiosity and motivation (Kang et al., 2009; Stahl and Feigenson, 2015), the LP hypothesis proposes the opposite. It posits that experiencing learning triggers an intrinsic reward, reinforcing the motivational state. Additionally, as mentioned in the previous paragraph, learning in a state of motivation will foster retention and learning. In sum, the LP hypothesis states that there is a closed feedback loop between learning and curiosity-driven intrinsic motivation (Oudeyer et al., 2016a).

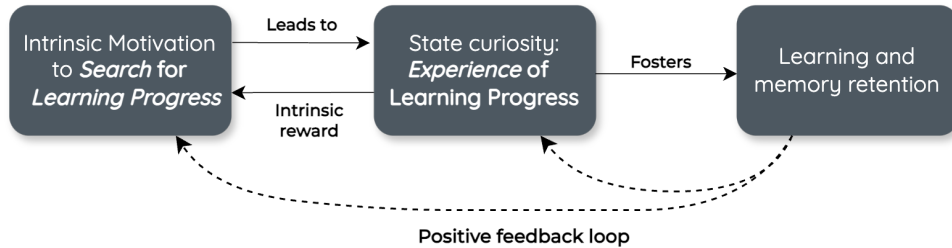


Fig 1.6. The Learning Progress Hypothesis (*adapted from Oudeyer et al., 2016a*)

1.2.4 Educational Technologies

In addition to developing theories on human learning and development, practical interventions have suggested that technologies can enhance instructional designs. For instance, the 2-sigma problem identified by Bloom, 1984 demonstrated that individually tutored students performed two standard deviations (sigma) better than those in conventional group instruction. Although later discussions reduced this effect size to 0.79-sigma (VanLEHN, 2011a), this reduction still underscores the significant impact of individualized training.

“If, by a miracle of mechanical ingenuity, a book could be so arranged that only to him who had done what was directed on page one would page two become visible, and so on, much that now requires personal instruction could be managed by print.”

Edward Thorndike, 1912

As Thorndike’s quote (1.2.4, Thorndike, 1912) illustrates, the concept of technology-enhanced learning predates computers. Early 20th-century psychologists and engineers, such as Sidney Pressey and later B.F. Skinner, proposed mechanical teaching machines based on the behaviorist principle that reinforcement strengthens associations between concepts. These machines delivered curricula in small, incremental steps, enabling self-paced learning with instant reinforcement for correct responses (Watters, 2023).

The advent of computers expanded these approaches. Early methods, inspired by Skinner, included Computer-Assisted Instruction (CAI) systems that provided feedback and hints. A notable example is PLATO (Programmed Logic for Automatic Teaching Operations), which supported 1,000 terminals and allowed teachers to design learning modules using a programming language. Despite this, most early systems employed a linear design, setting a predetermined route through the content (Yazdani and Lawler, 1986). The 1980s saw the emergence of Intelligent Tutoring System (ITS) with branched-path designs, offering individualized learning experiences (Self, 1998). CAI and ITS differ in focus: CAI

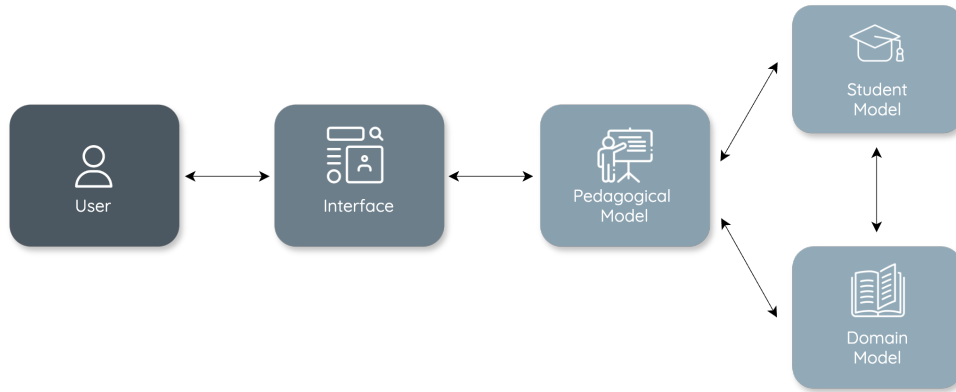


Fig 1.7. Core modules of an ITS.

are interaction-centered and address specific student groups' needs and tailors programs accordingly, while ITS are knowledge-centered and embody instructional principles to cover a range of teaching tasks (Chabay and Sherwood, 1992). More specifically, ITS models specify what to teach and how to teach it (Vandewaetere et al., 2011c), drawing from cognitive psychology, computer science, and AI (Chabay and Sherwood, 1992).

Adaptive systems operate within three adaptive loops (V. Aleven et al., 2016; Monterrat et al., 2017). The design loop uses student data to inform system design. For example, a mathematics course for third graders might tailor division problems based on previous performance data. The task loop selects tasks based on learner data, ensuring foundational skills like addition and multiplication are mastered before advancing to division. The step loop adapts tasks in real-time based on student responses, such as adjusting the number of digits in a division problem. For clarity, we will refer to these as macro-adaptive procedures (design and task loops) and micro-adaptive procedures (step loop) (E. L.-C. Law, Kickmeier, et al., 2008; Vandewaetere et al., 2011a).

Typically, adaptive systems like ITS include four core modules (Figure 1.7, Vandewaetere et al., 2011a): a learner model (describing the student's state), a domain model (representing the instructional content), a pedagogical model (encompassing teaching strategies), and an interface model (presenting content to learners). These modules enable various sources of adaptation, such as knowledge (e.g., Bayesian Knowledge Tracing (Pelánek, 2017a) or Cognitive Tutor (J. R. Anderson et al., 1995a)), learning strategies (El-Sabagh, 2021), cognitive profiles (Mampadi et al., 2011), metacognitive strategies (e.g., MetaTutor (Azevedo et al., 2010)), gamification (Monterrat et al., 2017), and motivation (Cordova and Lepper, 1996). These examples are not exhaustive but illustrate some common techniques.

We have discussed when and what to adapt, but the primary question remains: how to implement these adaptations effectively. While early CAI systems relied on rule-based expert systems (If X then Y else Z), modern Intelligent Tutoring System (ITS) utilize a variety of AI techniques. For instance, production-based systems like ACT-R (J. R. Anderson et al., 1997) are common for domain models, and Bayesian Knowledge Tracing is frequently used for student models to track progress and predict mastery levels. Pedagogical models often employ Reinforcement Learning (RL) (Doroudi et al., 2019), treating the student as an environment for pedagogical actions. In this framework, actions correspond to educational activities, and the system observes the outcomes, rewarding effective actions. These methods are just a few examples to give readers an idea of the important lines of research, and further details on these and other techniques will be provided in Chapters 2

and 4.

Starting from the issue of inter-individual differences in cognitive training and the existing methods to adapt programs, this section reviewed common techniques used in educational contexts to individualize training. The next section will integrate these different lines of work to outline the objectives and scientific context of this thesis.

1.3 Objectives, Scientific and Industrial Context of the Thesis

1.3.1 Collaboration between Bavelier Lab and the Flowers Team

This broad introduction has contextualized the scientific motivation behind this work. In this manuscript, two lines of research are proposed to be connected.

First, as previously discussed, there are unresolved questions regarding what constitutes an efficient CT program. In the Brain and Learning lab at University of Geneva ⁽⁹⁾, research has focused on identifying intrinsic features of AVG that could enhance training effects (F. Joessel, 2022a). However, several limitations were identified, particularly concerning the management of difficulty levels and tailoring them to participants in activities with numerous parameters. More generally, this addresses the challenge of accommodating individual differences and personalizing training trajectories effectively.

Second, educational sciences and technologies have long sought to address the issue of individualization. At Inria, the Flowers team, through the Kidlearn project, has proposed innovative ideas and algorithms to tackle this problem. Flowers is an interdisciplinary team dedicated to studying models of open-ended development and learning. One of their significant interests is developing educational technologies that personalize learning experiences. In this context, the Kidlearn project introduced an ITS with a novel algorithm called Zone of Proximal Development and Empirical Success (ZPDES) (Clement, 2018). The primary aim of this algorithm was to enhance motivation and learning in children's mathematical content by operationalizing the LP hypothesis. This initial objective was evaluated through several large-scale experiments where 7-8 year old schoolchildren learned to decompose numbers by manipulating money (Clement, 2018). Results showed that, compared to control procedures (e.g., predefined expert sequences), ZPDES led to higher levels of performance, more diverse learning trajectories, and increased motivation levels. Following these successful outcomes, the algorithm was tested on other populations, such as individuals with autism spectrum disorder (Mazon et al., 2023), and in different tasks, such as health education (Delmas et al., 2018). Several benefits were again observed with the use of the algorithm. It is therefore worth investigating whether the algorithm would yield similar benefits when transferred to other domains with different constraints.

A quick look at ZPDES

A detailed description of the algorithm is provided in Chapter 5; here, we aim to offer a quick and high-level overview of the key aspects of ZPDES implementation. First, ZPDES is based on the RL framework, with an instance of the algorithm created for each participant. It uses a micro-adaptive loop to provide optimal activities in real time. This loop works by proposing an activity, receiving feedback, updating the algorithm internal

⁹<https://www.unige.ch/fapse/brainlearning/>

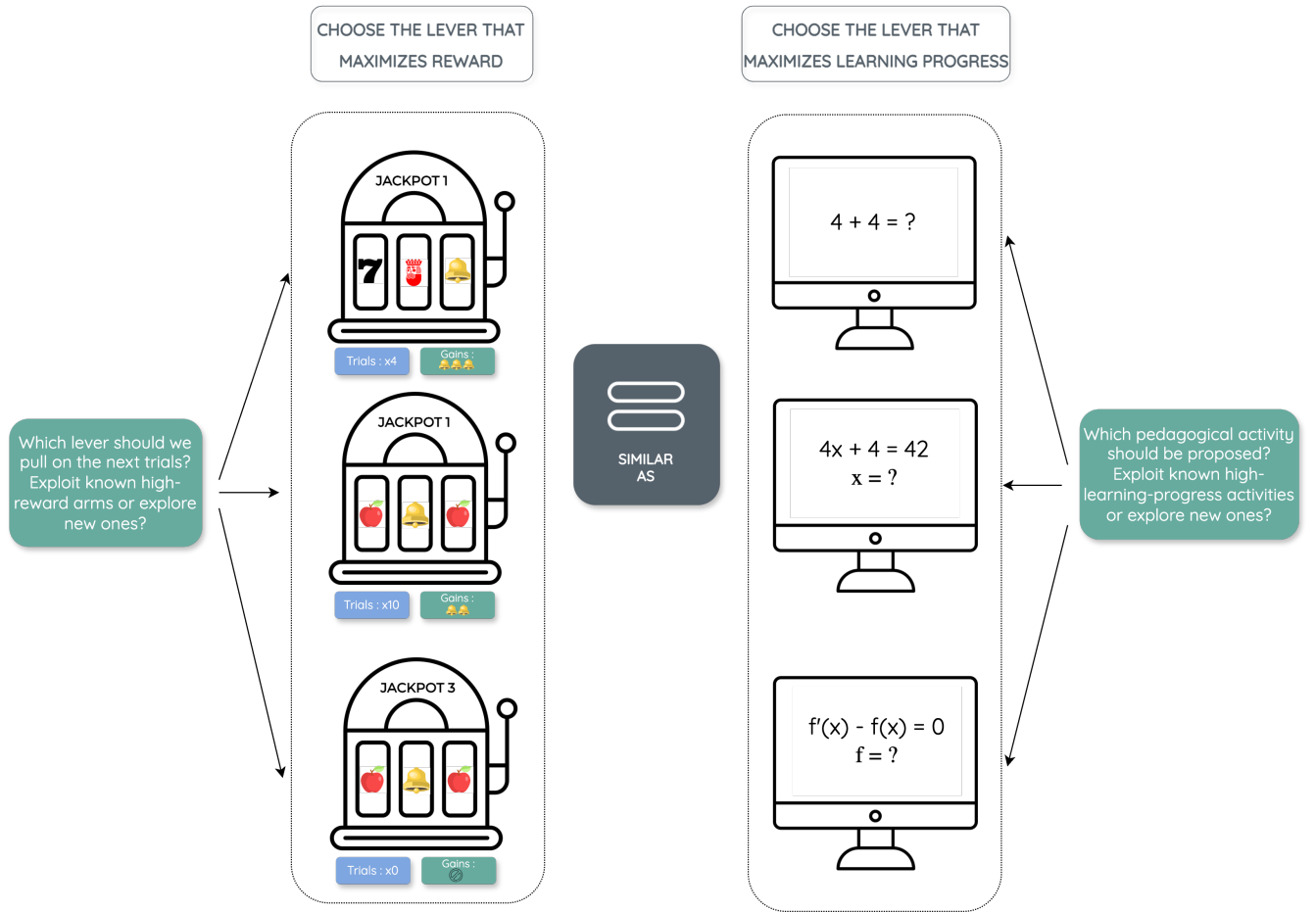


Fig 1.8. A visual representation of the multi-armed bandit problem, illustrating its application in both a classical casino setting and an educational context.

state, and then proposing a new activity. To operationalize the LP hypothesis, ZPDES targets activities with high learning progress. Identifying these activities is formulated as a multi-armed bandit problem. As illustrated in Figure 1.8(a), a multi-armed bandit scenario is analogous to a casino where multiple slot machines are available, and the player must determine which machine offers the highest expected gains. This scenario encapsulates a common situation in RL, known as the exploration/exploitation dilemma. Here, the choice is between continuing to play on known profitable machines (e.g., Jackpot 1 in Figure 1.8) or exploring potentially more profitable machines (e.g., Jackpot 3 in Figure 1.8). In an educational context, the slot machines represent pedagogical activities, and the gains represent learning progress. By framing the problem this way, ZPDES employs a multi-armed bandit algorithm to explore the space of possible activities, giving more opportunities to those with high learning progress. In addition to this LP-based sampling, ZPDES implements a model of ZPD to facilitate the exploration of the activity space. Therefore, as shown in Figure 1.9, the set of proposed activities evolves over time: when an activity is mastered, it is removed from the ZPD, and when the mean success rate for the available activities is sufficiently high, a new activity is added to the ZPD. In summary, the algorithm Zone of Proximal Development and Empirical Success derives its name from the two mechanisms it implements: restricting possible activities by modeling the **ZPD** and sampling activities with high learning progress estimated via **Empirical Success (ES)**.

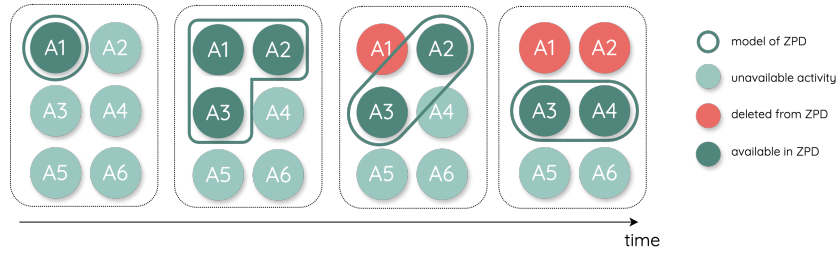


Fig 1.9. A visual representation of the evolution of the model of ZPD through time, implemented in ZPDES.

To summarize and conclude this section, our aim is to investigate whether algorithms driven by the search for LP can be connected to a CT program centered on the MOT task.

Industrial Context: Onepoint Company

In addition to the scientific context, this thesis is situated within an industrial framework through collaboration with Onepoint, a digital services company with over 3,500 employees. Onepoint is an innovative company continuously exploring new models. A few years ago, they established a research and development (R&D) team to contribute to scientific research in areas such as occupational psychology, cognitive psychology, software engineering, and AI. Although recent, there is already a wide variety of work, such as the development of a serious game to raise awareness about quality of life at work Massart et al., 2024, research on how training in reasoning tasks can enhance reasoning performance Franiatte et al., 2024, studies on the psychosocial impact of remote work during the COVID-19 context Massart and Chaumon, 2021, innovations in adaptive learning algorithms Vassoyan et al., 2023, advancements in explainable AI methods Lemberger and Saillenfest, 2024, or the use of multi-agent systems to assess maintainability in software development Bertrand et al., 2021. Building on this diverse body of work, the R&D team aims to establish a pool of experts to better guide their customers. As such, this thesis benefited from both the sponsorship and the numerous interactions with the industrial sector, enriching the research with practical insights and applications.

1.3.2 Main Research Questions

In 1895, Catharine Aiken, the American teacher who advocated for training her students' attention in the classroom, wrote a sentence that resonates with our current work:

“I found that innate curiosity which is shared in a greater or less degree by all, ambition, and a desire to excel would serve to stimulate the mental activities and concentrate the attention. These emotional states and tendencies, together with association, while none or all of them would create attention, would furnish the work of directing and holding the attention.”

Catharine Aiken, 1985

Building on Aiken's insight and the concepts introduced in the previous sections, this thesis explores the use of individualization algorithms to adapt CT interventions. The primary research question guiding this investigation is:

What is the impact of individualized cognitive training on cognitive performance and subjective experiences?

To further investigate this primary question, several sub-questions are addressed:

- What are the current strategies for personalizing cognitive training?
- How can the MOT task be tailored to participants, and what difficulty parameters are involved?
- How can the evolution of cognitive performance be assessed?
- Is an ITS based on the Learning Progress hypothesis (ZPDES) effective in the context of CT?
- How should learning trajectories and the evolution of subjective experiences be considered in evaluating the system?

1.3.3 Organization of the thesis

To address the research questions, this dissertation follows the format of a thesis by publication. The work is structured around five articles, progressing from a literature review to methodological considerations, and then culminating with experimental data and conclusions:

- In Chapter 2, we will systematically review the existing technologies used for individualized CT (article *in press*, preprint available in Adolphe et al., 2023). This chapter will provide a comprehensive examination of current methods and tools, setting the stage for our proposed interventions.
- In Chapter 3, we propose a narrative review of the MOT task, focusing on the impact of various parameters on performance and training outcomes (article *in press*). It will highlight key factors that influence effectiveness in CT. Together, Chapters 2 and 3 will provide an overview of existing knowledge and form the foundation for the empirical studies that follow.
- In Chapter 4, we will present the development and validation of a cognitive battery (published in *Frontiers in Psychology*, Adolphe et al., 2022a) designed to support and measure the effectiveness of the individualized CT interventions proposed in this thesis. This chapter will detail the methodologies and validation processes involved.
- In Chapter 5, we will present the results of applying our individualized procedures to a population of young adults (article *in prep*). We will analyze the data to assess the effectiveness and impact of the interventions.
- In Chapter 6, we will extend our study to older adults, presenting the outcomes of the individualized procedures in this demographic (article *in prep*).
- In Chapter 7, will synthesize the findings, discussing the implications of our research for the field of CT. It will also provide conclusions and suggestions for future research directions.

Part II

Litterature Review

Chapter 2

Exploring the Potential of Artificial Intelligence in Individualized Cognitive Training: a Systematic Review

Aims

The aim of this chapter is to provide an overview of existing individualization procedures used in CT. This review primarily aims to demonstrate how the AI-based procedure introduced in chapters 5 and 6 can be positioned in comparison with existing work.

Contents

2.1	Introduction	53
2.2	Material and method . .	61
2.3	Results	79
2.4	Discussion	84
2.5	Conclusion	87

Abstract

To tackle the challenge of responders heterogeneity, Cognitive Training research currently leverages AI Techniques for providing individualized curriculum rather than one-size-fits-all designs of curriculum. Our systematic review explored these new generations of adaptive methods in computerized CT and analyzed their outcomes in terms of learning mechanics (intra-training performance) and effectiveness (near, far and everyday life transfer effects of CT). A search up to June 2023 with multiple databases selected 19 computerized CT studies using AI techniques for individualized training. After outlining the AI-based individualization approach, this work analyzed CT setting (content, dose, etc), targeted population, intra-training performance tracking, and pre-post-CT effects. Half of selected studies employed a macro-adaptive approach mostly for multiple-cognitive domain training while the other half used a micro-adaptive approach with various techniques, especially for single-cognitive domain training. Two studies emphasized the favorable influence on CT effectiveness, while five underscored its capacity to enhance the training experience by boosting motivation, engagement, and offering diverse learning pathways. Methodological differences across studies and weaknesses in their design (no control group, small sample, etc.) were observed. Despite promising results in this new research avenue, more research is needed to fully understand and empirically support individualized techniques in cognitive training.

Collaborators

Adolphe Maxime (AM), Pech Marion (PM), Sawayama Masataka (SM), Delmas Alexandra (DM), Maurel Denis (MD), Sauz  on H  l  ne (SH), Oudeyer Pierre-Yves (OPY). ¹

This work has been submitted and is under review. It is also accessible as a preprint in Adolphe et al., 2023.

¹According to the Contributor Roles Taxonomy, collaborators contributed in various roles: Writing - Review & Editing (all); Conceptualization (AM, PM, SM, SH); Data Curation (AM, PM, SM, SH); Formal Analysis (AM, PM, SM, SH); Investigation (AM, PM, SM, SH); Methodology (AM, DM, MD, OPY, SH); Funding Acquisition (DM, MD, OPY, SH); Project Administration (DM, MD, OPY, SH).

2.1 Introduction

The repetitive and prolonged practice of specific cognitive activities, more often called Cognitive Training, is an umbrella concept with multiple dimensions and multiple issues. First of all, in the field of aging or neurocognitive rehabilitation, the hope of finding non-drug and non-invasive interventions is a path to be favored in first-line clinical care. Indeed, the presence of neurocognitive disorders or declines has a major impact on the comfort of life of the persons, and can lead to a decrease in autonomy, or even a slide towards a pathological condition (Depp et al., 2014a). Thus, many researchers have mobilized their workforce in the design of training or cognitive rehabilitation programs for older adults, for Mild Cognitive Impairment (MCI) patients (H. Zhang et al., 2019 ; Coyle et al., 2015; Reijnders et al., 2013; Silva et al., 2022), Alzheimer’s patients (Clare and Woods, 2003; Hill et al., 2017b; Kallio et al., 2017), Parkinson’s patients (Johansson et al., 2023; I. H. Leung et al., 2015a), or any patient with Acquired Brain Injury (ABI) (García-Rudolph and Gibert, 2014; Hallock et al., 2016a). Second, outside of these health issues, research on CT is growing to meet the needs of performance enhancement in certain activities: sports performance (Harris et al., 2018a), academic performance (Wollesen et al., 2022) or even professional performance (Lucia et al., 2021; Vartanian et al., 2016). Lastly, alongside the difficulties related to the restoration and enhancement of performance, CT constitutes a fundamental realm of exploration encompassing the study of learning mechanisms, their evolution, and their neural associations (Turnbull et al., 2022). Given the expansive nature of CT, which cover a diverse range of cognitive skills, interventions, as well as social and commercial implications, an open letter written by 70 researchers in 2014 brought attention to the challenge of inadequate compelling evidence in this complex and multifaceted field. In 2016, a response from 111 researchers acknowledged areas needing improvement while emphasizing the continued promise of various research directions. Subsequently, despite ongoing debate fueled by studies both supporting (Jaeggi et al., 2011, Lampit et al., 2014a) and challenging (Melby-Lervåg and Hulme, 2013b; Sala, 2017; Simons et al., 2016b) CT, research in this domain has witnessed heightened activity. Notably, the volume of publications on PubMed in 2016 surpassed the cumulative studies conducted in preceding years (Harvey et al., 2018), indicating a surge in scholarly interest and engagement. In the perspective of contributing to the improvement of this field, our Systematic Review (SR) explores adaptive methods of customizing the training program to each individual. This addresses the central challenge of managing the diversity in response to CT, which encompasses both inter-individual and intra-individual variability.

As described in Karbach et al., 2017a, Katz et al., 2016 and Guye, De Simoni, and von Bastian, 2017, prior cognitive performance, age, and education is a non exhaustive set of factors that influence the magnitude of the impact of the interventions. The compensation effect (greater CT-related improvement of participants with lower prior performance (Jaeggi et al., 2008a, Zinke et al., 2014)) and the magnification effect (greater CT-related improvement of participants with higher prior performance) are observed in many studies (Fu et al., 2020; Karbach et al., 2017a; Lövdén et al., 2012a). Thus, in order to maximize the likelihood of program response, many interventions proposed adapting the difficulty and content to participants. This adaptation can be implemented manually, before or during the program, by the designer or the health professional (Kelly et al., 2014). Utilizing prior knowledge of the participant’s progression and performance during training, these methods can also be implemented automatically (e.g Pedullà et al., 2016a). Classically, automatic approaches are based on a staircase procedure where the difficulty increases if

the participant successfully completes several activities in a row and decreases otherwise (Klingberg, 2010; von Bastian and Eschen, 2016). Originating in the field of psychophysics (Dixon and Mood, 1948, Cecala, 2016 and Treutwein, 1995), the use of staircase strategies for training has the advantage of bringing the participant to his maximum capacity and pushing him to exceed it. However, even if these so-called "adaptive" procedures are easy to deploy in computerized CT systems, they lack flexibility and responsiveness in their ability to individualize the procedure. First, they do not take account the whole learning trajectory followed by the participant (only some of the previous activities are considered for the calculation of the future activity) (Singh et al., 2022, Zini et al., 2022). This suggests that a participant who has temporarily dropped to a lower level of difficulty due to factors such as fatigue or inattention, will be presented with the same task as another participant who has reached their true limit of learning, and will have to invest an equivalent amount of time to regain their previous maximum level. Secondly, this strategy poses challenges in managing a substantial number of parameters concurrently as it becomes complex to infer the progression of difficulty when multiple parameters are altered simultaneously (Edwards et al., 2005). Thirdly, staircase strategies result in a limit around which participants oscillate until improvement is observed. As a consequence, since participants consistently encounter similar stimuli near the threshold, this pattern can generate a perception of repetition that may be demotivating, discouraging, and not conducive to effective training. While certain programs (NeuroTracker, n.d.) have suggested incorporating adaptive steps to update task difficulty, they still exhibit limited parameter involvement in controlling the difficulty. Consequently, the training activities' space remains underutilized for the learner, restricting the range of learning opportunities for progress. Finally, the inflexible structure inherent in the unique trajectory design dictated by the staircase strategy obstructs the integration of the abundant knowledge and theories available from diverse fields like education sciences and psychology. For instance, a notable drawback is its inability to accommodate various signals from learners, such as physiological measurements (EMG, EEG), posture, or interaction data (like clicks), which can be valuable for tailoring the choice of educational activities and gaining insights into how learners react to the curriculum they receive. Considering the limitations outlined above, this systematic review aims to emphasize novel approaches for tailoring interventions to individual participants' needs. Thus, the interventions incorporated into this review will be labeled as "Individualized Computerized CT," contrasting them with the majority of self-proclaimed "Adaptive Computerized CT" to support the aspiration of providing genuine personalization to each participant.

Beneath the inquiry into the variability of CT responses lies the fundamental question of how to assess the effectiveness of these interventions. Traditionally, CT effectiveness is evaluated in terms of the extent of impact with a short-term spectrum corresponding to local effectiveness (improvement in performance on tasks similar to those trained, i.e., near effect) and a broad spectrum corresponding to global effectiveness (improvement in performance on tasks not similar to those trained but involving common cognitive mechanisms and functions, i.e., far effect). This range of impact is expressed in terms of Near and Far transfer (NFT) (von Bastian et al., 2022b). The NFT effects are generally assessed using cognitive batteries (Adolphe et al., 2022a, Steyvers and Schafer, 2020a, Gronwall, 1977, Tombaugh, 2006, Willis et al., 2006) and allow the evolution of the participant's performance after training to be quantified. Research in this domain frequently concludes after establishing efficacy, without delving into the ecological transfer of training i.e the practical influence of training on real-life tasks (Simons et al., 2016b). This gray area can be attributed by the fact that the ecological validity of CT is difficult to objectify, except

with the use of assessments with a more ecological content or questionnaires in which participants are asked to self-report the improvements perceived in real life. These tasks or questionnaires often have methodological limitations (ecological content validity, and subjective bias (Schmiedek, 2016b)). In addition to these considerations of effectiveness measures, many SRs or reviews raise weaknesses in the level of evidence provided by the studies (e.g., von Bastian et al., 2022b). These weaknesses are related to the study design (i.e. presence of a control group (Boot et al., 2013), randomization of group assignment, blindness of researchers and participants, sample size, etc.) (von Bastian et al., 2022b) and the design of the interventions (nature and type of training task, dosage, etc.) (Lampit et al., 2014a). Echoing the reproducibility crisis of science, it is observed that some studies showed significant effects of CT, while others are unable to reproduce these results. Among the salient factors identified, the lack of standardization of the content used is highlighted by recent SRs (Pergher et al., 2020). Consequently, this review will give particular attention to the methodological decisions and the resulting conclusions, striving to provide a thorough depiction of the field’s status.

To the best of our knowledge, no SR has been proposed to identify the new generation of individualized CT and to analyze their impact in terms of near or far effectiveness. In compliance with the PRISMA standards, as illustrated in the flow diagram (Figure 2.1) and the checklist provided in the supporting information, this study aims to concentrate on interventions that offer more adaptable strategies, facilitating enhanced individualization of content. We are particularly interested in CT proposing either automatic individualization of multimedia content or of the difficulty of the task. Inclusion criteria for this review necessitate that strategies facilitate the tailoring of interventions to individuals or representative groups. Such strategies should enable the generation of individualized and optimized learning trajectories for each learner. Hence, this criterion for inclusion implies the utilization of automation strategies spanning different levels of intelligence, notably those grounded in artificial intelligence.

2.1.1 Research question

The current systematic review of the literature first asked what individualization strategies exist into computerized CT tools (sections descriptive results, Q1-Q2). Secondly, it questioned the motivations of researchers to produce this type of strategy, i.e., specific individualization goals targeted by the strategy (section Q1-Q2). Finally, it examined the effectiveness of the included studies in light of the quality of the evidence provided, i.e., study design and statistical power (sections Q3, Q4, Q5). The ultimate aim was therefore to establish an inventory of existing flexible adaptive strategies and their level of maturity to serve the field of CT.

2.1.2 Background

The development of adaptive methods in CT is mainly fed by two main research fields, i.e., the field of computerized CT and the field of intelligent tutoring systems (ITS) even if the contribution of the latter one is larger to those of the former (see for reviews, Vandewaetere et al., 2011a, Mousavinasab et al., 2021).

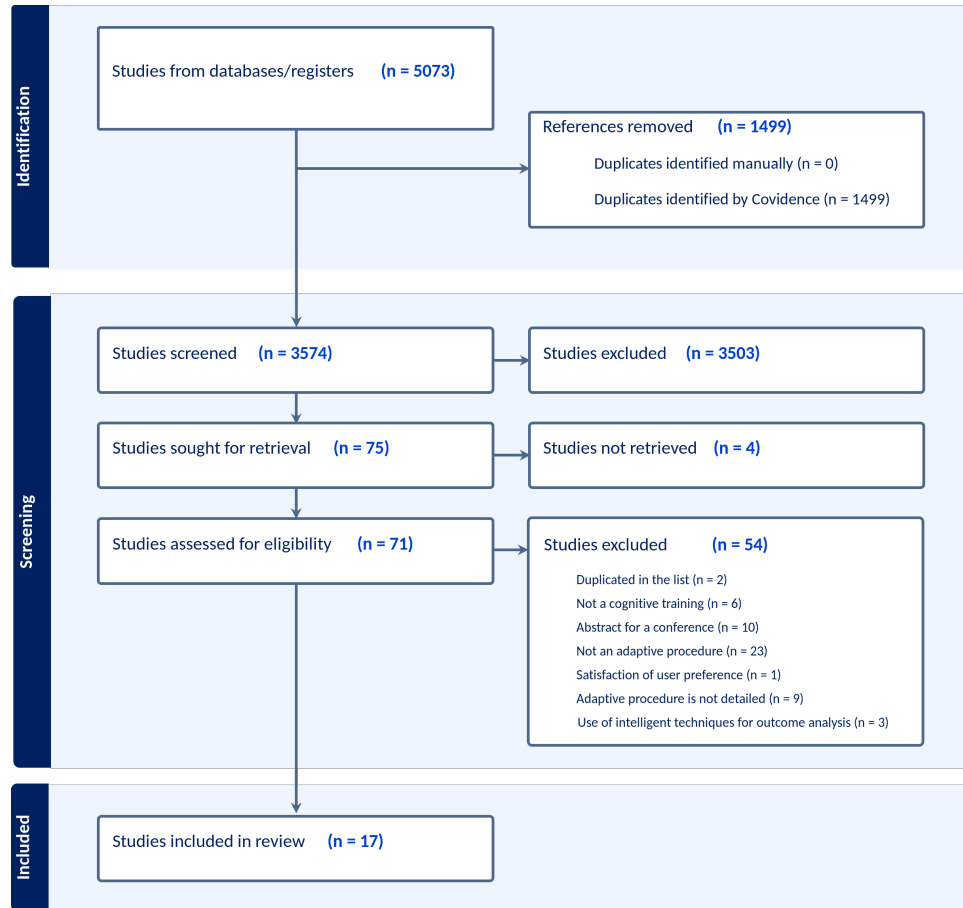


Fig 2.1. PRISMA Flow chart.

Insights from adaptive computerized CT research

This line of research has mainly contributed to exploring staircase methods for CT. Often, these methods consisted of the execution of graded exercises, whose difficulty increases gradually according to a set of predefined rules, considering the results the trainees achieve. Frequently, predefined rules are derived from expert knowledge. For instance, the exercises are typically structured hierarchically according to difficulty levels, and the progression between levels is primarily determined by predefined thresholds, often set at 70% of correct answers for each level of exercises. Hence, the staircase methods consist of a unique trajectory design of CT program, involving that all trainees follow a single path although at different speeds or with a different number of attempts. Several computerized CT systems for various CT purposes are based on this design of program personalization (Neri et al., 2021), such as Brainer (Brainer, n.d.), Neurotracker® (Vater et al., 2021a), RehaCom® (RehaCom, n.d.), CogniPlus® (CogniPlus, n.d.), HappyNeuron Pro® (Happyneuron, n.d.), Erica (Erica, n.d.), the Padua Rehabilitation Tool (PRT) Cardullo, 2017), MS Rehab (Gaspari et al., 2020), Cogni-Track (Tacchino et al., 2015) and CogniFit Personal Coach® (CogniFit, n.d.). In the majority of investigations that have contrasted adaptive strategies of this design with conventional approaches, a consistent finding has been the enhanced CT outcomes associated with adaptive strategies (as evidenced by studies like Klingberg, 2010; Pedullà et al., 2016a; Peretz et al., 2011)). Nevertheless, contrasting results have emerged in certain studies, exemplified by von Bastian and Eschen, 2016, which did not detect any advantages under adaptive conditions. Notably, this particular

study implemented adaptive adjustments between sessions rather than within the same session, which may account for the disparity in outcomes.

From a more distant standpoint, recent SR highlighted the promising findings from CT studies comparing gamified contents to non-gamified ones as a result of the critical role of trainee 's motivation and engagement in the individualization of CT (Lumsden et al., 2016b, Vermeir et al., 2020). Taken together, these overall results supported the added value of personalization of CT for fostering their outcomes. However, as mentioned above, the staircase methods have some limitations and are not really adaptive due to its single trajectory design, i.e, the system adapts the CT in the same manner for all trainees instead to specifically adapt the program to the trainee performance by creating a specific path into the program. Due to this strong limitation, most of existing computerized CT systems (e.g., HappyNeuron Pro®, Erica, MS-Rehab®) include a manual calibration for defining the initial level of exercise at the beginning of the CT and the successive tweaks of difficulty level across the CT (often done by the clinicians). However, as the trained tasks involve a significant number of parameters to determine the level of difficulty, manual calibrations become increasingly complex and numerous.

Insights from ITS research

Educational sciences have widely contributed to demonstrate that factors intrinsic to the learner (such as prior knowledge, emotional load, mental load or motivation) and extrinsic factors (such as all the variables related to the instructional design), are mediators of the efficiency of the learning functions. By nature, the effectiveness of CT is no exception to this observation and responds to similar factors. Hence, it seems natural to ask how effective instructional methods from the educational sciences can be transferred to the field of CT, and more particularly those providing an individualization of learning. Learning theories indicate that learning requires an appropriately sized “mismatch” – a gap between the cognitive capacity and the requirements of the external task that the cognitive system must adapt to in order to improve performance (Lövdén et al., 2010a). As a result, the evidence-based assets of individualized learning over one-size-fits-all educational approaches are today well documented (M. I. Deunk et al., 2015, Iterbeke et al., 2021). ITS offers a framework for the automated creation of curricula tailored to individual students. While there are multiple methods available to enrich and personalize educational content with ITS for each learner, the majority of systems are organized around three primary components (Vandewaetere et al., 2011a). Firstly, there is the aspect of adapting to the instructional source, which refers to what the system will tailor, including aspects like the learner's learning style ((Sun et al., 2007, Bunderson and Martinez, 2000)), existing knowledge (Koedinger and Corbett, 2006), or preferences (Ray and Belden, 2007). Secondly, there is the target of adaptive instruction, specifying what aspects will undergo adaptation. This could involve the content of the instruction (Sun et al., 2007) or the manner of presentation (Milne et al., 1997). Thirdly, the adaptive component functions as the intermediary, creating a pathway between the first two components. It dictates how to adapt a target to a source, which can be achieved through diverse methods. This last component, also called the tutoring module, is the engine generating a curriculum of training activities for learners in ITS. Adaptive feedback, hint, and recommendation-generating, navigation of the learning path, and presenting adaptive educational content constitute the core of this component (Carter and Blank, 2013). The contents' adapting to the learner's needs is the most relevant tutoring dimension of ITS for the individualization purpose into a CT program. In order to tailor content to individual learners, numerous ITS draw from

the concepts of the zone of proximal development (ZPD) (Vygotsky and Cole, 1978b) and the state of Flow (Csikszentmihalyi, 2000). These concepts are closely tied to the Goldilocks effect (Seitz, 2018, Jonge et al., 2012), wherein learning is optimal in tasks that strike a balance between simplicity and excessive challenge. Following them, many ITS aim to offer the learner pedagogical activities that are neither too difficult nor too easy with regard to their abilities, so that they can be engaged and progress in their acquisitions without being anxious or bored during the process. ITS can also suggest activities that may be challenging for the learner to solve independently, but become manageable with the assistance of hints or guidance from the teacher. According to this ZPD principle, the tutoring component classically integrates a performance threshold principle for exercise difficulty shift (often chosen around 70%) to maintain an average optimal learning trajectory (Seitz, 2018). Several signals or performance dimensions can be used to guide the generation of a curriculum: some ITS are interested in using an optimal emotional level (Khadimallah et al., 2020) or learning progress (Clement et al., 2013a, W. Ma et al., 2014) or both (Oudeyer et al., 2007).

Sorting Keys of AI techniques for content adapting to learner’s capabilities

On a macroscopic scale, the adaptability of Intelligent Tutoring Systems (ITS) can be categorized into two primary design approaches for managing learning curricula (Bartolomé et al., 2018). The first approach, known as the "linear design" which we prefer to label as the "unique trajectory design" within this review, encompasses all learners pursuing a singular route, albeit at varying paces or with distinct numbers of attempts. Such a design is similar to the staircase methods used in the field of CT. The second approach, termed the "branched-paths design" empowers each learner to pursue a unique path tailored to their individual needs. In this review, we opt to refer to it as the "Individualized" procedure. Consequently, this leads to diverse learning trajectories across learners, encompassing linear, non-linear (leaps and backtracking), or hybrid paths, thereby rendering "branched-paths designs" as truly personalized learning environments.

Moving on the mesoscopic scale, as described in Figure 2.2, the adaptability of content can manifest across various tiers: at the level of selecting activity types within the training called macro-adaptive learning or by manipulating the parameters of activities presented at each time step called micro-adaptive learning (E. L.-C. Law, Kickmeier-Rust, et al., 2008).

Finally at microscopic scale, various AI techniques have been used for the implementation of the content adapting to the learner’s needs. They can be broadly categorized into four main families of AI techniques employed individually or in combination (Chang et al., 2020, Mousavinasab et al., 2021):

- Condition-action rules-based reasoning traditionally refers to rule-based decisions (if X, then Y) that determined the outcome of adaptive instruction. Rules are set by the instructor prior to the learning process (e.g., rule-based expert system or semantic rule-based reasoning). In the context of CT, this category would encompass the staircase procedure as introduced previously.
- Probabilistic modeling and Bayesian networks refer to a set of techniques that rely on graphical model to encode probabilistic relationships between variables of interest. A key advantage of using them is that their structure is ideal for combining prior knowledge, which is often in causal form, with observed data. Into an ITS, prior knowledge consists of a stereotyped model based on the learner’s goals, tasks, and

interests, while observed data is extracted from the interaction between the learner and the environment. Bayesian techniques can also be used when data is missing, a common problem in the learning sciences.

- Machine learning techniques involve the use of algorithms and statistical models to enable computer systems to learn from data and improve their performance on a task without being explicitly programmed (see Badillo et al., 2020 for an introduction). In the context of ITS, machine learning techniques can be used to individualize the learning experience for each student by leveraging data collected during interactions with the system. The strategies can operate in two different modes: one is an incremental approach, where the model evolves during direct interactions with users, and the other is an offline method, which includes data collection, model development without real-time interaction, and its application to learners thereafter. They can be divided into several subfields:
 - Data mining refer to a set of techniques used to extract insights and knowledge from large datasets such as student interactions with the system or demographic data. These techniques involve analyzing the data to identify patterns and relationships that can be used to personalize the learning experience for each student. The extracted features can then be combined with decision-making modules to adapt the learning path and provide targeted support and guidance to the student. One example of a widely used data mining technique in ITS is clustering (García-Rudolph and Gibert, 2014). This method enables the identification of different groups of students based on their learning profile, needs, and preferences. By clustering students, ITS can create tailored learning paths that address the specific needs of each group, leading to more effective and efficient learning outcomes.
 - Artificial neural networks and deep learning (DL) techniques are a set of techniques inspired by the structure and function of the human brain and are designed to learn from large datasets of student interactions with the system. In ITS, they can be used to model student behavior and performance, predict future outcomes, and adapt the learning experience to the individual needs of each student.
 - Reinforcement learning (RL) is a type of machine learning in which an agent learns to make decisions in an environment by receiving feedback in the form of rewards or punishments. In the context of ITS, the RL agent can serve as an instructor and receive a reward based on the effectiveness of its pedagogical approach towards the student (see Doroudi et al., 2019 for a review). Numerous algorithms have been developed to tackle this challenge. One common strategy involves maintaining a tabular record of how effective a specific pedagogical activity is, quantified by the cumulative rewards it garners, when employed with a student possessing a particular skill level. Through an iterative process of proposing various activities, the agent seeks to determine the optimal actions that maximize its overall reward (see Q-learning algorithm in Zini et al., 2022 for an example). Another approach to address this challenge draws an analogy to a casino scenario featuring multiple slot machines. Within this metaphor, critical questions center on the selection of the most effective 'slot machines,' their optimal utilization frequency, and the establishment of a suitable sequence. In the educational context, these metaphorical 'slot machines' represent different

pedagogical activities, and their success is gauged by the extent of knowledge acquisition by the student. To tackle the 'exploration-exploitation dilemma' inherent in this context, various techniques are employed such as multi-armed bandit algorithms (see Clement et al., 2013a for an example).

- Natural language processing focuses on the interaction between computers and humans through natural language, including tasks such as text classification, sentiment analysis, and machine translation. ITS can use techniques such as text classification and sentiment analysis to understand students' written or spoken responses, enabling individualized feedback (see Nye et al., 2014 for a review).
- Evolutionary algorithms are a family of optimization algorithms that are inspired by the process of natural selection and evolution to solve complex problems. By treating the potential solutions as a population of individuals possessing diverse traits, these algorithms employ a fitness function in conjunction with an evolutionary process to deduce the optimal solution (see Pillay, 2020 for a review). In the ITS literature, these techniques have been employed in various ways such as learner performance prediction or design of learning environments.

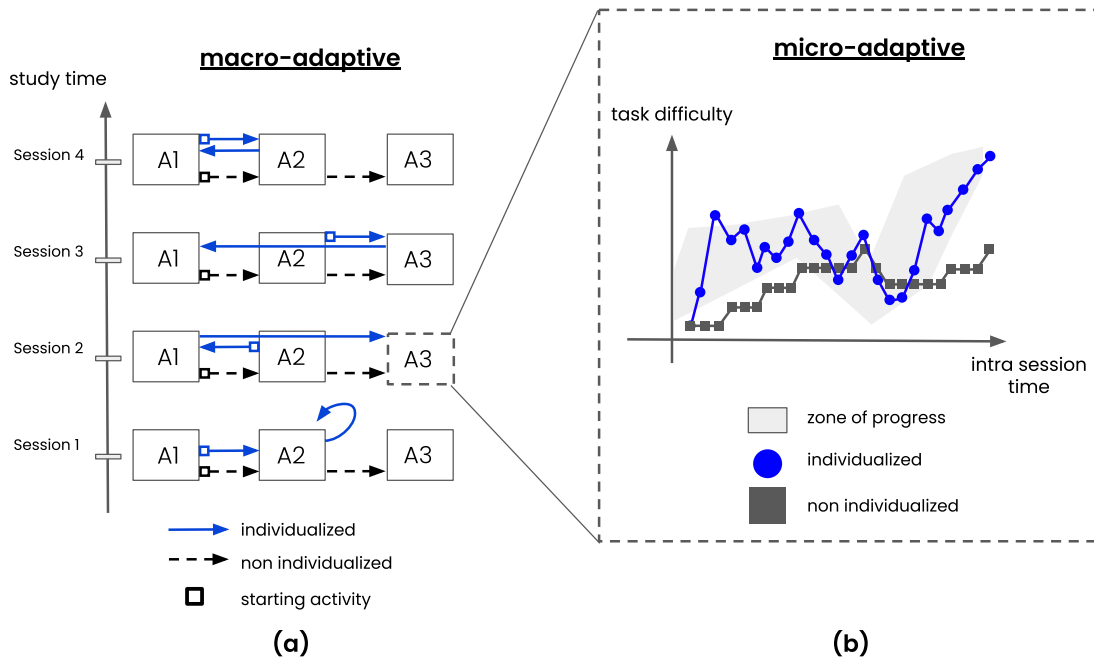


Fig 2.2. Illustration of macro and micro-adaptive strategies.

(a): Macro-adaptive strategy exemplified by two trajectories within a CT program (unique trajectory or individualized design) spanning sessions 1 to 4, each session offering three potential cognitive tasks (A1 to A3). Arrows depict task order for each session. Non individualized trajectory always propose same order A1, A2 and A3 while individualized path adapts the trajectory according to training objectives. (b): Micro-adaptive strategy demonstrated by two trajectories of task difficulty adjustments specifically for task A3 during session 2. The non-individualized trajectory relies on a staircase approach that falls short of identifying the optimal zone of progress when contrasted with the individualized procedure, which proves to be a more suitable fit.

Evaluation of AI techniques

The evaluation methods of individualized techniques into ITS are of two kinds, either formal or empirical (for review; see Soofi and Ahmed, 2019). Formal validations consist essentially in testing the system with simulations using learners' models for assessing the ITS behaviors in order to compare two or several AI techniques. Empirical validations are multiple-ways in terms of expected outcomes or study designs. The judgment criteria can be qualitative (i.e., experts or learners' feedback, learner experience questionnaires, etc.) or quantitative (i.e., learning performance, level of activities performed, etc.), or both. They can be based only on training phase (interaction data) or include pre- and post-training measurement, or both. Ensuring the validity of scientific research, whether validated formally or empirically, hinges upon the accessibility of both the dataset and the employed model. It is worth noting that a notable factor contributing to the reproducibility crisis is the limited access to comprehensive research materials. The study design varies from feasibility or pilot study (e.g. prototype testing with few users) to Randomized Controlled Trial (large sample of individuals, control group, pre-and post-training measurement). The gold standard for evidence-based adaptive ITS is to compare it with a control condition often consisting of unique trajectory design (e.g staircase procedure) on qualitative and quantitative measurements taking place before, during and after the training and providing insights on NFT effects of the training.

Operationalized research questions

Pertaining to the central inquiry of this work - "Do the emerging generations of individualized strategies hold promise for computer-based cognitive training?" - five specific operational research questions were formulated as follows:

- Q1. What AI Techniques have been employed in computerized CT, and what are the underlying research motivations driving their utilization?
- Q2. What are the domains of CT for which adaptive techniques have been designed?
- Q3. What populations are targeted and what are the characteristics of the CT settings?
- Q4. How effective are they in empirical CT studies? What effects are reported (NFT learning effects)? Are the effects dependent on characteristics of CT settings?
- Q5. What Type of validation have been conducted for these new generations of computerized CT?

2.2 Material and method

A pre-established protocol was formulated and officially registered with PROSPERO (registration number: 2021 CRD42021241515). The checklist of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was applied to guide the systematic review process (PRISMA). The COVIDENCE tool was also used to manage and organize the work.

2.2.1 Databases and Search query

The initial database searches were conducted between February and April 2023 using the following electronic databases to conduct the study: PsycInfo, Medline, ETHOS, EMBASE, IBSS, PubMed, IEEE Xplore, ACM Digital Library, Springer, Taylor Francis, Scopus, Education Resources Information Center (ERIC), ScienceDirect/Elsevier and EBSCO. In addition to the database searches, a hand search of relevant journals and gray literature were also conducted to ensure all relevant works were included in the review. According to the research question, we used the following query: TITLE-ABS-KEY ("Cognitive training") AND ALL ("Machine Learning" OR "reinforcement learning" OR "personalized*" OR "individualized" OR "intelligent tutoring system") AND NOT TITLE-ABS-KEY("Transcranial direct current stimulation").

2.2.2 Eligibility criteria

In this SR, we included all computerized CT studying individualized training that allows for differentiated learning paths in terms of content (type of exercises). No restrictions were set regarding the publication date, but the included studies had to be in English. Furthermore, no age or population criterion was used. Table 2.1 presents details of inclusion and exclusion criteria.

Table 2.1. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
The adaptive procedure relies on individual training performance and/or training experience.	The personalization only consists in satisfying user preference (e.g visual features, content type, gaming component) without adapting the learning path (e.g Quaglini et al., 2009)
The adaptive procedure is being tested on a dataset of CT, and the strategy is thoroughly described.	The adaptive procedure adheres to a "unique trajectory design" (e.g similar to staircase algorithms [Faria et al., 2019])
Machine learning techniques are used to predict participants' behavior (adherence, success, emotional state, skill level) or to directly compute the optimal next activity	The adaptive procedure is not detailed (e.g S. Hardy et al., 2015)
Machine learning techniques are used on training data from CT results, ECG, EEG, MRI, fMRI, wearable sensor data, and longitudinal training experience measures.	The intervention involves a form of neuromodulation (e.g., tDCS)
Neurofeedback and machine learning techniques are used for individualized programs	Intelligent techniques are used as tools for a better outcome analysis (e.g., effect size analysis) (e.g Vladisauskas et al., 2022)
Individualized techniques encompass both online strategies (where the participant engages while the model is developed) and offline strategies (involving data collection to build the model with no access to data during interactions)	Not a CT (e.g Kim et al., 2018)
English writing	Non-peer-reviewed papers, opinion pieces, or abstract conference papers

2.2.3 Screening and selection method

The screening phase was conducted on articles until February 2023. In total, 5073 papers were found, as presented in Figure 2.1. All duplicates were removed, which reduced the results to 3574. Papers were selected through an iterative process of filtering. According to our search strategy (inclusion/exclusion criteria), studies were first filtered on titles and abstracts resulting in 71 articles to go through to the next stage of full-article review. All the screening process was carefully evaluated by two authors. When there was uncertainty or disagreement among the reviewers, consensus was reached through discussion. If no consensus was found, a third review was designed for the final decision. The full-text review of the remaining papers results in 17 papers with 19 studies included for the systematic review. The main reasons for the exclusions are reported in the PRISMA flowchart (Figure 2.1).

2.2.4 Data extraction

To answer our five research questions, four coding sheets were developed for extracting the searched information.

To address Q1 and Q2, information regarding the AI approach (macro-, micro-adaptive or both) and the AI techniques used, as well as the targeted cognitive domains of CT were collected in (Table 2.3). For the AI techniques, 8 families were distinguished : condition-action rules-based reasoning, probabilistic models or bayesian networks (e.g Kalman Filters (KF), Hidden Markov Models (HMM)), data mining (e.g Regression, Clustering), neural networks or deep learning (e.g multi-layered perceptron (MLP), convolutional neural networks (CNN), Long Short Term Memory (LSTM)), reinforcement learning (e.g Q-learning, Actor-critic), natural language processing, evolutionary algorithms and recommendation systems. Regarding cognitive domains, global CTs (multiple cognitive domains) were distinguished from specific CTs addressing a single cognitive domain. We used the categorization of cognitive functions traditionally used in psychology, as follows: perception (visual, auditory, spatial, etc.), attention (selective, sustained, divided components), learning and memory (working, semantic, episodic, procedural), language (production and understanding), executive functions (inhibition, updating, and cognitive flexibility) and reasoning and problem solving (categorization, generalization, deductive and inductive inference, predictive and evaluative judgment).

Table 2.3. Overall descriptive results, AI techniques and Cognitive domains of CT for the selected studies.

Study	Date	Title	AI approach and AI Techniques	Cognitive domain of CT	Study design	Population Type	Sample Size
García-Rudolph and Gibert García-Rudolph and Gibert, 2014	2014	A data mining approach to identify cognitive NeuroRehabilitation Range in Traumatic Brain Injury patients	Macro-adaptive learning Decision tree objectives: Prediction of the optimal neurorehabilitation range	Multiple domains Attention, memory, language, executive functions	Non-randomised controlled trial	Clinical sample	n=327
Fermé et al. Fermé et al., 2020	2020	AI-Rehab: A Framework for AI Driven Neurorehabilitation Training - The Profiling Challenge	Micro and macro adaptive Data mining and belief revision engines objectives: Participant profiling	Multiple domains Attention, Memory (semantic, episodic), language (understanding), reasoning (categorization) and problem-solving (maze, navigation task).	Feasibility study (study protocol)	NA	NA
Xu et al. Xu et al., 2018	2018	Personalized Serious Games for Cognitive Intervention with Lifelog Visual Analytics	Macro adaptive Deep learning and clustering techniques Personalization of game content with lifelog visual content	Multiple domains Attention, memory, visuospatial and executive functions	Individual randomized trial (crossover study)	Non clinical sample	n=26
Reidy et al. Reidy et al., 2020	2020	Facial Electromyography-based Adaptive Virtual Reality Gaming for Cognitive Training	Micro adaptive Data mining and machine learning EMG data preprocessing and affect classification	Multiple domains Memory (episodic), executive and problem-solving functions	Non-randomised controlled trial (crossover study)	Non clinical sample	n=6
Kitakoshi et al. (a) Kitakoshi et al., 2015	2015	Cognitive Training System for Dementia Prevention Using Memory Game Based on the Concept of Human-Agent Interaction	Micro adaptive Reinforcement learning (bucket brigade algorithm) Difficulty level adjustment and break offering system	Specific domain Memory (episodic)	Non-randomised controlled trial (crossover study)	Non clinical sample	n=6
Kitakoshi et al. (b) Kitakoshi et al., 2020a	2020	A Study on Coordination of Exercise Difficulty in Cognitive Training System for Older Adults, study-1	Micro adaptive Reinforcement learning (bucket brigade algorithm) Difficulty level adjustment	Specific domain Memory (episodic)	Non-randomised controlled trial (crossover study)	Non clinical sample	n=5
NA = Not Applicable						Continued on next page	

Table 2.3: continued from previous page

Study	Date	Title	AI approach and AI Techniques	Cognitive domain of CT	Study design	Population Type	Sample Size
Kitakoshi et al. (b) Kitakoshi et al., 2020a	2020	A Study on Coordination of Exercise Difficulty in Cognitive Training System for Older Adults - study-2	Micro adaptive Reinforcement learning (bucket brigade algorithm) Difficulty level adjustment	Specific domain Memory (episodic)	Non-randomised controlled trial (crossover study)	Non clinical sample	n=5
Rathnayaka et al. Rathnayaka et al., 2021	2021	Cognitive Rehabilitation based Personalized Solution for Dementia Patients using Reinforcement Learning	Micro adaptive Reinforcement learning (Q-learning) Difficulty level adjustment	Multiple domains Attention, memory, language, executive functions	Non-randomised controlled trial	Clinical sample	n=56
Shen and Xu Shen and Xu, 2021	2020	Research on children's cognitive development for learning disabilities using recommendation method	Macro adaptive Recommendation system (collaborative filtering) Proposition of a curriculum based on the similarity between children performances and preferences	Multiple domains Attention, memory, language, executive function (flexibility), reasoning	Individual randomized controlled trial	Non clinical sample	n=30
Sandeep et al. Sandeep et al., 2020b	2020	Application of Machine Learning Models for Tracking Participant Skills in Cognitive Training - study-1	Micro adaptive Machine learning and deep learning (Hidden Markov Model, Kalman filters, LSTM) Prediction of performance evolution through training	Specific domain (Working) Memory	Feasibility study (framework description - secondary data analysis)	Non clinical sample	n=262 (Dataset)
Sandeep et al. Sandeep et al., 2020b	2020	Application of Machine Learning Models for Tracking Participant Skills in Cognitive Training - study-2	Micro adaptive Machine learning and deep learning (Hidden Markov Model, Kalman filters, LSTM) Prediction of performance evolution through training	(Working) Memory	Feasibility study (framework description + secondary data analysis)	Non clinical sample	n=177 (Dataset)
Wilms Wilms, 2011	2011	Using artificial intelligence to control and adapt level of difficulty in computed-based cognitive therapy	Micro adaptive Reinforcement learning (Actor-critic method) Difficulty level adjustment	Specific domain (Visual) Attention	Non-comparative Study (case study)	Clinical sample	n=1
NA = Not Applicable						Continued on next page	

Table 2.3: continued from previous page

[illegible]

For response to Q3, Table 2.4 aimed to collect descriptive data for each selected study in terms of population included, sample size, characteristics of CT design (content, dose, location). In addition, this sheet was also dedicated to Q4 as it relates to the effectiveness of AI based individualized computerized CTs according to several judgment criteria (intra-training performance, pre/post training effect, near/far effect, etc) (Table 2.4). In order to address Q4, we also developed a meticulously crafted scale to assess the presence of significant features that contribute to substantiating the effectiveness of the intervention. Indeed, as elucidated in (C. S. Green et al., 2019), CT interventions must incorporate significant supplementary elements to demonstrate their effectiveness. Therefore, the proposed scale assigns a rating ranging from 0 to 3 for various dimensions, including information related to dosage and location, intra-training performance measures, subjective evaluation, pre-post comparisons, quality of the cognitive evaluation employed, and follow-up assessment. By summing the scores for all items, each study was assigned a grade ranging from 0 to 11.

Table 2.4. Sample characteristics and characteristics of CT setup for the selected studies.

		CT features		Within - CT measurements		CT effectiveness assessment			Note
Study	Population	Content	Dosage	Location and training performance measures	Post and intra training subjective experience	Pre-post comparison	Cognitive measurement (near, far effect (NFT) and everyday life transfer)	Follow-up	Max = 11
García-Rudolph and Gibert García-Rudolph and Gibert, 2014	ABI and TBI participants	Multi-domain PREVIRNEC system : rehabilitation tasks (attention, memory, executive functions, language) - 115 tasks	Duration : ND Frequency : ND Location : Home	No	No	Yes (+1)	NFT: Standardized NAB (28 tasks covering language, attention, memory, learning and executive functions) - source ND (+1)	No	2
Fermé et al. Fermé et al., 2020 (<i>Study protocol</i>)	NONE	Multi-domain 5 modules about knowledge (memory of stories, cancellation, questions of general knowledge, image pairs), comprehension (association, categorization), application (mazes, navigation); analysis (visual memory, word search); evaluation (comprehension of contexts..) - no task	Duration : ND Frequency : ND Location : Home	NA	NA	Yes	NFT: MoCA (short-term memory, executive functions, visuospatial abilities, language, attention, concentration, working memory, temporal and spatial orientation)	NA	NA

ND = Not Documented ; NA = Not Applicable; NAB = Neuropsychological Assessment Battery

Continued on next page

Table 2.4: continued from previous page

		CT features		Within - CT measurements		CT effectiveness assessment			Note
Study	Population	Content	Dosage	Location and training performance measures	Post and intra training subjective experience	Pre-post comparison	Cognitive measurement (near, far and everyday life transfer (NFT))	Follow-up	Max = 11
Xu et al. Xu et al., 2018	older adults free of mental disease / dementia / MCI (age=63.7)	Multi-domain Puzzle games (memory, attention, speed, visuo-spatial and executive functions) - 8 tasks	Duration : 2 weeks Frequency : 10 mn / per week (with 4 specific games) Location : Home (+2)	Yes (user adherence and preference) (+1)	Hand-made questionnaires - (elicited enjoyment, content and gaming mechanism preference, perceived difficulty and attention level) (+1)	Yes (+1)	NFT: MoCA (short-term memory, executive functions, visuospatial abilities, language, attention, concentration, working memory, temporal and spatial orientation) (+2)	No	7
Reidy et al. Reidy et al., 2020	older adults free of mental disease / dementia / MCI (age=60 to 100)	Multi-domain Virtual Reality based tasks: virtual supermarket (working memory) and multi-room museum (episodic memory) tasks - 2 tasks	Duration : 30 mn Frequency : 2 sessions of 15 minutes per day Location : laboratory (+2)	No	Standardized questionnaire - gaming experience questionnaire (immersion, engagement, flow) (+2)	Yes (+1)	NFT: Standardized NAB (spatial memory, perception, attention/orientation, memory, fluency, language) - source ND (+1)	No	6
Kitakoshi et al. (a) Kitakoshi et al., 2015	older adults	Specific domain Memory game - 1 task	Duration: 6 weeks Frequency: at least 5 min on participants behalf - 2 weeks per condition Location : Home (+2)	Yes (learning path and self-selected dosage) (+1)	Hand-made questionnaires - (enjoyment, motivation, perceived difficulty) (+1)	No	No	No	4
Kitakoshi et al. (b) Kitakoshi et al., 2020a	older adults (age=79.2)	Specific domain Memory game - 1 task	Duration: 2 weeks Frequency: 10 min per day Location : Home (+2)	Yes (learning path) (+1)	Hand-made questionnaires - (motivation and engagement) (+1)	No	No	No	4

ND = Not Documented ; NA = Not Applicable; NAB = Neuropsychological Assessment Battery

Continued on next page

Table 2.4: continued from previous page

		CT features		Within - CT measurements		CT effectiveness assessment			Note
Study	Population	Content	Dosage	Location and training performance measures	Post and intra training subjective experience	Pre-post comparison	Cognitive Measurement (near, far and everyday life transfer (NFT))	Follow-up	Max = 11
Kitakoshi et al. (b) Kitakoshi et al., 2020a	older adults (age=79.2)	Specific domain Memory game - 1 task	Duration: 2 weeks Frequency: 10 min per day Location : Home (+2)	Yes (learning path and self-selected dosage) (+1)	Hand-made questionnaires - (motivation and engagement) (+1)	No	No	No	4
Rathnayaka et al. Rathnayaka et al., 2021	adults with dementia	Multi-domain D-care (attention and concentration, executive functions, language and memory skills) - 4 tasks	Duration : 1 month Frequency : ND Location : ND	Yes (learning path) (+1)	No	No	No	No	1
Shen and Xu Shen and Xu, 2021	children (age=10 to 11)	Multi-domain CogDaily (speed, memory, attention, flexibility, logic training) - 17 tasks	Duration: 2 weeks Frequency: about 15 min per day Location: Laboratory (+2)	No	No	Yes (+1)	NFT: Wechsler Memory Scale (processing speed and memory) (+2)	No	5
Sandeep et al. Sandeep et al., 2020b (Data collection)	Young adults (age=19.87)	Specific domain N-back training ("Tapback", "Recall" and "Recollect the study") - 3 tasks	Duration : 8-10 days Frequency: 16 to 20 sessions of 20 min with 2 sessions per day including a 10 min break between sequential sessions Location: Home	NA	NA	NA	NA	NA	NA

ND = Not Documented ; NA = Not Applicable; NAB = Neuropsychological Assessment Battery

Continued on next page

Continued on next page

Table 2.4: continued from previous page

		CT features		Within - CT measurements		CT effectiveness assessment			Note
Study	Population	Content	Dosage	Location and training performance measures	Post and intra training subjective experience	Pre-post comparison	Cognitive Measurement (near, far and everyday life transfer (NFT))	Follow-up	Max = 11
Sandeep et al. Sandeep et al., 2020b (Data collection)	Young adults (age=19.79)	Specific domain N-back training ("Tapback", "Recall" and "Recollect the study") - 3 tasks	Duration : 8-10 days Frequency : 16 to 20 sessions of 20 min with 2 sessions per day including a 10 min break between sequential sessions Location: Home	NA	NA	NA	NA	NA	NA
Wilms Wilms, 2011	young adult with ABI (age=53)	Specific domain VisATT (letter span and vision detection speed) - 1 task	Duration: 3 weeks Frequency: 30 min session per day Location: Home (+2)	Yes (learning path) (+1)	No	No	No	No	3
Solana et al. Solana et al., 2014	adults with cognitive decline (ABI)	Multi-domain Guttman Neuro Personal Trainer (GNPT, PREVIRNEC 2) (attention, memory, executive functions) - 95 tasks	Duration : 4 to 7 months Frequency : 2- 3 sessions of 1 hour per week with a number of total session of 60 Location: Home (+2)	Yes (learning path comparison) (+1)	No	Yes (+1)	NFT: Standardized NAB (attention, memory, executive functions) - source ND (+1)	No	5
Zini et al. Zini et al., 2022	young adults	Specific domain MS-rehab: (alternating attention and working memory) - 1 task	Duration : ND Frequency: 20 types of exercise per session Location: Home (+1)	Yes (learning path) (+1)	No	Yes (+1)	NFT: PASAT (processing speed, working memory, sustained attention) (+2)	Yes - only for near effect (trained task)	5

ND = Not Documented ; NA = Not Applicable; NAB = Neuropsychological Assessment Battery

Continued on next page

Table 2.4: continued from previous page

		CT features		Within - CT measurements		CT effectiveness assessment			Note
Study	Population	Content	Dosage	Location and training performance measures	Post and intra training subjective experience	Pre-post comparison	Cognitive Measurement (near, far and everyday life transfer (NFT))	Follow-up	Max = 11
Zebda et al. Zebda et al., 2022	young adults	Specific domain cooking game (visual attention, working memory) - 1 task	Duration : 1 day Frequency : 45 minutes, Location : Laboratory, (+2)	No	Handmade questionnaires - user engagement (semi-structured interview about perceived differences between conditions, likeability, positive and negative aspects) (+1)	No	No	No	3
Eun et al. Eun et al., 2022	older adults (60 to 80 and over)	Multi-domain 4 modules (attention, logic, response time and memory) - 6 tasks	Duration: 8 weeks Frequency: no limit and ND Location : Laboratory (+1)	Yes (intra-training performance) (+1)	Hand-made questionnaires - satisfaction (engagement, fun, subjective performance) (+1)	Yes (+1)	No	No	4
Tsiakas et al. Tsiakas et al., 2018	young (undergraduate and graduate students)	Specific domain NIH Toolbox Cognition Battery (Working Memory test with socially assistive robots-based approaches) - 1 task	Duration: ND Frequency: Data collection - 20 minutes (including a post session user survey) Location: Laboratory	NA	NA	NA	NA	NA	NA

ND = Not Documented ; NA = Not Applicable; NAB = Neuropsychological Assessment Battery

Continued on next page

Continued on next page

Table 2.4: continued from previous page

		CT features		Within - CT measurements		CT effectiveness assessment			Notes
Study	Population	Content	Dosage	Location and training performance measures	Post and intra training subjective experience	Pre-post comparison	Cognitive Measurement (near, far and everyday life transfer (NFT))	Follow-up	Max = 11
Book et al. Book et al., 2022 (<i>Study protocol</i>)	Mild Cognitive Impairment (MCI)	Multi-domain MAKSCog (attention, executive function, perceptual-motor, executive functions, perceptual motor, language) - 10 tasks	Study protocol: Duration : 6 months and open phase in which participants can freely continue to use the CCTs Frequency : at least 30 min per day, 3 days a week Location : Home	No	Standardized questionnaire - User Engagement questionnaire (attractiveness, perspicuity, efficiency, dependability, stimulation and novelty of software), Hand-made questionnaire of usability	No	No	No	NA
Singh et al. Singh et al., 2022	older adults mean age 72.6	Multi-domain The Mind Frontiers cognitive training (Working memory updating, switching, dual N Back, TowerOfLondon, PipeMania, FigureWeights VisualSpatial) - 7 tasks	Duration : first period of 12 weeks (5 days out of 7) and second period of 6 weeks Frequency : Data collection - 45 minutes a day for phase 1, no limit for phase 2 Location: Home	Data collection - Yes (learning path)	Data collection - Hand-made questionnaires - (technical competence, subjective cognition, perceived benefits)	NA	NA	NA	NA

ND = Not Documented ; NA = Not Applicable; NAB = Neuropsychological Assessment Battery

Finally, Tables 2.5 and 2.6 collected information for a SIGN analysis (Harbour and Miller, 2001) to assess the quality of study design in the field of individualized computerized CT. The SIGN ratings estimates the strength of available evidence provided by a study, based on the methodological design and the evaluation of possible biases. Regarding study designs, we considered various options as outlined in the SIGN guidelines. We included experimental studies, both with and without a comparison group. Studies with comparisons were classified into three possible variations: cluster randomized controlled trials, where randomization occurs at the group level; individual randomized controlled trials, where randomization occurs at the individual level; and non-randomized controlled trials, which involve no randomization. We also considered feasibility studies that proposed a descriptive framework. Some of these studies were supported by secondary data analysis and utilize existing datasets to extract valuable information to propose a descriptive framework. It is important to note that the objective of our SR is to provide an overview of the current state of the art and the level of maturity of individualized CT. Consequently, our criteria for study inclusion and exclusion were not restricted to particular research designs; in other words, we did not constrain the incorporation of studies with lower maturity, such as those lacking comparative analyses.

For each included controlled trial, we employed the SIGN methodology checklist, which presents a grading system ranging from 0 (not applicable) to 3 (well covered) for various items including participant assignment strategy, randomization, measurement types and validity, among others (see appendix 2.5). This assessment resulted in a final grade that evaluates the extent to which the study was conducted to minimize bias, with grades of (++) indicating high quality, (+) indicating acceptable quality, (-) indicating low quality, and (- -) indicating unacceptable quality. The SIGN methodology proves to be a highly efficient rating system for assessing the quality of methodologies used in the included studies. Therefore, to compare the results of the SIGN analysis with the scores on our specifically designed scale, Table 2.5 displays a comparison between the SIGN risk of bias assessment, our customized evaluation of intervention quality, and the conclusions made by the authors.

Table 2.5. Risks of bias, proof level rating and authors conclusions.

Study	How well was the study done to minimize bias?	Is the overall effect due to the study intervention?	Note on the custom scale	Summarise the authors' conclusions. Add any comments on your own assessment of the study, and the extent to which it answers your question and mention any areas of uncertainty raised above.
García-Rudolph and Gibert García-Rudolph and Gibert, 2014	+	No	2	The authors introduced two methods for identifying neurorehabilitation (NRR) in patient samples. Both methods prioritize difficulty level and dosage to enhance rehabilitation effects. Clinical validation on a limited sample (n=10) favors the sectorized annotated plan strategy for optimal difficulty targeting.
Fermé et al. Fermé et al., 2020	- -	Not applicable	Not applicable	The authors proposed a framework to personalize the treatment of a cognitive rehabilitation tool. Their framework involves determining participants' cognitive profiles and employing a belief revision system for continuous cognitive level updates. This strategy aims to sustain an ideal difficulty level and motivation, yielding optimal rehabilitation outcomes.
Xu et al. Xu et al., 2018	+	Yes	7	The authors proposed an individualized serious game for cognitive training by employing a daily life recording strategy and intelligent techniques to incorporate visual lifelogs into training. The results indicate moderate effects on user adherence (significant difference in playing frequency on a Wilcoxon signed-rank test, $p=0.049$, Hedge $g=0.39$) and engagement (significant difference on paired t-test, $t(25) = 3.410, p = 0.001$) in favor of the personalized strategy. However, cognitive improvements were not observed ($p=0.691$) (paired t-test, $t(25)=-0.5, p=0.691$). While this marks a promising step towards optimal content individualization, the crossover design prevents efficacy assessment for cognitive improvement.
Reidy et al. Reidy et al., 2020	+	Yes	6	The authors proposed a VR based CT and used intelligent strategies to automatically extract and classify affects from EMG data. Qualitative feedback analysis suggests that the individualized condition enhances feelings of competency and appropriate challenge. However, the study's limited sample size and crossover design preclude assessing cognitive improvement efficacy.
Kitakoshi et al. (a) Kitakoshi et al., 2015	+	No	4	The authors assessed the impact of a difficulty adjustment reinforcement learning algorithm (DA) and a break offering system (DABO). revealed higher enjoyment in the control condition and greater motivation in the DABO condition. Questionnaires indicated the DA algorithm offered suitable difficulty for most participants. Learning path analysis indicated appropriate difficulty levels in the DA condition. Nonetheless, the study's small sample size and crossover design hindered efficacy assessment for cognitive improvement.

Continued on next page

Table 2.5 – continued from previous page

Study	How well was the study done to minimize bias?	Is the overall effect due to the study intervention?	Note on the custom scale	Summarise the authors' work and conclusions. Add any comments on your own assessment of the study, and the extent to which it answers your question and mention any areas of uncertainty raised above.
Kitakoshi et al. (b) Kitakoshi et al., 2020a	-	No	4	The authors proposed a personalized CT of memory through reinforcement learning. A preliminary study assessing the impact of 2 structures of the activity space (9 difficulty vs 13 difficulty levels) favored a lower number of difficulty levels. Interviews revealed inter-subject variability in optimal activity space perception, suggesting further investigation (study 2 below).
Kitakoshi et al. (b) Kitakoshi et al., 2020a	-	No	4	The authors proposed individualized memory CT through reinforcement learning, comparing low-number (9) and high-number (13) difficulty level activity structures. Questionnaire analysis indicated that the high-number condition required less effort and allowed longer play sessions. Learning path and success rate analysis indicated high-number difficulty levels was better suited for difficulty adjustment algorithms. However, limited sample size and crossover design impeded assessing cognitive improvement efficacy.
Rath-nayaka et al. Rath-nayaka et al., 2021	- -	No	1	The authors proposed an individualized cognitive rehabilitation based on a reinforcement learning algorithm (Q-learning). The intervention group exhibited performance improvement across all proposed cognitive activities. Notably, the study lacks information about cognitive performance, subjective questionnaires, or inter-group comparisons.
Shen and Xu Shen and Xu, 2021	++	Yes	5	The authors proposed a recommendation algorithm for personalized cognitive training. Pre-test comparisons in cognitive performance showed no difference between groups (independent t-test, $t(15)=1.4$, $p>0.05$ for processing speed and $t(15)=-1.02$, $p=0.32$ for memory quotient). Post-test cognitive performance analysis revealed improvements only for the intervention group in processing speed (paired sampled t-test, $t(15)=-2.62$, $p=0.02$) and in memory quotient ($t(15)=-2.60$, $p=0.02$).
Sandeep et al. Sandeep et al., 2020b	- -	Not applicable	Not applicable	The authors aimed to compare machine learning algorithms (Hidden Markov Models (HMM), Kalman filter (KF), and Long Short Term Memory (LSTM)) for predicting participant skill levels. Using data from a cognitive training intervention with the Recall game, history-driven HMM demonstrated better fit than HMM with a universal transition matrix (RMSE=5.6%). Both HMM-based models effectively predicted skill levels. KF and LSTM estimated performance and skill levels but with weaker accuracy (RMSE=18.83% and 9.34% respectively).
Continued on next page				

Table 2.5 – continued from previous page

Study	How well was the study done to minimize bias?	Is the overall effect due to the study intervention?	Note on the custom scale	Summarise the authors' work and conclusions. Add any comments on your own assessment of the study, and the extent to which it answers your question and mention any areas of uncertainty raised above.
Sandeep et al. Sandeep et al., 2020b	- -	Not applicable	Not applicable	The authors aimed to compare machine learning algorithms (HMM, KF, and LSTM) for predicting participant skill levels. The dataset included diverse learning trajectories from a Recollect cognitive training intervention. HMM with a universal transition matrix displayed better fit (test RMSE=12.54%) than history-driven HMM. Both HMM-based models effectively predicted skill levels. Study results differed from Study 1, revealing sensitivity to algorithm choice during initial data generation (difficulty adjustment procedure during initial intervention). KF and LSTM estimated performance and skill levels with less accuracy (RMSE=31.52% and 18.77% respectively).
Wilms Wilms, 2011	-	No	3	The author introduced a difficulty-adjusting reinforcement learning algorithm (actor-critic) for cognitive training. While the algorithm adapted difficulty levels, the study design precluded drawing conclusions about the approach's effectiveness.
Solana et al. Solana et al., 2014	+	Yes	5	The authors presented a clustering-recommendation strategy for individualized cognitive rehabilitation sequences. Comparisons of selected tasks and difficulty levels chosen by the intelligent strategy and the manual planning showed significant differences ($p < 0.001$). No cognitive improvement disparities were observed between the two planning methods ($p = 0.34$).
Zini et al. Zini et al., 2022	++	Yes	5	A reinforcement learning algorithm (SARSA) was proposed for individualized cognitive training. Results showed that participants started with homogeneous pre-test scores (2-tailed T-test, $p = 0.42$) and both groups improved after training (2-tailed paired T-test, $p = 1.7 \times 10^{-5}$ for group intervention and $p = 0.02$ for group control). Intervention participants using the RL algorithm showed greater cognitive improvement than control group (2-tailed T-test, $p = 4 \times 10^{-4}$). Learning trajectories indicated no significant success rate differences (2-tailed T-test, $p = 0.56$). On all trained tasks, the intervention group completed fewer activities on average than the control group. Follow-up evaluations demonstrated no between-group performance differences (task 1: $p = 0.33$, task 2: $p = 0.06$). An additional experiment with a modified RL algorithm (fine-tuned policy) showed no cognitive improvement differences but required fewer activities.
Zebda et al. Zebda et al., 2022	-	No	3	The authors proposed individualized cognitive training via robot interactions using reinforcement learning (Q-learning). Multiple case studies highlighted participants' successful identification of the adaptive condition, with semi-structured interviews emphasizing participant enjoyment.
Continued on next page				

Table 2.5 – continued from previous page

Study	How well was the study done to minimize bias?	Is the overall effect due to the study intervention?	Note on the custom scale	Summarise the authors' work and conclusions. Add any comments on your own assessment of the study, and the extent to which it answers your question and mention any areas of uncertainty raised above.
Eun et al. Eun et al., 2022	- -	No	4	This study introduced individualized cognitive training based on participant skill levels, utilizing a LSTM model for dynamic difficulty adjustment. The intervention group exhibited improved quality of life, certain geriatric depression test components, and mini-mental status examination results. Pre-post cognitive performance comparison showed significant improvement in all cognitive activities (except one) (repeated measure ANOVA, $t=2.76$ $p=0.006$ for memory training, $t=5.94$, $p=0.00$ for vision adaptation, $t=10.4$, $p=0.000$ for icon training, $t=5.423$ $p=0.000$ for graph training). The study design did not allow for separating the personalized procedure's impact from the training program itself.
Tsiakas et al. Tsiakas et al., 2018	- -	Not applicable	Not applicable	The authors introduced socially assistive robots for cognitive training (CT), which tailor learning by monitoring task engagement and performance. Their approach involves modeling artificial participants, training reinforcement learning (RL) models, and assessing them in a virtual environment. Results indicate RL models effectively generate distinct policies for various user profiles.
Book et al. Book et al., 2022	- -	Not applicable	Not applicable	This study suggests an individualized cognitive training design based on performance prediction through logistic regression. However, no data is provided to support the proposal (study protocol).
Singh et al. Singh et al., 2022	- -	Not applicable	Not applicable	The authors presented data augmentation techniques and deep-learning strategies (CNN, LSTM, CNN-LSTM) for predicting adherence to cognitive training. Model fitting showed successful training and prediction on the dataset, with approximately 75% accuracy, AUC, and F-score.

2.3 Results

2.3.1 Descriptive results

The systematic review processed seventeen papers including nineteen studies (12 in journals and 7 in proceedings articles) that have been published from 2011 to 2022 (Table 2.3). Almost 70% (n=13) of included papers were published during the last three years demonstrating the relatively low maturity of the field. Six studies are not empirical in nature and provide from either a study protocol (n=2, Fermé et al., 2020 and Book et al., 2022) or a feasibility study (formal validation) evaluating new methods on existing datasets (n=4, Sandeep et al., 2020b, Singh et al., 2022 and Tsiakas et al., 2018). According to the SIGN methodology for study design (Harbour and Miller, 2001), eleven used a controlled trial (n=8 non-randomised controlled trial and n=3 Individual randomized controlled trial) and two proposed either a case study (Wilms, 2011) or a non-comparative study (Eun et al., 2022). All the controlled trials included an active control group, either using a between-subject or within-subject design. During the intervention, the participants engaged in the same training as the intervention group, but without any adaptive procedure. It is noteworthy that none of the studies included passive control conditions where no intervention was implemented. Among the total of nineteen studies, approximately 70% (n=13) aimed to assess CT with non-clinical samples, while the remaining 30% (n=6) had a rehabilitative objective and investigated clinical samples. When assessing the research conducted on actual populations, the average sample size was 85, and the median sample size was 20. Nevertheless, within the three studies with the largest participant pools, (García-Rudolph and Gibert, 2014) presented two cohorts consisting of n=123 and n=327 individuals. Notably, the individualized CT was exclusively examined in a subgroup of n=10 participants within the treatment condition. Factoring in this information, the mean sample size adjusted to 60, with the median sample size reduced to 10.

2.3.2 Q1 & Q2. What Type of AI Techniques have been used in the field of computerized CT? What are the Subject/Domains of CT for which adaptive techniques have been designed?



Main results:

- A quarter of the included strategies employed a macro-adaptive approach, all of which were for multi-domain cognitive training.
- The majority of papers presented micro-adaptive procedures, either for predicting the next optimal learning activity directly or for extracting patterns to inform optimal pedagogical decisions.
- Half of the micro-adaptive procedures targeted a single, cross-cutting cognitive function, while the remaining half employed a multi-domain approach.

Among the nineteen papers examined, only 26% of them (n=5) put forth the utilization of a macro-adaptive procedure to customize the intervention. For example, (García-Rudolph and Gibert, 2014) aimed to improve the understanding of optimal learning objectives.

Specifically they used visual annotated plans and decision trees techniques to identify the range of difficulty known as the "neurorehabilitation range" (NRR). Other approaches suggested tailoring the entire curriculum in advance through a recommendation system that leveraged participant similarities (Xu et al., 2018) or by employing clustering techniques to identify cognitive profiles (Shen and Xu, 2021). Another proposal involved directly customizing the visual content of cognitive activities by utilizing automatic extraction of relevant images from daily visual logs (Xu et al., 2018).

The 74% (n=14) of remaining papers used a micro-adaptive approach with different strategies. These studies can be broadly categorized into two groups. The first category encompasses eight studies that primarily concentrated on directly predicting the next optimal activity by tailoring the difficulty level or the game content. Reinforcement learning methods were commonly used, with three different algorithms employed: Q-learning (n=4), Bucket brigade (n=3), and Actor-critic (n=1). Additionally, Eun et al., 2022 proposed a method utilizing deep learning, particularly Long Short-Term Memory (LSTM). The second category comprises six intelligent methods designed to extract valuable information from collected data, facilitating the generation of optimal pedagogical decisions. In all studies within this category, the choice of the next activity is based on expert hand-designed heuristics or algorithms. It is worth noting that among the studies in this category, four of them are feasibility studies without evaluating a real population. One approach aims to predict participants' performance on the next activities based on their previous trajectory. For this purpose, (Sandeep et al., 2020b) proposed Bayesian techniques such as hidden Markov models and Kalman filters, as well as deep learning utilizing LSTM. Another strategy involved employing machine learning techniques, specifically logistic regression, to predict participant performance (Book et al., 2022). In addition, (Singh et al., 2022) utilized deep learning algorithms (LSTM and CNN) directly to infer the probability of dropout in the next activities. Finally, (Reidy et al., 2020) suggested using deep learning techniques to extract useful information from EMG data.

As demonstrated by Figure 2.3, half of the studies (n=10) adopted a multiple cognitive domain approach for designing the CT. Among the nine studies focusing on single domain; the targeted functions were attention (Wilms, 2011) or working memory (Sandeep et al., 2020b, Zini et al., 2022), i.e., cognitive functions that are seen as cross-cutting to many other cognitive functions or activities, and are therefore expected in a CT to improve a large number of cognitive domains. It is noteworthy that studies using specific domain training used mostly a micro-adaptive approach with RL techniques. Consequently, these observations indicate that micro-adaptive strategies are preferred for the single cognitive domain CTs while macro-adaptive strategies are preferred for CTs with multiple cognitive domains.

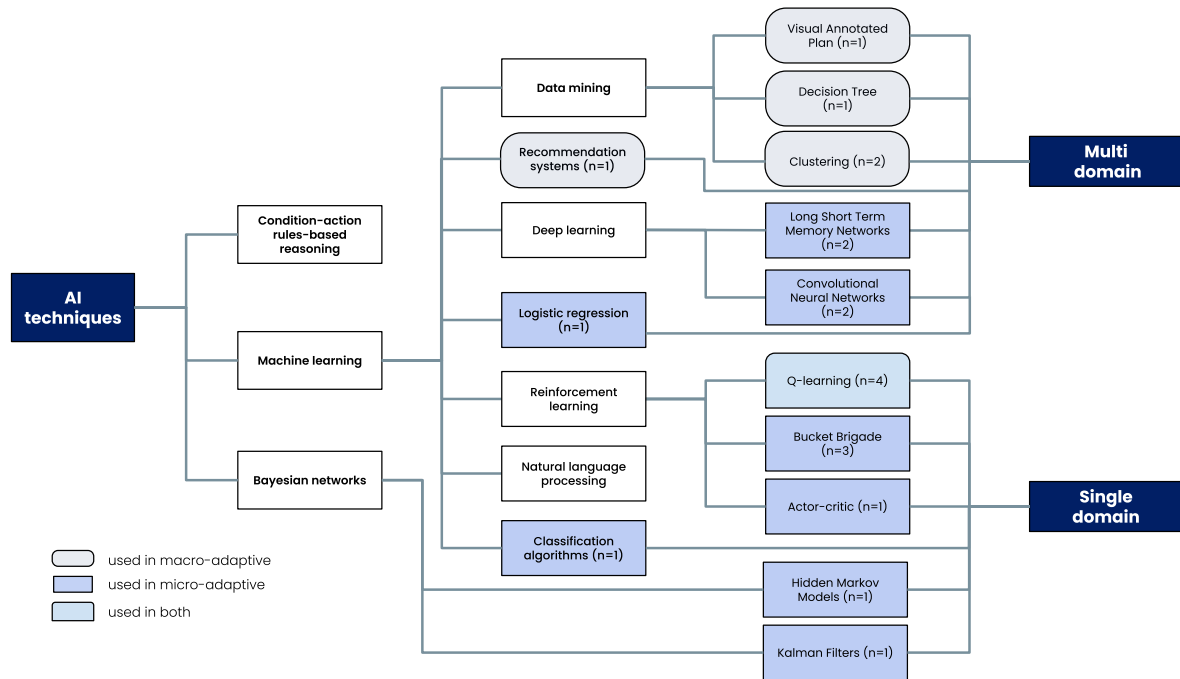


Fig 2.3. Distribution of AI techniques depending on type of CT studied (multi or single domain)

2.3.3 Q3. What populations are targeted and what are the characteristics of the CT design?



Main results:

- The majority of studies involved non-clinical adult populations.
- Experimental designs varied widely, with many conducted remotely, and no clear patterns emerged regarding intervention duration, frequency, or assessment strategies.

Among the thirteen studies with non clinical samples, twelve of them included adults ($n=7$ with older adults and $n=5$ with young adults) for whom specific domain CTs ($n=7$) were performed rather than multiple-domain ones ($n=4$). The only study including children performed a multiple-domains CT (Table 2.4). Among the six studies with a rehabilitative purpose, three of them included young adults with acquired brain injury (ABI) or traumatic brain injuries (TBI), two of them involved older adults with dementia or mild cognitive impairment (MCI). For these two types of clinical samples, the multiple-cognitive domains approach has been widely used (80%, $n=4$). The remaining study (Fermé et al., 2020) proposed a general framework that is agnostic to a specific population. Taken together, the selected studies mirrors well the two distinct literature, where CTs are often single-domain by targeting a cross-cutting function (attention, working memory) (e.g., Harvey et al., 2018), and cognitive rehabilitation programs are rather multiple-domains, as this intervention design has been shown to be more clinically effective than single domain interventions (e.g., Cicerone et al., 2019, L. Nguyen et al., 2019). In relation to the CT settings outlined in Table 2.4, the majority of interventions were carried out

remotely at participants' homes (n=9), while others took place in laboratory settings (n=4), and information was not provided for (n=2) cases. The time duration of the CT varied significantly, ranging from lengthy periods of seven months to brief sessions of only thirty minutes. However, the most commonly reported duration was two weeks (n=4), and in some studies, information regarding the duration was not available (n=4). Among the studies that documented the CT dose (n=10), there was substantial variation observed, with session duration ranging from five minutes to one hour per day. Cumulative sessions encompassed a wide range, from 30 minutes (n=1) to over 600 minutes (n=3), often with intermediate duration averaging around 140 to 215 minutes or 2 hours and 30 minutes to 3 hours and 30 minutes (n=5). A total of 48% of the studies (n=9) intended to document the training effect throughout the experiment using objective measures of performance or participants' subjective experiences related to the intervention. Regarding the assessment of participants' subjective experiences (n=8), the majority of studies (n=6) relied on manual evaluations (non standardized measurements). The subjective evaluations were related to several dimensions such as engagement, game preferences, motivation or perceived difficulty. In (n=6) studies, pre- and post-intervention comparisons of cognitive performance were conducted.

2.3.4 Q4. How effective are they in empirical CT studies? What effects are reported (NFT and everyday life transfer effects)? Are the effects dependent on the CT design (content, dose, location) and the targeted sample?



Main results:

- In one study, no distinctions were found between groups in pre-post assessments, highlighting that an automated individualized procedure exhibited equivalent efficacy to a manual approach.
- Two studies exhibited more substantial cognitive enhancements in post-test measurements, specifically in near-transfer measures.
- Several studies utilized non-comparative or cross-over designs, making it challenging to differentiate the training's impact in pre-post assessments.
- Five studies exhibited varying learning trajectories through intra-training measures, while six showed subjective differences in motivation, engagement, and play frequency between individualized and control groups.

Multiple dimensions were considered to present the effectiveness of empirical CT studies. Firstly, out of the seven studies that aimed to evaluate the progression of cognitive performance using pre-post assessments, three interventions (Eun et al., 2022; Reidy et al., 2020; Xu et al., 2018) employed either a crossover or a non comparative design, making it challenging to distinguish the impact of the control procedure from the individualized approach on cognitive performance. In the other hand, no significant differences in cognitive enhancement were found between the intervention and control groups in Solana et al., 2014, indicating that the automated procedure's effectiveness matches that of the manual approach across a comprehensive neuropsychological assessment battery. Furthermore, by

ensuring group homogeneity during the pre-test, both Shen and Xu, 2021 and Zini et al., 2022 demonstrated that the personalized approach resulted in more pronounced cognitive changes concerning measures of near transfer. Zini et al., 2022 also conducted a follow-up evaluation on the trained task but were not able to see any difference in performance between groups. Furthermore, García-Rudolph and Gibert, 2014 observed a significant improvement of performance for a small subset of participants treated with an optimal difficulty level. It is noteworthy that the 3 interventions showing significant changes were all using different CT programs (multi and single domain, different dosage, laboratory and at home based, population of healthy young adults, children and ABI patients...) and different cognitive evaluations.

Another aspect leveraged to assess the impact of the proposed intervention was to observe quantitative intra-training measures. First, two studies (Rathnayaka et al., 2021, Wilms, 2011), presented an increase in the performance on the trained task as a proxy for cognitive evolution. Then other authors showed how the individualized procedures affected the learning path proposed. Kitakoshi et al., 2015, 2020a; Solana et al., 2014 performed a comparative analysis of the learning trajectories of the non-adaptive control group and the treatment group, revealing notable differences in the patterns of learning. Moreover, analysis of quantitative intra-training observations revealed differences in the schedule of activity proposed: Xu et al., 2018 showed a significant increase in the self-management of playing frequency with the individualized game compared to the non personalized but no significant difference in intensity (average sessions length). Additionally, Zini et al., 2022 demonstrated a significant disparity in the number of episodes played, indicating that the individualized procedure facilitated greater cognitive improvement in a shorter period of time.

To gain insights into the impact of the intervention, subjective measures were also employed. Firstly, Reidy et al., 2020 utilized the Game Experience Questionnaire (IJsselstein et al., 2013) demonstrating that the individualized procedure positively influenced the participants' sense of competence. The intervention also led to a better-suited level of difficulty, as evidenced by an increase in flow and a decrease in the feeling of challenge. Then Eun et al., 2022 showcased various positive impacts of individualized CT. Participants reported an improvement in subjective health condition and overall quality of life. Moreover, there was a reduction in certain items of the Geriatric Depression Scale Short Form (GDSSF-K, Leshner and Berryhill, 1994) and a positive change in the Mini-Mental State Examination (MMSE, Tombaugh and McIntyre, 1992). It is worth noting that these results were not compared with an active control trial. Additionally, three other studies utilized custom-made questionnaires to assess subjective performance. Xu et al., 2018 found that participants using the individualized procedure experienced higher enjoyment, which was further substantiated by qualitative feedback obtained through interviews. Kitakoshi et al., 2015 revealed that the individualized intervention fostered greater motivation to use the system and maintained a suitable difficulty level. Furthermore, Zebda et al., 2022 interviews indicated that the individualized procedure was perceived as more stimulating and engaging.

2.3.5 Q5. What Type of Validation have been conducted for these new generations of computerized CT?



Main results:

- In accordance with SIGN ratings, two studies were rated as (++), five as (+), four as (-), and eight as (- -).
- The mean score on the customized scale was 4.1 out of 11.
- Increased scores on the customized scale corresponded to higher ratings on the SIGN rating scale.

Based on the SIGN rating, it was found that out of the nineteen studies examined, only 2 received the highest score (++) (Shen and Xu, 2021, Zini et al., 2022). Five studies received an acceptable score (+). Within this group, two randomized controlled trials (García-Rudolph and Gibert, 2014, Solana et al., 2014) focused on clinical samples, and the reason for not receiving a (++) grade was due to issues related to participant randomization and blinding. The remaining three studies (Xu et al., 2018, Reidy et al., 2020, Kitakoshi et al., 2015) received a (+) grade primarily because of their implementation of a crossover design.

Among the 19 studies, four studies received a (-) score. Two of these studies (Kitakoshi et al., 2020a) utilized a crossover design but lacked important information in their reports (see Appendix 2.5), while the other two studies (Wilms, 2011, Zebda et al., 2022) did not include any control group. Additionally, eight studies received a (- -) score. Six of these studies were either proposing a study protocol or conducting a feasibility study. The remaining two interventions (Rathnayaka et al., 2021, Eun et al., 2022) were assigned a (- -) score due to the absence of a control group and a lack of important information (see Appendix 2.5).

Studies that obtained a (++) score achieved an average score of 5 on the customized scale, while studies with a (+) score had an average score of 4.8. For studies receiving a (-) score, the average score was 3.5, and for studies with a (- -) score, the average score was 2.5 (whenever applicable). These findings emphasize a noticeable correlation between risk evaluations and the number of standards fulfilled in CT research. Furthermore, the results demonstrate that the majority of studies (n=12) did not meet the acceptable criterion of the SIGN methodology (+), and none of the studies fulfilled all the standards outlined by the customized scale. The average score across all studies was 4.1.

2.4 Discussion

This SR explored the wide array of AI techniques employed to enhance individualized CT. To begin with, the deployment of macro-adaptive strategies, which may draw from participant resemblances or the formation of cognitive profiles, facilitates the utilization of existing knowledge in the development of individualized schedules for cognitive tasks. These approaches prove particularly valuable when implementing multi-domain CT, especially when multiple cognitive processes are involved in numerous activities. As evidenced in various cognitive rehabilitation studies (García-Rudolph and Gibert, 2014, Fermé et al., 2020), health professionals often face challenges in selecting the most suitable sequence

of activities. Consequently, employing macro-adaptive strategies that can leverage data from large cohorts presents a promising avenue for developing effective interventions. Furthermore, as suggested by (Singh et al., 2022), these methods offer insights into the mechanisms underlying improvements and adherence to the interventions. Nevertheless, macro-adaptive procedures, by their inherent nature, do not entail direct adaptation of the content and difficulty levels of individual tasks. Instead, they yield more intricate outcomes that require comprehensive analysis. Consequently, given the still exploratory state of the field, the majority of studies reviewed here have focused on tailoring single or a few training tasks using a micro-adaptive approach. Micro-adaptive procedures propose to use previous interaction with the user to personalize the learning trajectory. Most modern AI techniques leverage collected data from the training path and thus fits particularly well with the CT paradigm where many short episodes are played. While this task may appear less challenging than planning a complete curriculum in advance, it requires data-efficient strategies to identify and suggest activities with appropriate dynamics for tailoring the path to each participant’s needs. As a result, many studies employing deep learning or machine learning techniques are still undergoing formal validation and are currently in the feasibility study stage, being tested solely on previous data and not yet evaluated on real participants (Singh et al., 2022). Reinforcement learning paradigm, where the artificial teacher, or system, proposes activities based on the participant’s previous interactions looks like a particularly good fit for that purpose but also has its limitations: to enhance data efficiency, most strategies rely on tabular approaches, which in turn restrict the number of parameters available for adaptation. Moreover, for several studies of this SR, a two-stage time consuming strategy is commonly employed where a first teacher policy is being trained on a group of participants and is then fine-tuned for each participant (e.g., Zini et al., 2022). Finally, a third family of strategies based on recommendation algorithms show promise but also require sufficient pre-collected data to achieve efficiency in personalizing the training experience.

Additionally, it is noteworthy to observe that most micro-adaptive strategies propose a personalization based on the difficulty of the cognitive tasks. As proposed by Vygotsky and Cole, 1978b, Csikszentmihalyi, 2000, Ryan and Deci, 2017, the key idea is to propose an optimal difficulty in order to foster training gains and motivation and is tightly connected with the optimal cognitive challenge (Lövdén et al., 2010a). For that purpose, while many studies primarily focus on choosing the correct parameter set, certain approaches suggest modifying the content according to participants’ visual cues. This alternative approach to customizing training harmonizes effectively with Mayer’s Cognitive Theory of Multimedia Learning and his personalization principle (R. Mayer, 2017). Along this line of customization, the adaptation of interactions through assistive robots (Zebda et al., 2022), chatbots (Kitakoshi et al., 2015) or virtual reality (Reidy et al., 2020) is likely to be another key factor for participant engagement. In this direction, it is conceivable that recent advancements with large language models will enable better dialogic adaptation, potentially impacting motivation and engagement (Abdelghani, Wang, et al., 2023). However the issue of reproducibility becomes increasingly significant when incorporating complex data-driven strategies. Ensuring the transferability of models and reproducibility of experiments raises a challenge in the absence of provided code or dataset accessibility across the included papers. This lack of transparency is of growing criticality for research reproducibility. Consequently, the field of AI frequently encounters a black box scenario, which hampers reproducibility efforts. In the context of CT and its human stakes, it is important to understand the methods used to individualize the training path for each trainer, and if these are not sufficiently transparent, they must at least be traced or documented as

predictors of targeted intra-training mechanics.

The findings of the present SR highlighted the current state of individualized CT as a field with relatively low maturity. Following the recommendation put forth by C. S. Green et al., 2019, there exists an urgent requirement to clarify the objectives of each study within the CT community. To achieve this, C. S. Green et al., 2019 proposed a distinction among several categories: feasibility, aimed at "testing the viability of a particular paradigm or project"; mechanistic, focused on "identifying the mechanism(s) of action of a behavioral intervention for cognitive enhancement"; efficacy, with the goal of "validating an intervention as the primary cause of cognitive improvements beyond any placebo or expectation-related effects"; and effectiveness, concerned with evaluating whether a given intervention "achieves the desired and predicted positive impact, often involving real-world outcomes". Notably, none of the studies encompassed in this review employed such terminology, yet it becomes obvious that the majority of interventions are currently positioned at the feasibility or mechanistic study stages. This observation is reinforced by the predominant focus of these interventions on non-clinical populations, specifically targeting young adults. Moreover, it is essential to note that very few studies adhered to the gold standard of Randomized Controlled Trials (RCTs). While RCTs have certain limitations, such as the need for stable, long-term interventions spanning several years to establish robust scientific evidence, they remain a crucial benchmark for evaluating interventions (Mohr et al., 2015).

Specifically in the context of individualized interventions, mere observation of favorable and definitive outcomes arising from an individualization algorithm in the context of pre-post training effects is insufficient. What is imperative is the ability to elucidate its impact on the active cognitive mechanisms underpinning the training process, and subsequently, to establish a coherent connection between these mechanisms and the resultant effectiveness. A deep understanding of the causal relationships existing between the behaviors governing individualization and the intricate mechanics of training, as well as their collective impact on training effectiveness, stands as an essential foundation for the advancement of these emerging computerized cognitive therapies. To attain this level of understanding, the incorporation of judicious supplementary evaluations holds utmost significance. These assessments should aim to gain a comprehensive understanding of algorithm behavior, allowing researchers to gauge the effectiveness and adaptability of the individualized interventions. Moreover, the integration of subjective questionnaires can help evaluate participants' motivation and engagement levels, providing valuable insights into their experiences and receptiveness to the intervention. Such subjective metrics also possess the potential to shed light on how the customization of training can serve as an efficient mechanism for enhancing participants' commitment to the program, thereby potentially mitigating the unfortunate phenomenon of attrition, which regrettably tends to manifest, particularly among older adults or clinical cohorts, who nonetheless manifest a demonstrable necessity for the training regimen (Depp et al., 2014a).

The field of individualized CT, as depicted in this SR, mirrors the broader literature on CT, which is characterized by methodological and empirical weaknesses in assessing intervention effectiveness leading to controversy among experts (Allaire et al., 2014). This review highlights significant heterogeneity in methods, cognitive domains, dosage, and study populations, aligning with findings from other studies appealing for a greater compliance with more rigorous methodological standards. Unlike prior meta-analyses that presented mixed results regarding the dose-dependency of training effects in CT (e.g. at least 10 sessions for Kelly et al., 2014 or 3 or fewer sessions in Lampit et al., 2014a), our study does not provide evidence supporting a particular direction. Additionally, while previous

studies (e.g., L. Nguyen et al., 2019) have indicated that multi-component training may exhibit greater efficacy compared to single-component training, half of the investigations included in this review primarily concentrate on attention or working memory functions. The emphasis on these functions is justified by their crucial cross-cutting role in everyday activities and their vulnerability to impairment in various cognitive pathologies (Depp et al., 2014a). Lastly, as emphasized in several systematic reviews (Harvey et al., 2018; Lampit et al., 2014a, and the reanalysis of 2018 Webb et al., 2018c), the definition of a suitable cognitive battery that assesses NFT and ecological transfer significantly influences the measured outcomes and the conclusions drawn. Notably, improvements in certain cognitive domains might not manifest when assessed using different cognitive tasks (see Webb et al., 2018c). Furthermore, it's essential to keep in mind that enhancements observed in a specific cognitive domain do not necessarily guarantee true transfer, as evidenced by the case of verbal memory training and its effects on neuropsychological tests (Harvey et al., 2018). The studies included in this review underscore the substantial diversity in assessment methods and the limited availability of approaches to evaluate broader ecological transfer.

2.5 Conclusion

The present systematic review puts forth a range of potential methodologies to better address interindividual differences and offers captivating prospects for the future development of the field. The hypothesis of heightened engagement and motivation found support in the limited number of studies that investigated this aspect. Further investigations are necessary to validate whether AI strategies can truly empower each participant's cognitive potential, and then ensure CT benefits for all. Although additional research endeavors adhering rigorously to methodological standards are still required, the first results appear promising. In line with this drive for progress, a notable observation emerged during the course of this systematic review: the number of included papers nearly doubled, particularly in the year 2022. This indicates a growing interest in individualized cognitive training and underscores the optimistic outlook for the field's future.

Appendix

Risk of bias

For the evaluation of potential bias within each controlled trial incorporated in this comprehensive review, we initially employed the Cochrane checklist to ascertain the underlying study design of each individual study. Subsequently, with regard to each dimension outlined in Table 2.6, a grading scale ranging from 0 (not applicable) to 3 (adequately addressed) was employed to assign scores. In instances where specific criteria were not explicitly stated, a grade of 1 was attributed. In situations involving crossover studies, a score of 2 was designated for the criterion pertaining to participant similarity between the control and intervention groups. Furthermore, for studies utilizing a crossover design, a score of 2 was ascribed to the concealment method criterion. In cases where researchers employed handcrafted questionnaires despite the existence of standardized alternatives, the assigned grade was automatically reduced by 1. The ultimate assessment presented in this table is established through the following categorizations: studies scoring 0 criteria as poorly addressed are denoted as (++), while those with 1 to 3 criteria marked as poorly addressed receive a rating of (+). Conversely, studies where more than 3 criteria are deemed inadequately addressed are indicated as (-), and non-comparative studies are indicated with (- -).

Table 2.6. Grading scale results.

Study	Appropriate and clearly focused question	Randomized assignment to treatment group	Adequate concealment method	Double blind allocation	Homogeneity between groups	Only difference is treatment	Standard, valid and reliable measures	Percentage of dropouts	Intention to treat analysis	Valid multi-sites comparison	Grade
García-Rudolph and Gibert García-Rudolph and Gibert, 2014	3	1	2	1	3	3	3	2	3	0	+
Xu et al. Xu et al., 2018	3	3	3	1	1	3	2	2	3	0	+
Reidy et al. Reidy et al., 2020	2	2	3	1	1	3	2	2	3	0	+
Kitakoshi et al. (a) Kitakoshi et al., 2020a	2	2	3	1	1	3	2	2	3	0	-
Kitakoshi et al. (b) Kitakoshi et al., 2015	2	2	1	1	1	3	2	2	3	0	+
Shen and Xu Shen and Xu, 2021	3	3	3	3	3	2	2	2	3	0	++

Continued on next page

Table 2.6 – continued from previous page

Study	Appropri- ate and clearly focused question	Random- ized assign- ment to treatment group	Ade- quate conceal- ment method	Double blind al- location	Homogene- ity between groups	Only dif- ference is treatment	Standard, valid and reliable measures	Percent- age of dropouts	Inten- tion to treat analysis	Valid multi- sites compar- ison	Grade
Solana et al. Solana et al., 2014	3	1	3	1	3	3	3	2	3	0	+
Zini et al. Zini et al., 2022	3	3	3	3	3	3	2	2	3	0	++
Zebda et al. Zebda et al., 2022	3	2	2	1	1	3	1	2	3	0	-

Chapter 3

Broadening the Lens: A Narrative Review of Parameters Shaping Multi-Object Tracking Performance and Training Efficacy

Aims

After reviewing the existing techniques for individualizing a CT, we now focus on the specific task we will use for training: the MOT. This chapter provides an overview of current theories and knowledge about this task, particularly examining the literature through the lens of the manipulated parameters.

Contents

3.1	Introduction	92
3.2	The MOT underlying theories	94
3.3	Focus on MOT training	108
3.4	Outcomes from MOT practices in Real-Life	114
3.5	Conclusion	117

Abstract

The MOT task, known for its dynamic nature and reliance on controlled attention, serves as a core activity in computerized CT. Despite its popularity, the cognitive mechanisms underlying this specific attention-based CT and its outcomes (near and far transfer effects) are still poorly defined, and lacking comprehensive documentation in literature. To move forward, our purpose is a focus on the adjustable parameters (number of targets, target's speed, etc.) into the MOT task for providing new insights on MOT-based CT across neurotypical and non-neurotypical trainees. Three aims were delineated : (1) Identifying the cognitive processes influenced by each adjustable parameter of the MOT task ; (2) Determining which specific parameters, when progressively modified during repeated MOT practice (i.e., CT), yield the most significant improvements in MOT performance ; and (3) Assessing which improvements in MOT performance lead to the most effective transfer effects (i.e., CT efficacy). This final step included also the MOT outcomes in real-world settings. This comprehensive approach helps clarify the role of MOT in cognitive enhancement strategies and its potential for broader cognitive benefits.

Collaborators

Pech Marion (PM), Adolphe Maxime (AM), Oudeyer Pierre-Yves (OPY), Sauzéon Hélène (SH).¹

¹According to the Contributor Roles Taxonomy, collaborators contributed in various roles: Writing - Review & Editing (all); PM and AM contributed equally as co-first authors; Conceptualization (PM, AM, SH); Data Curation (PM, AM, SH); Formal Analysis (PM, AM, SH); Investigation (PM, AM, SH); Methodology (PM, AM, SH); Funding Acquisition (PM, OPY, SH); Project Administration (PM, OPY, SH).

3.1 Introduction

Cognitive Training (CT) interventions have gained popularity as cost-effective and easy-to-administer options for individuals across all age groups. These non-pharmacological interventions typically involve the repetitive and extensive practice of computer-based cognitive tasks, with the assumptions that improvements in trained tasks will transfer to non-trained tasks that are cognitively similar (near transfer effect) or related (far effects including real-life activity assumed as tapping on trained cognitive processes). However, the empirical evidence regarding the CT effectiveness is weak (Lampit et al., 2014c; Simons et al., 2016c; Webb et al., 2018a). While most CTs induce a large and robust performance gains in the trained or closely related-tasks (near effect) across various ages (e.g., Von Bastian & Oberauer, 2013, these training effects often fail to transfer to untrained tasks (far effect) or to improvements in everyday life functioning (ecological transfer effect) (De Simoni & von Bastian, 2018; Guye & von Bastian, 2017). This discrepancy raises questions about the underlying mechanism of CT and the conditions necessary for genuine transfer effects to occur.

Despite of this, some studies have reported positive outcomes of an ecological transfer to everyday life functioning, particularly when the CT specifically taps on executive functions such as attentional control or perceptivo-cognitive skills (i.e., processing the most important information at the right time to make accurate decisions) (Binder et al., 2016). These positive results have been shown in diverse populations including athletes (Romeas et al., 2016b), drivers (Bowers, Anastasio, et al., 2011), professional (Vater et al., 2021b), children (Bertoni et al., 2019a; Peng & Miller, 2016b; Tullo, Bertone, et al., 2018), older adults (Ballesteros et al., 2020b). Importantly, the inclusion of attentional control demands, such as divided attention tasks or controlled shifting tasks, has been one key factor linked to these successful outcomes (e.g., Bowers, Anastasio, et al., 2011, suggesting that the nature of the cognitive tasks used in CT is a critical ingredient for achieving genuine gains.

As exemplified, the Multi-Object Tracking (MOT) task, a complex dynamic task focusing on attentional control, has been widely attractive in the field of computerized CT. As illustrated in Figure 3.1, in a typical MOT task, observers are initially presented with several identical objects, a subset of which is then highlighted to indicate their status as targets. Following this, all objects, once again appearing identical, move independently and unpredictably around the display. At the end of the movement phase, observers must identify the original targets (B. J. Scholl, 2009). Figure 3.1 shows two variants of the MOT task.

Within MOT, several parameters can be manipulated such as the **number of targets**, the **number of distractors**, the **speed of the object's movements**, the **spatial distance between targets** and the **tracking duration**. The attentional control increases as the value of each parameter increases, especially as parameter combinations include high values (A. Holcombe, 2023; Vater et al., 2021b). This multidimensional parameterization of attentional control makes the MOT task an ideal candidate for a progressively refined training program. As a result, the MOT practice as CT has demonstrated tangible benefits in daily activities such as driving and gaming, or yielded to cognitive enhancements in areas like executive function and working memory. Despite its widespread use, the cognitive underpinnings of such attention-based CT remain unclear and not well documented in terms of their relationships with transfer effects (Strobach & Karbach, 2021; Vater et al.,

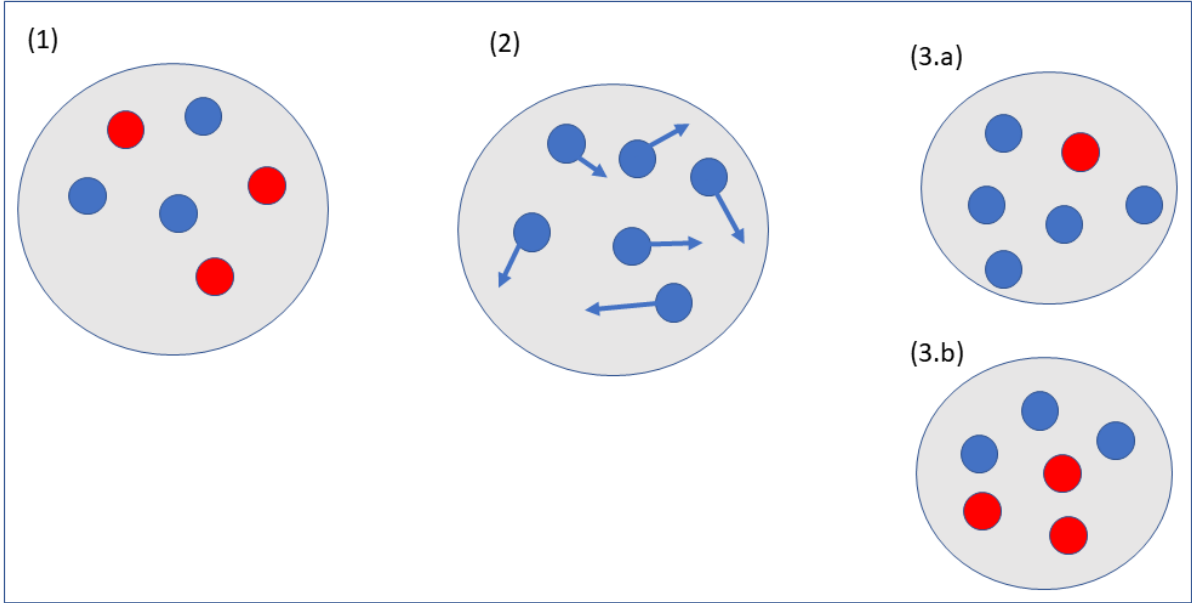


Fig 3.1. An example of a variant of the Multi-object tracking task (1) Three discs are briefly colored red to mark them as tracking targets. (2) All items look the same and move around randomly on the screen. (3) (a) At the end, one item is highlighted and the participant reports if it was a target or not. (3) (b) At the end of the motion phase, the participant must recognize the initial discs identified as targets. Animations of many different variants of this task can be viewed at or downloaded from <https://perception.yale.edu/Brian/demos/MOT.html>

2021b) for a critical review on the marketed Neurotracker program). While previous reviews, such as those by Meyerhoff et al. (2017b) or A. Holcombe (2023), have delved into the existing body of knowledge on the MOT task, this narrative review aims to examine the underlying mechanisms of MOT through the prism of task parameter manipulation. Hence, our goal is three-fold: (1) to map the cognitive processes associated with adjustable MOT parameters; (2) to identify the associations between the CT variations of MOT parameter(s) and the best improvements in MOT performance; and (3) to identify which MOT performance enhancement(s) yields the best transfer effects. For each step, we will assess how the effects of parameter manipulation in neurotypical populations align with those in neurodiverse groups, including individuals with attentional and executive disorders.

As such, our first contribution is fundamental in nature as it synthesizes findings on classical MOT in neurotypical and non-neurotypical individuals in order to identify the cognitive mechanisms at work in MOT according to the parameters manipulated and individual specificities (type of cognitive impairment). The second contribution, with a more applied focus, compiles evidence on the use of MOT tasks for CT. It examines how different task parameters and individual cognitive characteristics affect training outcomes, while also exploring the broader implications of training through the examination of near, far and ecological transfer effects. Operationally, we aim to answer to the three following questions:

- **Q1.** (a) How do changes in parameters affect performance in MOT tasks, and what theories explain these effects? (b) Which insights are provided neurodiverse populations on this question?

- **Q2.** (a) How does cognitive training using MOT tasks affect MOT performance (near transfer effect), and does changing MOT task parameters influence these effects? (b) How does MOT Practice affect cognitive functioning (near to far transfer)? Do neurodiverse conditions impact these outcomes?
- **Q3.** (a) What practices-related changes in MOT performance are linked to real-world transfer effects in neurotypical individuals? (b) Do neurodiverse conditions alter these effects?

3.2 Cognitive underpinnings of MOT task according to adjustable parameters

Two main families of approaches have made complementary contributions to our understanding of the mechanisms involved in the MOT task. The first is the classical analytical approach of cognitive psychology, which consists in flushing out the cognitive mechanisms by manipulating the parameters of the task (e.g., the number of targets). The second is a global approach, which seeks to uncover mechanisms by studying the correlation between MOT performances and tests specifically assessing cognitive functions (e.g., working memory tests).

3.2.1 Q1. (a) How do changes in parameters affect performance in MOT tasks, and what theories explain these effects?

Analytical approach: MOT parameters manipulations

Many studies have explored the MOT task as a phenomenon by examining parameters that influence performance and the underlying cognitive processes (Meyerhoff et al., 2017b). Key parameters identified as either facilitators or barriers to MOT task success include the **number of targets**, their **speed**, the **tracking duration** and the **perceptual characteristics** of targets within the task environment.

A non exhaustive set of studies manipulating MOT parameters as a way to better understand tracking mechanisms is proposed in Figure??.

Numbers of targets. The quantity of targets to be tracked is a pivotal parameter in MOT task performance. This parameter has been one of the first manipulated by Z. W. Pylyshyn and Storm (1988b) to assess the FINST theory. The FINST (fingers of instantiation) model (e.g., Z. Pylyshyn, 1994) suggests that the brain assigns visual spatial indexes (FINSTs) to a limited number of objects in the visual field. These indexes act as perceptual "hands" or "fingers" attaching to objects and allowing their tracking without the need for detailed attention or conscious recognition of their features. This theory relies on low-level early vision processes, enabling tracking to be carried out "pre-attentively", without involving memory representations. If the flash highlighted a tracked target, subjects were instructed to respond as quickly and accurately as possible by pressing a key. The foundational study demonstrated that participants could successfully track **four to five targets with approximately 85% accuracy** while the increment in targets to be tracked progressively demands attention. As a result, it was hypothesized that periodic attentional effort was needed to reactivate index binding. While tracking mechanisms were

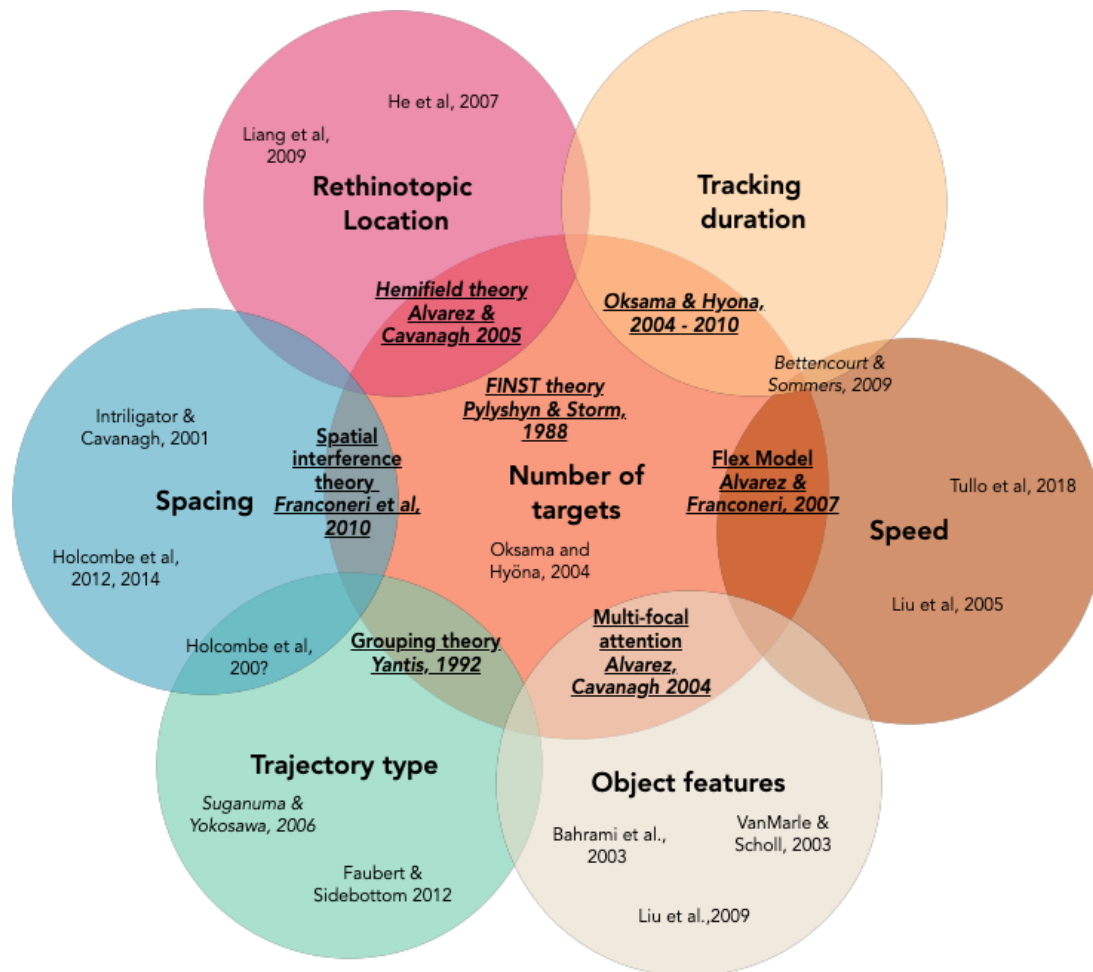


Fig 3.2. A non exhaustive set of studies manipulating MOT parameters as a way to better understand tracking mechanisms

initially described as preattentive in earlier studies (Z. W. Pylyshyn & Storm, 1988b), this term was later substituted by “preconceptual” in subsequent research (Z. W. Pylyshyn, 2001b). This change in terminology helps clarify the cognitive mechanisms involved in the task, especially noting that focused attention may become important as the number of targets increases (Meyerhoff et al., 2017b; Z. W. Pylyshyn, 2001b). Building on Z. W. Pylyshyn and Storm (1988b) initial findings, subsequent studies, including those employing Alvarez’s FLEX model Alvarez and Franconeri (2007a) and Horowitz and Cohen (2010), have consistently observed the set size effect. Specifically, Oksama and Hyönä (2004a) found that while keeping tracking time constant, there were linear (explaining 87.6% of the variance), quadratic (6.6% variance explained), and cubic (5.3% variance explained) decreases in performance as set size increased. The linear trend suggests that tracking might not be purely parallel as Z. W. Pylyshyn and Storm (1988b) described, because performance should remain stable until the limit of visual indices is reached. The nonlinear trends suggest the possibility of shifting the attentional spotlight from one target to another, indicating that attention allocation might degrade over time. This supports the debate on serial versus parallel processing and introduces the concept of **multifocal attention**. This concept investigates the potential for multiple attentional beams to be directed towards different objects simultaneously (Cavanagh & Alvarez, 2005).

Speed The speed parameter denotes the pace at which targets and distractors move across the display. As demonstrated by (G. Liu et al., 2005), tracking accuracy declined systematically with increases in object speed. As such, this observation raises the question of a size limited tracking capacity, typically limited to - the magical number (Meyerhoff et al., 2017b) - 4 targets (Z. W. Pylyshyn & Storm, 1988b). Alvarez and Franconeri (2007a) observed a significant negative correlation between the speed threshold required for accurate tracking and the number of concurrently tracked objects. Specifically, the decrement in speed threshold exhibited a logarithmic relationship with the increment in target quantity. Furthermore, participants were capable of tracking as many as eight targets, provided the speeds were adjusted to their individual tracking abilities. The finding that the tracking resource might be more adaptable than previously thought is evidenced by the decrease in speed thresholds as the number of targets increases, suggesting a balance between the number of objects and the complexity of tracking them. This led Alvarez and Franconeri (2007a) to propose the FLEX model of MOT, which posits that tracking capacity is managed by a flexible, continuous pool of resources that varies depending on the difficulty of tracking each object individually. Bettencourt and Somers (2009b) built on this concept but noted that at lower speeds and shorter tracking durations, objects tend to stay close to their initial positions, implicating visual short-term memory rather than tracking processes. They tested this by examining a wider range of speeds and set sizes with prolonged tracking times, and found that performance still deteriorates at higher speeds, supporting the concept of a flexible tracking resource. In a study exploring cognitive load, both **the number of targets** and **their movement speeds** were found to be **critical factors**. Tullo, Faubert, and Bertone (2018a) examined how abilities in 3D Multiple Object Tracking (3D-MOT) correlate with fluid reasoning intelligence, also considering working memory. Results showed that individuals who performed well in tracking multiple objects also exhibited superior fluid reasoning abilities, which were not linked to verbal intelligence. The study suggests using average tracking speed as a measure to assess 3D-MOT ability and its association with fluid reasoning intelligence, highlighting the **cognitive demands** in tasks requiring **focused attention**.

Tracking duration The tracking duration significantly impacts the difficulty of tracking in the MOT task. Oksama and Hyönä (2004a) observed that the greatest decline in performance occurred between 5 and 9 seconds (error percentage from 9.0 to 19.5%), whereas performance plateaus from 9 to 13 seconds (error percentage from 19.5 to 21.6%). The tracking performance seems susceptible to the duration effects. This observation did not fit with the pre-attentive view of Z. W. Pylyshyn and Storm (1988b), where tracking duration is expected as having no effect on MOT performance. Temporal dynamics in MOT have been manipulated through different methods of target presentation. Liang et al. (2022) utilized a simultaneous-sequential paradigm, originally proposed by Eriksen and Spencer (1969), to study MOT performance. This method involves tracking all targets at once (simultaneous condition) or focusing on a subset while the others remain static (sequential condition). The study found that with two target objects, tracking accuracy was significantly higher and reaction times were shorter in the sequential condition compared to the simultaneous condition. However, with four targets, performance improved in the simultaneous condition, suggesting the involvement of multi-focus attention mechanisms. When the number of targets increased to six, the sequential condition showed better performance than the simultaneous condition. These findings indicate that the **optimal tracking** strategy **varies** with the **number of targets**. With two targets, the task might

be managed within a single focus of attention, while four targets likely necessitate the distribution of attention across multiple foci. At six targets, the resource demands may exceed what can be efficiently managed by multiple foci, causing a shift back to reliance on single-focus attention. These results highlight the complex mechanisms of attentional tracking and suggest a strong link between the flexible functions related to working memory and dynamic spatial visual attention. By controlling the working memory load in the sequential condition, it was observed that tracking more targets requires high-level, non-automatic processes associated with executive functions. Thus, manipulating these parameters could potentially enhance cognitive functions such as shifting, updating, and inhibition, which are crucial for successful task completion.

Perceptual characteristics in terms of spatial configurations Yantis (1992b) introduced a theory suggesting that the visual system groups individual target objects into a higher-order visual representation, which aids in tracking multiple moving objects. This process involves two key stages: **pre-attentive group formation**, driven by the characteristics of stimuli, and **intentional group maintenance**, which requires focused attention. The group formation is governed by Gestalt laws, which help simplify complex scenes into more manageable shapes, whereas group maintenance involves continuous updates to the representation of moving elements, demanding higher cognitive processes like mental rotation and controlled attention. Studies by H. M. Fehd (2010) support this by showing that observers tend to focus on the invisible centroid of objects, a strategy that improves tracking efficiency by reducing the need to attend to individual objects. The theory also emphasizes the role of perceptual grouping factors such as adherence to the Gestalt law of common fate and the initial configuration of objects, which significantly influence tracking capabilities. Complex configurations that require more demanding mental transformations can hinder tracking performance. Moreover, it's observed that tracking is more effective when objects that disappear reappear at their last known position rather than at a new location predicted from their trajectory. An extension of this research by Suganuma and Yokosawa (2006) explored how synchronized trajectories between targets and distractors affect MOT performance. They found that performance improved significantly in conditions where the objects lacked a defined spatiotemporal relationship (random condition), compared to scenarios where target-distractor pairs moved together in close proximity (chasing condition) or exhibited coordinated movements with offsets (trail condition).

Another spatial feature affecting MOT is the **consideration of depth**, such as by varying the freedom of movement of objects across different depth planes. Initially deemed cognitively demanding (Faubert & Sidebottom, 2012), tracking 3D objects across various depth planes has been shown to be easier than tracking on a single depth plane in 2D (Cooke et al., 2017; Dünser & Mancero, 2009; Viswanathan & Mingolla, 2002). The additional depth information in 3D may facilitate object discrimination, thereby enhancing tracking performance. However, the advantage of 3D over 2D in MOT accuracy is influenced by the shape of the reference frame, especially at high object speeds (G. Liu et al., 2005). These findings underscore how spatial features and movement dynamics collectively impact MOT performance.

In MOT tasks where objects repel each other, the term **"spacing between objects"** denotes the minimal distance maintained between them. Tracking becomes more difficult when targets and distractors are in close proximity due to the challenge of isolating individual elements in crowded conditions. This phenomenon, known as "crowding,"

hinders the ability to distinguish a target among nearby objects. According to Bouma’s Law, crowding issues are mitigated if objects are spaced more than half their distance from the center of vision (Bouma, 1970). However, this guideline adjusts near the edges of our visual field and serves more as a general rule than an absolute (Gurnsey et al., 2011). Several factors contribute to this issue. First, the limits of attentional spatial resolution play a pivotal role. While our eyes can discern fine details, our attentional system perceives closely situated similar objects as a group or texture rather than as distinct items. This perceptual grouping makes it challenging to select an individual target from a cluster of distractors, even if they are visually distinguishable (He et al., 1997). Secondly, in densely populated visual fields, there is a tendency to generalize or average features, which complicates distinguishing between targets and distractors (Parkes et al., 2001; Pelli et al., 2004). The debate continues on whether tracking near distractors relies solely on enhancing targets without attending to distractors (“push-only” models) or involves active suppression of distractors (“push-pull” models) (Bettencourt & Somers, 2009b). Evidence has shown that **crowding negatively affects MOT performance**. Based on these insights, Franconeri et al. (2010) proposed the “spatial interference theory,” suggesting that MOT utilizes parallel processing with potentially unlimited trackers, where performance degrades due to object spacing rather than changes in speed, duration, or number of targets. Their findings indicated that tracking abilities remain consistent with stable object spacing, regardless of variations in other parameters. Conversely, A. O. Holcombe and Chen (2013) refined this view by showing that within the same visual hemifield, both increasing speed and adding more targets can impair performance, even with significant object separation. These findings suggest a finite nature of tracking resources and challenge the hypothesis of unlimited trackers, indicating a limited capacity to monitor distinct entities, thereby highlighting the significant impact of crowding and spacing on MOT performance.

Attentional resolution varies across different parts of the visual field, with studies suggesting it is finer (able to discern smaller details) in the lower visual field compared to the upper (He et al., 1997). In the context of MOT, Alvarez and Cavanagh (2005) demonstrated that MOT performance is influenced by how targets are distributed across the visual hemifields. Their study found that when targets are confined within a single hemifield, increasing their number from two to four significantly reduces tracking accuracy. However, **when the increase in targets is evenly distributed across both hemifields-either one in each or two per hemifield-tracking performance does not suffer**, suggesting a level of hemifield independence in MOT tasks (Alvarez & Cavanagh, 2005; Delvenne, 2005). Alvarez and Cavanagh (2005)’s findings underscore the notion of hemifield independence in attentional tracking. They propose that each visual hemifield (left and right) has an independent capacity for attentional processing, indicating that the left hemisphere of the brain can track objects in the right hemifield and vice versa, with each hemisphere managing attention for its respective hemifield independently. This model challenges the traditional view of a unified global attentional capacity for object tracking. Instead, it suggests that attention is dynamically and flexibly distributed across the visual field, not anchored to fixed points but able to expand, contract, and shift to encompass moving objects effectively. This concept emphasizes the adaptability and spatial variation of attentional resources in the visual field, pivotal for understanding and improving strategies in MOT tasks.

Perceptual characteristics in terms of visual features Specific visual characteristics of targets, such as color or shape, significantly influence the success of MOT tasks, particularly if these features change during task execution (Bahrami, 2003; T. Liu et al., 2009). The retention of feature values is generally less robust than location information, which can negatively impact tracking performance. For instance, targets that move like fluid substances, changing texture or appearance, pose additional challenges by complicating the determination of their precise location (vanMarle & Scholl, 2003). This complexity highlights the sensitivity of attentive tracking processes to dynamic visual information. In studies of target enhancement or distractor suppression mechanisms, it has been observed that tracking performance decreases as the size of the objects being tracked reduces (Bettencourt & Somers, 2009b). The Multiple Identity Tracking (MIT) task, as explored by Oksama and Hyönä (2004a), requires participants to track two types of stimuli: drawings of familiar objects (e.g., a coat, lobster, rocking chair) representing different semantic categories and pseudo-objects, which are object-like items without a known identity. This approach suggests that individuals track the identity-location bindings of multiple objects in a serial manner, with attentional focus closely linked to eye movements (Corbetta & Shulman, 1998; Deubel & Schneider, 1996). Currently, there is a lack of studies that examine the impact of various object attributes (e.g., semantic content like faces versus simple color attributes) on MOT performance. The Object-File Theory proposed by Kahneman et al. (1992a), which posits that in a dynamically changing visual scene, temporary memory representations- metaphorically termed "object files"- are essential. These "object files" compile different types of information (location, feature, semantic) about objects, although they are primarily identified by the spatiotemporal properties of the objects, not by their featural or semantic attributes. The theory also suggests that there is a limited capacity to maintain multiple "object files" open simultaneously, emphasizing the challenges associated with managing multiple dynamic objects in visual tracking tasks.

Summary The variety of factors influencing performance in MOT tasks underscores the complexity of such a task and highlights the ongoing questions about the mechanisms behind tracking (see in Figure ??). Since the initial conceptualization of MOT by Z. W. Pylyshyn and Storm (1988b), which introduced the idea of tracking driven by pre-attentive stimuli, numerous studies have adjusted parameters. A central debate in the field concerns the nature of the tracking resources used: whether tracking involves a limited number of discrete slots or a flexible, continuous pool of resources allocated per target. Research often utilizes manipulations of speed, number of objects, and tracking duration to explore these theories. Performance generally declines with increases in speed, number of targets, or tracking duration, with notable discontinuities in these declines. These findings lend support the theory of a more non-continuous resource pool. Additionally, studies focusing on the spacing between objects and their trajectories have indicated that crowding negatively impacts tracking performance. These findings led Franconeri et al. (2010) to propose the spatial interference theory, which suggests that tracking uses a limited discrete set of trackers and that performance declines are due to either noise from closely spaced targets or attentional capture by nearby distractors. Yet, further experiments by A. O. Holcombe and Chen (2013), which extended the range of crowding scenarios, demonstrated that performance still suffers from additional targets or increased speed even without spatial interference, suggesting limitations in tracking resources. Other investigations have examined the retinotopic location of objects, showing

decreased performance when additional targets are added within the same hemifield, which points to independent tracking mechanisms within each hemifield (Alvarez & Cavanagh, 2005; Liang et al., 2022). The debate over whether tracking is performed in parallel or serially continues as well. While some theories suggest that a single attentional spotlight may shift among targets serially, the proposition of **multifocal attention** implies that several foci can simultaneously track multiple targets. Research, including studies by Liang et al. (2022) and the MIT tasks by Oksama and Hyönä (2004a), indicates that factors like the number of targets and the level of attentional detail required can influence whether tracking in MOT tasks is done serially or through multiple focus points. This active debate highlights the evolving nature of MOT research, which explores various theories on whether tracking resources are limited and discrete or flexible and continuous, and whether tracking occurs in parallel or in series. This complexity shows the depth of understanding needed to fully comprehend the cognitive processes involved in MOT. What we see is that these tracking mechanisms are part of a larger system that connects to various brain functions like attention (in its various forms), short-term/working memory, and possibly even more complex functions like executive control (like inhibiting or switching tasks)(cf. Table 3.1)

Global approach: MOT relationships with others cognitive functions

Few researchers have investigated how different **attentional paradigms** interrelate within the context of the **MOT task** (Adolphe et al., 2022b; Eayrs & Lavie, 2018c; Huang et al., 2012a; Skogsberg et al., 2015b; Treviño, Zhu, Lu, et al., 2021). Huang et al. (2012a) tested a large sample of 257 individuals, on both MOT and other tasks that engage attention (Meyerhoff et al., 2017b; Vater et al., 2021b). They found a strong correlation between MOT performance and various tasks involving quick judgments of static visual stimuli. Correlations ranged from moderate to strong (between 0.5 and 0.7) with MOT performance for tasks assessing selective attention, such as conjunction visual search, spatial configuration search, span of enumeration, rapid color identification post-masking, symmetry detection, reaction time to color stimuli, short-term visual memory, and change detection. Weaker correlations were noted between MOT performance and Raven’s intelligence test and tasks measuring suppression or avoidance of interference, like the Stroop task, attentional capture, and inhibition of return. The findings suggest that while some attention-related tasks are closely related to individual differences in MOT performance, the overall pattern remains complex. Treviño, Zhu, Lu, et al. (2021) expanded on earlier work by Huang et al. (2012a) to see if commonly used attention tests align with current cognitive psychology and neuroscience paradigms. They examined the concept of the “general attention factor” with a study involving 636 participants who completed an online battery of tests on TestMyBrain.org, including six experimental tests (MOT, Flanker Interference, Visual Working Memory, Approximate Number Sense, Spatial Configuration Visual Search, and Gradual Onset Continuous Performance Task) and eight neuropsychological tests (Trail Making Test versions (TMT), Digit Symbol Coding, Forward and Backward Digit Span, Letter Cancellation, Spatial Span, and Arithmetic). An exploratory factor analysis conducted on a subset of 357 participants revealed a five-factor structure: (1) attentional capacity, (2) Search, (3) Digit Span, (4) Arithmetic, and (5) Sustained Attention. Specifically, The first factor (1) comprised three experimental paradigms: MOT, Visual Working Memory, and Approximate Number Sense and two neuropsychological tests (Digit Symbol Coding and Spatial Span) as measure

of speed processing and spatial working memory, respectively. Interestingly, the findings suggest that Digit Span and Arithmetic tests should be classified as attention tests. Five tests were identified as assessing attention: Digit Symbol Coding and Spatial Span evaluate attentional capacity, while TMT-A, TMT-B, and Letter Cancellation assess search or attention-shifting ability (Treviño, Zhu, Lu, et al., 2021). Converging with these results, Eayrs and Lavie (2018c) explored visual capacity by increasing **perceptual load** to induce "inattention blindness" phenomena. They hypothesized that visual capacity is linked to our ability to count a limited number of items quickly (subitizing skill). To test this, they assessed visual perception using four tasks: change blindness, load-induced blindness, MOT, and subitizing capacity, measured as the number of items that could be reported in parallel in an enumeration task. Moreover, working memory was evaluated using several scales (OSPAN for operational span, RSPAN for reading span, and SSPAN for symmetry span) with over 200 participants aged 18 to 64 years. The findings confirm a common limit in perceptual capacity for visual detection across the four tasks. Specifically, they found that individuals who could quickly count more items also showed greater accuracy in detecting changes and peripheral stimuli while focused on another task. Confirmatory factor analysis suggests a moderate correlation between the MOT task and the working memory factor (0.40), and high correlation with the perception factor (0.61). These results demonstrate that perceptual capacity consistently affects tasks involving perceptual load, emphasizing its crucial role in attentional processing across various task demands. Another study (Adolphe et al., 2022b) developed an open-source cognitive test battery to assess attention and memory. This battery includes seven tasks: MOT, enumeration, go/no-go, load-induced blindness, task-switching, working memory, and memorability. Their findings showed a relationship between **working memory** performance and **MOT**, aligning with previous studies by (J. Allen, 2006) and (Lapierre, Cropper, & Howe, 2017). The results indicate that specific, underlying cognitive abilities can be detected within broad cognitive test batteries and that these abilities are relevant across various visual tasks. In particular, underlying factors were linked to the connection between MOT and enumeration (as noted by C. S. Green & Bavelier, 2006c, between MOT and load-induced blindness Eayrs & Lavie, 2018c, and between MOT and working memory J. Allen, 2006; Lapierre, Cropper, & Howe, 2017).

The lack of understanding of the cognitive processes impacting the success of this task can contribute to the scarcity of research integrating the MOT task into CT programs, despite the promising results reported for such CT (Vater et al., 2021b). Finally, although the involvement of various cognitive processes in the MOT task has been evident in neurotypical individuals, we propose to explore in the following section whether variations in performance among non-neurotypical individuals can offer additional insights into the cognitive processes engaged by the MOT task.

Box 1: Takeaway messages on MOT

Numerous studies in visual and attentional research have employed the cognitively multi-determined MOT task:

- The FINST Model's pioneering concept (Z. Pylyshyn, 1994) identified MOT as a process potentially guided by **pre-attentive stimuli**, using a mechanism that tracks multiple objects without detailed attention or conscious recognition. The brain assigns visual spatial indexes to a limited number of objects in the visual field.
- The Grouping Theory (Yantis, 1992b) shed light on the visual system's capacity to simplify tracking by **unitizing individual targets into a cohesive visual entity**.
- The FLEX Model (Alvarez & Franconeri, 2007a) introduced the idea of a **malleable pool of attentional resources that adjusts dynamically to the demands of tracking complexity**.
- The Spatial Interference Theory (Franconeri et al., 2010) brought attention to how performance could wane when targets are too close, due to noise and attentional disruption.
- A. O. Holcombe and Chen (2013)'s work underscored the limitations of tracking resources, countering the notion that spatial interference alone affects tracking accuracy.
- The Multifocal Attention Theory (Cavanagh & Alvarez, 2005) explored the possibility that multiple attentional beams can be directed towards **different objects simultaneously**, enhancing the understanding of how attention is distributed in MOT tasks.
- The correlation-based studies (*global approach*) revealed a strong to moderate bond with the visual processing, selective, sustainable and divided attention, as well as with working memory.

Taken together the findings from analytical and global approaches stresses that MOT does not tap on a monolithic cognitive function but a complex dynamic interplay of visual processing, attentional resources, and and working memory, shaped by both the intrinsic properties of the objects being tracked and the overarching conditions of the task.

3.2.2 Q1. (b) Does the performance of neurodiverse populations offer clues about the fundamental processes of MOT task performance?

The MOT task has also been applied in conditions where attention might be impaired or deficient shedding light on the cognitive mechanisms involved in the task.

Number of targets and tracking duration In clinical populations, a study by Alnawmasi and Khuu (2022) investigated how mild Traumatic Brain Injury (TBI) affects the controlled ability to allocate and maintain visual attention on multiple moving targets. This study measured sensitivity- the rate of correct target detection considering both correct hits and false alarms—and reaction time under various conditions involving different numbers of targets, tracking durations, and distractor dots. Adult participants included individuals with mild TBI and control subjects matched by age, gender, and IQ. Findings revealed that as the number of targets increased, both groups experienced a decrease in target identification accuracy, with a significantly steeper decline for the TBI group. The increase of distractor dots also reduced task accuracy more drastically in the TBI group than in the controls, likely due to an increased sensitivity to crowding, indicating possibly lower spatial resolution or limited attentional resources in the TBI group. Moreover, the TBI group showed poorer detection performance as tracking duration increased, suggesting that prolonged tracking demands more sustained attention. The reduced spacing required a more focused attentional spotlight, demanding additional cognitive resources, which were presumably scarcer in TBI participants. Consequently, the study demonstrated that individuals with mild TBI face greater challenges in maintaining attention on multiple moving objects, especially under conditions of increased tracking load and distraction.

Perceptual characteristics and speed Research involving young, non-neurotypical populations with Attention Deficit Hyperactivity Disorder (ADHD) has identified difficulties in both dividing and sustaining attention during tasks. For instance, Peng and Miller (2016b) showed in a meta-analysis that computerized attention training significantly enhances attention more than traditional interventions by researchers or teachers. Stubbart (2016) used the MOT task to examine cognitive processes impacted by ADHD, leveraging the task’s dynamic nature to better mimic real-world attention demands. Past research suggested that ADHD-related behavioral difficulties originate from issues with sustaining, selecting, and dividing attention (Peng & Miller, 2016b). However, traditional attention tasks have only modestly correlated with behavioral patterns noted by parents and teachers, as highlighted in studies by Barkley (1991) and Nigg et al. (2005), and Jonsdottir et al. (2006). In response, Stubbart (2016)’s study implemented the MOT task, which closely resembles real-world scenarios. The specific variant used was the “Catch the Spies” game (Stubbart et al., 2023), which includes immediate and delayed report conditions. The study involved children with ADHD and typically developing children, matching them in a task where they had to follow three target spheres out of eight within a virtual 3D cube. The task also incorporated a 3D biological motion perception test, where participants were asked to discern the direction of point-light walkers depicted as moving left or right. Adjustments were made to the speed of the spheres using a staircase procedure to vary difficulty levels. The session, lasting 1.5 hours, began with an assessment of cognitive function using the WASI-II and established baseline attention through the Conners CPT. Participants then completed both the 3D MOT and 3D biological motion

tasks, with the order counterbalanced among participants. Results suggested that MOT tasks are developmentally suitable for children with ADHD, as task accuracy was similar between the groups. Moreover, MOT task performance did not correlate with attention problem ratings from clinical measures, indicating that the behavioral symptoms critical for diagnosing ADHD might not be linked with difficulties in tracking moving objects. This suggests a potential reevaluation of how MOT task duration could be adjusted to better address challenges in sustained attention, rather than only monitoring moving objects. When considering dyslexic individuals, manipulating the perceptual characteristics in MOT tasks is crucial due to their reading difficulties. These issues underscore the importance of spatial layout parameters, which require visuospatial skills and updating capabilities linked to executive functions. This adjustment can influence selective attention and potentially alleviate crowding effects. In populations known to have attentional deficits, studies often employ the MOT task alongside other cognitive tasks to provide a more holistic understanding of the cognitive mechanisms involved. Bowers, Anastasio, et al. (2011) focused on the cognitive demands essential in dynamic settings, such as driving, where attention is critical. They used the MOT task not only to measure basic attentional capacity but also to explore its association with higher cognitive functions like fluid intelligence and executive functions among both young and older adults, including those with Mild Cognitive Impairment (MCI). By adjusting the tracking speed according to participant performance, the study leveraged the dynamic nature of the MOT task, requiring participants to continually adapt and engage a wide array of cognitive processes. The findings suggest that MOT task performance is linked to driving ability, demonstrating that tracking multiple objects at varying speeds taps into cognitive areas essential for safe driving. Additionally, the moderate correlation between MOT and UFOV scores underscores the importance of divided attention in both tasks. However, the unique demands of the MOT task, such as tracking moving objects at adjustable speeds, seem to activate additional cognitive functions not as directly measured by the UFOV task. This indicates that the MOT task, especially in its brief format developed for this study, may provide a more detailed assessment of dynamic attentional capabilities and executive function than traditional divided attention tasks. For older adults and those with MCI, the MOT task proved effective in predicting driving performance, highlighting its sensitivity to subtle cognitive deficits that may affect driving safety. Skogsberg et al. (2015b) explored whether there is a singular attention faculty or distinct processes that govern this faculty by examining individual differences in voluntary visual attention capabilities across a series of 11 representative tasks. Among 222 randomly selected participants, 129 completed eight visual attention tasks, including central and peripheral focusing, global attention, spatial and object-based shifting, MOT, rapid reengagement, and object vigilance. An additional 93 participants and 35 individuals with high ADHD traits completed three more tasks—attentional grouping, controlling motion, and spatial vigilance—bringing the total to 11 tasks. The results showed four clusters organized into **two functional dimensions**: one contrasting **spatiotemporal attention** (such as multiple-object tracking, spatial shifting, and controlling motion) **against global attention** (such as global attention, object-based shifting, and attentional grouping), and the other contrasting **transient attention** (such as rapid reengagement) **against sustained attention** (such as spatial and object vigilance). These findings indicate that attention capabilities can be divided into two main dimensions, suggesting that these different types of attention may rely on separate and loosely connected mechanisms. The results also imply that mechanisms of voluntary visual attention may be selectively enhanced or diminished due to genetic, experiential, or pathological factors. To deepen our understanding of how the brain manages

MOT, we suggest looking at it through a different lens: by examining the performance changes on MOT related to its practice within the context of CT.

Table 3.1. Overview of MOT task parameters and their implications.

Parameter manipulated	Values	Observed performances	Cognitive function	Theoretical model
Object number	8 objects (3 or 4 targets) (Z. W. Pylyshyn, 2001b; Z. W. Pylyshyn & Storm, 1988b), 2,4,6 objects (Oksama & Hyönä, 2008) 1 to 8 targets (Alvarez & Franconeri, 2007a)	Performance decreases as the number of objects increases	Attention Dynamics: Constraint of available resources with periodic attentional effort. Focused attention Selective attention	FINST Visual spatial indexes (Z. W. Pylyshyn & Storm, 1988b): limited number of objects in the visual field FLEX Model (Alvarez & Franconeri, 2007a)
Objects speed	0°/s to 42°/s (Alvarez & Franconeri, 2007a), 0.06 cm/s to 544 cm/s (Tullo, Bertone, et al., 2018)	Speed and Performance Relationship: Performance decreases as speed increases and increases as speed decreases. Tracking accuracy declines systematically with increases in object speed	Visual short term memory Focused attention	Results converge towards FLEX model (Alvarez & Franconeri, 2007a). Tracking capacity governed by a continuous pool of resources (Bettencourt & Somers, 2009b)
Tracking duration	5, 9, 13 s (Oksama & Hyönä, 2008)	Tracking Performance Over Time: Performance declines from 5 to 9 seconds (error percentage increases from 9.0% to 19.5%). Same performance (plateau effect) from 9 to 13 seconds (error percentage ranges from 19.5% to 21.6%). Difficulty maintaining attention on multiple objects over extended periods.	Flexible functions of working memory and dynamic spatial visual attention (< 5 s). As the number of targets increases, reliance on high-level, non-automatic processes tied to executive functions grows (> 5 s): shifting, updating and inhibition	MOT Theory Comparison: (Oksama & Hyönä, 2008) findings contradict (Z. W. Pylyshyn & Storm, 1988b)'s pre-attentive view, which posits that tracking duration does not affect MOT performance
Perceptual characteristics	Spacing between objects Circle: 0 cd/m ² ; diameter = 8 pixels Cumulative distance: 1° ≈ 18 pixels Configuration: 4 pairs on corners of an imaginary 300-pixel square 2 pairs centered 60 pixels above and below the fixation point (Franconeri et al., 2010) Task Duration: From 1.5 to 6 s Direction Changes: random direction changes (clockwise or counter-clockwise) Timing of changes randomly determined from a rectangular distribution	Impact of Object Proximity: The closeness of objects has a deleterious impact, known as crowding	Flexibility, Cognitive load, Executive function such as updating	Spatial Interference Theory (Franconeri et al., 2010): parallel processing with a potentially unlimited number of trackers. Tracking capabilities remain stable with consistent spacing, despite variations in speed or duration Contrasting Views (A. Holcombe, 2023): within the same hemifield, both speed augmentation and target addition impair performance, even with substantial separation. Highlights the finite nature of tracking resources, challenging the unlimited trackers hypothesis.
Continued on next page				

Table 3.1: continued from previous page

Parameter manipulated	Values	Observed performances	Cognitive Process	Theoretical model
Objects features	Trajectory (H. Fehd & Seiffert, 2008) Dot Positions: Randomized to prevent overlap with each other or frame Dot Movement: Set number of pixels per frame Speeds tested: 0.5 (3 pixels/s), 1 (6 pixels/s), 2 (12 pixels/s), 3 (18 pixels/s), 4 (24 pixels/s) Movement retinotopic (Suganuma & Yokosawa, 2006) Motion: Perceptually smooth. Speed Range: 0.5 to 4.5 deg/s, varied randomly except during critical phases Velocity Change: Between -1 and +1 deg/s Direction Change: Random, up to 90°, with each item maintaining a constant velocity or direction for 400 ms to 2 s Trajectory: Boundary Interaction	Trajectory Modifications and Tracking Performance: Alterations in tracking ability when target and distractors pursue each other or move uniformly. Better performances when objects reappear at their last known position rather than a new location based on prior movement Highlighting the importance of maintaining consistent trajectories Tracking accuracy decrease when increasing target numbers from 2 to 4 within a single hemifield Tracking performance stable: increasing number of targets across both hemifields Performance declines depending on the movement and distribution of targets across hemifields	Group Maintenance in Dynamic Environments: Continuous updating of representations of moving elements Mental rotation, working memory and controlled attention for tracking Dynamic Distribution of Attention: Attention is not anchored to specific points but is dynamically and flexibly distributed across the visual scene. Continuous resource that can expand, contract, and shift to cover moving objects	Grouping Theory (Yantis, 1992b): Common movements serve as cues for forming global object representations (Suganuma & Yokosawa, 2006) to emphasize the role of motion cues in enhancing group perception
	Appearance of objects	Color, shape (Bahrami, 2003; T. Liu et al., 2009) fluid-like texture (vanMarle & Scholl, 2003) impact MOT success	Memory Types for Familiar Objects: Semantic memory: general knowledge about familiar objects Episodic memory: Recollects specific events, experiences with familiar objects	Hemifield Theory (Alvarez & Cavanagh, 2005): each visual hemifield (left and right) has an independent capacity for attentional processing Multi-focal Attention (Alvarez & Cavanagh, 2005): Simultaneous allocation of attention to multiple locations or objects within the visual field

3.3 Performance Changes in MOT related to its repeated practice

3.3.1 Q2.(a) How does cognitive training using MOT tasks affect MOT performance (near transfer effect), and does changing MOT task parameters influence these effects?

In a neurotypical sample, a study by Parsons et al. (2016a) proposed a training based on the Neurotracker protocol presented in Box 2 3.3. Specifically, training consisted of 10 sessions of 45 min to 1 hour spanning over 5 weeks.

As outlined in Box 23.3, a final speed threshold score is assigned at the end of each block (set of 20 trials). The session's overall performance is then calculated as the average across all such blocks. Significant enhancements on MOT performance were observed from the initial to the final session in both the intervention group. The performances of the 3D-MOT training group significantly surpasses that of the non-active control group ($p < .01$). Authors assumed that success in the transfer of training was feasible through two mechanisms involved in the MOT task, which are **isolation and overloading** (Parsons et al., 2016a). Isolation refers to the number of functions called upon for the task which should be limited and consistent (Parsons et al., 2016a), and overloading is a function that involves engaging it beyond its current capacity. The authors claimed that an effective training of any function requires an overloading necessary to induce adaptation (in the brain: neuroplasticity). In this study, and subsequent others using the Neurotracker tool, this is accomplished by adjusting the **speed parameter** of each trial to ensure the task remains sufficiently challenging (Parsons et al., 2016a). Consistent with these findings, Harris, Wilson, Crowe, and Vine (2020) observed similar improvements in MOT performance, albeit with slight modifications to the training protocol. In a study involving 84 young adults, participants were divided into four groups: an untrained control group, a standard NeuroTracker training group as outlined in Box 2 3.3, an abbreviated NeuroTracker training group with shorter training sessions, and a portable NeuroTracker group using computer-based training. A 2 (first session, last session) $\times$ 4 (groups) ANOVA revealed a significant main effect of time ($p < 0.001$), indicating that practice led to enhanced performance. These findings were replicated in a study involving older adults both with and without subjective cognitive decline (Musteata et al., 2019a). Participants underwent 14 sessions of 25-30 min spanning over seven weeks of 3D-MOT training (following protocol in Box 1). The results showed an improvement on average speed thresholds scores through the training sessions. Details regarding the performance progression within the control group were not reported.

Same results were observed with non neurotypical populations. By varying the protocol in Box 2 3.3, with 3 targets to track instead of 4, Tullo, Faubert, and Bertone (2018a) were able to show that students with neurodevelopmental disorders (e.g., with Autistic Spectrum Disorders- ASD or ADHD) improved their speed thresholds with training. Results have shown a 41% improvement in the trained group (3D MOT) ($p < .001$) (Tullo, Faubert, & Bertone, 2018a). A trajectory analysis of performance improvement showed that if all trained groups were able to improve in the task, reduced progress rate were observed with the active or passive control group. In another context of developmental disorders such as dyslexia, the MOT task has shed light on certain cognitive processes involved in reading, such as selective attention (Bertoni et al., 2019a). Reading difficulties are thought to originate in capacities related to visual crowding and spatial attention (Bosse et al., 2007;

Facoetti & Molteni, 2001; Franceschini et al., 2013; Zorzi et al., 2012). Indeed, difficulties in visual crowding, which can limit our ability to identify individual stimuli when multiple objects are displayed in close proximity (see Gori & Facoetti, 2015; Pelli et al., 2004; Whitney & Levi, 2011, and selective attention deficits were observed. To investigate this phenomenon, Bertoni et al., 2019a conducted several experiments by offering variants of the MOT task to young children with dyslexia and compared the results with children without reading difficulties. The results highlighted that dyslexic children exhibited excessive crowding and a deficit in selective attention but showed a reduction in crowding and an increase in reading speed, improving selective attention capabilities after a training phase (Bertoni et al., 2019a). In the same study (first experiment), the spatial layout of targets (with four different orientations) was manipulated according to 4 different orientations : upward, downward, rightward, or leftward (chance level = 25%). The target-to-distractor spacing was measured as the center-to-center distance and was equal to 2.2°, 2.5°, or 2.8°. Also, targets and distractors were letters, and in another experiment, used versions of the MOT in the form of video games. The main results highlighted that dyslexic children had excessive crowding and a deficit in selective attention, but after a training phase, a reduction in crowding was observed Bertoni et al., 2019a.

Improvements in MOT performance, as evidenced by the increase in speed threshold for a constant number of targets or by the expansion of tracking capacity (the highest number of targets that can be accurately tracked), are consistently reported across various studies. Also, such enhancements are observable in both neurotypical and neurodiverse groups. Nevertheless, these outcomes merely demonstrate that training influences performance on the specific task trained. Considering that the ultimate aim of cognitive training is to achieve transfer effects to untrained tasks, ranging from closely related to more distant ones, the following section will delve into this issue.

3.3.2 Q2.(b) How does MOT Practice affect cognitive functioning (near to far transfer)?

To evaluate the advantages of CT, researchers typically use a variety of neuropsychological tests (Musteata et al., 2019a). These tests are designed to measure the subtle effects of the training. The scope of these assessments ranges from near transfer, indicating progress in tasks that are similar but different from the training activities, to far transfer, which shows improvements in a wider array of cognitive tasks. Although this classification is useful for examining learning outcomes, it is crucial to acknowledge its multidimensional aspect in practical settings, as emphasized by A. O. Holcombe and Chen (2013), and to carefully implement this framework. Parsons et al. (2016a) were pioneers in examining the impact of CT based on MOT tasks on attention, working memory, and visual processing speed. They utilized functional brain imaging to study these effects in a neurotypical population, specifically twenty university-aged students divided into either a training group or a non-active control group. Neuropsychological tests and brain function correlations were measured using quantitative electroencephalography (qEEG). The researchers anticipated observable quantitative changes in brain function that would align with the cognitive functions being tested. Specifically, they expected an increase in beta waves, related to attention processes, and gamma waves in the occipital cortex. Their results confirmed these hypotheses, showing enhanced attention and associated decreases in 2 to 11 Hz slow-wave activity along with increases in beta waves. Changes in the gamma band were also noted in the occipital cortex, indicating that ten sessions of 3D-MOT training could improve

attention, visual information processing speed, and working memory. Musteata et al. (2019a) found that integrating the 3D-MOT task into CT could enhance memory, processing speed, and attention in older adults, both with and without subjective cognitive decline. The study, conducted over 14 sessions, showed that the experimental group performed better in episodic memory tasks (retrieving the previous encoded abstract wordlist, recall words) and cognitive flexibility tasks (generating words) compared to the control group. Although no lasting effects were observed on working memory and executive functions at a 5-week follow-up but episodic memory improvements were noted. These results must be interpreted with caution due to unknown activities of participants post-intervention. Harris, Wilson, Crowe, and Vine (2020) aimed to evaluate the effects of 3D-MOT training on a 2D-MOT task (near transfer), N-back task (mid-transfer), and tasks involving route recall and audio monitoring (far transfer). Their findings were mixed and somewhat contrasted with those of Parsons et al. (2016a) and Musteata et al. (2019a). In the near-transfer task, they observed learning effects over time but no significant group differences, suggesting no strong transfer effects from training with NeuroTracker (3D-MOT) to 2D-MOT tasks. However, in the N-back task (mid-transfer), a significant interaction between group and time indicated that the training group showed greater improvements than the control group. No improvement was found in the far-transfer tasks, highlighting challenges in achieving transfer effects to significantly different tasks. This aligns with earlier research by Strong and Alvarez (2017a), who found no transfer effects when varying either the motion type (i.e translational or rotational trajectory) for the retinotopic location (upper or visual field) of MOT tasks, raising questions about the effectiveness of such training. In a study by Tullo, Bertone, et al. (2018), significant attentional improvements were observed in participants with neurodevelopmental disorders (e.g., ASD or ADHD) following MOT training. The study included participants aged 6-18, divided into three groups: a Neurotracker intervention group (3D-MOT), a control group engaged in a math-based game, and a usual treatment group. Over five weeks, they underwent 15 training sessions. The CPT-3, a measure of attention, was administered before and after the training along with other general intelligence measures to assess the effects of training. The results indicated that only the 3D-MOT group showed significant improvements in attention, with a marked enhancement in post-training CPT-3 scores compared to baseline. This improvement demonstrates the efficacy of the task in boosting attention in this population. The Neurotracker group's results also showed a significant enhancement compared to the control group in both Neurotracker and CPT-3 performance, suggesting a near transfer to far (executive) functions. These findings align with other studies indicating possible near transfers to attentional functions and far transfers to executive functions following MOT training (Tullo, Bertone, et al., 2018). However, the issue of transfer remains underexplored due to the infrequent use of comprehensive neuropsychological evaluations in cognitive training programs and the significant heterogeneity in the assessments used to evaluate the effects of MOT training (Vater et al., 2021b). Another research approach involves designing cognitive training that combines MOT with other tasks. This method is often used in recreational video games, as noted by C. S. Green and Bavelier (2008) and Boot et al. (2008). C. S. Green and Bavelier (2008) reviewed evidence showing that playing action video games (AVGs) leads to improvements in perceptual, attentional, and cognitive skills. Specifically, AVGs have been demonstrated to enhance functions ranging from perceptual skills to cognitive flexibility, including significant improvements in attentional abilities. A. Joessel (2022) conducted a study to explore how the unique features of AVGs, such as high pacing and significant demands on both divided and focused attention, contribute to cognitive enhancements. This study reviewed previous research on the impact

of video game play on cognition and suggested that the distinctive combination of these demands in AVGs facilitates cognitive improvements. It emphasized the MOT task as critical for adjusting these cognitive demands and highlighted its potential integration into AVGs for developing effective CT interventions. In this context, A. Joessel (2022) investigated the cognitive benefits of AVGs through a feasibility study with 263 online participants. Before and after a 12-hour training session on one of four game variants incorporating dual-task MOT to target different cognitive demands were evaluated. The Pacing Condition (1) used low-demand MOT training with constant speed and targets, paired with an adaptive banner detection task. The Focused Attention Condition (2) combined low-demand MOT training with an adaptive banner discrimination task. The Divided Attention Condition (3) paired adaptive MOT training with a low-demand banner detection task, gradually increasing object speed. The Combined Condition (4) integrated adaptive MOT, adaptive banner discrimination, and adaptive timing to challenge both focused and divided attention. The cognitive assessment battery included tasks to assess attentional control: an MOT task (near transfer), (1) UFOV task for spatial attentional control (far transfer), and (2) an Attentional Blink task for temporal attentional control (far transfer). (3) Short-term memory task similar to those developed by Luck and Vogel (1997a) and Vogel et al. (2001) (far transfer). (4) The N-back task was administered only at post-test. The study found that these game variants led to cognitive transfer to another MOT task, showing significant improvement over a control group with no game interaction.

In another context, such as the sports field, CT based on the MOT task has also shown positive results. This is particularly evident when training programs, which often involve dual-task exercises, include dual-task MOT tasks. For example, a study by Jeunet et al. (2020) developed an attentional training program to improve the skills of soccer goalkeepers, combining neurofeedback with tasks such as Covert Visual Spatial Attention (CVSA) and MOT to simulate dynamic visual attention. The study found that enhancements in CVSA performance correlated with better outcomes in the MOT task, thus supporting the effectiveness of CVSA training in strengthening cognitive abilities relevant to MOT tasks.

3.3.3 What about a non neurotypical population ?

Research has highlighted that Low Vision (LV) often correlates with a significant reduction in attention towards peripheral visual fields, despite no diagnosed impairments in these areas (Ludt & Goodrich, 2002; Tadin et al., 2012). This could indicate higher-level attentional deficits in individuals with LV Tadin et al. (2012). To mitigate these deficits, some studies have utilized training based on MOT principles, focusing on dynamic attention to objects in both peripheral and central visual fields. One such study, conducted by Nyquist et al. (2016a), employed a Dual-Task MOT paradigm to assess the effects of perceptual training on children with low vision. This study compared three different training regimens: a conventional AVG, a novel modified Multi-Attentional Tracking (MAT) task, and a control non-action video game (NAVG). Participants, divided into three groups, underwent ten training sessions before retesting their performance. The MAT task incorporated standard MOT tasks but added a motion discrimination task that required detecting and discriminating unpredictable peripheral motion stimuli. The task parameters started with a variable number of objects (4 to 6), with 2 to 5 designated as targets. Two independent QUEST staircase procedures adjusted the velocity of objects and the duration of the dual tasks, with set size adjustments based on velocity thresholds. Training impact was evaluated through pre- and post-training psychophysical assessments focused

on both central and peripheral visual fields. Motion perception was measured across various visual field locations before and after training, supplemented by a natural visual search task. The training effects were re-evaluated after 12 months. The study's findings indicated significant improvements in the MAT task after training, enhancing both tracking capacity and discrimination skills. Comparisons from pre- to post-tests showed that the Dual-Task MAT and AVG groups achieved more substantial gains in visual attention than the control group, with the dual-Task MOT group's performance mirroring that of the AVG group. Notable far-transfer effects were observed in tasks requiring attention to peripherally presented stimuli. These included improvements in single-target motion direction discrimination, multi-target direction comparison, and spatial crowding tolerance, areas where both AVG and MAT training boosted performance. Significant advancements were also seen in the visual search task. Overall, the improvements from MAT training were comparable to, or in some instances better than, those from AVG training, especially in tasks where enhancements were noted. This study underscores the potential of perceptual training to impact peripheral vision significantly. While incorporating the MOT task into CT programs shows promise, it is still uncommon. Research does point to significant benefits from this training, including near transfers (improvements in tasks similar to the training) and far transfers (improvements in tasks that are not directly trained).

In pathological contexts such as MCI, some studies (Bowers, Anastasio, et al., 2011; Musteata et al., 2019a) suggest that manipulating parameters related to dynamic attention and processing speed could be beneficial. It is plausible that for conditions like frontotemporal dementia, which primarily affects attention and inhibition systems, integrating MOT training with inhibition tasks could improve performance by requiring updates in dynamic attention—a capability often diminished in these patients. Similarly, for early-stage latent Cerebrovascular Disease, characterized by deficits in fronto-subcortical circuitry affecting attention, memory, and executive functions, tailored MOT training programs designed to address these specific deficits could prove advantageous. This nuanced understanding of cognitive dynamics enables the development of customized MOT training strategies for the elderly, aimed not just at preserving cognitive flexibility but also at strengthening areas vulnerable to age-related decline. Implementing strategic MOT tasks that challenge and subsequently enhance the cognitive functions most affected by aging offers a practical approach for sustaining cognitive health and ensuring functional independence in later years.

Summary As we have seen, several methods exist for assessing performance evolution and managing training difficulty in MOT training. In relation to the analytical approach to the MOT task, speed appears to be the most frequently used parameter, particularly for evaluating near transfer effects such as attentional function, visual processing, and short-term memory. To this end, most studies identify the maximum speed at which a participant can track a predefined number of targets with a minimum level of accuracy, or they assess both the highest speed and the greatest number of targets a participant can simultaneously track. Most study results seem to converge on an improvement in MOT task performance over time (Bertoni et al., 2019a; Harris, Wilson, Crowe, & Vine, 2020; Parsons et al., 2016a; Thompson et al., 2010; Tullo, Bertone, et al., 2018). When the global approach is adopted (see Q1), the expected effects are more related to far transfer, often involving neuropsychological tasks that target working memory and executive functions. However, the current question is to explore transfer effects in real-life situations.

Box 2: Common experimental design with NeuroTracker task training

3D-MOT or NeuroTracker:

Focusing on MOT as a training task, this review examines multiple studies utilizing either the commercial NeuroTracker tool or a variant of the 3D-MOT. To give readers an overview of usual experimental designs, we selected four articles that were prominently featured on the NeuroTracker website (<https://www.neurotrackerx.com/scientific-studies>) to illustrate typical experimental protocol (Faubert, 2013b, Parsons et al., 2016a, Musteata et al., 2019a, Romeas et al., 2016b).

Apparatus

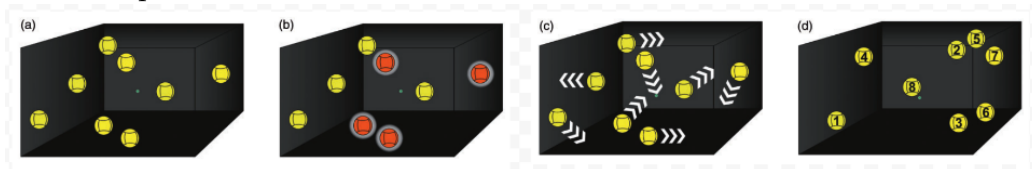
Participants either stands or sits inside or in front of an augmented reality environment with a virtual scene projected on the frontal and sometimes lateral walls (for instance the EON IcubeTM, a 7x10x10 feet room Romeas et al., 2016b or a CAVE a 8x8x8 feet room Legault and Faubert, 2012c). The task is practiced between 1 and 2 meters from the frontal display Legault and Faubert, 2012c with stereoscopy generated by the use of active shutter glasses (for instance the CrytalEyes 4s (RealD) Romeas et al., 2016b).

Task

In the CORE mode of Neurotracker, participants usually have to track 4 targets (colored in red) among 4 distractors (in yellow). After an initial presentation of object (typically around 2s) (a), an indexing phase lasts around 1 second where targets are highlighted with a halo (b). Then, objects move linearly in the 3D space without occlusion for 8s (c). Objects are indexed with numbers and participants have to verbally recall the number of targets initially presented (d). Training sessions are typically structured in several blocks of 20 trials Parsons et al., 2016a or 8 minutes. Complete training last around 15 sessions separated by break days Faubert, 2013b.

Difficulty adjustment and performance estimation

Difficulty is adjusted through a 1up-1down procedure on speed. Staircase steps are usually set to 0.05log. After each block, staircases are reset and performance on the session is computed as the mean of the final state of all staircases.



Neurotracker protocol, *image taken from Romeas et al., 2016b*

3.4 Outcomes from MOT practices in Real-Life

3.4.1 Q3.(a) What practices-related changes in MOT performance are linked to real-world transfer effects in neurotypical individuals?

To start with, in the sports domain, studies focusing on the ecological effects of MOT-based training have extensively examined perceptual-cognitive skills using two main approaches: the expert performance approach and the cognitive component skill approach (Fleddermann et al., 2019). The expert performance approach assesses athletes' perceptual-cognitive expertise using sport-specific stimuli and contexts, such as decision-making, attention, and memory skills in simulated or authentic sport settings. For example, elite athletes performed faster and more accurately than non-athletes or semi-athletes in these real-life settings (Mann et al., 2007). Conversely, the cognitive component skill approach investigates basic perceptual-cognitive skills in a non-sport-specific context, exploring how these fundamental skills correlate with sports expertise. For instance, elite athletes outperform non-athletes in basic cognitive tests, such as processing speed and various attention paradigms (Voss et al., 2010). These methodologies are crucial for elucidating the link between sports expertise and enhanced perceptual-cognitive skills. Given the recognized importance of perceptual-cognitive expertise (Furley & Wood, 2016) and its correlation with sports expertise, there is an increasing emphasis on developing and training these skills. However, relatively few studies, only 16 according to (Zentgraf et al., 2017a), have investigated the development of perceptual-cognitive expertise through Perceptual-Cognitive Training (PCT) interventions in this field. One notable example is the study by Faubert and Sidebottom (2012), which demonstrated task-specific practice effects among athletes in team sports such as soccer, rugby, and ice hockey following 15 sessions of 3D-MOT. The results suggested that professional athletes have a greater capacity to learn complex, dynamic tasks compared to novices. Additionally, the study by Romeas et al. (2016b) investigated the impact of MOT-based training on decision-making skills in soccer. This study assessed three fundamental skills—passing, dribbling, and shooting—before and after a training protocol involving 3D-MOT or 3D soccer videos. The experimental and active control groups underwent 10 training sessions, while the passive control group received no specific training. Post-training results showed improved decision-making accuracy in passing for the 3D-MOT trained group, a finding supported by players' self-assessments using a visual analogue scale. Lastly, Fleddermann et al. (2019) explored the effectiveness of generic PCT for elite volleyball players, incorporating MOT with physical activities. This eight-week program included twice-weekly sessions of 3D-MOT training combined with volleyball-specific motor tasks conducted in a laboratory setting. Post-training assessments involved tests on the 3D-MOT, four near-transfer cognitive tests- sustained attention (d2-R), memory span (KAI-N), working speed (KAI-N), and processing speed (Zahlenverbindungstest, ZVT)- and a far-transfer, lab-based volleyball-specific blocking test. The results indicated significant improvements in processing speed and sustained attention (near transfer effects) among volleyball experts, but no significant improvements in working memory or far transfer effects. Interestingly, the study also highlighted potential drawbacks of limited resources in dual-task scenarios, such as reduced performance in activities like jumping. This suggests that optimizing the interference between cognitive demands and motor execution might enhance overall performance. This idea aligns with Wickens' multiple resource model (2002) (Wickens, 2002), which attributes

performance enhancements more to increased processing speed of critical task information than to direct improvements in decision-making. Moreover, performance improvements observed in Romeas et al. (2016b) could be due to the automatization of procedural tasks, such as dribbling, which reduces cognitive load and allows for more efficient processing of complex scenes. This potentially improves decision-making compared to a control group without training. Although these improvements may seem small, they represent a significant shift in how athletes manage cognitive tasks by reducing cognitive load and enhancing information processing efficiency. However, the efficacy of ecological transfer effects from training remains uncertain. Fleddermann et al. (2019) noted that current training methods often lack the variability and unpredictability seen in real sports settings. A potential solution could involve introducing more unpredictability into MOT training by varying task parameters such as speed, orientation, and number of targets. Despite the practical implications, no studies have yet adopted this specific approach. Furthermore, Fleddermann et al. (2019) discussed but did not delve into how different athletes respond to such training, noting that elite athletes might have a smaller margin for improvement compared to novices, and current training programs generally do not offer varied difficulty levels tailored to different athlete groups, often only adjusting the speed parameter of MOT tasks.

In the area of aging research, Legault and Faubert (2012c) investigated whether training older individuals on the 3D-MOT speed task could improve their perception of Biological Motion (BM), a socially significant task. This study included three groups: an experimental group, an untrained group, and a visual perceptual training group. Participants received training on the 3D-MOT task weekly, with the BM task introduced during the fifth week, over five consecutive weeks. The training modified the virtual size and speed of the spheres, while maintaining a constant target number of three for recall in each trial. The stimuli involved a walker and a mask. The findings showed that effective training transfer occurred only in the experimental group, whereas the control group (no training) and the training control group (spatial contrast) showed no difference. The study suggests that this type of training could help older adults process socially relevant stimuli, such as human movements in crowded or sporting environments. Two key factors for the ecological transfer of perceptual skills in older adults are the **size of the visual field** and the **speed** thresholds. Subsequent research by Legault and Faubert (2012c) indicated that training on the UFOV, which includes aspects of both visual field and speed, has a positive effect on the retrospective driving abilities of older individuals. However, these findings are based on a correlational study, as highlighted by Legault and Faubert (2012c).

Summary These findings underscore the potential of attentional training via MOT to not only enhance specific attentional skills but also to generalize these improvements to various cognitive areas (Peng & Miller, 2016b). MOT practice indeed leads to transfers in complex ecological tasks such as dual tasks involving body-cognition coordination, whether in sports Fleddermann et al. (2019) or driving Legault and Faubert (2012c). Indeed, the transfer effect, where training in one cognitive domain positively influences performance in unrelated tasks, is a primary focus of current research. Hence, the benefits of attentional training can extend beyond the tasks initially targeted, thereby offering a broad enhancement of cognitive functions. Additionally, variations in training protocols often focus mainly on the speed parameter in the MOT tasks. However, these studies face limitations, such as the complexity involved in creating an ecological task that is both specific to the target domain and sufficiently variable to reflect real-world conditions.

Another issue is the lack of standardized measures before and after interventions across different studies in this area. Furthermore, despite recognition of interindividual variability in responses in the scientific literature, it is not consistently accounted for in cognitive training protocols. This neglect could result in either underestimating or overestimating the effects of interventions, or in the deployment of training programs that are not tailored to diverse participant profiles.

3.4.2 Q3.(b) Do neurodiverse conditions alter these effects?

While post-training improvements on tasks are often restricted to immediate effects, adjusting the parameters of MOT can induce significant changes in the mode of information processing, which may shift between automatic and controlled modes (B. A. Anderson, 2018). This variation is particularly evident between neurotypical populations and those with distinct characteristics. A study by Pothier et al. (2015) involving three age groups (young adults, young-old, and old-old) demonstrated the adverse effects of performing an MOT task while walking. The findings indicate that certain procedural and automatic processes deteriorate with age, leading to reduced flexibility and lower performance in dual-task situations. Additionally, in a different study by Meyer (2019a) on reading acquisition, it was found that some processes, which are automatic for neurotypical children, require conscious effort for those with learning difficulties. B. A. Anderson (2018) notes that in neurotypical individuals, processes that are initially voluntary (activated during early learning stages) can become less conscious over time. This transition facilitates achieving task goals and reduces the need for explicit control in certain situations, a shift that does not consistently occur in non-neurotypical populations. Bertoni et al. (2019a) conducted a study on children with developmental dyslexia (DD) that involved manipulating the speed of targets and spatial conditions during MOT training. This training used action video games (AVGs) that share characteristics with MOT tasks, emphasizing: (1) high speeds of moving objects; (2) significant perceptual, cognitive, and motor demands necessary for accurate motor planning; (3) unpredictability in timing and spatial arrangements; (4) a focus on peripheral processing. Children with DD were randomly assigned to either AVG training or NAVG training (Franceschini et al., 2013, 2017; Gori & Facoetti, 2015). These training conditions specifically enhanced peripheral processing, the speed of processing multiple transient events, and perceptual load, thereby improving visual-spatial attention mechanisms and their neural networks (see Bavelier, Achtman, et al., 2012; Bediou et al., 2018c; Föcker et al., 2018). The results showed that just 12 hours of AVG training could reduce visual confusion and enhance reading speed in children with DD, significantly impacting clinical practices. Improvements in reading were noted particularly in speed, without compromising accuracy. These findings align with selective enhancements in processing speed for grapheme-to-phoneme conversion, previously observed in some AVG training studies (e.g., Franceschini et al., 2017; Gori et al., 2016). Another experiment confirmed that reductions in visual confusion and improvements in reading speed occurred only in DD children who effectively engaged in AVG training, as evidenced by improved game scores Bertoni et al., 2019a; Franceschini & Bertoni, 2019. These outcomes suggest that improvements in visual confusion and reading speed are contingent upon enhanced visual-spatial attention mechanisms, as indicated by AVG scores during the training Bertoni et al., 2019a. These observations indicate that the parameters manipulated in MOT have varied effects depending on the groups being studied. This is particularly relevant for individuals with specific disorders such as reading difficulties, where a controlled approach is preferable due to the lack of automatic processes usually strengthened through

repetition, as highlighted by B. A. Anderson, 2018. Considering the complex interaction between cognitive functions and the attentional system, incorporating MOT training within cognitive enhancement programs for both aging and pathological contexts shows promise.

3.5 Conclusion

The research on MOT task engages a broad spectrum of individuals, from youth (Bertoni et al., 2019a) to the elderly (Legault & Faubert, 2012c), and those with various health conditions, proving relevant in fields like sports (Fleddermann et al., 2019), academia, and healthcare (Nyquist et al., 2016a). Its adaptable parameters (i.e., modifiable task features) facilitate the study of cognitive mechanisms and their links with clinical profiles, especially in attentional and executive disorders. For instance, adjusting the number of distractors can modulate the cognitive load for individuals with dyslexia, enhancing our understanding of their visual and attentional processes. However, the broader application of these insights, particularly in CT programs, is limited by the infrequent exploration beyond basic parameter adjustments (such as the speed and the number of targets). Despite its potential, the integration of MOT into CT is still in its initial stages. Studies often focus on dual tasks (e.g., MOT combined UFOV), motor activities, or game-based tasks) to closely match the real-life activity with the trained task. However, these studies generally lack a large set of neuropsychological assessments to fully understand the cognitive outcomes (Bowers, Anastasio, et al., 2011; Fleddermann et al., 2019). This underscores the need for a more robust methodological approach to effectively connect MOT with cognitive improvements. Attentional paradigms incorporating MOT have shown promising correlations with attentional and executive functions, as well as memory capacities (Adolphe et al., 2022b; Eayrs & Lavie, 2018c; Huang et al., 2012a; Skogsberg et al., 2015b; Treviño, Zhu, Lu, et al., 2021), supporting the argument for including MOT in CT programs. This aligns with findings by Shanon (2018), demonstrating the efficacy of CT in enhancing attentional skills. However, the diversity in assessment tests raises questions about the consistency and validity of skill transfer, an issue highlighted by Vater et al. (2021b) and further elaborated by Von Bastian and Oberauer (2013) due to varied training outcomes. Adding to this diversity in cognitive evaluations are psychometric issues, such as studies claiming to assess working memory with tests that actually evaluate short-term memory functions (Scharfen & Memmert, 2021a). Another significant limitation is the prevailing use of linear training models like the staircase method in many CT programs (**Adolphe et al.; 2022 in press**), highlighting a critical shortfall: the lack of personalized training approaches and consistent cognitive function measures across studies. This one-size-fits-all strategy, although manageable for minor adjustments, falls short in accommodating individual learning trajectories or the variability noted in studies by A. Joessel (2022) and Nyquist et al. (2016a). Moreover, not accounting for individual differences and the adaptability of task difficulty can lead to either magnifying effects (suggesting that individuals already performing at a high level will benefit most from CT) or compensatory effects in certain subjects (where those with lower performance may benefit more from CT due to their larger room for improvement). The outcome of CT remains ambiguous. Specifically, it is not always clear whether the training serves to automate attentional processes (such as processing speed and grouping) or to enhance executive strategies (such as shifting and body-cognition coordination), or potentially both (Karchach & Unger, 2014). Consequently, this raises questions about the relevance of

the training depending on the sample of individuals it is intended for. An overlooked yet critical aspect in the narrative review of MOT's application across education and health sectors is the role of participant engagement and motivation. Engagement is paramount in areas where personal progress heavily depends on the individual's active participation. Future research should explore how varying MOT parameters not only affect cognitive outcomes but also influence engagement levels. Understanding this relationship is vital for designing MOT-based interventions that not only target cognitive improvements but also ensure sustained participation through intrinsic motivation and engagement. This focus on engagement could significantly enhance the efficacy and applicability of MOT in real-world settings, particularly in educational and healthcare contexts where motivation plays a crucial role in success (Moen et al., 2018a). Looking ahead, research must pinpoint which MOT enhancements facilitate the most meaningful skill transfers, especially in contexts of attentional and executive challenges. Emphasizing intra-training learning paths and MOT's real-world applicability will be key. The quest for MOT's ecological validity, crucial for ensuring its real-life relevance, requires a deep dive into how environmental complexities influence task performance, advocating for scenarios that mirror actual life situations as proposed by J. M. Ericson and Beck (2013), Stubbert (2016), and Lochner and Trick (2014). This exploration is vital for MOT's future application in daily activities and broader CT paradigms.

Part III

Experimental work

Chapter 4

An Open-Source Cognitive Test Battery to Assess Human Attention and Memory

Aims

After reviewing existing works on individualized cognitive training and the literature on the MOT task, this chapter aims to introduce a cognitive battery designed to assess the impact of the MOT training presented in chapters 5 and 6. Consequently, the cognitive battery discussed in this chapter forms a cornerstone of the subsequent evaluation.

Contents

4.1	Introduction	121
4.2	Materials and Methods .	124
4.3	Results	131
4.4	Discussion	139
4.5	Conclusions	142

Abstract

Cognitive test batteries are widely used in diverse research fields, such as cognitive training, cognitive disorder assessment, or brain mechanism understanding. Although they need flexibility according to their usage objectives, most test batteries are not available as open-source software and are not be tuned by researchers in detail. The present study introduces an open-source cognitive test battery to assess attention and memory, using a javascript library, p5.js. Because of the ubiquitous nature of dynamic attention in our daily lives, it is crucial to have tools for its assessment or training. For that purpose, our test battery includes seven cognitive tasks (multiple-objects tracking, enumeration, go/no-go, load-induced blindness, task-switching, working memory, and memorability), common in cognitive science literature. By using the test battery, we conducted an online experiment to collect the benchmark data. Results conducted on two separate days showed the high cross-day reliability. Specifically, the task performance did not largely change with the different days. Besides, our test battery captures diverse individual differences and can evaluate them based on the cognitive factors extracted from latent factor analysis. Since we share our source code as open-source software, users can expand and manipulate experimental conditions flexibly. Our test battery is also flexible in terms of the experimental environment, i.e., it is possible to experiment either online or in a laboratory environment.

Collaborators

Adolphe Maxime (AM), Sawayama Masataka (SM), Delmas Alexandra (DM), Maurel Denis (MD), Sauz  on H  l  ne (SH), Oudeyer Pierre-Yves (OPY). ¹This work has been published in Frontiers in Psychology (Adolphe et al., 2022a).

¹According to the Contributor Roles Taxonomy, collaborators contributed in various roles: Writing - Review & Editing (all); Conceptualization (AM, SM, SH); Data Curation (AM, SM, SH); Formal Analysis (AM, SM, SH); Investigation (AM, SM, SH); Methodology (AM, SM, DM, MD, OPY, SH); Funding Acquisition (DM, MD, OPY, SH); Project Administration (DM, MD, OPY, SH).

4.1 Introduction

Cognitive abilities such as attention or memory are essential for our daily life. Researchers have measured these abilities for many years to elucidate human cognitive mechanisms, diagnose various mental disorders, and evaluate cognitive training effects. Previous studies in cognitive science have generally investigated a specific task using various stimulus parameters to understand the underlying mechanisms in detail Baldauf and Desimone, 2014; Luck and Vogel, 1997b; Maunsell, 2015. On the other hand, works in the cognitive diagnosis and training literature utilize test batteries, including various cognitive tasks, to evaluate the individual’s diverse cognitive state C. S. Green and Bavelier, 2003a; Hosokawa et al., 2019; Kueider et al., 2012a; Lampit et al., 2014b; Steyvers and Schafer, 2020b. For instance, researchers in cognitive training studies leverage a cognitive test battery before and after training to estimate how their intervention affects the cognitive state. Since the purpose of cognitive test batteries generally needs to cover a variety of cognitive abilities, such as vision, memory, auditory, or logical reasoning Folstein et al., 1975b; Nasreddine et al., 2005, each task includes a small number of stimulus parameters to keep the experimental time short. However, if researchers focus on specific cognitive abilities in cognitive training investigations, e.g., visual attention or memory, such limited parameters can be insufficient to evaluate cognitive states because complicated cognitive processes mediate each ability, as explored in the cognitive mechanism investigations.

The present study aims to develop an online open-source test battery to leverage the two research directions, i.e., cognitive mechanism understanding and test battery assessment. Specifically, while investigating various parameters for each task, as in the investigation of cognitive mechanisms, we have explored the relationship across diverse cognitive tasks, as in the studies of test batteries. We consider that one of the difficulties in developing such extension in the previous literature is mainly related to the proprietary nature of existing cognitive assessment software. Indeed, either classic or computerized, most cognitive batteries are commercial Connors et al., 2000; Hosokawa et al., 2019; Kraus and Breznitz, 2009; Mielke et al., 2015; Preiss et al., 2013 and the researchers do not have flexible control over the parameters of the program. While this constraint allows researchers to share a common standard framework, it does not easily allow the work to be extended to new goals. Since the trend of experimental environments quickly changes depending on the technology development, flexibility and openness of the software are essential to ensure that the test battery is used over a long period. For instance, there has recently been a great demand to investigate online experiments. Some recent cognitive training studies also utilize online training. To evaluate the effect of such training, one needs to evaluate the cognitive ability using an online test battery. Since our test battery uses a browser-based platform, using a javascript library, p5.js, experimenters can flexibly launch the experiment under various environments referring to its source code.

Our test battery includes seven cognitive tasks: multiple object tracking, enumeration, load-induced blindness, go/no-go, task switching, working memory, and memorization. Our purpose is to create a test battery for cognitive training studies that focuses on the specific capacity of visual attention and memory rather than multiple cognitive domains such as auditory, linguistic, and logical reasoning tasks, as in previous work Soreq et al., 2021; Steyvers and Schafer, 2020b. In particular, we selected tasks related to a multiple object tracking (MOT) task measuring dynamic attention Cavanagh and Alvarez, 2005a; Culham et al., 1998. MOT is a cognitive task in which participants are required to track multiple moving objects simultaneously in a cluttered scene. Because such tracking abilities

are essential in daily situations, many cognitive training works utilize MOT as a training task (cf. Vater et al., 2021c) for various participant populations, such as young adults Harris, Wilson, Smith, et al., 2020; Legault and Faubert, 2012a, older adults Legault and Faubert, 2012a; Legault et al., 2013b, professional athletes Faubert, 2013a, and video game players Benoit et al., 2020. For instance, Legault et al., 2013b used a 3D MOT task for training and showed that the training efficiency for healthy older adults was similar to younger adults. Based on a task related to object tracking abilities, we used the taxonomy of Barnett and Ceci, 2002b to build the assessment tool. On the content dimension, we chose the enumeration and load-induced blindness tasks for the near transfer tasks related to MOT, as some previous work has shown their performance correlation Eayrs and Lavie, 2018b; C. S. Green and Bavelier, 2006b. For the far transfer tasks, we used other attention tasks, i.e., go/no-go and task switching tasks. On the memory dimension, as some have shown, the contribution of working memory abilities to MOT performance R. Allen et al., 2006; Lapierre, Cropper, and Howe, 2017, we used spatial working memory and memorability tasks as near and far transfer tasks, respectively. The choice of these tasks also allows for the evaluation of a transfer on the dimensions of the type of outcomes (e.g., accuracy, reaction time) as well as on the specificity (e.g., single and dual tasks) of the tasks. Because cognitive training studies can use our test battery as a pre/post evaluation, we consider that each participant can complete all tasks within an hour and a half.

The tasks have been intensively investigated in visual attention and memory literature. The multiple-object tracking task measures participants' tracking ability (Figure 2.1b) Bettencourt and Somers, 2009a; Cavanagh and Alvarez, 2005a; Z. W. Pylyshyn and Storm, 1988c; Vul, Frank, et al., 2009; Zhong et al., 2014. The task difficulty depends on multiple factors such as the number of targets, the number of distractors, or the object speed. The enumeration task measures human counting ability (Figure 2.2a) C. S. Green and Bavelier, 2003a, 2006b; Trick and Pylyshyn, 1993. This task asks participants to count flashed multiple objects. Depending on the counting number, it has been known that observers show different cognitive performances. Specifically, for smaller numbers of items (e.g., 2-4), participants can count them effortlessly and quickly, as called "subitizing." In contrast, it has been considered that the counting efficiency decreases for larger numbers of items (e.g., more than 5), which we used in our experiments. The load-induced blindness task measures the divided attention and the useful field of view (Figure 2.2b) Dye and Bavelier, 2010; Eayrs and Lavie, 2018b; Macdonald and Lavie, 2008. This task asks participants to perform a dual-attention task, both foveal and peripheral detection tasks. Since this task requires peripheral target detection, it is related to the works of the useful field of view (UFOV) Edwards et al., 2018. The UFOV is generally hard to measure using online experiments because it needs a large visual angle. However, using a dual-task that needs divided attention can narrow the field. The go/no-go task measures the ability to distinguish between relevant and irrelevant information (Figure 2.3a) Bokura et al., 2001; Conners et al., 2000; T. M. Mani et al., 2005; Nash et al., 2013. This task requires participants to attend a cue and to answer if the following stimulus is the target or not. This task is also known as a standard cognitive test, called the cued continuous performance test (CPT) Conners et al., 2000. The task-switching task measures the flexibility of selective attention (Figure 2.3b) Meiran, 1996; Monsell, 2003; Monsell et al., 2003a. This task requires participants to shift their attention between different goals. A lot of paradigms have been suggested to measure flexibility Monsell, 2003. We chose the task-cueing paradigm in the task-switching paradigms, where the task was unpredictable, and a task cue appeared before and with the stimulus Monsell et al., 2003a because it is easier to present to online participants.

The working memory task here indicates the spatial span task (the Corsi Block Tapping Task) and is to measure spatial short-term memory capacity (Figure ??a) Berch et al., 1998a; Soreq et al., 2021.

Another advantage of our test battery is that we compare a recent cognitive task, the memorability task, with traditional ones used in the test battery literature. The memorability task measures the ability to memorize natural scene images (Figure ??b) Bylinskii et al., 2015a, 2021; Isola et al., 2011b; Khosla et al., 2015. This task was initially proposed in the computer vision community to explore what intrinsic image features are memorable for human participants. However, not only the intrinsic image factors but also human cognitive factors mediate this task performance. Specifically, even if an image has identical intrinsic factors, memorability can change with how observers pay attention to it Bylinskii et al., 2015a; Mancas and Le Meur, 2013. Although some studies show the contribution of cognitive factors to memorability, it is still unclear how it relates to diverse human cognitive abilities. Thus, the inclusion of the memorability task to our test battery can contribute to either cognitive mechanism understanding or diverse test battery development. For the cognitive mechanism understanding, our investigation can clarify what kind of underlying cognitive abilities mediate memorability performance by comparing other cognitive task performances. Moreover, for the test battery development, including a cognitive task using natural scene images is needed to assess the ecological validity of cognitive training because most tests use artificial stimuli. The difficulty of using natural scene images is how to control the task difficulty as it has to be controlled constantly using different natural images. An advantage of the memorability datasets is that the previous works of the memorability share the experimental data Bylinskii et al., 2015a, 2021; Isola et al., 2011b; Khosla et al., 2015, and we can extract constant difficulty images from the datasets.

In the following Materials and Methods section, we describe each task in detail and how to collect the benchmark data. After discussing our benchmark data in the Results and Discussion section, we show the data availability. In summary, our investigation includes the following features:

- RQ1- We suggest an online open-source cognitive test battery including a wide variety of attention and memory tasks with various stimulus parameters.
- RQ2- Experimenters can flexibly run it in various environments (online or in the laboratory) using a web browser.
- RQ3- Our test battery captures diverse individual differences and can evaluate them based on the latent cognitive factors.
- RQ4- It is flexible in expanding stimulus conditions and adding new tasks because all source codes and data are available. Besides, we prepare a playground to test our cognitive tasks to support users' understanding of the task procedure in the following link (https://github.com/mswym/cog_testbattery).

4.2 Materials and Methods

4.2.1 Participants

Fifty naïve participants, aged from 21 to 71 (median=25, mean=29.0, standard deviation=11.6) years old, engaged in the experiments. All gave informed consent approved by the Operational Committee for the Evaluation of Legal and Ethical Risks (OCELER).

4.2.2 Apparatus

The benchmark data acquisition was conducted online using a web browser with a javascript library, p5.js (<https://p5js.org/>). Each participant accessed the web server hosted in our laboratory and engaged in the tasks. The platform in our experiment was organized by a python web framework, Django (<https://www.djangoproject.com/>). The informed consent and the schedule management of the two-days experiment were controlled using the platform. Our test battery, including instructions and practice trials for each task, was implemented with the javascript library p5.js. Users can run our test battery either with our Django platform or separately using a shared webserver. All codes and data are available from the following link: https://github.com/mswym/cog_testbattery.

4.2.3 General procedure

Our experiment has been conducted over two days (median time between pre and post assessment: 1 day and 16 hours). Participants registered for the experiment on the first day and reported the monitor size in cm or inch. We asked them to use the same monitor across days and to see the monitor from a distance of 50 cm. We extracted the monitor pixel size they used and defined the visual angle based on the information, as common in online experiments Q. Li et al., 2020. During the experiment, a virtual character provided the guidelines on how to use the experimental platform (Figure 4.1a). Including the character, the platform was implemented as a gamified interface to keep participants' motivation high. Clement et al., 2013b; Hosokawa et al., 2019; Lumsden et al., 2016a.

4.2.4 Stimuli and procedure of each task

This subsection describes the stimuli and procedure of each task. We decided the stimulus parameters of each task based on previous cognitive science works and our preliminary investigation on a browser-based investigation.

Multiple-object tracking task (Figure 4.1b)

Our task procedure followed Bettencourt and Somers, 2009a because it fits to conduct online experiments efficiently while exploring the tracking ability for the numbers of targets and the target speed conditions. In our experiment, either the target or distractor number was five. We controlled the task difficulty by changing the target speed in 1, 4, and 8 degrees/s. The diameter of each disk was 1.2 degrees. On each trial, five of the ten discs were briefly highlighted in red for one second to show they were the targets to track. After that, ten objects started to move for eight seconds. The moving direction was determined randomly at first and bounced at the corner of the square canvas of 12 x 12 degrees. We allowed the occlusion between the objects. Participants' task was to remember the target

discs and track these positions until they stopped. They answered the target position by clicking the buttons placed on the final object positions.

Enumeration task (Figure 4.2a)

This task procedure followed C. S. Green and Bavelier, 2006b in which the authors compared an enumeration task with the ability of MOT. Each trial began with the presentation of a fixation pattern in the center of a middle gray background. After 1000 ms, a set of white circles was presented for 50 ms. The number of circles was selected from 5, 6, 7, 8, or 9 in a pseudo-random order. The diameter of each circle was 0.5×0.5 degrees, and its color was white. The circles were presented within a diameter of 5 degrees in the background. The position of the circles was not overlapped in the region. The participants were asked to count these circles and answer how many circles were presented using a slider bar. Each stimulus condition was tested 20 times for each observer.

Load-induced blindness task (Figure 4.2b)

The load-induced blindness procedure followed Eayrs and Lavie, 2018b, in which the load-induced blindness ability was compared with MOT. On each trial, after presenting the fixation pattern of 1000 ms, participants viewed a 50 ms presentation of a cross target with four Gabor stimuli. After 950 ms, they were asked to answer which of the lines was longer using mouse clicking. Then, they were asked to answer which of the four Gabor stimuli had the enhanced contrast by clicking one of the four buttons. They were asked to correctly answer at least the foveal task. If not, their response to the peripheral task was not recorded. The length of each cross pattern was 0.5 or 1.0 degrees, and the vertical or horizontal line was randomly selected for the longer one. The color of the cross was black. For the Gabor stimuli, the standard deviation of the Gaussian envelope was 0.7 degrees. The spatial frequency and the orientation of the grating were 2.2 cycles/degree and 0.0 degrees, respectively. The mean luminance was set to the background color (i.e., middle gray). The enhanced contrast of the target was 0.8, and the others were 0.4. The Gabor stimuli were presented at a distance of 3 (near condition) or 6 (far condition) degrees from the center position on the screen. Each stimulus condition was tested 20 times for each observer.

Go/no-go task (Figure 4.3a)

The ten single digits (from 0 to 9) were used as the stimuli. We decided the stimulus presentation time based on T. M. Mani et al., 2005, though we cannot strictly control the presentation time due to browser-based experiments. Each trial began with the 1000 ms presentation of the fixation point. Then, each digit was presented one by one for 50 ms with the interstimulus interval (ISI) of 950 ms. The digit stimuli were presented within an area of 1.5 degrees squares. Participants were asked to focus on the number "7" and answer whether the number after the "7" was the "3" or not. If the number after "7" was "3" (Go trial), they had to press the key "J" as soon as possible. If not (No-go trial), they were asked to keep not responding. The probability of Go/No-go trials was 50/50 %. Each stimulus condition was tested 18 times for each observer.

Task-switching task (Figure 4.3b)

We used the task-cueing paradigm in the task-switching paradigms, where the task was unpredictable, and a task cue appeared before and with the stimulus Monsell et al., 2003a. We designed our original stimulus patterns to make the cues and tasks clearer for online participants. A digit from the set 1–4, 6–9 was used for the target stimuli. Participants’ tasks changed with the task cue. When the task cue was the blue diamond-shaped background, participants had to answer whether the target digit was odd/even by using the key "F"/"J," respectively. In contrast, when the task cue was the red square background, they had to answer whether the target digit was higher/lower than five by using the key "F"/"J," respectively. Each trial began with the presentation of the task cue. After the cue presentation of 650 ms, a target digit was displayed. The size of the background rectangle was 4.9 degrees on each side. The target digit was shown in the center of the background within an area of 1.5 degrees squares. After the participant’s response, a blank screen of 1000 ms was presented. We used the first 20 trials as practice ones. Each stimulus condition was tested 30 times for each observer.

Working memory task (Figure 4.4a)

We used a typical procedure of computerized Corsi Block Tapping tasks (e.g., Soreq et al., 2021). On each trial, sixteen light gray squares were displayed in a four-by-four grid. One of these squares was sequentially flashed with a reddish color for 900 ms. The order of the flashes was randomized for each trial. After the flash presentation, participants were asked to repeat the sequence by clicking on the squares in the same forward order. The size of each square was 2.0 degrees on each side. The number of flashes was selected from the set of {4,5,6,7,8} in a pseudo-randomized order. Each stimulus condition was tested 12 times for each observer.

Memorability task (Figure 4.4b)

The memorability task measures human memory performance for natural scene images (Figure 4.4b) Bylinskii et al., 2015a, 2021; Isola et al., 2011b; Khosla et al., 2015. Our experiment extracted images with intermediate memorability scores from the FIne-GRained ImageMemorability (FIGRIM) dataset Bylinskii et al., 2015a because it has been shown that cognitive factors are more effective for these images. Each trial began with the presentation of a natural scene photograph for 1000 ms. During the presentation, participants were asked to remember each photograph and answer whether the photograph is presented twice or not, by pressing the key "J" as soon as possible. After each presentation, participants received feedback if the response was correct or not for 1400 ms. There were two blocks for the memorability task. On each block, participants viewed a set of 120 images within a specific natural scene category, "bedroom" or "kitchen," in the dataset. We chose the image of hit rates on the interval [0.60,0.70]. Forty images were the targets and displayed twice for each block. Forty images were the fillers and displayed once. Eight of the targets were presented with a long distance of 100–109 images between an image and its repeat. Thirty-two of the targets were presented with a short distance of 2-5 images.

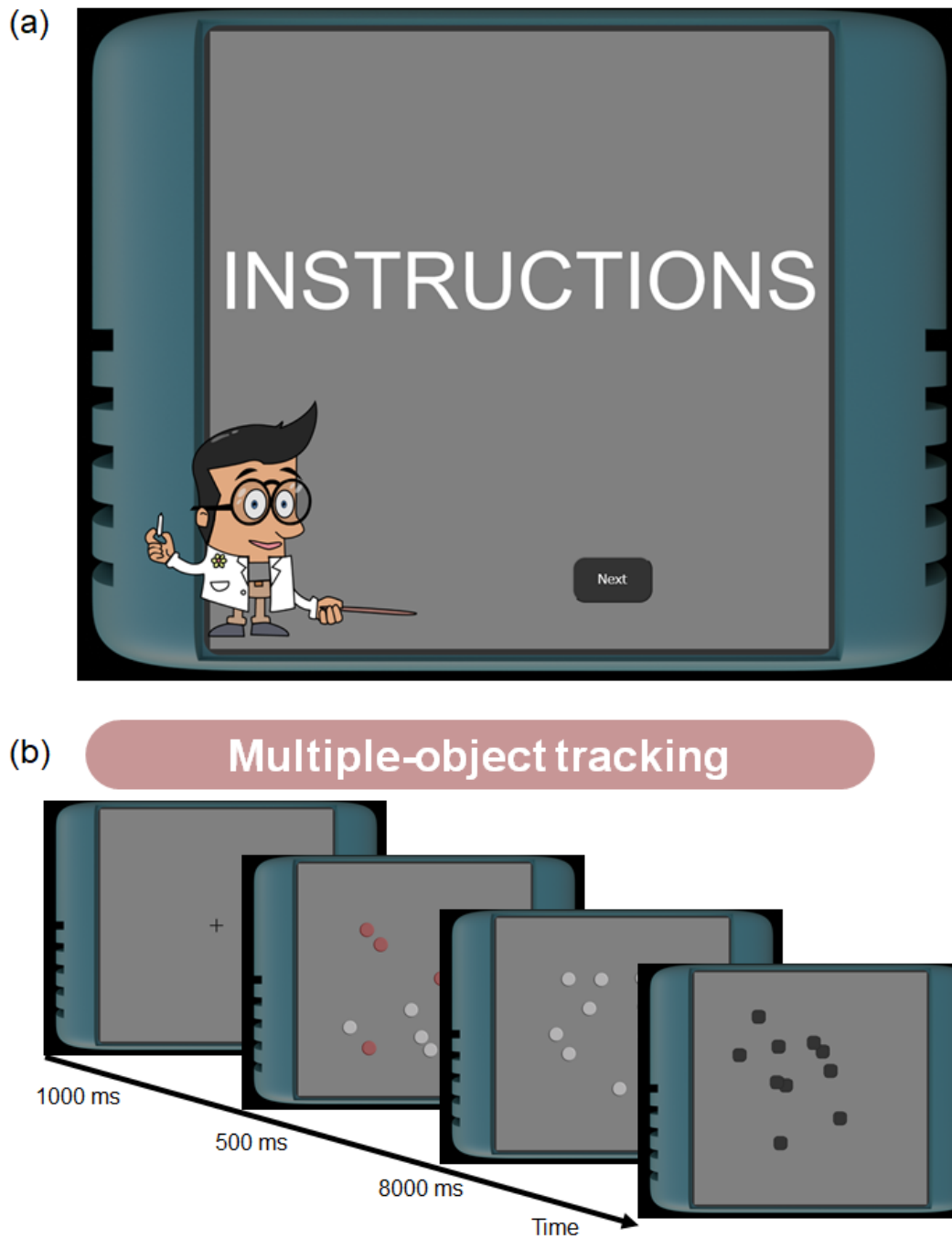


Fig 4.1. (a) interface example of the test battery. (b) Stimuli and tasks in the multiple-object tracking (MOT) task. After presenting the fixation point, five of the ten discs were briefly highlighted in red to show they are the targets to track. Then, ten objects started to move for eight seconds. Participants answered which are the target objects by clicking black boxes after the moving scene ended.



Fig 4.2. (a) Stimuli and tasks in the enumeration task. After presenting the fixation point, a brief flash of multiple white circles was presented. Participants answered how many circles were shown by using a slider. (b) Stimuli and tasks in the load-induced blindness task. Participants were asked to perform a dual-task, answering the length of the gazing point and the contrast of the images presented in the surroundings. After showing the fixation dot, a cross target with four Gabor stimuli was briefly presented. Participants first answered which of the lines was longer using mouse clicking. Then they answered which of the four Gabor stimuli had the enhanced contrast by clicking one of the four buttons.

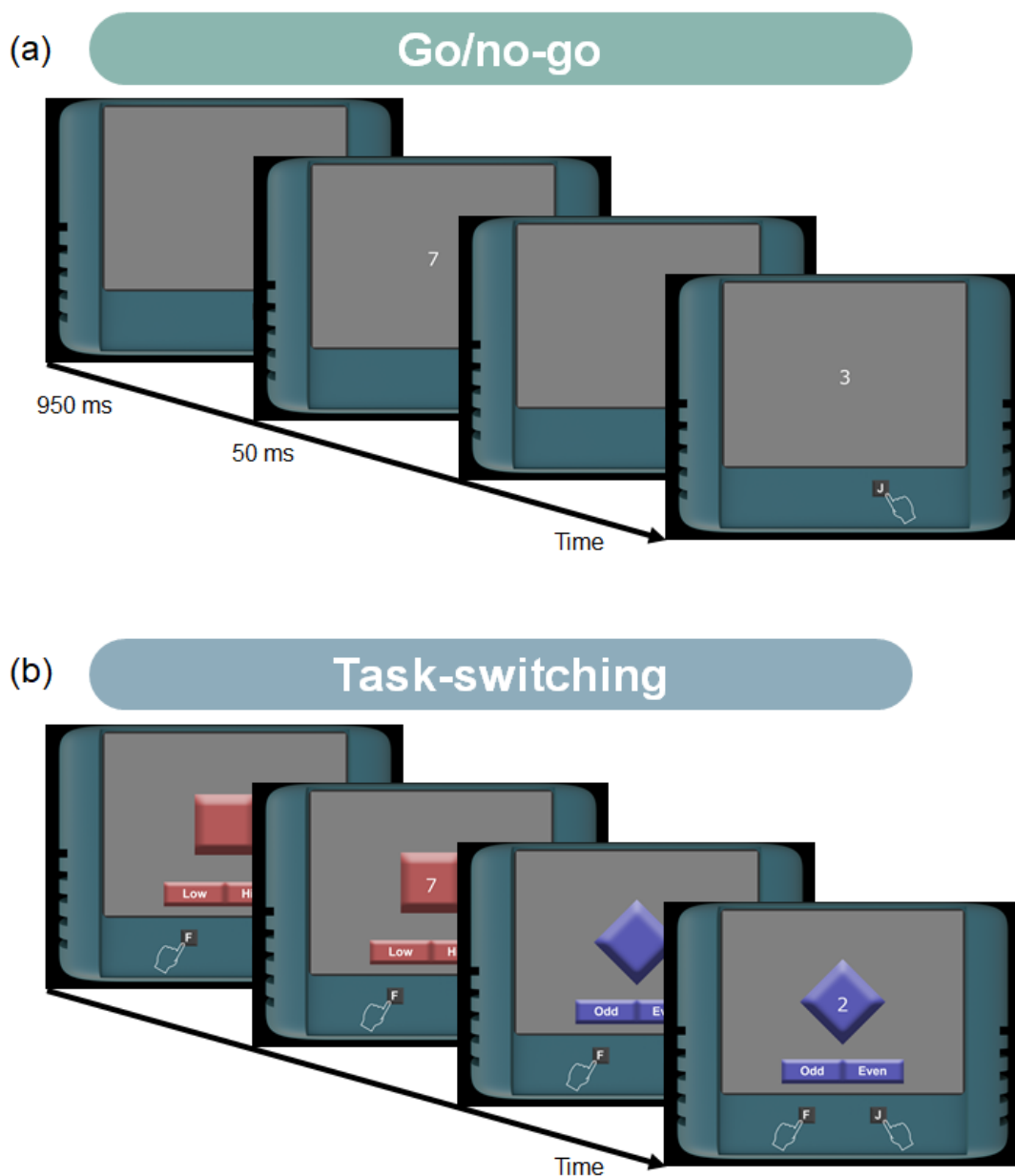


Fig 4.3. (a) Stimuli and tasks in the go/no-go task. On each trial, a digit was briefly presented one by one. Participants were asked to focus on the number "7" and answer whether the number after the "7" was the "3" or not. If the number after "7" was "3," they had to press the key "J" as soon as possible. (b) Stimuli and tasks in the task-switching task. Participants' tasks changed with the task cue. When the task cue was the blue diamond-shaped background, participants had to answer whether the target digit was odd/even. In contrast, when the task cue was the red square background, they had to answer whether the target digit was higher/lower than five.

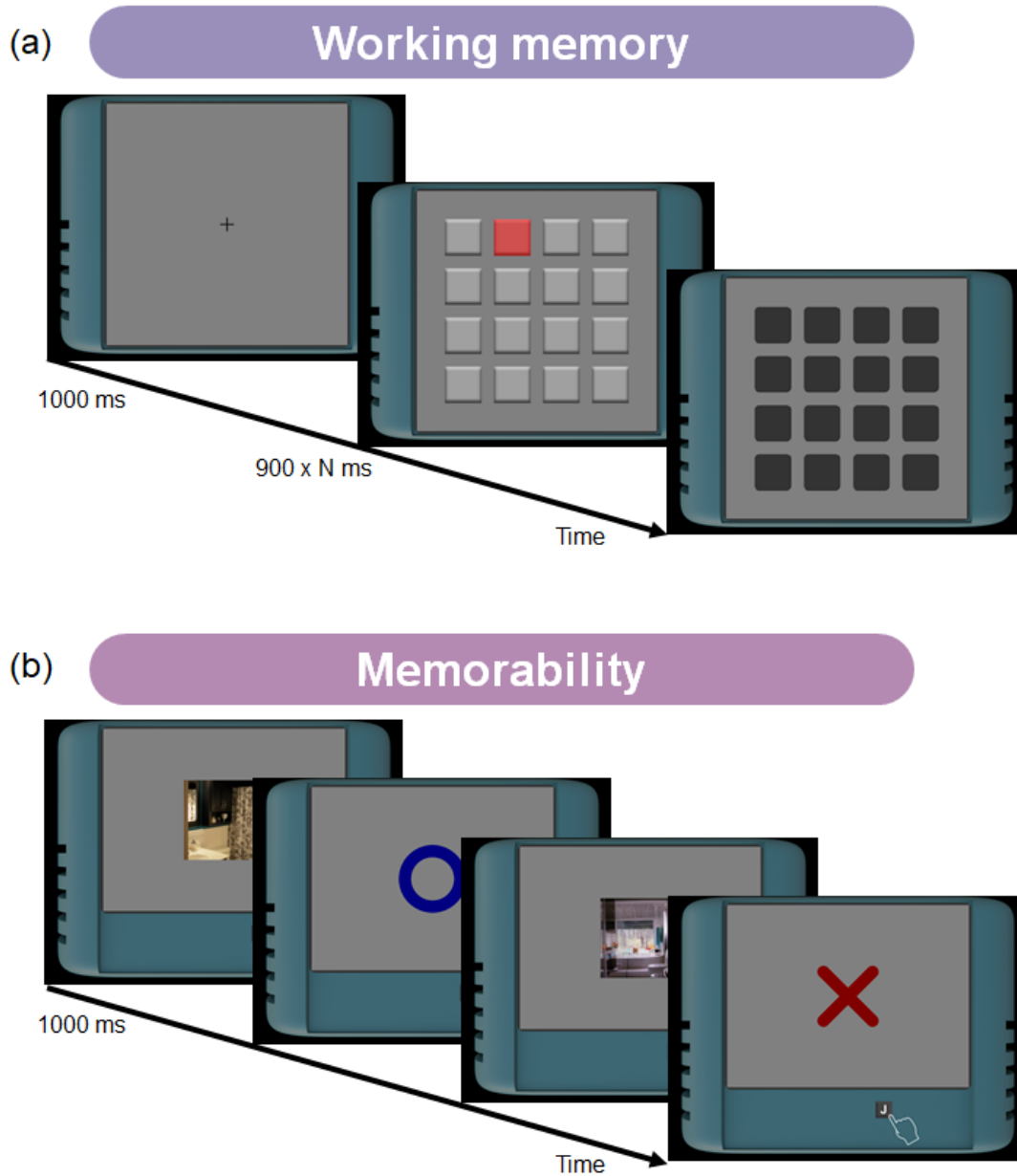


Fig 4.4. (a) Stimuli and tasks in the working memory task. On each trial, one of sixteen squares was sequentially flashed with a reddish color briefly. After the presentation, participants answered the sequence by clicking on the squares in the same order. (b) Stimuli and tasks in the memorability task. On each trial, a natural scene photograph was presented one by one. Participants were asked to remember each photograph and answer whether the photograph is presented twice or not.

4.3 Results

Our test battery includes multiple parameters in each task, and each participant engages in all tasks. This experimental design enables us to evaluate cognitive tasks either within each task in detail or across different diverse tasks. In terms of cognitive test batteries, the evaluation contributes to understanding the effect of parameters on each task, as multiple cognitive abilities can mediate each task. In terms of cognitive mechanism understanding, it contributes to connecting the understanding of each task with other tasks' performances.

This section first describes the basic performance of our cognitive test battery to confirm whether our stimulus manipulation can capture diverse individual differences and how parameter differences affect the performance. Then, as in other cognitive test batteries, we summarize the reliability and validity on the tasks. We evaluate the reliability of our test battery by analyzing the cross-day performance consistency. In the third section, we analyze the latent cognitive factor and evaluate the validity of our tasks to measure human attention and memory. We discuss our latent factors compared with the previous works in cognitive test batteries and cognitive sciences in Section 4 to clarify the position of our test battery.

4.3.1 Basic performance

Figure 4.5 shows the results of basic performances for each task. Different panels show different tasks. Blue circles of each panel indicate the individual performance of the response probability or reaction time for each task condition. The thin green line connects each individual performance across different stimulus parameters. We analyzed the data using Bayesian statistical methods to estimate the mean parameters (accuracy and reaction time) and their 95% credible intervals, shown in the red squares and error bars in Figure 4.5 Andrews and Baguley, 2013; Makowski et al., 2019. Our main motivations to use bayesian inference were the access to credible intervals and to the a posteriori distribution (not only to a point estimate) Kruschke, 2021a. We performed the model inference by Hamiltonian Monte Carlo with the NUTS sampler using PyStan. The simulation parameters of the chain and the iteration were 4 and 10,000, respectively. We estimated the accuracy parameter per task and per condition for each participant by using the binomial distribution as the likelihood and with the uniform distribution on the interval $[0,1]$ as prior for the probability of success per trial. The estimated accuracy was averaged across observers for each sampling and calculated the 2.5% and 97.5% percentiles of the distribution (i.e., 95% credible interval). For the reaction time estimation, we estimated the parameter per task and per condition for each participant by using the the normal distribution as the likelihood and the uniform distribution on the interval $[0,1000]$ as a prior. As in the accuracy estimation, we estimated the averaged mean reaction times and their 95% credible intervals.

We calculated the correct response probability (accuracy) of each stimulus condition for the enumeration, the load-induced blindness, the multiple-object tracking, and the working memory tasks. The accuracy for the multiple-object tracking corresponds to how many objects participants could track, as for the ordinate of the multiple-object tracking tasks in Figure 4.5. For the memorability and go/no-go tasks, we defined the hit rate (HR) and the false alarm rate (FAR) according to the previous works. The HR for the memorability indicated the correct response probability for the images presented for the second time. The HR for the go/no-go task indicated the correct response probability for the go trials. The FAR for the memorability and go/no-go tasks meant the wrong

response probability for the images presented for the first time and the wrong response probability for the no-go trials, respectively. We also evaluated the reaction time (RT) for the trial on which observers correctly responded. For the task-switching task, we used the switching cost metric in addition to the accuracy of the switch and non-switch trials. The switching cost refers to the reaction time difference between the switch and non-switch trials. The positive switching cost indicates that participants took more cognitive load for the switching condition. We evaluated this either the odd/even or large/small condition.

Since a cognitive test battery aims to measure personalized cognitive state, it needs to cover diverse individual differences. Our results showed the large individual difference in the accuracy on the enumeration, the multiple-object-tracking, the load-induced blindness, the working memory, and the memorability tasks (blue circles in Figure 4.5). In addition, the relative individual performance was not consistent across different stimulus conditions for some tasks. For instance, the individual trends for the enumeration and working memory tasks, depicted by the green lines in Figure 4.5, show complex interactions depending on the stimulus parameters.

For the go/no-go and the task-switching tasks, the response probability of HR/FAR and the response accuracy were saturated on most participants, but the reaction time and the switching cost time showed large individual differences, respectively.

Although our results showed large individual differences, the overall performance across participants, shown in the red squares and error bars in Figure 4.5, changed with the stimulus condition on each task, consistent with previous works. The task performance on the enumeration task decreased as the target number increased Trick and Pylyshyn, 1993. For the multiple-object-tracking, the task accuracy and the averaged tracking number decreased as the object speed increased Bettencourt and Somers, 2009a. For the load-induced blindness task, regardless of the condition (near or far) most participants showed misdetection Eayrs and Lavie, 2018b. The switching cost was positive for large/small or odd/even task type Monsell et al., 2003a. The working memory task performance was also decreased with the target number participants remembered Berch et al., 1998a. For the memorability task, the HR was decreased when the target interval was long (> 100) compared to when the target interval was short Khosla et al., 2015.

4.3.2 Reliability across days

We calculated the reliability across two experiment days. The purpose of the analysis is to understand how repeating the set of tasks affects the performance as if the test battery is used as the pre/post assessments of cognitive training. For this purpose, we need to understand the reference performance of repeating the tasks without training to evaluate how much the training improves cognitive ability.

Each participant engaged in the same tasks for two days in our experiment. We evaluated the test-retest reliability across the days with two traditional metrics in the cognitive test battery literature and one analysis based on Bayesian statistics. First, we calculated the performance correlation between the days. The Pearson's correlation coefficients (r) of the accuracy for each task were as follows: (1) multiple-object tracking; 0.89, (2) enumeration; 0.81, (3) load-induced blindness; 0.52, (4) go/no-go; 0.95, (5) task-switching; 0.56, (6) working memory; 0.89, and (7) memorability; 0.66. Second, we conducted the Bland-Altman analysis across days Bland and Altman, 1986 (Figure 4.6). The Bland-Altman plot visualizes the performance differences across the days as a function of the mean performance. Each plot indicates each participant colored with age. We averaged each participant's accuracy across different conditions in each task. Results

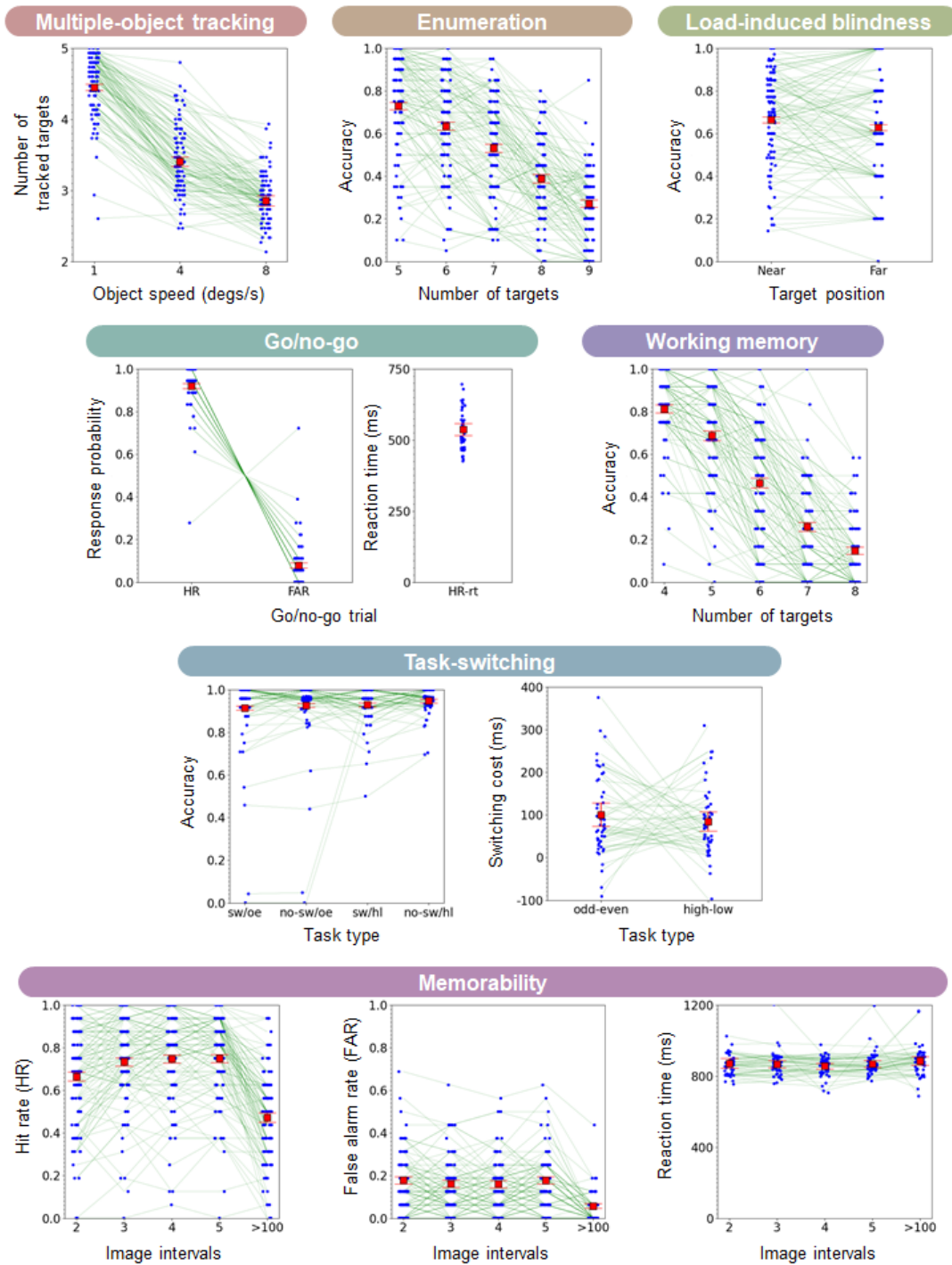


Fig 4.5. Results of all tasks. The response probability or reaction time is shown for each task. The horizontal axis of each panel indicates the stimulus conditions. The small blue circle depicts the individual performance. The thin green line connects each individual performance across different stimulus parameters. The red square and error bars show the mean probability and 95 % credible intervals calculated from Bayesian statistical simulation.

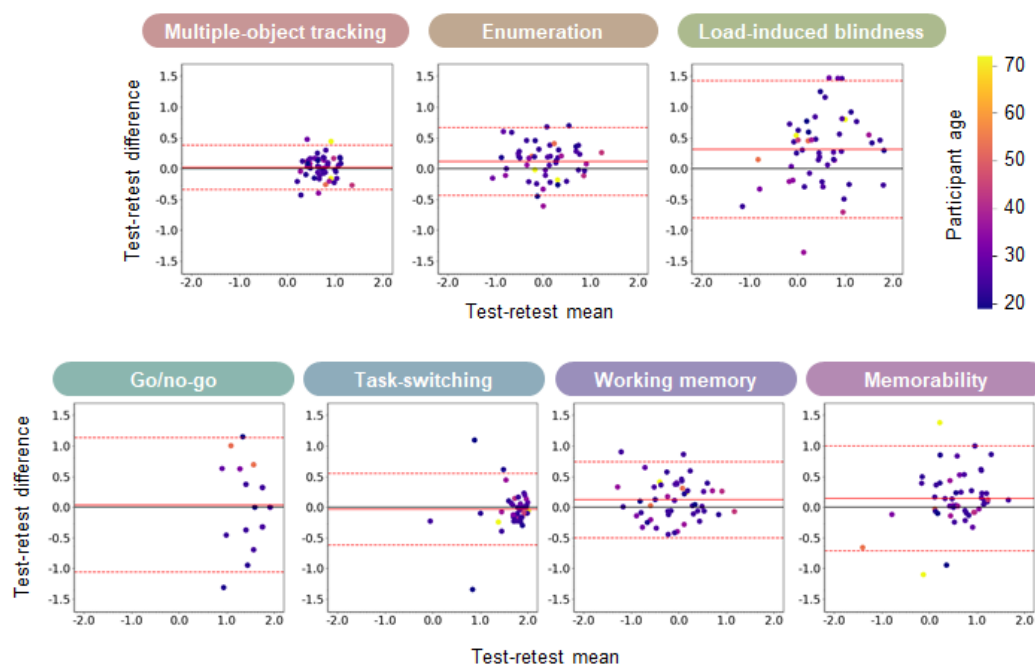


Fig 4.6. Bland-Altman plot for the accuracy data. Each participant’s performance difference was plotted as a function of the mean performance of the two days. Different colors indicate different participant ages. The accuracy data is converted to the inverse normal cumulative distribution function, as in the latent factor analysis. The positive difference means that the second day performance is better than the first day. The solid red line indicates the mean difference across participants, while the dashed red line indicates the ± 1.96 SD of the differences.

showed that some participants showed performance improvement (the positive value in the test-retest difference), but we could not observe a clear trend of age on the improvement.

Next, to evaluate these test-retest effects statistically, we analyzed the posterior distribution differences of test-retest performances using the Bayesian analysis. Figure 4.7 shows the posterior distribution differences between the first and second-day performance. We first estimated the posterior distributions of 10,000 samples of each day by Bayesian statistical methods described above and took the difference of the two days. We subtracted Day 1 from Day 2 for the accuracy distribution and Day 2 from Day 1 for the reaction time distribution to make the training effect positive. We focused on how much the mean difference of each distribution and the 95% credible interval (i.e., highest density interval) deviated from the zero of each difference distribution. The more the distribution deviates to the positive direction, the better the second-day performance is than the first-day one. Results showed that the 95% credible intervals included the no difference point for 32 out of 39 conditions. The credible interval deviated from the point for the long interval and one short-interval condition in the memorability task, the small number conditions in the enumeration task, and the near and far conditions in the load-induced blindness task. Even for these conditions, the mean distribution difference, indicating the effect size of repeating the task, was relatively small (less than 0.1 probability e.g., less than 2 trials per session for load-induced and enumeration task). These findings suggest that the task performance does not improve simply by repeating the tasks twice, and therefore, the test battery is appropriate for the pre/post assessment for cognitive training to evaluate how

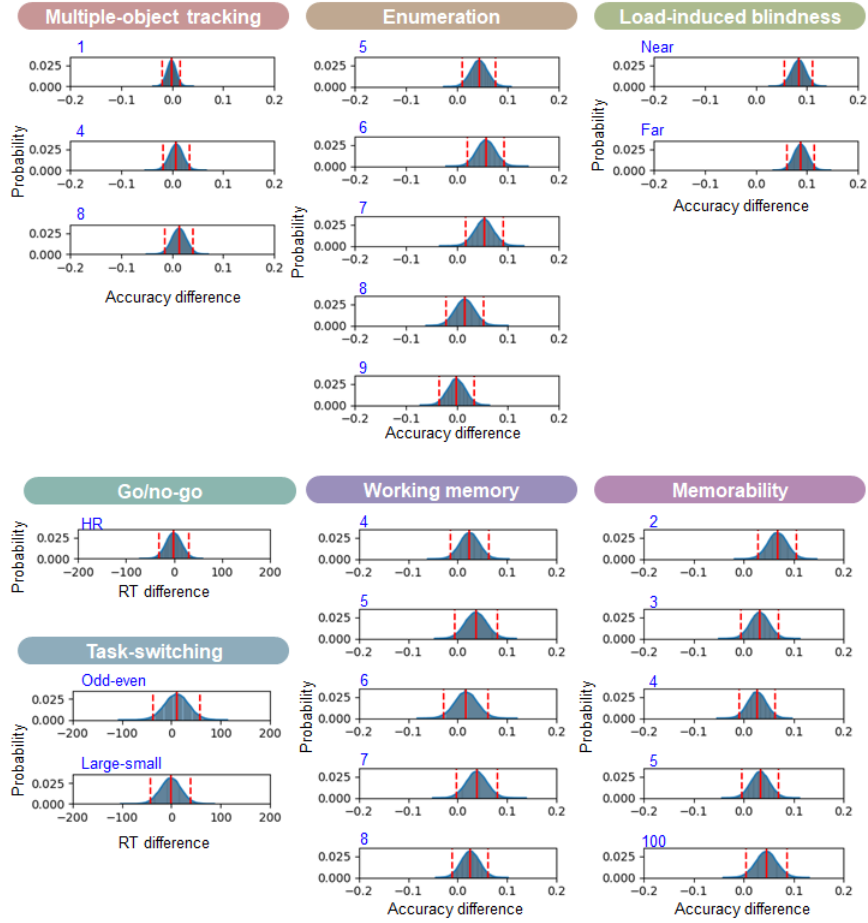


Fig 4.7. Cross-day performance difference. Each panel shows the probability density of the posterior distribution difference between the first and second-day performance (Day 1 -Day 2 for the RT difference and Day 2 - Day 1 for the accuracy difference). The positive value means the better performance in Day2 for either RT or accuracy differences. The vertical red solid and dotted lines indicate the mean and 95% credible intervals, respectively. The condition names put on the left-top for each panel with blue color correspond to the ones shown as the abscissa in Figure 4.5. We show here cross-day performance for the parameters used in the latent factor analysis.

much the training was effective for participants' cognitive ability.

4.3.3 Latent factor analysis

Our cognitive tasks captured the large individual difference, but there remains a question about what internal cognitive factors mediate these differences. To explore the factors, we conducted the latent factor analysis, as in cognitive test battery validation Vermeent et al., 2020. We first transformed the probability data using the inverse normal cumulative distribution function to deal with it for continuous decompositions. We converted the zero and one probability according to the total trial number (i.e., corrected the zero value to $1/2N$ and the one value to $1-(1/2N)$, where N is the total trial number) to avoid the infinity value of the transformation Macmillan and Kaplan, 1985. After data normalization, subtracting variable means from each observation and scaling it using variable standard

deviations, we applied principal component analysis (PCA) to the data. We used 23 variables extracted from the task conditions, shown in Figure 4.7, and 100 participants' data of the first and second days for observations.

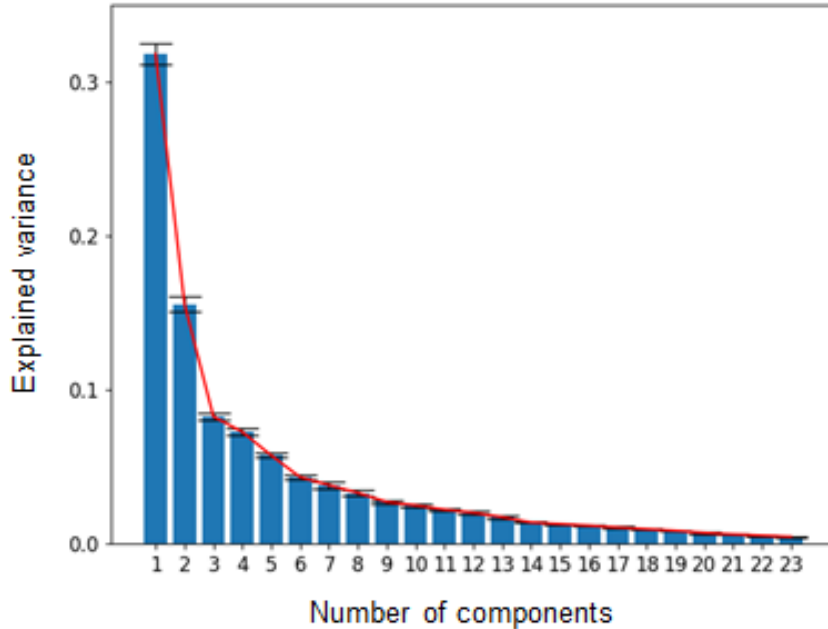


Fig 4.8. Scree plot of the latent factor analysis.

Figure 4.8 shows the explained variances in PCA as a function of the number of components. Based on the plot, we extracted the six components because the cut point shows an "elbow" point L. H. Nguyen and Holmes, 2019. When including the six components, the total explained variance was over 70 %, and each point after six only explains the variance of less than 4%. By using the six factors, Figure 4.10 visualizes the loading of each component and the hierarchical clustering based on the latent factor similarity between different task conditions. The first component showed negative for reaction data of the go/no-go and task-switching tasks and positive for the accuracy data for the other tasks. While the smaller value in the reaction time means a fast (better) response, the larger value in the accuracy means better performance. Therefore, the first component can be associated with a general ability factor, the shared ability across different cognitive tasks to solve them Steyvers and Schafer, 2020b.

The latent factors after the first one in PCA are constrained by the orthogonality of the input parameters. To interpret the factors conservatively, we also conducted the independent component analysis (ICA), where the latent factor orthogonality is not constrained by the input parameters. We used the FastICA implemented in the python scikit-learn library with the six parameters. The results showed that the hierarchical clusterings of the ICA were similar to PCA (Figure 4.10). It is noteworthy that the different conditions within the same task were not always clustered in near categories. For instance, these hierarchical clustering analyses showed that the larger target numbers in the enumeration tasks (8 and 9) were separated from the smaller numbers. Similar trends to this separation were observed in the MOT task.

For either PCA or ICA, the loading results showed the factors to which the memorability

tasks contributed, correlated with the load-induced blindness tasks and the large numbers of the enumeration task. For instance, components 1 and 2 in PCA have the loading of the same sign from the memorability tasks and the load-induced blindness tasks, and similar trends can be found in component 2 in ICA (Figure 4.10). Besides, component 3 in PCA and component 3 in ICA shows the correlated loading from the long-interval memorability task and the large numbers of the enumeration task (Figure 4.10). Besides, the loading results for ICA showed the factors to which the MOT contributed, correlated with the enumeration, the load-induced blindness, and the working memory tasks. For example, the MOT, the enumeration, the load-induced blindness, and the working memory tasks contributed to component 1 (Figure 4.10b). Also, the MOT, the enumeration, the working memory, the go/no-go, and the memorability tasks contributed to component 3 (Figure 4.10b). Figure 4.9 showed the individual participant distribution of the first and second PCA latent components. Each plot is colored according to each participant's age. All participants' data and their basic attributes (i.e., age) are shared in our repository for users to review their future works.

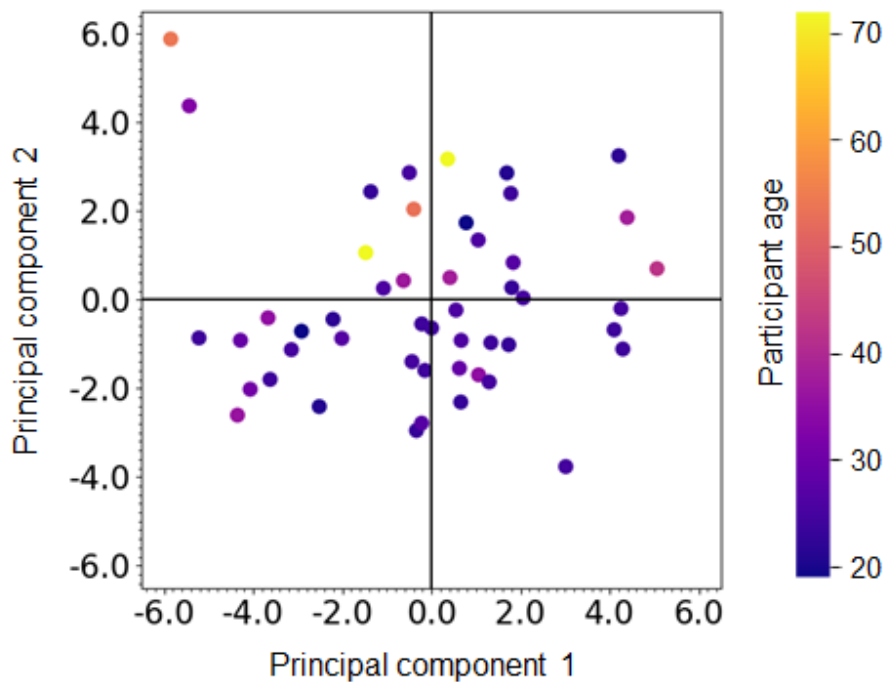


Fig 4.9. Individual data for PCA components 1 and 2. Different colors indicate different participant ages

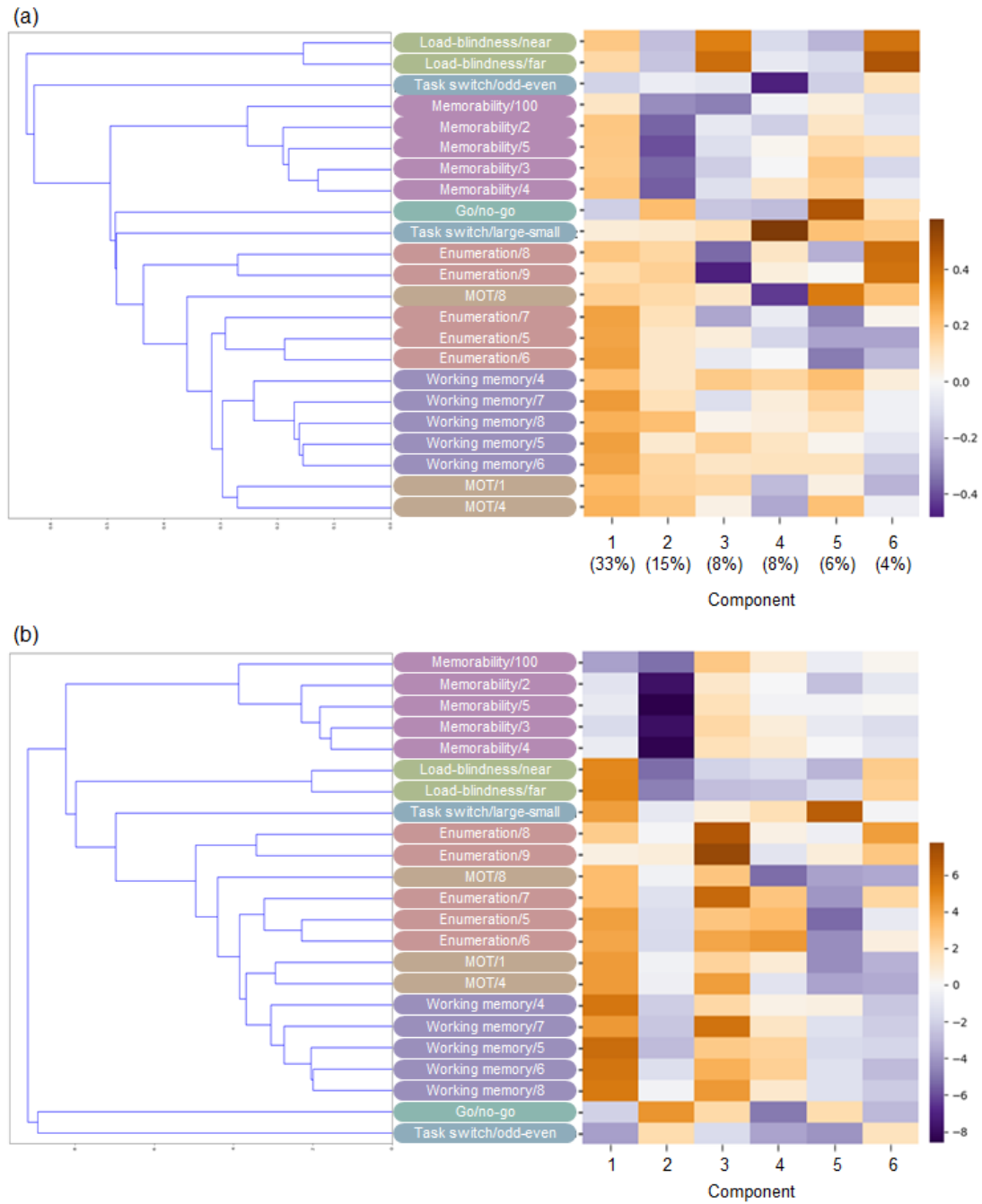


Fig 4.10. The hierarchical clustering and the loading in (a) PCA and (b) ICA. The components in PCA are numbered in order of the magnitude of the explained variance, as in the legend. The component order in ICA is arbitrary because the analysis does not have the priority of the order.

4.4 Discussion

The objective of the study is to create and evaluate the cognitive test battery to measure diverse human attention and memory. The test battery includes seven cognitive tasks: multiple-object tracking, enumeration, load-induced blindness, go/no-go, task-switching, working memory, and memorability. The results of the basic performance show systematic shifts according to the task difficulty and suggest that our test battery covers diverse individual differences. The reliability analysis shows that the task performance across different days is highly similar to each other. Cross-day reliability is essential to use this test battery for cognitive training because learners engage in the cognitive assessment before/after their training, and thus the assessment is optimal if the task performance without training is not affected. In addition, the latent factor analysis showed what internal cognitive factors mediate the individual differences. Specifically, the results suggest that a general ability across all tasks and some task-specific ability underlie the cognitive test battery performance.

These latent factors are consistent with the previous behavioral and neurological findings in cognitive science literature. Many works with a large-scale cognitive task have also reported the shared ability across the visual attention and memory tasks Panichello and Buschman, 2021; Steyvers and Schafer, 2020b. For instance, Steyvers and Schafer, 2020b investigated the behavioral performance with a large-scale cognitive test and analyzed the latent factors using probabilistic PCA. They found that a general ability factor mediates across all tasks, including visual tasks like ours. Panichello and Buschman, 2021 recently suggest from their neurological investigation that the prefrontal cortex works as a domain-general controller for attention and memory tasks. In addition to the general ability, the domain-specific components are also discussed in previous findings Friedman et al., 2008; Larrabee, 2015; Miyake et al., 2000a; Nakai and Nishimoto, 2020; Panichello and Buschman, 2021; Vermeent et al., 2020. For instance, Friedman et al., 2008 investigated latent factor analysis of executive function and suggested that updating and shifting function mediates the task performance in addition to a common cognitive factor. The working memory and the visual-spatial processing are separate but related factors Larrabee, 2015, and the tasks related to these factors could be separately represented in the brain, in addition to overlapped common representation LaBar et al., 1999.

In the cognitive test battery context, the domain-specific latent components are tightly connected to specific tasks. For instance, the working memory factor loads to only span tasks in the validation study of a computerized cognitive test battery Vermeent et al., 2020. In contrast, most of our latent factors were not task-specific, e.g., we did not see the factor only affecting our spatial span task and we found the multiple factors affecting the same task. In addition, it is noteworthy that our results are consistent with previous cognitive science works. For instance, some studies have shown that the working memory performance is related to the MOT R. Allen et al., 2006; Lapierre, Cropper, and Howe, 2017, consistent with our results about the components 1 and 3 of the ICA analysis (Figure ??b). These findings suggest that our latent factors capture more focused cognitive abilities than general cognitive test batteries, overlapped across multiple visual tasks.

MOT tasks are common in the cognitive training literature Harris, Wilson, Smith, et al., 2020; Legault and Faubert, 2012a; Vater et al., 2021c, as we are also conducting such a training project, and it is important to understand to what extent MOT training effects propagate to various cognitive abilities. Evaluating how various task performance is related to MOT abilities in our cognitive test battery contributes to understanding such training

transfer in cognitive training works. Our latent factor analysis showed that the general cognitive ability mediates the MOT performance, including other tasks. Furthermore, consistent with previous works, we found the latent factors contributing to the MOT and the enumeration (Figure ??b, components 1 and 3) C. S. Green and Bavelier, 2006b, the MOT and the load-induced blindness (Figure ??b, component 1) Eayrs and Lavie, 2018b, and the MOT and the working memory (Figure ??b, components 1 and 3) R. Allen et al., 2006; Lapierre, Cropper, and Howe, 2017.

The memorability task has been originally suggested in the computer vision literature, and it is not clear about the relationship with classic cognitive tasks. Previous studies have mainly investigated the task in terms of intrinsic image factors driving humans' image memorizing. However, it has also been shown that cognitive factors mediate the task, especially for the intermediate difficulty we used in our memorability test. Some brain imaging and neurophysiological studies also suggested the neural basis of cognitive contributions Bainbridge and Rissman, 2018; Jaegle et al., 2019; Mohsenzadeh et al., 2019. Specifically, Mohsenzadeh et al., 2019 used a high-resolution-spatiotemporal brain imaging technique with combining fMRI and MEG measurements and recorded the brain activity during the memorability task. They compared brain responses between high and low memorability images and showed that both early visual processing and later cognitive processing mediates the difference between high and low memorability. The present finding can contribute to understanding these processing. Our latent factor analysis showed that the factors including the memorability task are mainly related to the accuracy of the enumeration task with higher difficulty and load-induced blindness task. It has been suggested that the target misdetection in the load-induced blindness and enumeration is due to inattentional blindness over perceptual capacity and can be a different process from the working memory ability Bredemeier and Simons, 2012; Eayrs and Lavie, 2018b. Our finding suggests that the cognitive processing mediating the memorability task is misdetection of the intrinsic image factors in an image due to the inattentional blindness, rather than failure of keeping image contents using working memory ability. This finding is also consistent with the previous result that the memorability performance depends on the eye-gaze position.

Unlike conventional cognitive test batteries, we did not extract a single threshold or slope of a psychometric function for each task but used multiple stimulus parameters' performance for the latent factor analysis. When researchers measure a single threshold for a specific stimulus direction, they implicitly assume that a single cognitive mechanism mediates the task along with the stimulus parameters they controlled. In other words, they assume that participants with the threshold of better performance are superior in a specific cognitive ability. However, this is not always the case if multiple visual mechanisms mediate the task dimension. Consistent with the notion, our latent factor analysis showed that the identical stimulus parameter is not always categorized in the same cluster (Figure ??). Also, the basic performance results showed that the individual trend is highly complex on each task (green lines in Figure ??). The finding suggests that complex interaction lies on the cognitive mechanisms depending on stimulus parameters even in the same task.

We determined the stimuli and procedure of our cognitive tasks by following previous works about cognitive mechanisms of attention and memory. These works tend to overlook individual differences presumably due to small numbers of participant sampling, but our study showed diverse performance differences across individuals for all tasks. For instance, in the task-switching task, the switching cost largely depends on individuals. For some participants, the difference between switching and non-switching trials is more than 200 ms on average, but there are few differences for other participants. When researchers

investigate cognitive ability using one specific task, large individual differences are unknown factors making interpretation difficult. However, as in our latent factor analysis, when the same participants engage in multiple tasks, the large individual difference in one task can be a clue for understanding cognitive mechanisms for another task. This notion suggests that our test battery can also be used for the investigation of cognitive mechanisms as a benchmark evaluation of each participant. Some experiments for cognitive mechanism understanding are hard to collect many participants, e.g., brain imaging experiments. If researchers conduct a new investigation with our test battery, the individual differences in the new experiment can be more understandable.

One limitation of our investigation is that we do not strictly control the observer attributes when recruiting participants. One typical attribute affecting cognitive performance is the age of participants. Aging affects various aspects of cognitive abilities. For instance, it has been shown that the capacity of tracking objects in MOT tasks decreases for older participants Legault et al., 2013b; Sekuler et al., 2008; Trick and Pylyshyn, 1993. Legault et al., 2013b used a 3D MOT task, called the Cave Automatic Virtual Environment (CAVE), and showed that healthy older adults have lower tracking ability than younger adults, but that training with a 3D MOT task improves the tracking ability of healthy older adults in a similar learning function to younger adults. In addition to MOT, other cognitive tasks such as visuospatial attention T. Curran et al., 2001; Greenwood et al., 1993 or working memory Salthouse, 1994 depend on the age of participants. Furthermore, it has been known that other observer attributes such as the level of expertise in sports Faubert, 2013a or gaming Benoit et al., 2020; C. S. Green and Bavelier, 2006a affect cognitive abilities. One needs to separate the participant group according to the targeting attribute to investigate the effect of each attribute on cognitive performance. Although our investigation does not control the population, we analyzed how the performance of different age participants is distributed in our cognitive tasks (Figures ?? and ??). Further investigation is needed to elucidate the effect of observer attributes.

Another limitation is that our online experiment is not strictly controlled in stimulus presentation and response collection compared to laboratory experiments. For instance, the reaction time can be potentially inaccurate due to participants' environment setting because the accuracy depends on the response input device. However, recent studies have suggested that the reaction time measured in web experiments can be comparable with lab experiments Armitage and Eerola, 2020; de Leeuw and Motz, 2016; Hilbig, 2016. In our experiment, we only measured the reaction time by the keyboard input device, not by the mouse clicking (or touch clicking), and restricting the device contributes to decreasing the measurement distortion Armitage and Eerola, 2020. Besides, our reaction time data was comparable with previous findings in lab environments. For instance, the reaction time of the memorability task with intermediate memorability scores in a lab experiment is around 900 ms, which is consistent with our current results Vö et al., 2017. Based on these findings, we believe that using reaction time as a metric for our test battery is acceptable.

In addition, we did not apply the gamma correction according to each monitor's characteristics during our online experiment. One needs a photometer to conduct the gamma correction strictly for each monitor, which cannot be available in online experiments. A way for online experiments is to correct the nonlinearity based on participants' responses using a grating chart, but it could be affected by the quality of participants' responses. We did not apply such user-based correction and presented stimuli without the gamma correction. Previous studies in visual perception and cognition literature have shown that the performance in online experiments can be comparable to that in strictly controlled laboratory experiments for some visual tasks Bylinskii et al., 2015a; Sasaki and Yamada,

2019; Sawayama et al., 2022. For example, the memorability task is conducted both in online and laboratory experiments Bylinskii et al., 2015a. Some studies have suggested that the contrast sensitivity performance could be comparable under sufficient repetition for each condition between online and laboratory experiments Sasaki and Yamada, 2019 and that suprathreshold contrast discrimination with large contrast differences could be stable across online and laboratory experiments compared to blur discrimination tasks for natural object stimuli Sawayama et al., 2022. However, the optimal presentation, especially for the Gabor stimuli in the load-induced blindness task, is to use a linearly corrected monitor. The way of presentation can be critical when users conduct our test battery for some populations that have reduced contrast abilities, e.g., older adults. It has been known that contrast sensitivity is worse for older adults than younger adults because aging changes the optical properties of the eyes Owsley, 2016. When one does not strictly control the stimulus presentation, the effect of such front-end properties can not be evaluated appropriately. Therefore, if users conduct our test battery for such populations in a non-controlled online experiment, they should be extra careful when interpreting the results of the load-induced blindness task to understand whether the obtained performance is due to cognitive abilities or the front-end properties. One additional control for the load-induced blindness in an online experiment might be to conduct a contrast discrimination task without the attention load of the foveal length judgment to confirm whether participants could discriminate the contrast differences without divided attention.

It is noteworthy that we share all source codes and data to conduct the cognitive assessment experiment from our repository (https://github.com/mswym/cog_testbattery). Not only can users conduct our experiment as we did on their own server, but also they can do it more flexibly. One use-case is to conduct our test battery on a shared server. Another case is to conduct it in the laboratory environment. In this case, users can strictly control the monitor size and viewing distance and run the experiment using a web browser.

4.5 Conclusions

In summary, we suggest an online open-source cognitive test battery including the seven cognitive tasks: multiple-object tracking, enumeration, load-induced blindness, go/no-go, task-switching, working memory, and memorability. Our test battery can flexibly be used either online or in laboratory experiments with a web browser. Our benchmark test shows that it captures diverse individual differences and can evaluate them based on latent cognitive factors. Besides, our results suggest a novel finding that the cognitive factor mediating the memorability task is the ability related to inattentional blindness rather than working memory.

Chapter 5

AI-Enhanced Multi-Object Tracking: Toward Individualized Cognitive Training

Aims

The previous chapters have reviewed existing literature on strategies to adapt CT and highlighted the specificities and advantages of MOT. Additionally, we introduced a cognitive battery designed to reliably assess cognitive performance. This chapter aims to integrate all these components and present the data collected during the evaluation of a new AI-based individualized MOT-based training.

Contents

5.1	Introduction . .	144
5.2	Material and method	150
5.3	Results	163
5.4	Discussion . . .	171
5.5	Conclusion . . .	177

Abstract

This study evaluates the effectiveness of an AI-based individualized adjustment procedure versus a traditional staircase method in modulating task difficulty for cognitive training using the Multi-Object Tracking (MOT) task. Conducted as a randomized controlled trial with 72 healthy young adults, it examines the impact of these methodologies on learning trajectories, task performance progression, and participants' subjective experiences, including cognitive load, engagement, and motivation. The AI-based method, which employs a multi-armed bandit algorithm informed by learning progress and a zone of proximal development model, facilitated engagement with a broader spectrum of tasks, enhancing the variety of learning experiences. Performance assessments both within the training context and across a comprehensive cognitive battery of seven tasks revealed improvements in all participants for both the trained task and related tasks, with the AI group demonstrating broader cognitive enhancements not seen in the control group. Subjective assessments reported a decrease in cognitive load over time, indicating efficient learning; however, the AI-based training was perceived as more demanding, leading to lower engagement and motivation scores. Although further investigation is needed to understand these observations, these results highlight the capabilities of AI-driven individualized training methods to enhance cognitive training effectiveness and suggest that they can offer significant advantages over traditional methods.

Collaborators

Adolphe Maxime (AM), Pech Marion (PM), Sawayama Masataka (SM), Clement Benjamin (CB), Joessel Freya (JF), Bavelier Daphné (BD), Delmas Alexandra (DA), Maurel Denis (MD), Oudeyer Pierre-Yves (OPY), Sauzéon Hélène (SH). ¹ This work is currently *in prep* and represents a preliminary version of a future article.

¹Conceptualization (AM, SM, CB, JF, BD, SH, POY); Data Curation (AM, PM); Formal Analysis (AM, PM); Investigation (AM, PM); Writing - Review & Editing (all).

5.1 Introduction

Cognitive tasks aimed at exploring the complexities of human cognition have long been instrumental in shedding light on the intricate workings of the mind. Among these, the Multi-Object Tracking (MOT) task stands as a critical paradigm in cognitive science for probing the mechanisms of visual attention and cognitive processing (Z. W. Pylyshyn and Storm, 1988a). This task, which requires the simultaneous tracking of multiple moving targets amid distractors in a dynamic way, serves as a pivotal tool for understanding how attentional resources are allocated and managed in scenarios that mirror the intricacies of real-world environments (B. Scholl, 2009).

Previous work have demonstrated that MOT performance correlates with outcomes in tasks that demand high perceptual load, focused attention, and working memory capabilities (Eayrs and Lavie, 2018a; Huang et al., 2012b; Skogsberg et al., 2015a; Treviño, Zhu, Lu, et al., 2021). Furthermore, this correlation extends to practical performance metrics in everyday activities, such as driving (Bowers, Anastasio, et al., 2011), video gaming proficiency (C. Green and Bavelier, 2006; Trick et al., 2005), performance of elite athletes (Wierzbicki et al., 2023), or for academic skills among children (Peng and Miller, 2016a, Bertoni et al., 2019b, Tullo, Faubert, and Bertone, 2018b), highlighting the task’s broad applicability and relevance across various contexts. Therefore, there is a growing interest in discerning whether these connections are more accurately described as causal relationships rather than mere correlations. This involves reevaluating the task as a proactive intervention strategy, rather than solely a means of diagnostic or evaluation.

This idea suggests examining the MOT task as a viable cognitive training (CT) tool designed to improve a broad array of skills. Some lines of evidence have already highlighted the possibility to enhance biological motion perception (Legault and Faubert, 2012b) or more specialized abilities such as improving passing precision in elite soccer athletes (Romeas et al., 2016c). However, as echoed in the broader literature of cognitive training (Melby-Lervåg and Hulme, 2013b, Webb et al., 2018b, Luis-Ruiz et al., 2020), there are today mixed evidences in the effectiveness of MOT training and the possibility of transfer toward untrained or daily activities (Vater et al., 2021d). Several prior works have demonstrated that when transfer occurred it was limited to a narrow range of tasks (Strong and Alvarez, 2017b, Harris, Wilson, Smith, et al., 2020, Scharfen and Memmert, 2021b).

One plausible explanation for these disparate outcomes is the significant variation in individual responses to such interventions. While some participants exhibit marked improvements, others derive minimal or no benefit, highlighting the challenge of high inter- and intra-individual variability (Katz et al., 2021, van der Donk et al., 2017, Könen and Karbach, 2015). In this work, we posit that, particularly in the context of the MOT task, tailoring the level of difficulty to the individual’s specific characteristics is crucial for effective training outcomes. While prior studies have demonstrated that adaptive cognitive training yields greater learning benefits compared to non-adaptive methods (Pedullà et al., 2016a; Plass et al., 2019a), the majority have depended on a staircase algorithm to adjust difficulty across various cognitive tasks. Although staircase methods are effective for gradually increasing challenges to match the learner’s level, they typically prescribe a singular developmental path, offering limited customization, and struggle to manage multiple parameters simultaneously (as discussed in Adolphe et al., 2023 for a comprehensive analysis of this issue). However, the rich body of literature indicates that MOT performance is influenced by multiple factors, including the number of targets (Z. W.

Pylyshyn, 2001a), speed (A. O. Holcombe and Chen, 2012), tracking duration (Oksama and Hyönä, 2004b), and motion trajectory types (B. J. Scholl et al., 2001; Suganuma and Yokosawa, 2006). Despite this, much of the existing research has predominantly concentrated on a singular aspect: stabilizing the number of targets while employing speed adjustments through staircase procedures to tailor task difficulty. Therefore, it is pertinent to recognize the MOT task as a complex activity capable of engaging diverse cognitive processes or routines, contingent upon the specific parameters employed. Acknowledging this perspective, the abundant evidence demonstrating MOT performance variation across these parameters presents a valuable opportunity for deeper exploration and application.

Against this backdrop, this study addresses the issue of inter-individual differences by proposing a new individualized MOT training manipulating a larger set of parameters than previously studied. This tailored approach aims to adapt the difficulty of the task to align with the unique cognitive profiles of individuals, offering a solution to the limitations observed in standard training programs. By adopting a personalized approach to training difficulty, based on individual learning progress, we also aim to provide a more effective and engaging training experience (Clement et al., 2015, B. Clément et al., 2024). This work assessed 74 healthy young adults in a randomized controlled trial, contrasting MOT training employing our individualized approach with traditional training using a staircase method. Thus, in addition to showcasing the efficacy of an agnostic difficulty adjustment procedure that dynamically adapts to each participant’s evolving capabilities, this study aims to enhance the applicability of MOT in fostering attentional improvements and cognitive gains. Focusing on personalized training paradigms, we aim to maximize the potential of MOT tasks, contributing to a more effective and engaging cognitive training landscape. This work will be articulated around the 3 following questions:

- RQ1- **Training differences:** How do learning trajectories differ when using our individualized AI-based adjustment procedure compared to a traditional staircase method?
- RQ2- **Impact on cognitive performance:** How does the individualized AI-based procedure affect the progression of task performance in the training activity and in overall cognitive performance as assessed by a broader set of cognitive tasks?
- RQ3- **Impact on subjective feeling:** How does our individualized AI-based procedure affect participants’ self-reported experiences, including cognitive load, engagement, and motivation?

5.1.1 Background

Cognitive training effect and the MOT task Over the past four decades, the Multi-Object Tracking (MOT) task has been a focal point for an expanding array of research inquiries, each delving into its intricate mechanisms and contributing to a burgeoning debate on cognitive attentional processes. As outlined by Meyerhoff et al., 2017a, the evolution of MOT research is marked by a proliferation of theoretical frameworks, each providing unique insights into the task’s underlying complexity. Initially, Z. W. Pylyshyn and Storm, 1988a and Z. W. Pylyshyn, 2001a posited that tracking is governed by a pre-attentive mechanism facilitated by a visual indexing system (FINST), capable of engaging with a limited number of targets—typically around four in the neurotypical population. However, this perspective was soon contested by studies manipulating trajectory patterns; notably, Yantis, 1992a demonstrated that performance could be enhanced through perceptual

grouping, such as when objects form virtual polygons or share a common fate, challenging the simplicity of the initial model. The intricacies of the MOT task were further unraveled through varied manipulations of object speed and target quantity. While Z. W. Pylyshyn and Storm, 1988a, Yantis, 1992a, and Kahneman et al., 1992b supported a model of parallel processing within a unique attentional focus, Cavanagh and Alvarez, 2005b argued for a model of simultaneous monitoring through multiple, distinct foci of attention. The debate was further enriched by the introduction of the multi-identity tracking paradigm, which suggested either a parallel or a sequential shifting of attention among targets depending on the spatial resolution needed to track and identify the targets (Oksama and Hyönä, 2004b, J. Li et al., 2019b). Finally, manipulations of inter-object spacing, tracking duration and the frequency of object interactions highlighted the elasticity of tracking capabilities, moving the conversation beyond fixed limits ("the magical number 4") to a more adaptive, contextually influenced framework as suggested by the Flex model (Alvarez and Franconeri, 2007b), the spatial interference theory (Franconeri et al., 2010) or some Bayesian computational models (W. J. Ma and Huang, 2009; Vul, Alvarez, et al., 2009).

On top of the theoretical debates around the mechanisms underlying MOT's performance, additional research has adopted a more holistic approach, aiming to position the MOT task within a wider cognitive context. This approach elucidates the task's relationship with diverse cognitive and perceptual capabilities by employing cognitive assessments. Initially, MOT was found to correlate with tasks linked to attention, as shown in the study by Skogsberg et al., 2015a, where performance in MOT was associated with tasks that require the rapid shifting of attention across space (spatial shifting task) and the ability to bias attention towards motion perception (motion control task). Furthermore, Bowers, Anastasio, et al., 2011 identified significant correlations between MOT performance and the Useful Field Of View task (UFOV), a task that assesses visual attention Woutersen et al., 2017. Subsequently, an analysis by Eayrs and Lavie, 2018a of performance across various cognitive and perceptual tasks, such as enumeration and change blindness, showed strong correlations between MOT skills and tasks reliant on perceptual abilities like subitizing. Additionally, their exploratory factor analysis suggested that the MOT task not only requires perceptual skills but also taps into working memory components. This finding is consistent with the results from Huang et al., 2012b, Treviño, Zhu, Lu, et al., 2021, and Adolphe et al., 2022a, who also observed correlations between MOT performance and tasks related to spatial span and visual working memory, further illuminating the multidimensional nature of the MOT task within cognitive processing.

Building upon the recognition of MOT's reliance on a spectrum of cognitive and perceptual skills and its dynamic alignment with real-world scenarios, a number of studies have ventured into exploring the correlation between MOT proficiency and practical performance in everyday tasks. First, Bowers, Anastasio, et al., 2011 demonstrated that MOT could predict driving abilities in an older adult population (above 50 years old). Then, delving into expert groups, research showed that athletes, with handball players (Wierzbicki et al., 2023) as an example, outperformed their non-athlete counterparts in MOT tasks. This trend extends to the realm of digital gaming, where action video game players, both adults and children (C. Green and Bavelier, 2006; Trick et al., 2005), exhibit enhanced MOT performance compared to non-players. The study by C. Green and Bavelier, 2006 is particularly significant as it establishes a causal link between the routine practice of action video games and marked improvements in MOT capabilities. While this study illustrates how video game practice can enhance MOT performance, the question remains open as to whether training in MOT tasks could reciprocally improve video game performance, or more broadly, if such practice could translate into improvements in daily

activities and tasks.

The application of the MOT task in cognitive training has garnered significant attention following its introduction, particularly with the advent of commercial platforms like NeuroTracker (Vater et al., 2021d). A first consistent finding across the literature is the systematic enhancement in MOT performance after training - mostly always measured by enhanced speed threshold - in healthy young adults (Parsons et al., 2016b, Parsons and Faubert, 2021) and elder adults (Legault et al., 2013a) and across diverse populations including those with multiple sclerosis (Harenberg et al., 2021), mild traumatic brain injury (Corbin-Berrigan et al., 2018), visual impairments (Nyquist et al., 2016b), as well as in and neurodiverse groups encompassing individuals with learning disabilities and ADHD (Tullo, Faubert, and Bertone, 2018b), each showing notable improvements in MOT tasks albeit with varying baselines and progression rates. However, comprehensive analyses, including systematic reviews, depict a varied landscape concerning the effectiveness of training programs based on MOT (Vater et al., 2021d, Zentgraf et al., 2017b).

Enhancements in performance at near tasks following MOT training have been noted in several studies. For instance, Parsons et al., 2016b reported improvements in healthy young adults on tasks that test sustained and selective attention (such as the integrated visual auditory continuous performance test) and working memory (using the Wechsler Adult Intelligence Scale WAIS-III). Similar 'near transfer' effects, where training benefits transfer to closely related tasks, were identified by Harris, Wilson, Smith, et al., 2020 in a n-back task, by Michaels et al., 2022 in the UFOV task, and among elite soccer players by Scharfen and Memmert, 2021b in a go-nogo task. This effect was also observable in older adults, as demonstrated by Musteata et al., 2019b, who found improvements in episodic memory, working memory, cognitive flexibility, and processing speed immediately after training and at a 1-month follow-up. However, while these findings suggest the potential for near transfer effects from MOT training, the evidence remains mixed. For example, Strong and Alvarez, 2017b explored performance after training on tasks that were close variations of the trained task, altering aspects such as motion type or retinotopic location. Their findings revealed limited transfer, particularly when both motion type and retinotopic location were modified, indicating that further research is necessary to fully understand the extent and conditions of transfer effects from MOT training.

The debate intensifies when considering the far or ecological transfer effects of MOT training, which concerns the application of training outcomes to broader, real-world tasks. Advocates for MOT training highlight its benefits, demonstrated by improvements in specific skill sets, particularly in sports. For instance, studies have shown enhanced passing accuracy in soccer players (Romeas et al., 2016c) and better detection of biological motion (Legault and Faubert, 2012b), suggesting that MOT training can have broad and meaningful impacts. However, this optimistic view is met with caution due to findings from other research indicating more modest results. For example, Moen et al., 2018b reported no significant enhancements in executive functions among athletes from diverse sports disciplines, Harris, Wilson, Smith, et al., 2020 found an absence of far transfer effects in tasks like route recall among healthy young adults, and Komarudin et al., 2021 observed limited transfer effects in elite basketball players. These mixed outcomes contribute to paint a complex picture of the efficacy of MOT training in real-world applications.

Intelligent Tutoring Systems Evidence from developmental and educational psychology suggests that development and learning can be enhanced. One key approach is to align the complexity of training with the learner's current level of mastery and cognitive state.

Lev Vygotsky's early research highlighted instances where students could accomplish tasks typically beyond their reach with just minimal assistance, suggesting the existence of a potential developmental stage beyond their current one (Vygotsky and Cole, 1978a). From this observation, he proposed the concept of zone of proximal development (ZPD), which encompasses tasks that are challenging yet achievable, slightly surpassing the learner's present capabilities. Selecting tasks from this zone ensures an optimal level of difficulty. These foundational ideas have influenced various educational learning models, such as Cognitive Load Theory (Sweller et al., 2019), and is linked to the Goldilocks effect from cognitive training research (Seitz, 2018), which posits that learning is most effective when tasks provide just the right level of challenge, neither too simple nor too difficult.

Derived from these theoretical frameworks, numerous educational strategies have been explored. Research in education has revealed that the learning process is significantly influenced by both intrinsic elements (such as prior knowledge, emotional state, cognitive workload, and motivation) and extrinsic elements (including various instructional design aspects). Tailoring the learning curriculum to individual needs has been identified as a crucial component of an effective training strategy (BLOOM, 1984, M. Deunk et al., 2015). Nonetheless, the application of individualized training, especially in classroom settings, presents practical challenges. Educational technologies have emerged as a potential solution to this obstacle. Intelligent Tutoring Systems (ITS) aim to facilitate student learning by offering a customized sequence of educational activities. These systems typically consist of three interconnected modules: the domain model, which outlines the learning activities; the learner model, which characterizes the student's emotional and cognitive states; and the instructional model, which combines the domain and learner models to establish pedagogical guidelines for selecting appropriate activities for a student in a specific state (Vandewaetere et al., 2011b).

Since Skinner's introduction of teaching machines in 1961 (Skinner, 1961), a variety of systems have been developed (J. R. Anderson et al., 1995b; Conati et al., 2002; Graesser et al., 1999). Studies have shown that with sufficiently detailed customization, Intelligent Tutoring Systems (ITS) can match the effectiveness of one-on-one human tutoring (VanLEHN, 2011b). According to Bartolomé et al., 2018, there are two primary design approaches for structuring learning curricula. The first, known as "linear design," dictates a single, unified path for all learners, albeit allowing for variations in pace and the number of attempts. This approach often employs staircase algorithms and is praised for its scaffolding capabilities, aiding in gradually increasing learning complexity (Sampayo-Vargas et al., 2013, Kalyuga and Sweller, 2005, F. Joessel et al., 2023, Choi et al., 2008, Verniani et al., 2024). While effective in navigating learners towards an ideal level of challenge, these methods typically depend on basic heuristics to evaluate student performance. Consequently, numerous studies have explored more sophisticated techniques for modeling student performance, such as Knowledge Tracing methods Corbett and Anderson, 1994, and its variations (Diard et al., 2010, Meyer, 2019b), to better align challenges with the learner's current abilities.

The second approach, known as "branched-paths design," allows learners to embark on individualized learning trajectories tailored to their specific needs. This methodology has seen various implementations, with many researchers drawing inspiration from the learning progress (LP) hypothesis (Oudeyer et al., 2016b). This hypothesis is a modern reward-learning framework that integrates the concepts of ZPD and flow, establishing a link between intrinsic motivation and learning progress. It posits a beneficial cycle where students are more engaged and learn effectively in areas of personal interest and intrinsic motivation, and the subsequent learning progress provides an internal reward,

further fueling the motivation to learn for its own sake. This concept also intersects with Self-Determination Theory (Ryan and Deci, 2017). While these ideas have been explored in the machine learning field, the Zone of Proximal Development and Empirical Success (ZPDES) algorithm introduced by Clement et al., 2015 leverages a multi-armed bandit algorithm driven by learning progress to navigate through activities within a ZPD framework. Since its introduction, the ZPDES algorithm has been advantageous in teaching mathematical concepts to neurotypical children (Clement, 2018), children with ASD (Mazon et al., 2023), and has been applied in various contexts such as motor learning (Sungeelee et al., 2024), korean language learning (Mu et al., 2021), health literacy (Delmas et al., 2018), or even managing the difficulty level in a VR car driving experience (Ropelato et al., 2017). Building upon the foundational work by Clement et al., 2015, numerous algorithmic enhancements have been introduced. These improvements largely maintain the LP-driven multi-armed bandit mechanism for activity selection but innovate on the dynamics of ZPD evolution. For example, Shabana et al., 2022 in their ITS named Curriculum Tutor, introduced a change point algorithm to better ascertain activity mastery. Further developments include the automatic design of the activity space to reduce dependence on expert knowledge (Annabi and Nguyen, 2023 Ropelato et al., 2017), and proposals by Mu et al., 2017, Mu et al., 2018, Mu et al., 2021 for identifying the initial ZPD, incorporating forgetting mechanisms for learned activities to suggest re-engagement with potentially forgotten tasks. Azeiteiro and Lopes, 2019 also proposed to leverage similarities between students to better evaluate learning dynamics. Lastly, B. Clément et al., 2024 suggested incorporating learner choice into the exploration of the activity space.

Individualized cognitive training Wider research has underscored the importance of adaptive content in CT. For instance, variables such as age and baseline performance levels are predictors of CT outcomes (Karchach et al., 2017a, Katz et al., 2016, Guye, De Simoni, and von Bastian, 2017). Commonly observed in CT interventions are magnification and compensation effects, where respectively, those with initially higher or lower performance levels show greater benefits (Bürki et al., 2014b; Traut et al., 2021b). Consequently, numerous studies have shown that adaptive training methods result in more significant improvements compared to non-adaptive approaches (Neri et al., 2021; Pedullà et al., 2016a). While many studies have implemented adaptive training using staircase techniques, a recent systematic review (Adolphe et al., 2023) highlighted 19 studies that introduced a variety of alternative methods. These strategies can generally be classified into single-domain or multi-domain CT—that is, training that concentrates on either one task or several tasks, respectively—and they incorporate either micro or macro adaptive training, meaning they personalize parameters at the task level (micro) or select activities throughout the training program (macro). According to the review by Adolphe et al., 2023, about half of the studies utilized multi-domain CT combined with macro-adaptive strategies, whereas the other half primarily adopted micro-adaptive training, usually focusing on single-domain tasks. Additionally, a crucial insight from this review is the nascent stage of these methodologies, with the majority of the referenced studies being in preliminary or feasibility phases and characterized by small sample sizes.

The present study falls into the category of offering a single-domain CT with a micro-adaptive approach. In other words, our aim is to offer a personalized trajectory based on a MOT training by adapting the content on a trial basis. As proposed in Adolphe et al., 2023, micro-adaptive methodologies employed can broadly be divided based on two main

goals. The first goal encompasses methods designed to forecast performance or engagement without directly determining the subsequent activity to be introduced. Examples include performance prediction by Sandeep et al., 2020a, deep learning applications by Eun et al., 2022 and Singh et al., 2022, and logistic regression analysis by Book et al., 2022. With these strategies, the method does not yield the next activity but rather infer information leverage to make a choice. In the other hand, the second goal encompass strategies that directly predict the next activity to propose. Notably, most approaches rely on Reinforcement Learning (RL) techniques, a framework particularly well-suited for this kind of task, as noted by Doroudi et al., 2019. Specific examples include the use of the Bucket Brigade algorithm by Kitakoshi et al., 2015, 2020b, Q-learning by Rathnayaka et al., 2021, Tsiakas et al., 2018 and Zini et al., 2022, Actor-Critic methods by Wilms, 2011 or the algorithms presented in this work.

5.2 Material and method

5.2.1 Schedule

As illustrated in Figure 5.1, all participants conducted the online experiment during 12 days (two blocks of five consecutive days with two days off at mid experiment). A cognitive assessment was completed on the first and on the last day. Following previous results in Bediou et al., 2018a, 2023b; Lampit et al., 2014a on training dosage, participants practiced the MOT task for a total of 8 hours with a maximum training duration of one hour per day. To better fit with participants schedule, it was possible to divide by two sessions of 30 minutes per day and freely choose time of the day to perform the training. After or between training sessions participants were also tasked to answer questionnaires about their feeling in practices.

5.2.2 Participants

Participants were recruited with Prolific² with the only filter to be aged between 18 and 65. Everyday, participants engaged in the experiment were positioned in a new Prolific session. Payments increased throughout the week as a bonus for completing the sessions till the end. Participants not participating within schedule were alerted by message. After the first session of evaluation, participants were pseudo-randomly assigned to 2 possible groups: control or intervention (named ZPDES group in the rest of the manuscript). Age and attention deficit were controlled in order to get homogeneous groups.

5.2.3 The training task: Multi-Object Tracking

Following the design in (F. Joessel, 2022b), the objective of each training trial was to track multiple targets, while simultaneously ignoring irrelevant distractors. Initially, the targets were shown moving for a set duration (1000 milliseconds). This was followed by a period of immobility for all objects (1000 milliseconds), ensuring the participant identified all the targets to be tracked. After the brief pause the targets altered to resemble the distractors in both shape and color. Subsequently, after a variable duration of few seconds of tracking, all objects ceased movement. Participants were then tasked with identifying

²www.prolific.com

the original positions of the targets by clicking on them. Prior to each training session, a detailed step-by-step tutorial was provided for guidance.

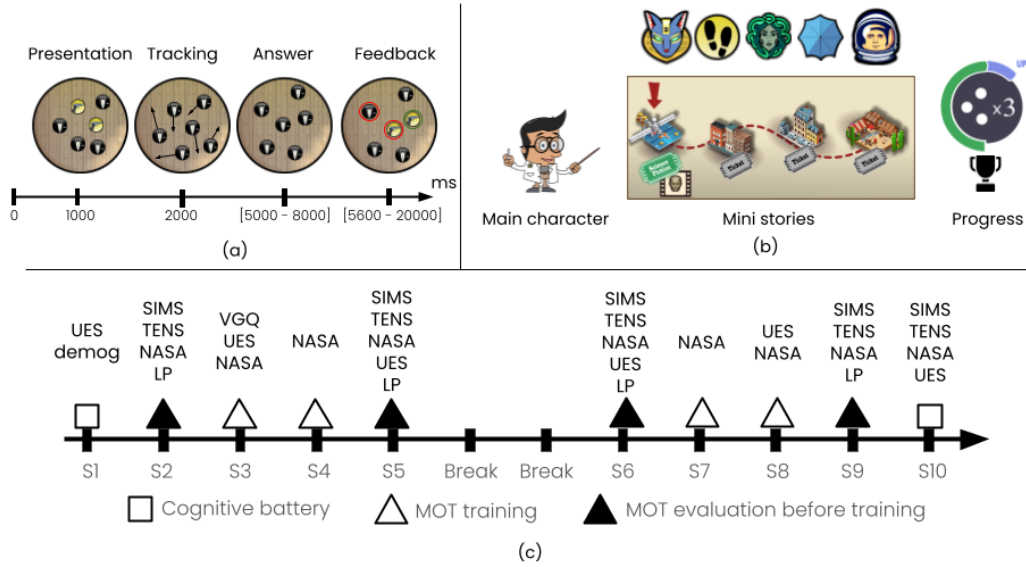


Fig 5.1. (a) The MOT task. (b) Several visual snapshots of our intervention. (c) Schedule proposed to participants

As illustrated in Figure 5.1, the visual characteristics varied daily, depending on the session. The shapes, colors of objects and background, determined during a preliminary study, were carefully chosen to ensure that contrast did not affect performance. As task variability appears to be a key element to foster participant motivation and long term learning (Raviv et al., 2022), the gamified elements were introduced in a random manner throughout the week (see section 5.2.6 for details). The objects were animated within a circular area, maintaining a distance ranging from a minimum of 2° to a maximum of 8° of visual angle, to facilitate peripheral vision engagement. At the center of this area, a fixation cross was consistently displayed. The size of each object was standardized to 1° of visual angle (F. Joessel, 2022b). Participants were instructed to keep a distance of 50 cm from the screen, and the physical size of their screen was collected before the experiment. Using these parameters, visual angles were adjusted for each participant to ensure uniformity.

As outlined in following sections, the task's complexity is dynamically regulated by an AI algorithm, adjusting it on a per-task basis. Five key parameters are manipulated: the number of objects to track or targets, the speed of all objects, the tracking duration, the spacing between objects, and the answer duration. The selection of these parameters was informed by their demonstrated impact on performance, as cited in (chapter 3). The number of targets is chosen to range between 2 to 7. To ensure at least 2 more distractors than targets, the total count of objects is fixed to 16. The initial positions of the objects are randomly assigned. Their movement paths are linear, incorporating a Gaussian random deviation relative to the current motion vector. Speed variations range from $2^\circ/\text{s}$ to $5^\circ/\text{s}$. To prevent overlap, a variable repulsion zone, ranging from 1° to 0.3° , is maintained around each object. The duration for tracking these objects can vary between 3 and 10 seconds. Participants are required to identify the initially presented targets by clicking on them. The duration allotted for responding varies between 6 and 12 seconds.

5.2.4 AI-based individualization procedure

To tailor the learning path to the specific needs of each participant, an AI-based virtual instructor dynamically selects the most suitable sequence of activities over time. Selecting an activity should be understood as choosing parameter values impacting the type of activity, the task dynamics or the task difficulty. This study utilizes five such parameters for the MOT task (number of targets, speed, spacing between objects, tracking duration and answer duration).

The "Zone of Proximal Development and Empirical Success" (ZPDES) algorithm - our virtual instructor - is interacting with the participant through a reinforcement learning framework augmented by human feedback (Clement et al., 2015). Reinforcement learning (Subramanian et al., 2022) is a method in which an agent observes the state of an environment, in this case, the previous responses of a human participant and acts upon it, here by suggesting a new activity. This decision alters the environment's state, specifically the skill level of the participant, and the agent then receives feedback, or a reward, that gauges how effective the action was in that particular state. In our context, the reward is the participant's learning progress, and the agent's goal is to maximize this progress.

ZPDES approaches the identification of an optimal teaching sequence as a multi-armed bandit problem. Analogous to a gambler at a casino facing multiple slot machines, the challenge is to identify which machines yield the highest gains. Then, in our teaching scenario, the key task is to ascertain which MOT activities (comparable to the slot machines in the casino analogy) are most beneficial. To achieve this, a balance must be struck between exploration, which involves experimenting with new activities, and exploitation, which focuses on utilizing activities already known to improve learning. To address this optimization challenge, the ZPDES algorithm assumes that an activity with high learning progress should be presented more often. Then, as described in next paragraph, the algorithm also proposes a model of the ZPD as a way to limit the exploration of the activity space.

Note: To provide a general sense of the algorithm main ideas, a simplified version is presented in the following sections. The detailed and exact description is provided in Appendix 5.5.

For a given participant, let's consider an expert able to track the history of answers for any activity in the finite and discrete set of all possible activities. By updating the history after each presentation of an activity, the expert is able to update its learning benefits, known as the quality of the given activity a . If all activities have a uniform non zero quality at start, the quality of an activity - denoted by $q(a)$ - is updated through the update rule:

$$q(a)^t \leftarrow \beta q(a)^{t-1} + (1 - \beta)r(a)^t \quad (5.1)$$

Where β is an hyperparameter in $[0, 1]$ that weight the previous quality estimates relative to the most recent reward signal. Here $r(a)^t$ corresponds to the reward associated with the current activity. As highlighted in Oudeyer et al., 2016b, the learning progress can be used as an efficient signal toward interesting pedagogical activities as feeling of learning progress has been proven to foster learning, memory retention and motivation. As such, the reward associated is an estimation of the learning progress for the last activity sampled:

$$r(a)^t = \left| \sum_{k=t-D/2}^t \frac{h_{a,k}}{D/2} - \sum_{k=t-D}^{t-D/2} \frac{h_{a,k}}{D/2} \right| \quad (5.2)$$

Here, $h_{a,k}$ corresponds to the results of activity a in history at time k . Thus, $r(a)^t$ quantifies the reward associated with engaging in activity a at time t , calculated as the absolute difference between the average success rate of the most recent $D/2$ time steps and the preceding $D/2$ time steps within a defined window of size D . Importantly, the absolute learning progress is utilized here as it enables the promotion of activities showing both increasing and decreasing progress, the latter potentially indicating instances of forgetting. For our MOT task, the result of an activity is computed as the F1-score, that is the harmonic mean of precision and recall (see equation 5.3). Precision in our scenario is a metric sensitive to the number of distractors as it corresponds to the number of target retrieved over the total number of objects clicked. Recall in our scenario corresponds to number of targets retrieved over the number of targets to retrieved. In our settings, the F1-score is more relevant than traditional metrics like ENOT (B. J. Scholl et al., 2001) because participants can click on all scene objects, necessitating consideration of distractor clicks.

$$\text{F1-SCORE} = \frac{2}{\frac{1}{\text{PRECISION}} + \frac{1}{\text{RECALL}}} \quad (5.3)$$

Finally, in order to sample a new activity, qualities of all activities are transformed into probabilities by following a 2 steps procedure. First, to encourage exploration, a randomly drawn exploration term is added:

$$q(a)' = (1 - \gamma)q(a) + \gamma\xi_u \quad (5.4)$$

Where ξ_u is randomly sampled from a uniform distribution in $[0,1]$ and γ is an hyperparameter in $[0,1]$.

Second, qualities are normalized and utilized for sampling:

$$p(a) = \frac{q(a)'}{\sum_{a_k} q(a_k)'} \quad (5.5)$$

On top of the multi-armed bandit algorithm, to mitigate the explosion of possible combination of parameter values, prior knowledge of the training task is leveraged to streamline the exploration of the activity space. Drawing on the concept of the ZPD, the range of possible activities is narrowed to a subset containing only optimally pedagogical activities. In modeling this ZPD, only a select group of activities are 'activated' — meaning they are available for proposal and monitored by the expert. When the average success rate, computed as the mean F1-score over last trials, of these activated values surpasses a predefined threshold (typically 70% in this study), the ZPD expands, activating a new value for more challenging activities. Conversely, a value is deactivated once its success rate exceeds a different predetermined threshold (90% in this study). These rules ensure that the ZPD consistently aligns with an optimal challenge level.

It is noteworthy that the ZPDES algorithm facilitates a more adaptable exploration of the activity space. Unlike the staircase method, which provides a single predetermined route toward the activity space, the ZPDES algorithm can generate a unique trajectory for each participant. Another perspective on this difference is that the staircase algorithm offers a single degree of freedom in difficulty management (increasing or decreasing along a

predetermined path), whereas ZPDES offers a much larger set of options. Further details are provided in the related Appendix (5.5).

Staircase procedure To compare the individualized procedure with a traditional method, we employed a staircase algorithm as a control mechanism to regulate the difficulty level of the training. Specifically, we used a 2up/1down algorithm, which escalates in difficulty after two successes and decreases after one failure. In contrast to the previous method (F. Joessel, 2022b) that used a 3up/1down approach, we opted for this less restrictive configuration to facilitate quicker progression in the staircase. Additionally, to maintain a comparable activity space between the two procedures, a hierarchical structure was implemented. The four parameters (speed, spacing, tracking duration, and answer duration) were managed using a first staircase named S1. To avoid large variation in difficulty, steps in S1 were managed one parameter at a time (in a circular manner). As such, it required 8 consecutive successes for all S1 parameters to advance by one step. Once all parameters reached either their minimum or maximum threshold on S1, the staircase algorithm pertaining to the number of targets, named S2, was updated. Subsequently, all values of S1 were reset to their highest or lowest possible settings.

5.2.5 Cognitive battery

A cognitive evaluation was conducted before and after the MOT training. Seven cognitive tasks from Adolphe et al., 2022c were proposed. In order to be able to detect near to far transfer effect, the choice of the tasks was conducted to evaluate on tasks that showed a correlation in performance and tap on similar cognitive dimension as the MOT (near effect) but also on different control tasks (far effects). We employed a comprehensive set of tests to broadly assess cognitive functions as follows: measures of selective and divided attention (UFOV, Loadblindness, Enumeration), cognitive flexibility (task switching), inhibition (Go/NoGo), working memory (Corsi), and episodic memory (memorability). Duration of the evaluation was planned to last 1 hour and 30 minutes. All tasks are freely accessible for tests³ and open-sourced⁴. For each task, performance metrics were calculated separately for each difficulty condition and also averaged across all conditions.

Multi-Object Tracking - Near-Transfer Evaluation Task For the evaluation, MOT task was inspired by Bettencourt and Somers, 2009c and differed from the training phase on some settings. The decision to use a similar yet slightly varied MOT task was influenced by Strong and Alvarez, 2017b, which demonstrated that the transfer of training to a similar task was reduced when the MOT test involved only changes in both motion types and retinotopic locations. Participants task was to remember a set of targets discs moving around a set of distractors objects. Targets and distractors were respectively presented as red and white discs of 1.2° radius. Presentation of the targets lasted 1s and the tracking phase 8s. The available scene was a square canvas of 12x12° where objects followed linear trajectories and bounced on borders. Occlusion was allowed between objects. Participants had to click on objects to recall targets. Task difficulty was controlled by changing speed (1, 4, and 8 degrees/s) and the number of targets (3, 5). All combinations of speed and the number of targets resulted in six difficulty conditions, each presented nine times in a random order, leading to a total of 54 trials. The dependent variable employed to evaluate

³https://flowers-ol.bordeaux.inria.fr/flowers_demo

⁴<https://github.com/flowersteam/cognitive-testbattery>

performance was accuracy, defined as the ratio of correct trials (where all targets were retrieved) to the total number of trials.

Useful Field of View (UFOV) The UFOV test (Ball et al., 1988, L. Hoffman et al., 2005) is used to assess selective and divided visual attention. This task has shown significant correlation with the MOT task in Bowers, Anastasio, et al., 2011. The experimental design used in this study is an implementation of the description provided in Yung et al., 2015b. Participants were tasked to discriminate a central target stimulus and to detect the radial location of a peripheral stimulus. Central target stimuli were 1° smiley flashing in center of the screen with either long or short hair. Peripheral stimulus was 1° stars taking possible 8 radial directions at 7° eccentricity. On all available other radial direction squared distractors were positionned at 7°, 5° and 3° eccentricity. Position of peripheral stimulus and choice of smiley hair type was randomized across trials. Stimuli time presentation was controlled with a 3-down, 1-up staircase procedure where stimuli presentation time was decreased after 3 correct (i.e., both central and peripheral tasks were correctly completed) consecutive trials and increased after 1 failure. Staircase step size was set to 2 frames before the 3 first reversals in the staircase and decreased to 1 frame after. Conditions to end the task were: if participant reached the ceiling (99 frames) or floor duration (1 frame) for 10 consecutive trials, if participant completed 8 reversals in staircase, or if participant reached a maximum of 72 trials, whichever occurred first. The dependant variable employed to evaluate performance was the the average display duration proposed over the last 5 attempts of the staircase.

Load-Induced Blindness The Load-Induced Blindness (LIB) task assesses an individual's attention system's ability to manage visual stimuli under varying perceptual loads, investigating the boundaries and resource allocation of visual attention. Latent factor analysis in Eayrs and Lavie, 2018a demonstrated that MOT and LIB tasks are correlated and tap into a common perceptual capacity component. Participants were asked to engage in a dual task: a discrimination task on central vision and a detection on peripheral vision. Initially, each trial began with a fixation pattern for one second, followed by a central cross displaying four Gabor patterns for 50 milliseconds. The central vision task required identifying the longer line in the cross, either vertical or horizontal. For peripheral vision, the task involved spotting the Gabor pattern with increased contrast. Participants were instructed to primarily complete the central vision task accurately; if they failed, their responses to the peripheral task were disregarded. The lengths of the cross patterns were set at either 0.5° or 1.0°, and the orientation of the longer line within each cross was determined randomly, being either vertical or horizontal. The Gabor patterns were presented at distances of either 3 degrees (representing a nearer condition) or 6 degrees (indicating a farther condition) from the center of the screen, thus creating two levels of task difficulty. Each participant experienced twenty trials in each of these stimulus condition. The dependent variable employed to evaluate performance was accuracy, defined as the ratio of correct trials (where both discrimination tasks are correct) to the total number of trials.

Enumeration In the enumeration task, participants were tasked with counting a set of circles presented for a brief duration. This task primarily assesses subitizing capacity, which has been identified to correlate with MOT performance in Eayrs and Lavie, 2018a. Each trial began with a fixation pattern against a middle gray background, lasting for

1,000 milliseconds. Following this, a set of white circles, each $0.5 \times 0.5^\circ$ in diameter, appeared for a short span of 50 ms. These circles were distributed across a 5° diameter area on the screen, arranged to avoid any overlap. The participants' objective was to count the number of circles and record their count using a slider bar. The level of difficulty in the task was modulated by altering the count of circles, which varied from 5 to 9 in a pseudo-random sequence. Each level of difficulty was presented 20 times, cumulating in a total of 100 trials. The dependent variable employed to evaluate performance was accuracy, defined as the ratio of correct trials (where the dot count is correct) to the total number of trials.

Corsi Block Tapping In the Corsi Block Tapping task Berch et al., 1998b, participants were required to memorize and replicate a sequence of squares displayed on a 4x4 grid. This task primarily assesses spatial working memory. Previous works indicated that the MOT task not only loaded on perceptual components but also on constructs related to working memory C. Green and Bavelier, 2006, Eayrs and Lavie, 2018a. Each square in the sequence flashed in red for 900 ms. The sequence was randomly determined for every trial. Following the sequence presentation, we used the forward version of the task where participants were required to reproduce the sequence by clicking on the squares in the same order as they appeared. The dimensions of each square were 2.0° on each side. The task's difficulty varied, with the sequence length ranging from 4 to 8 flashes, chosen in a pseudo-random manner. Each sequence length was tested 12 times per participant, totaling 60 trials for each observer. The dependent variable employed to evaluate performance was accuracy, defined as the ratio of correct trials (where the sequence is correctly retrieved) to the total number of trials.

Task-Switching The task-switching activity employed in this study draws inspiration from the task-cueing framework as described in Monsell et al., 2003b. This task primarily assess cognitive flexibility. Participants were presented with digits ranging from 1-4 or 6-9 and were required to perform two different tasks based on the cue provided. If a blue diamond-shaped background was displayed, they needed to identify the target digit as odd or even, using the "F" or "J" keys respectively. Alternatively, if the task cue was a red square background, the task was to determine whether the target digit was greater or less than five, again utilizing the "F" or "J" keys. The procedure for each trial started with the display of the task cue for 650 milliseconds, followed by the appearance of a target digit. The rectangle serving as the background for the cue measured 4.9° on each side, with the target digit centrally placed within a 1.5° square area. A blank screen of 1,000 milliseconds succeeded the participant's response. The first 20 trials were designated for practice. The evaluation phase consisted of 96 trials, with 48 switch and 48 non-switch trials presented in a pseudo-randomized order. For each trial, the digit displayed was randomized from 1 to 9, excluding 5. The dependent variables used to assess performance included accuracy, defined as the ratio of correct trials (where the response aligns with the objective) to the total number of trials, and the switching cost, which refers to the reaction times for trials involving objective task reconfiguration or where two consecutive task cues differ.

Go-NoGo The Go-NoGo task, inspired by the work of T. Mani et al., 2005, is a specialized variant of the Continuous Performance Test and primarily assess inhibition. In this task, participants were required to observe a sequence of digits that were presented rapidly and to respond specifically when the digit '3' followed the digit '7'. In instances

where '3' succeeded '7' (referred to as Go trials), participants needed to press the "J" key promptly. Conversely, if a different number followed '7' (termed NoGo trials), participants were instructed to refrain from responding. Each trial commenced with a fixation point displayed for 1,000 milliseconds. Subsequently, each digit appeared individually for a duration of 50 milliseconds, separated by an interstimulus interval (ISI) of 950 milliseconds. The digits were shown within a 1.5-degree square area. The number of Go and NoGo trials was equal, with 25 of each. The dependent variables used to evaluate performance included accuracy for GO trials, false alarm rate (hits on NoGo trials), and reaction times on GO trials.

Memorability The memorability task measures human memory performance for natural scene images (Isola et al., 2011a; Bylinskii et al., 2022). Stimuli are images with intermediate memorability scores from the FIne-GRained ImageMemorability (FIGRIM) dataset (Bylinskii et al., 2015b). This task was used as a secondary task to assess visual memory. Participants were assigned the task of observing a series of natural scene photographs and identifying any picture that appeared twice by promptly pressing the "J" key. Each photograph was shown for 1000 milliseconds, followed by a feedback period of 1400 milliseconds indicating whether the response was correct (a blue circle) or incorrect (a red cross). Photographs, falling under the category "bedroom" or "kitchen," were selectively chosen from the dataset based on their hit rates, which fell within the range of 0.6 to 0.7. The task was randomly presented twice to each participant during the evaluation phase, each time with two different versions of the same task. Each version contained 120 images in total. In each version, 40 images were displayed twice. The task's difficulty was adjusted based on the interval between the two presentations of the same image. The two versions differed in the range of images used to manage the gap. In the short-range difficulty version, the gap between two identical stimuli varied between 2 to 5 images. In the long-range difficulty version, this interval ranged from 100 to 109 images. The order of stimuli presentation was randomized across participants. The dependent variables used to assess performance included accuracy and reaction times for hit trials (second presentation of a stimulus), and the false alarm rate (first presentation of a stimulus).

5.2.6 Online experiment management tool and Gamification

The whole intervention was conducted online through a newly open-sourced web application⁵. Participants autonomously created their account and connected on the plan schedule. Following the experimental plan, when an activity was available - for participants consenting to provide an email address - a message was sent to warn the participant to practice.

To enhance motivation and engagement, participants were guided through the intervention by a character representing the experimenter (Figure 5.1). Additionally, participants followed mini-stories throughout the experiments. Mini-stories order was randomly set for each participant. The visual content changed depending on the progress in each mini-story. Each session of training started with a map explaining the mission of the day and the progress in the experiment. The incorporation of mini-stories was driven by the intent to engage participants and to introduce more variability in the training, while ensuring coherence across sessions. After each trial, a feedback was provided telling how many targets and distractors were retrieved. Participants were also able to access a progress

⁵<https://github.com/madolphe/Flowers-OL>

panel summarizing the number of success for different level of difficulties. Within this feedback interface, difficulty levels were only represented by the number of targets. This was done to avoid confusion with combination of parameters and because the number of targets was the most explicit parameter to communicate to participants. Additionally, to help participants better estimating their learning progress, a progress bar was added for each level of difficulty. This bar was described to participant as a "skill level". For the control group, the skill level for each difficulty (i.e., each number of targets) was computed as the maximum level achieved within the S1 staircase (i.e., speed, spacing, tracking duration, and answer duration). For the ZPDES group, the skill level for each difficulty was computed as the maximum activity level opened in the ZPD. All visual contents were either originally created or generated with a text-to-image generative model⁶.

5.2.7 MOT Intra-training evaluation

Several hidden evaluations through the intervention were used to assess and compare performance between groups. Four sessions of evaluation were proposed to the participants at the beginning of sessions 1, 4, 5 and 9. While participants were unaware of the evaluation, it provides reliable and comparable performance assessment between groups throughout the intervention on the trained task. The evaluation set was composed of 12 activities from all possible combinations of number of targets (2,4,6), speed (low or high) and spacing (low or high) randomly proposed 4 times each.

5.2.8 Subjective questionnaire

As highlighted in Adolphe et al., 2023, subjective assessments are a complementary useful tool to understand impact of training especially in the context of disentangling the effect of individualization. Questionnaires found in Table 5.1 were proposed throughout the week (following schedule in Figure 5.1) in order to assess the evolution of the participant feeling.

5.2.9 Data analysis procedure

The data analysis procedure was structured around our three research questions. Initially, to address RQ1, we focused on visualization and metrics concerning learning trajectories. Next, RQ2 was evaluated using a three-level analysis of the cognitive battery. Lastly, RQ3 was examined using mixed-effects models on questionnaire responses. All the scripts used for the analysis are publicly available at https://github.com/flowersteam/zpdes_mot_results_analysis.

Demographic data

The demographic information, including age, gender, and prior experience, was automatically collected and sourced from the Prolific platform. To analyze continuous demographic variables, Bayesian t-tests were implemented using the Python Pingouin library (Valat, 2018). For categorical variables, Chi-squared tests were conducted to assess the distributional differences among the discrete demographic factors.

⁶<https://www.crayon.com/>

Table 5.1. Summary of Questionnaires used. Details of all questionnaires can be retrieve in Annexe 5.5.1

Questionnaire name	Objectives
Demographic data	Control for usual demographic information (age, gender, study duration...).
Video-Game Questionnaire	Control for video game player profile. https://www.unige.ch/fapse/brainlearning/vgq
Self-assessment of attention capabilities	Control for attention deficits. https://add.org/wp-content/uploads/2015/03/adhd-questionnaire-ASRS111.pdf
NASA-TLX	Study cognitive load evolution. Hart, 2006
User Engagement Scale (UES)	Study engagement evolution. O’Brien et al., 2018
Technology based experience of need satisfaction interface (TENS)	Study competence, autonomy and relatedness evolution. Peters et al., 2018
Situational Motivation Scale (SIMS)	Study type of motivation through week. Guay, Vallerand, and Blanchard, 2000
Learning Progress	Study the estimated learning progress through the week.

Pre-post cognitive performance analysis

As previously described, cognitive performance was assessed using a cognitive battery comprising seven tasks, each with multiple difficulty levels. To explore the impact of our training on cognitive performance (Research question 2), we segmented our analysis into three levels of granularity. First, we examined the performance evolution in the observed variable space; in other words, we assessed the metrics’ evolution for each task. While this initial analysis level is crucial for understanding the training’s impact on various metrics, it is challenging to comprehend the collective evolution of participants across different metrics. Second, we assessed performance evolution in a non-observed latent space. Through Principal Component Analysis (PCA), we aimed to explore performance evolution in a lower-dimensional space representing the cognitive skills necessary to perform the tasks in the battery. Third, we examined an aggregate score by averaging all metrics to derive a broad cognitive ability score. Overall, our objective is to provide multiple levels of analysis that offer different insights into the impact of our training.

Level 1: Observed Performance Analysis For the observed variables, performance metrics primarily included accuracy and reaction time, with some tasks incorporating additional metrics specific to their requirements, such as switching cost in task switching or false alarm rates in Go/NoGo or memorability tasks. We applied Bayesian estimation to analyze all metrics of interest, utilizing its ability to integrate prior knowledge of the task and tailor the model accordingly. Following the procedure described in Kruschke, 2013, we implemented this approach separately for each group (control and ZPDES) and for each testing session (pre-intervention and post-intervention). Consequently, we obtained the posterior distributions of the metrics of interest for each task, group, and session.

As detailed in Kruschke, 2013, these posterior distributions were used to compute the differences between groups (control and ZPDES) both before and after the intervention, and to evaluate changes within each group from pre- to post-intervention. All distributions (within-group changes, between-group differences at pre- or post-intervention, and between-group change differences) were then utilized to assess performance evolution.

From these distributions, two metrics were explored: the probability of effect and the Savage-Dickey Density Ratio (SDDR). The probability of effect is assessed by finding the maximum between $P(\text{Param} > 0 \mid \text{Data})$, which evaluates the probability that the posterior distribution of the parameter shifts to the right (above zero), and $P(\text{Param} < 0 \mid \text{Data})$, which evaluates it shifting to the left (below zero). This calculation thus identifies the probability of detecting any significant effect, regardless of its direction. The SDDR is employed to compare the estimated distribution with a null distribution, which signifies the absence of an effect. This method is highlighted by Wagenmakers et al., 2010 as being directly equivalent to computing the Bayes factor (BF) when the models being compared are nested and involve a scalar parameter of interest. For instance, in this work, when analyzing the posterior distribution of the difference between pre- and post-test scores for a specific group, the SDDR is used to compare the null hypothesis (i.e., no difference) against an alternative hypothesis (i.e., difference is not null). The use of BF, as delineated by Jeffreys, 1998, provides a graded scale of evidence: BFs between 1 and 3 indicate moderate evidence, BFs between 3 and 10 represent substantial evidence, and BFs exceeding 10 signify strong evidence in favor of the alternative hypothesis. For consistency in terminology, all references to SDDR results in the remainder of this paper will be designated as BF.

The two selected metrics offer complementary features. The decision to utilize the probability of effect was motivated by its independence from underlying assumptions and its ability to convey our confidence in observing an effect after collecting data, although it is insufficient for hypothesis testing. In contrast, the BF facilitates a comparison between the posterior distribution and a reference distribution representing the null effect. This method, however, relies on a subjective criterion to define the null effect. For better clarity, Box 1 provides an overview of the key concepts of Bayesian analysis.

Details of the models used for each task can be found in Appendix ???. All simulations were conducted using the Markov Chain Monte Carlo (MCMC) method, specifically employing the No-U-Turn Sampler (NUTS M. D. Hoffman and Gelman, 2011) implemented in the python library PyMC4 (Abril-Pla et al., 2023). We ran 4 chains of 10,000 iterations for each simulation, including 1000 burn-in iterations. As recommended in Kruschke, 2021b, convergence was systematically checked (observation of both sampled MCMC traces and metrics such as autocorrelation).

Level 2: Latent Space Analysis To evaluate the performance evolution at the latent factor level, we employed Principal Component Analysis (PCA) as a way to examine performance in a latent, low-dimensional space. In our data processing steps, we handled missing data by replacing any missing values with the average value of the corresponding variable. Additionally, we standardized the data by converting each data point into a z-score, which involves scaling the data so that it has a mean of 0 and a standard deviation of 1, ensuring comparability across different variables. PCA was conducted on 24 features, consistent with the prior work of Adolphe et al., 2022c (including a metric for the UFOV task), using a combined dataset of pre- and post-test results. Finally we adhered to the Kaiser criterion, retaining principal components (PC) with an eigenvalue greater than 1

and investigate differences between groups at pre-test and evolution within groups from pre to post test using respectively bayesian unpaired and bayesian paired sampled t-tests from the python library pingouin (Vallat, 2018).

Level 3: Aggregate Cognitive Scoring To evaluate performance on a single general performance score, we computed an aggregate of all observed metrics (same set of metrics used for the PCA). To make features evolve in the same direction we first took the inverse of reaction times metrics. Then, all metrics were converted to z-score and average. Difference of performance at pre-test and evolution from pre to post test was tested with bayesian unpaired and bayesian paired sampled t-tests.

Subjective questionnaire analysis For all questionnaires, due to the session count surpassing two, linear mixed models were employed to process the responses. These models included fixed effects for both groups and time, as well as a random baseline effect to account for individual participant variances. After evaluating various models, we selected one that achieves parsimony while still accommodating the random differences among participants. PyMC4 was utilized to fit these models, and comprehensive details regarding the model specifications can be found in the appendix, specifically in Section 5.5.1.

Box 1: Key Concepts of our Bayesian Analysis

Bayesian key ideas The Bayesian perspective seeks to shift the traditional emphasis from the likelihood of the observed data under a given model $p(data | model)$ — a cornerstone of the frequentist approach—to the probability of the model given the observed data $p(model | data)$. This paradigm enables a probabilistic understanding of model parameters that integrates prior knowledge with the empirical data. In the context of hypothesis testing, Bayesian analysis doesn't evaluate the likelihood of the data given fixed parameters, but instead focuses on the parameters of interest themselves.

Model example Consider, for example, the parameter of accuracy - designated as p - of the load blindness cognitive task. Accuracy, in this scenario, is quantified as the ratio of correct responses to the total number of trials. Within a Bayesian framework, p can be modeled as the probability parameter of a binomial distribution — the distribution that represents the number of successes in a sequence of independent experiments. The goal here is not merely to estimate a single point value of p but to calculate the posterior distribution of p . The posterior distribution reflects all potential values for p , where the probability density of each value is determined by how well it aligns with the observed data, updated according to our initial understanding as expressed in the prior distribution. In this scenario, prior information is incorporated by defining that without collecting observation, p follows a beta distribution with specific parameters. This prior choice is informed by the beta distribution's natural suitability for modeling probabilities (confined between 0 and 1). This prior choice allows for the incorporation of knowledge about the modeling situation and should be considered independently for each variable to be modeled. Parameter estimation occurs individually for each group at each time point, with each group having its distinct beta distribution for drawing p parameters during testing times. Figure 5.2 displays the fitted model. In the diagram, elliptical nodes represent random variables: shaded nodes indicate directly observed variables in our dataset, such as trial number of trials (n) or group (g), while unshaded nodes are parameters estimated through Bayesian inference, such as beta parameters.

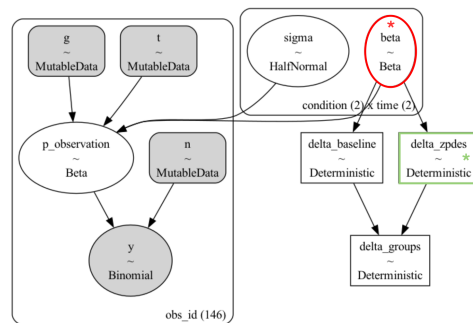


Fig 5.2. Illustration of the beta-binomial model

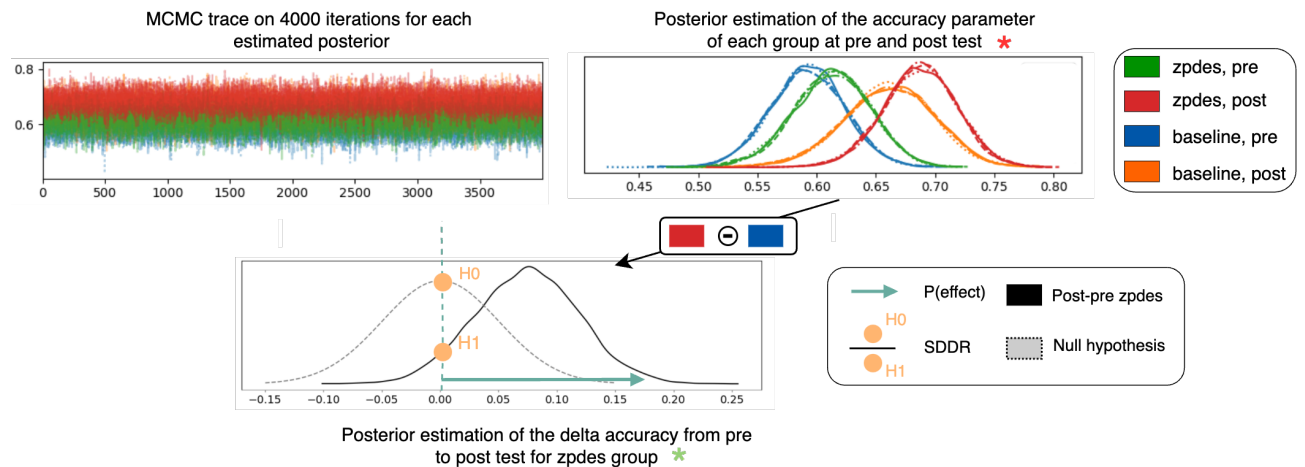


Fig 5.3. Illustration of the methodology used to derive the posterior post-pre evolution difference and the computation of the SDDR

Fitting model and use While the posterior distribution of this scenario can be analytically determined (cf “Front Matter”, 2015), most models require approximation methods like the Monte Carlo Markov Chain (MCMC). MCMC samples iteratively to approximate posterior distributions, as demonstrated in Figure 5.3 (left top row), for each group's distribution. Figure 5.3 (right top row), visualizes these distributions, comparing the ZPDES and control groups before and after intervention. Differences between estimated parameters are analyzed, with Figure 5.3 (bottom row) showing the delta between pre and post-test results in the ZPDES group. Two inferential heuristics can then be used: the probability of effect and the Savage-Dickey density ratio (SDDR), which evaluate the likelihood and size of significant changes, respectively, assessing the cognitive intervention's impact.

5.3 Results

5.3.1 Participants

A total of 72 participants were recruited via Prolific and assigned to either the control group ($n=35$) or the ZPDES group ($n=37$) through a pseudo-random process. Age comparisons between the groups showed no significant differences ($t(70.98) = 1.08$, $p = 0.28$, $BF = 0.39$). Similarly, a chi-square test indicated no significant variation in gender distribution across the groups ($\chi^2(1) = 0.36$, $p = 0.55$). Additionally, the proportion of ADHD profiles did not show significant differences between the groups ($\chi^2(1) = 2.29$, $p = 0.12$). Finally, an analysis of online study expertise, based on the number of prior Prolific experiments participants had engaged in, also found no significant group differences ($t(55.17) = -0.4$, $p = 0.66$, $BF = 0.26$).

5.3.2 RQ1- How do learning trajectories differ when using our individualized AI-based adjustment procedure compared to a traditional staircase method?

To examine the comparison of learning trajectories, we suggest segmenting the analysis to concentrate on three properties of a curriculum. The first attribute, denoted as **activity space exploitation**, concerns the strategy’s capability to fully utilize the entire activity space effectively, meaning choosing activities across the entire space while steering clear of infeasible areas. The second characteristic, referred to as the **sequential coherence** feature, examines the logical consistency of the suggested sequence. Lastly, the third characteristic assesses the **diversity** of the trajectory, questioning the variability within the proposed curricula. For the remainder of the manuscript, we refer to “activities” as the choice of a value for all five parameters (number of targets, speed, spacing between objects, tracking duration, and answer duration), which corresponds to a position in a five-dimensional space named the activity space.

RQ1.1, RQ1.2 - Activity space exploitation and sequential coherence Qualitative analysis of handpicked participants trajectories was a first step toward understanding differences between the AI based procedure and the staircase algorithm. Figure 5.4 depicts the relative time allocation across different types of activities and the activity transitions for the two best performing participants from each group. It is noteworthy that the participants experienced varying lengths of trajectory during the training sessions. This is because the sessions were time-bound (rather than being defined by a predetermined number of trials), enabling participants to engage at their own pace. Additionally, the duration of tracking was intentionally varied as one of the parameters to adjust the task’s difficulty level.

As expected, the trajectory of the control participant tended to cluster around particular regions of the activity space, a pattern also observed in additional trajectories presented in the Appendix. In contrast, the participant in ZPDES group experienced a more diverse set of different activities, notably progressing to more targets at a quicker pace during training. The pattern of transitions also varied markedly among participants. Governed by the staircase algorithm, control participants were required to achieve success at higher difficulty levels for all 4 parameters in S1 before advancing from n targets to $n+1$ (in S2), often resulting in a plateau in the variety of activities presented. In contrast, the

ZPDES algorithm navigated more rapidly through the number of targets and was adept at identifying niches of challenge and progress—as evidenced by the large dots at the 2-target level in Figure 5.4 (b), which are situated at particularly challenging values for speed, spacing, and other parameters.

In addition to visualizing trajectories, we conducted a quantitative analysis. Firstly, concerning RQ1.1, the increased exploration by the ZPDES group was further substantiated through a comparison between groups regarding the total number of distinct activities engaged in throughout the intervention. The findings indicated that participants in the ZPDES group engaged in a significantly wider variety of activities (mean=875.82) compared to the control group (mean=88.76, independent t-test difference between groups: $t(38.12)=17.19$, $p<0.001$, $BF\gg1000$). This outcome was expected, as the staircase method is more restricted, whereas the ZPDES algorithm can follow a more flexible predefined route of difficulty evolution. To control for different trajectory length explaining this, we looked at that the number of activity played and at idle time. Unpaired t-test revealed no significant difference between groups (trajectory length: $t(72.98)=0.92$, $p=0.36$, $BF=0.35$ and idle time: $t(69.54)=-0.9$, $p=0.36$, $BF=0.34$) revealing a similar duration of exposure for each group. Secondly, in relation to RQ2.2, we examined the coherence of the sequences of activities by analyzing the diversity of activities within a sliding window of 50 activities. Zpdes group showed again a larger number of different activities on this sliding window (mean=20.06) compared to control group (mean=10.38, with difference $t(38.12)=17.19$, $p<0.001$, $BF\gg1000$).

RQ1.3 Diversity In addition to participant-specific visualization, group-wide trajectory features were extracted. Figure 5.5, (a), compares the evolution through time of the proportion of participant practicing in each of the number of targets level. In line with the participant-specific analysis, it can be observed that the ZPDES group was able to practice on higher levels of number of targets and with higher variability in the activities being proposed. Chord diagrams of Figure 5.5, (b) further confirmed that ZPDES group experienced diverse transition. Additionally, It is noteworthy that most important transition in both ZPDES and control groups are within the same activity level. If this feature is natural for the staircase algorithm, it has been set up for the ZPDES algorithm to avoid to random trajectories and to make learning path more coherent.

RQ1- Summary To summarize our findings related to RQ1, the investigation into participants’ learning trajectories, from both quantitative and qualitative perspectives, revealed significant differences in three key characteristics evaluated. Regarding activity space exploitation, evidence showed that the ZPDES participants explored a larger portion of the activity space, indicating a more extensive investigation of potential activities. As a byproduct, in terms of trajectory variability, the ZPDES group experienced a wider variety of activities, highlighting a greater diversity in their learning journey. Finally, in terms of sequential coherence, both the ZPDES and staircase methods facilitated the development of logically structured trajectories, progressing from simpler to more complex activities.

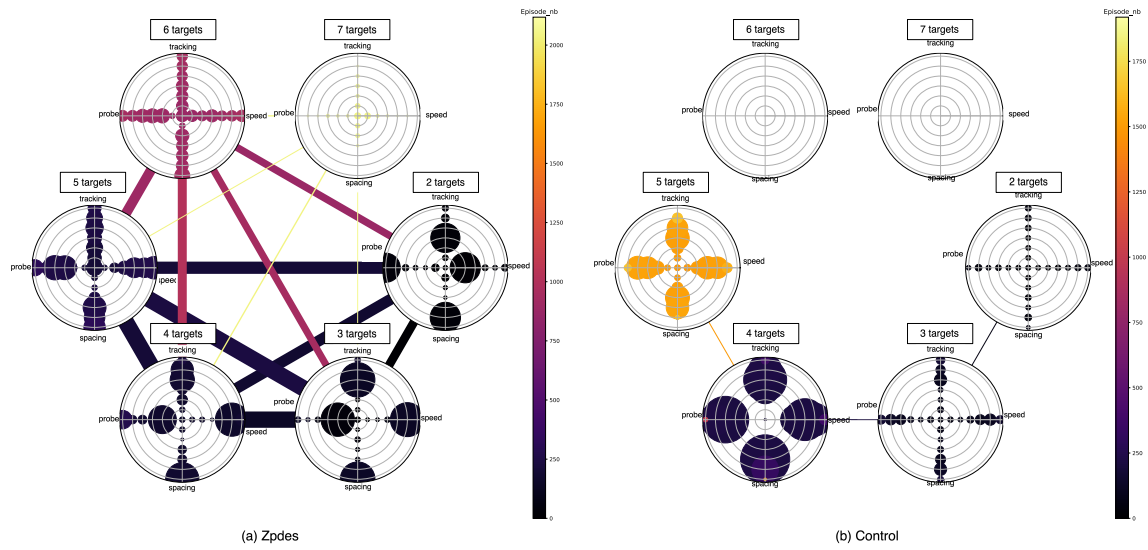


Fig 5.4. Two handpicked participant from ZPDES (a) and control (b). For each participant, the activity space is segmented into six discrete levels based on the number of targets, ranging from two to seven. For a given level of number of targets, each other parameter — speed, tracking duration, response time, and spacing — is symbolized by dots of varying sizes; the size correlates with the parameter’s relative frequency of proposition to participants. Larger dots indicate more frequent selection of that parameter value for practice. Transitions between different target difficulty levels are depicted by lines of varying thickness, where the thickness reflects the frequency with which participants experienced changes between levels. A thicker line denotes a more common transition. Color coding is utilized to indicate the sequence of the parameters’ first occurrence or the initial transition between levels; each color corresponds to a different index in the sequence of presentation or transition experienced by the participant.

5.3.3 RQ2- How does the individualized AI-based procedure affect the progression of task performance in the training activity and in overall cognitive performance as assessed by a broader set of cognitive tasks?

Evolution of performance in the training task Figure 5.6 illustrates the evolution of performance through the training for each of the 12 activities in the evaluation set. Significant performance improvement through time was detectable in 3/12 activities with $BF \gg 1$ and trends of improvements was found in 5 other activities (posterior probability of seeing an effect above 90% but with $BF < 1$). Ceiling and floor effect was present for respectively easiest and hardest activity. On all activities of the evaluation, no significant main effect of group was detected. Only one activity (four targets, high speed and low spacing i.e activity index 6 in the figure) presented a significant interaction effect with ZPDES improving at higher rate through time ($HDI=[0.00, 0.04]$, $BF=1.75$). Overall analysis showed an improvement of both groups on the trained task.

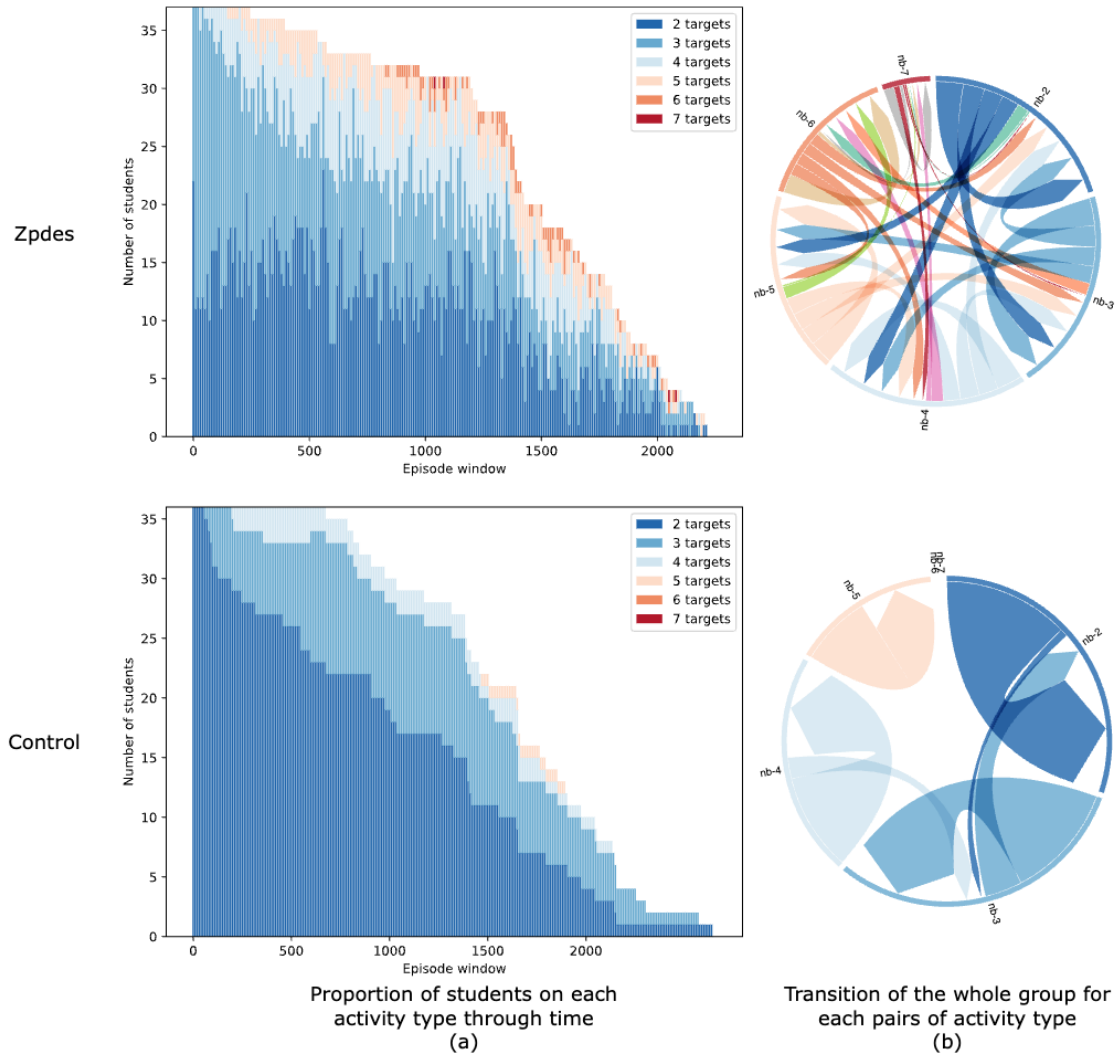


Fig 5.5. (a) Proportion of the population on each different number of targets (2 to 7 targets) through time. The total number of participants diminished over time as individuals experienced varying numbers of trials throughout their training.
 (b) Number of transition between activities in each group.

Evolution of performance in the cognitive battery: focus on tasks

Both groups showed improvements in accuracy measure for MOT, loadblindness and UFOV tasks Figure 5.7 illustrates the Bayesian analysis outcomes, highlighting differences in pre- and post-intervention performance for each group. In terms of accuracy metrics (a), an enhancement in cognitive abilities was noted in the MOT task for both groups, with ZPDES showing a notable increase (within group difference $P(\text{effect})=0.99$, $BF=63.54$) alongside the control group (within group difference $P(\text{effect})=0.97$, $BF=6.13$). Similar progress was detected in load blindness for both ZPDES (within group difference $P(\text{effect})=0.95$, $BF=3.36$) and the control group (within group difference $P(\text{effect})=0.91$, $BF=2.58$). The UFOV task (threshold metric) also showed a small improvement for both groups (ZPDES within group difference $P(\text{effect})=0.58$, $BF=1.28$ and control within group difference $P(\text{effect})=0.50$, $BF=1.21$)

Both groups showed no improvements in accuracy for enumeration, working memory and taskswitch tasks No credible post-pre within-group changes were

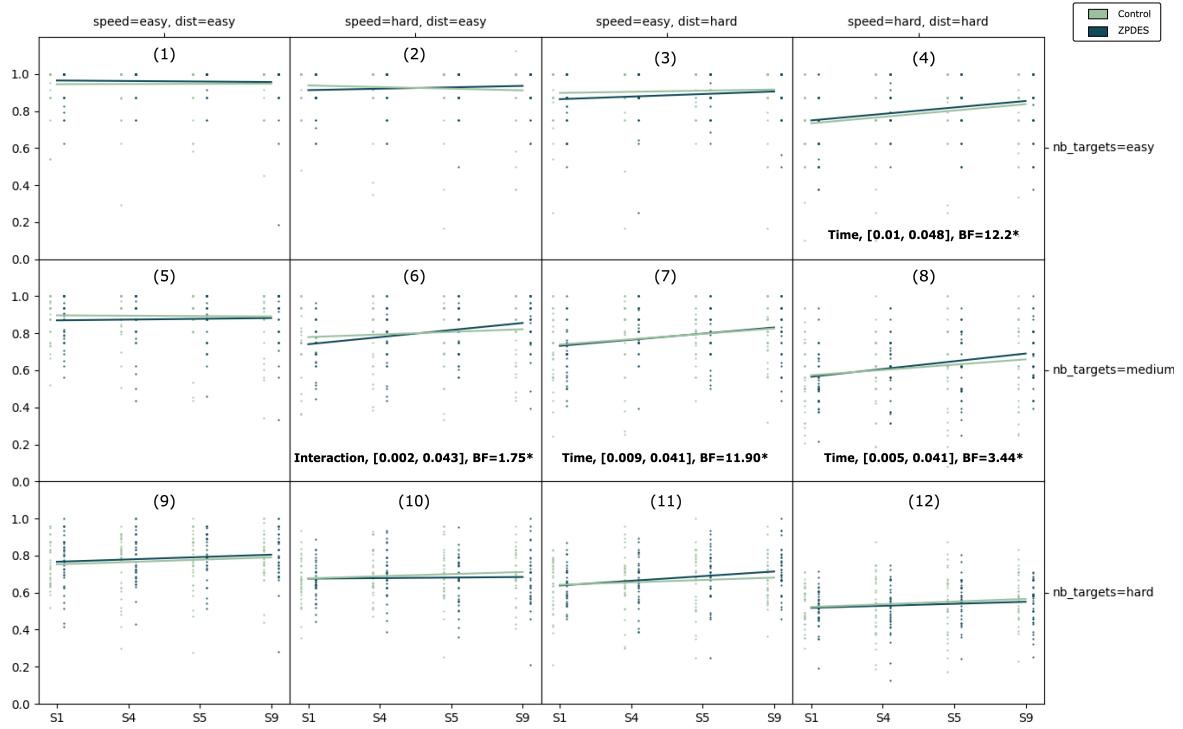


Fig 5.6. Performance evolution across the four evaluation sessions for each of the 12 activities in the testing set is organized such that rows correspond to a fixed number of targets (easy, medium, and hard, with 2, 4, and 6 targets respectively), and columns correspond to fixed pairs of speed and spacing.

observed in both groups for enumeration and working memory tasks. However, pre-tests between-groups differences revealed that the ZPDES group marginally outperformed the control group in the two tasks (enumeration between-group initial difference HDI=[-0.02,0.12], BF=1.71 and working memory between-group initial difference HDI=[-0.06,0.11], BF=1.05). These differences persisted in post-test (enumeration between-group final difference HDI=[-0.03,0.12], BF=1.47 and working memory post-test difference HDI=[-0.05,0.11], BF=1.20).

A notable pre-test difference was also observed in task switching (between-group initial difference HDI=[0.03,0.10], BF=1080), with only the control group exhibiting a weakly credible trend of progress (control within-group difference $P(\text{effect})=0.90$, BF=0.98). However, a ceiling effect noted in the pre-test for the ZPDES group (initial performance HDI=[0.96, 0.98]) may account for the lack of progression.

Only the ZPDES group showed improvements in accuracy for memorability and Go/No-go tasks Regarding memorability and go/no-go tasks, only the ZPDES group demonstrated significant enhancements. For memorability tasks, the ZPDES group improved (ZPDES within-group evolution $P(\text{effect})=0.94$, BF=2.75), unlike the control group, without any initial disparity between groups. In go/no-go accuracy, an initial pre-test difference was identified between groups (between-group initial difference=[0.01, 0.16], BF=9.36), which remained consistent at post-test (between-group final difference HDI=[0.05,0.18], BF=327.09), yet only the ZPDES group showed significant evolution ($P(\text{effect})=0.98$, BF=3.63) compared to a modest change in the control group ($P(\text{effect})=0.6$, BF=0.99).

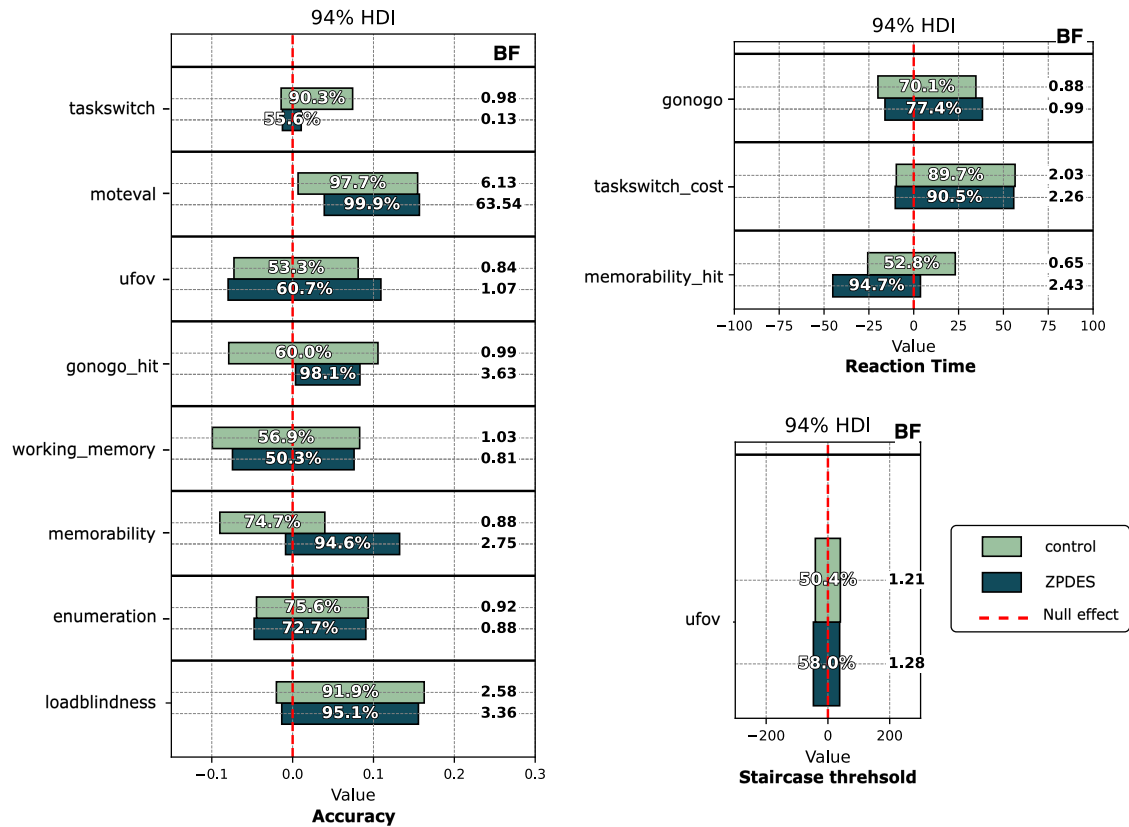


Fig 5.7. Pre-post evolution for accuracy (left) and RT (right) metrics. Probability in each bar represents the probability of effect. Corresponding BF is displayed right to the bars.

Mixed results in RT depending on tasks Regarding the reaction time (RT) metrics, enhancements were observed in both groups for the switching cost of Taskswitch (ZPDES within group difference $P(\text{effect})=0.90$, $BF=2.26$ and baseline within group difference $P(\text{effect})=0.89$, $BF=2.03$). As with the accuracy metric, in the memorability, only the ZPDES group demonstrated a significant performance enhancement (within-group difference $P(\text{effect})=0.94$, $BF=2.43$). For the Go/No-Go task, although a pre-test difference in reaction time performance was noted (RT Go difference= $[-54.99, -0.55]$, $BF=4.12$), this difference persisted at post-test (difference= $[-49.68, 3.89]$, $BF=2.71$) without significant evolution, indicating that the initial disparities remained stable over time.

In sum, the training impacted both groups in MOT, loadblindness, UFOV and switching cost. It only enhanced performance of ZPDES group in memorability (accuracy and RT) and go/nogo (accuracy). It did not reveal any strong evidence in enhancing performance in enumeration, working_memory.

Evolution of performance in the cognitive battery: focus on latent skills

Following Kaiser Criterion, 7 principal components were kept for latent factor analysis explaining 69.10% of variance as illustrated by Figure 5.8. Underlying constructs hold by each factor is further discussed in section 5.4. For all principal components, no difference were found between groups at pre-test with all BF being smaller than 0.5 (Table 5.3). Regarding pre-post evolution, differences between groups were observed. First, both groups showed an highly probable improvement from pre to post test on PC3 with $BF_{01}=73.58$

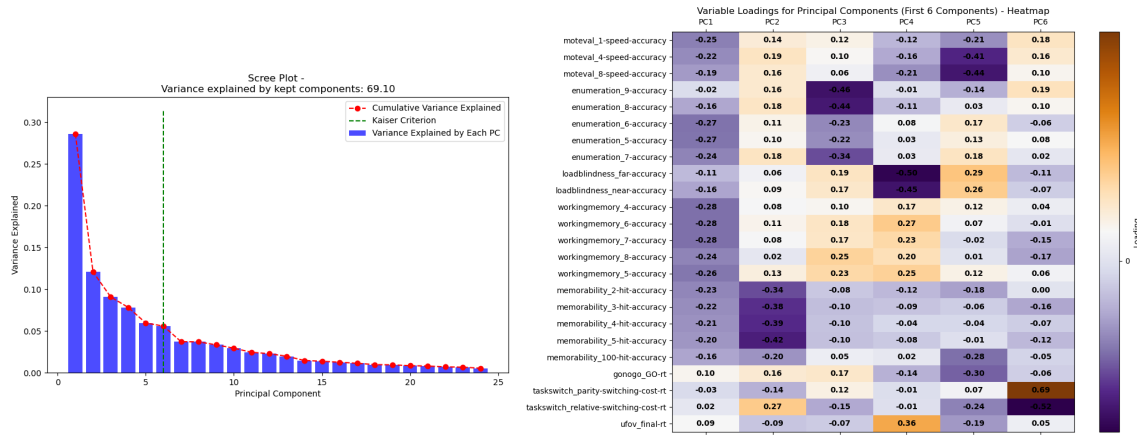


Fig 5.8. PCA results: scree plots (left) and Principal Components contribution (right)

and BF01=17.11 for ZPDES and control respectively. Then, improvements were observed only in the ZPDES group across two principal components namely PC0 (BF01=7.26) and PC4 (BF01=3.97), with no corresponding changes in the control condition. In sum, ZPDES group improved performance on more latent principal component than the baseline group (3 vs 1).

Component	Group	Pre-test Difference				Post-Pre Evolution			
		d	BF01	t-stats	p-value	d	BF01	t-stats	p-value
PC1	ZPDES	0.18	0.317	-0.793	0.43	0.34	7.265	2.970	0.005
	baseline	0.14	0.42	1.368	0.180	0.14	0.42	1.368	0.180
PC2	ZPDES	0.152	0.29	0.650	0.51	0.066	0.189	-0.370	0.714
	baseline	0.41	0.84	-1.857	0.072	0.41	0.84	-1.857	0.072
PC3	ZPDES	0.266	0.42	-1.136	0.260	0.045	0.185	0.311	0.758
	baseline	0.083	0.23	0.734	0.468	0.083	0.23	0.734	0.468
PC4	ZPDES	0.123	0.273	0.526	0.600	0.50	73.582	3.917	0.000
	baseline	0.43	17.11	3.341	0.002	0.43	17.11	3.341	0.002
PC5	ZPDES	0.254	0.40	1.088	0.280	0.487	3.97	2.694	0.011
	baseline	0.329	0.856	1.870	0.070	0.329	0.856	1.870	0.070
PC6	ZPDES	0.155	0.236	-0.790	0.434	0.155	0.236	-0.790	0.434
	baseline	0.232	0.366	1.249	0.220	0.232	0.366	1.249	0.220
Global Score	ZPDES	0.381	20.50	3.41	0.001	0.381	20.50	3.41	0.001
	baseline	0.19	1.12	2.034	0.049	0.19	1.12	2.034	0.049

Table 5.3. Corrected summary of pre-test differences and post-pre evolution for each principal component and group, with significant p-values in bold for improved readability.

Evolution of performance in the cognitive battery: focus on global score No significant difference between groups was observed at pre-test (BF01=0.56). However, both groups showed improvement from pre-test to post-test (ZPDES group: BF=20.50; control group: BF=1.12). While, at post-test, the difference between the groups was not significant (BF=0.332, t=0.8, p=0.39), the effect size of evolution was greater in the ZPDES group (d=0.381 vs d=0.19).

RQ2- Summary To summarize our findings related to RQ2; the training enhanced cognitive performance for both groups, as evidenced at multiple levels: our global score, dimension PC3 of PCA, and at the task level for MOT, loadblindness, and switching cost. Additionally, it specifically improved performance in the ZPDES group, as observed in the latent level analysis with dimensions PC1 and PC6, and at the task level for task memorability and gonogo.

5.3.4 RQ3- How does our individualized AI-based procedure affect participants' self-reported experiences?

NASA-tlx Most components of the NASA-TLX, except for physical demand, displayed a slight but consistent decrease over time, indicating a general reduction in cognitive load (Mental Demand: HDI=[-0.38, -0.15], BF=333; Temporal Demand: HDI=[-0.28, 0.03], BF=0.26; Performance: HDI=[-0.35, -0.03], BF=50; Effort: HDI=[-0.33, -0.08], BF=2.27; Frustration: HDI=[-0.333, -0.023], BF=0.86).

A clear group effect was evident particularly in Mental Demand and Effort, where the ZPDES group showed higher ratings compared to the control group, as indicated by positively skewed posterior distributions (HDI=[-0.112, 1.945], BF=2.22 and HDI=[-0.11, 1.97], BF=2.27, respectively). In contrast, Physical Demand was lower in the ZPDES group (HDI=[-2.11, -0.07], BF=4.14).

Furthermore, a significant Group*Time interaction was observed within the Performance component (HDI=[0.10, 0.53], BF=6.67). This reflects a distinct pattern where the ZPDES group, starting with lower self-ratings, demonstrated more pronounced improvements over time. Although an interaction was noted in the Performance component, indicating a possible Group*Time interaction across the broader Load Index, the evidence for such an interaction remains weak (HDI=[-0.070, 1.025], BF=1.02). This suggests that any potential interaction effect in the general load index might primarily stem from the interaction found in the Performance component.

SIMS Regarding the type of motivation throughout intervention, a significant group effect was present across all components, with the ZPDES group scoring lower in Intrinsic Motivation (HDI=[-1.15, -0.27], BF=18.41), Identified Regulation (HDI=[-1.27, -0.38], BF=107.76), External Regulation (HDI=[-1.242, -0.361], BF=138.69). The Self-determination index as described in Annex ??; was consistent with this observation with ZPDES group scoring lower than the control (HDI=[-3.42, -0.37], BF=11.82).

Additionally, there was a significant decrease over time for both groups in Amotivation (HDI=[-0.136, -0.028], BF=1.48), and a notable reduction in External Regulation as well (HDI=[-0.129, -0.026], BF=1.24). Intrinsic motivation and internal regulation showed no significant time effect. No significant interaction Group $\times$ Time was reported.

UES Across all metrics of the user engagement scale, a significant group effect was consistently observed, with the ZPDES group scoring lower across all dimensions of the questionnaire. The group effects showed a notable lower level in Focused Attention (HDI=[-0.69, -0.01], BF=1.25), Perceived Usability (HDI=[-0.84, -0.19], BF=14.11), Aesthetic (HDI=[-1.07, -0.32], BF=53.76), and Reward (HDI=[-0.697, -0.061], BF=2.45). The overall engagement score also reflected a significant lower level for the ZPDES group (HDI=[-0.70, -0.20], BF=19.75). All components showed strong evidence with Bayes Factor ranging from 1.25 to 100. Aesthetic was the only component to exhibit a negative trend over

time, though with a Bayes Factor below 1 (HDI=[-0.11, -0.01], BF=0.7), suggesting weak evidence for this trend. No interaction effects were reported.

TENS For the TENS questionnaire, a significant group effect was found for the Autonomy component where ZPDES group experienced a lower sense of autonomy (HDI=[-0.67, -0.07], BF=2.41). There were no substantial group effects for Competence or Relatedness. In evaluating the effects over time, a discernible decrease in perceived Competence was observed (HDI=[-0.08, -0.02], BF=1.48). No significant time effects were noted for Autonomy or Relatedness. No interaction Group $\times$ Time were reported.

Feeling of Learning Progress For the learning progress questionnaire, a main effect of the group was observed for the "Hard Feasible Zone" component (HDI=[-1.084, -0.136], BF=4.34), suggesting that participants in the ZPDES group self-reported that the number of targets they could track with difficulty was lower than that of the control group. Regarding a main effect of time, all components posterior distributions showed a possible slight positive shift (difficulty expectation: mean=0.046, easy feasible zone: mean=0.046, hard feasible zone: mean=0.02). However, the HDIs for these effects all contained 0 and the Bayes Factors were low (difficulty expectation: BF=0.09, easy feasible zone: BF=0.13, hard feasible zone: BF=0.04), indicating that while there is a suggestion of a possible increase over time, the evidence is not strong and there is substantial variability not accounted for by the time effect alone. No significant interaction Group $\times$ Time was reported.

RQ3 - summary To summarize the findings related to RQ3; initially, both groups exhibited similar responses across various components of the questionnaires. There was a reduction in cognitive load (Nasa-TLX), amotivation, and external regulation (SIMS), as well as perceived competence (TENS). Subsequently, a main effect of group was observed, with the ZPDES group experiencing higher levels of effort and mental demand, and lower scores in self-determination index (SIMS), engagement (UES), initial competence (LP), and autonomy (TENS). Finally, a group*time interaction was found in perceived performance, where the ZPDES group demonstrated greater self-perceived improvements than the control group.

5.4 Discussion

In this study, we compared the effectiveness of an AI-based individualized procedure with the traditional staircase method in managing task difficulty within a cognitive training utilizing the MOT task. Our initial research question focused on identifying differences in the curricula of each group. Through the exploration and visualization of learning trajectories, we observed notable differences between the two strategies. Participants in the ZPDES group engaged with a wider range of the activity space, encountering a greater variety of tasks. This was evident from multiple visualizations that also confirmed both methods effectively escalated task difficulty from simpler to more challenging levels. Our second research question investigated whether differences in trajectories resulted in varying impacts on performance. Initially, we assessed the training's impact on performance within the training context itself and noted improvements in MOT performance for both groups, with no significant differences between them. Subsequently, we evaluated the training's

impact on a cognitive battery comprising seven tasks that varied in their proximity to the MOT task. While both groups showed improvement in MOT, load blindness, UFOV, and switching costs only the ZPDES group showed enhancements in tasks such as memorability and go/no-go. This broader improvement was further substantiated in a latent space derived through PCA, where the ZPDES group improved on more cognitive components than the control group (3 PC for ZPDES vs. 1 for the control). Finally, our last research question explored the impact of training on subjective measures assessed through questionnaires. We found evidence that cognitive load decreased through time, an indicator of learning efficiency. However, differences emerged between groups: the ZPDES training was perceived as more effortful than that in the control group. Additionally, engagement, sense of autonomy, and motivation scores were lower in the ZPDES group. Notably, despite generally lower self-assessed competence (e.g., tracking fewer targets with difficulty), the ZPDES group reported a greater sense of performance improvement over time (Group $\times$ Time interaction) than the control group. This overview sets the stage for a detailed examination of each research question, analyzing the specific impacts of our training methodologies on cognitive performance and subjective experiences.

RQ1 - Training trajectories The staircase algorithm is an easy-to-implement procedure to determine threshold such as performance level in a learning context. Originally used in psychophysics (Dixon and Mood, 1948; Treutwein, 1995), it offers an efficient way to gradually move a learner from easy to hard task. By construction, the staircase algorithm generates a predefined learning path where there can be a unique route of progress. By contrast, the ZPDES algorithm makes less assumptions about the learning path and generate very different trajectories personalized to participants (as highlighted in Figure 5.5). Given that, while the 2 groups encountered different learning paths, both progressed on the intra-training evaluations or on the cognitive battery pre-post assessment. As such, we showed in this work that a variety of learning paths is available to enhance cognitive performance of trained or related tasks.

Additionally, while the staircase method effectively scaffolds learning (F. Joessel, 2022b), it is limited by focusing on a singular learning objective at any given time (i.e., mastering the highest step of the staircase with 75%) which results in limited variability in the learning curriculum. In contrast, participants in the ZPDES were offered activities across various progress niches (5.4), engaging in a wider array of tasks. Previous studies have indicated that variability in training is crucial for enduring learning in motor control training (Sungeelee et al., 2024, Shea and Morgan, 1979), to enhance attention and engagement (Raviv et al., 2022), and to regularize training to prevent overfitting in machine learning scenarios (C. F. G. D. Santos and Papa, 2022). Therefore, it is likely that the broader improvement observed in the ZPDES group is attributable to the increased diversity and variability in the learning trajectory.

Another key observation from the analysis of learning trajectories is the significant presence of plateaus in the learning paths of the control group participants. These plateaus indicate that participants in the control group faced extended periods where the tasks were excessively challenging, resulting in prolonged duration without learning progress. This mismatch between task difficulty and participant capacity is a desirable property (Lövdén et al., 2010b). However, this behavior presents two disadvantages: firstly, the plateaus result in a limited range of tasks, as participants spend extended periods on the same challenging activities (oscillation between achievable and unfeasible activities); secondly, the swift progress observed immediately following these hard zones implies that

the increase in task difficulty is not always gradual. Notably, after participants overcame these difficult phases, the subsequent activities were found to be less challenging, enabling rapid advancement through the staircase. This last observation illustrates the challenge of defining an adequate prior performance space (as proposed in previous work Diard et al., 2010) for this type of activity. Moreover, our results confirmed that the performance was impacted by the mutual evolution of the difficulty parameters used (as in Alvarez and Franconeri, 2007b or A. O. Holcombe and Chen, 2012).

Conversely, the ZPDES algorithm, while facing similar issues was more flexible in the way it handled multiple parameters at the same time. The use of a model of ZPD with the objective to find learning progress is an alternative to linear staircase procedure that provides more variability and avoid roadblocks in learning objectives. Here, the multi-armed bandit algorithm also offers a more data efficient strategy in comparison with other RL-based strategies that usually consider relatively small activity spaces (Zini et al., 2022, Wilms, 2011, Rathnayaka et al., 2021, Tsiakas et al., 2018).

Several difficulty MOT parameters In the context of MOT task training, the ability to simultaneously manage multiple parameters is critically important, as the task can be adapted to engage various perceptual and cognitive processes. Highlighted by Alvarez and Franconeri, 2007b, multiple parameters can collectively affect key performance-related properties. For instance, the number of objects, their speed, and the tracking duration all impact the number of interactions (i.e., or repulsion) between object, points of heightened attention load during the tracking. Therefore, manipulating multiple parameters could be crucial for enhancing training outcomes (chapter 3).

Previous studies, such as Eayrs and Lavie, 2018a, have established a correlation between MOT task performance and perceptual capacity. This suggests that adjusting various parameters such as object spacing, speed, or visual features—essentially controlling the perceptual load during MOT training—could maximize learning gains and enhance perceptual capacity. This concept extends beyond merely boosting perceptual capacities to potentially improving other cognitive factors (as identified in additional works like Eayrs and Lavie, 2018a; Legault et al., 2013a; Parsons and Faubert, 2021; Romeas et al., 2016c; Skogsberg et al., 2015a; Tullo, Guy, et al., 2018) through the manipulation of different sets of parameters.

Thus, regardless of the underlying mechanisms driving performance changes, manipulating several parameters simultaneously appears to be a promising strategy for achieving broader learning gains in cognitive training generally and MOT training specifically. Currently, most MOT-based training focuses solely on manipulating objects speed (Legault and Faubert, 2012b; Parsons and Faubert, 2021; Parsons et al., 2016b; Tullo, Guy, et al., 2018). This work represents a preliminary step toward integrating more parameters into experimental designs, and it is hoped that future research will continue in this direction.

RQ2 - Impact of training on objective performance Our results demonstrated that both groups improved over time on the MOT evaluations conducted during training. Improvement in MOT performance through training was anticipated, as observed in several previous studies (Parsons et al., 2016b, Romeas et al., 2016c). Broadly considering all intra-training evaluation tasks, both groups exhibited similar improvements, except for one evaluation task (with four targets, high speed, and high spacing) where the ZPDES group showed greater performance enhancement. Given the between-group differences observed in the pre-post cognitive battery, the small difference between groups in the

intra-training MOT evaluation was unexpected. However, the small number of activities in the testing set (12), coupled with the limited number of repetitions (4 for each), might account for the negligible difference between the groups.

The performance improvement measured in the pre-post battery of this study aligns with previous research on MOT training impact with diverse perceptual, attentional and memory tasks (Parsons and Faubert, 2021, Legault and Faubert, 2012b, Corbin-Berrigan et al., 2018). First, the aggregate performance score showed an improvement from pre to post tests for both groups. Second, this study provides evidence that both groups were able to enhance their performance on related but untrained tasks such as loadblindness, UFOV and taskswitch tasks. However, unlike previous findings, this study did not find evidence of performance enhancement in enumeration or working memory tasks (Parsons et al., 2016b).

Latent factor analysis indicated that both groups improved on a single latent factor (PC3). Detailed examination of the loadings of this component revealed strong negative contributions from enumeration tasks and the memorability task, as well as positive contributions from MOT, load blindness, and working memory tasks. Given these positive and negative contributions, this PC might represent a cognitive factor related to dynamic visual information processing and attentional resource allocation. Indeed, the positive contributors (MOT, workingmemory tasks, and loadblindness) all involve managing and manipulating information in a dynamic and often spatially complex environment, requiring sustained attention and working memory. Conversely, the negative contribution from tasks like memorability and enumeration suggests that this PC is less associated with static visual memory and the rapid, effortless perception of quantities, which are more automatic, less resource-intensive aspects of cognition. As highlighted in the previous section, this enhancement in a specific "skill" or cognitive dimension can be discussed concerning participants' training trajectories. For control participants, most of their training focused on activities with a low number of targets but with very high speed, low spacing, and extended tracking duration. This focus helps explain the performance enhancement, particularly in resource-consuming and long-lasting visual dynamic processing tasks.

Our study also found that only participants in the ZPDES group showed improvements in the memorability task (both accuracy and reaction times) and the go/nogo (hit accuracy) task. Once again, this outcome may be considered in the context of the specific training trajectories provided to these participants. For instance, enhancements in memorability might stem from exposure to a larger number of targets at very low speeds and varied tracking durations. Indeed, extended tracking durations could place higher demands on memory capabilities or necessitate strategies that involve longer retention periods. A potential explanation for the lower effect in the control group is that high tracking duration was always coupled with a high load on other parameters, such as high speed or low spacing between objects. Furthermore, the go/nogo tasks, which demand inhibition, might be enhanced by trials featuring high speeds and minimal spacing between objects. Once again, even though participants were confronted with high-speed activities, the load was never concentrated on this single parameter. Several theories or model of transfer could potentially explain these results. First, as suggested by Gathercole et al., 2019, specific trial conditions might challenge participants with tasks that require new cognitive routines, routines that could be reused and potentially explain the transfer of skills to other tasks (a phenomenon less observed in the control group in this study). Another possible explanation is derived from the "learning to learn" hypothesis, suggesting that training enhances the ability to learn faster in new situations (Bavelier, Green, et al., 2012). A key computational aspect of this hypothesis is the enhancement of probabilistic inference,

meaning an improved consideration of previously collected data. This enhancement has been particularly observed in contexts such as action video games or attentional and perceptual training (Bejjanki et al., 2014; Gozli et al., 2014; von Bastian et al., 2022a; R.-Y. Zhang et al., 2021a). Consequently, although further research is required to confirm this, it can be posited that the diverse exposure inherent in the ZPDES group's trajectories may have amplified this feature. As such, these hypotheses highlight the significance of the difficulty adjustment procedure, where multiple difficulty parameters significantly influence training outcomes.

Improvements in PC1 and PC6, exclusive to the ZPDES group, with no comparable effects seen in the control group were also observed. PC1 is interpreted as a general factor influencing overall performance in our battery. It is noteworthy that, when using an aggregate score (a linear combination with equal weight for each task), both groups showed enhancements from pre- to post-test. The exclusive enhancement in PC1 for the ZPDES group (a linear combination with varying weights maximizing projected variance) suggests that the ZPDES group experienced more comprehensive improvements than the control group. Upon closer examination of PC6, which is heavily influenced by MOT tasks under high perceptual load and cognitive flexibility from task-switching tasks, this component may be interpreted as reflecting 'dynamic divided attention'. The specific training trajectories followed by the ZPDES group could, again, account for these enhancements. For example, tasks with a high perceptual load (i.e., a large number of targets) at various speeds (often at very low speeds) necessitate increased divided attention, potentially facilitating transfer to tasks that also demand this type of attention.

RQ3 - Self-reported subjective feeling As noted in Adolphe et al., 2023, incorporating questionnaires is an effective method for assessing the impact of training. For instance, this approach is particularly beneficial for evaluating metacognitive experiences, which are known to influence the regulatory functions of cognitive processes involved in training activities (Efklides, 2006b). Metacognitive experiences, such as individuals' perceptions of their own learning and cognitive efforts, provide valuable insights into how learners monitor and adjust their strategies based on their subjective assessments of cognitive load and task difficulty. In this study, we corroborated the "objective" learning progress by probing every week different meta-cognitive dimensions and documented a decrease in self-reported cognitive load over time.

Significant differences were also evident between the groups. The mental demand and effort were consistently higher in the ZPDES group, which suggests that the participants experienced more challenging learning paths, indicating that this type of training may result in more effortful trajectories. Since the learning benefits appeared to be greater in the ZPDES group, as measured by cognitive battery tests, this raises questions about the staircase algorithm's ability to appropriately adjust the difficulty level and what that optimal difficulty level might be. Moreover, this observation can be connected to Cognitive Load Theory (CLT) (Paas et al., 2004, Tricot, 1998), which distinguishes between intrinsic cognitive load—related to the inherent complexity of the training task regardless of its presentation—and extraneous cognitive load (Sweller, 2011), which concerns the manner in which the task is presented (instructional designs). CLT suggests that an optimal cognitive load (maximization of intrinsic part associated to a minimization of extrinsic part) promotes learning and that appropriate variability in training increases intrinsic load, which can enhance learning (provided that sufficient working memory resources are available Paas and Van Merriënboer, 1994).

In addition to cognitive load, differences in engagement and motivation were particularly noteworthy, with the ZPDES group showing lower levels compared to the control group. These results were unexpected and contradict the premises of the ZPDES algorithm. According to the learning progress hypothesis (Oudeyer et al., 2016b), it was anticipated that exploration guided by learning progress would enhance intrinsic motivation. Thus, these findings prompt questions about the effect of cognitive load on motivation levels. It is possible that higher cognitive loads could adversely affect intrinsic motivation levels. Initial discussions in CLT considered motivation primarily as a precursor of engaging in training, suggesting that sufficient motivation is necessary to engage in an effortful training. As such, learners would make anticipatory judgments about the perceived attributes of the instructional design and would invest mental effort accordingly (Salomon, 1983). However, this perspective has been challenged in previous research (cf Poupard et al., 2022 or Feldon et al., 2019 for a review). For example, Feldon et al., 2018 explored the effects of manipulating extraneous cognitive load on self-efficacy (i.e., motivation) and found that a group with well-designed instructions (i.e., lower extraneous cognitive load) exhibited higher self-efficacy in post-tests. While these results are intriguing, the nature of the relationship between motivation and cognitive load remains unclear. It is conceivable that the relationship is bidirectional: mental effort is invested when there is sufficient motivational belief, but if results are inadequate, this mental effort could decrease motivation. Further research is necessary to understand how these observations might be integrated into ITS. For instance, prior research Ten et al., 2021a has suggested considering success rate, a proxy for cognitive load, as a complementary objective in activity selection.

Furthermore, despite participants in the ZPDES group experiencing higher objective learning benefits after training, their subjective perception of learning progress was lower or similar as of the control group. This discrepancy may align with the significantly lower ratings observed in the ZPDES group for the competence component, or with the question related to tracking the maximum number of targets with difficulty in the LP questionnaire. These lower ratings are particularly noteworthy given that a comparison of objective performance between groups in the intra-evaluation did not reveal any differences. It is important to note that MOT's activity space is characterized by significant nonlinearities in difficulty evolution. For instance, performance has been observed to decrease quadratically and cubically with increases in either the number of targets or the tracking duration (Oksama and Hyönä, 2004c). Therefore, it is plausible that ZPDES exploration occasionally proposed unfeasible activities or open certain activities too quickly, which might have led ZPDES participants to poorer subjective estimations of their learning progress.

Additionally, it is important to note that although the ZPDES group self-rated their competence lower, an interaction effect on the performance component was observed where ZPDES participants self-reported an increase in performance throughout the training, which was not observed in the control group. This implies that although ZPDES participants may have underestimated their competence (i.e., their general abilities independent of recent tasks), they were nonetheless able to recognize a genuine improvement in performance (i.e., their success in specific sessions) over time. This finding highlights a potential mismatch between self-perceived competence and actual performance gains. This discrepancy can be linked to the previously unobserved connection between motivation and objective progress, particularly evident in the ZPDES group. It is conceivable that much of the progress achieved in MOT tasks involves the automatization of processes used in these tasks, progress that may not be accessible to participants' metacognition. Therefore, even when objective progress is evident, subjective recognition of this progress may be lacking. An alternative explanation would be that the adaptive procedure - by always proposing

challenging activities - limits the number of positive feedbacks received by participants which hinder a correct estimation of competence (as suggested in Sansone, 1986, Deci et al., 1999, Hung and Seitz, 2014a). Additional research is necessary to disentangle the various influences of all dimensions, but this could be a crucial aspect to consider in the development of future difficulty adjustment algorithms.

Limitations and perspectives This work is currently in its feasibility stage, and there are several areas that could be enhanced in future research. As indicated in the previous section, efforts can be made to boost motivation and engagement. For example, all participants in our study began at the same initial ZPD. Drawing on research such as that by van der Velde et al., 2021 and Pliakos et al., 2019, which addresses this 'cold start problem,' could help more swiftly place participants within an optimal initial ZPD. Additionally, to better manage cognitive load levels, we might incorporate strategies from earlier studies that suggest methods to better recognize mastery (Diard et al., 2010, Watson, 2017, Shabana et al., 2022) or address the forgetting of mastered activities (Mu et al., 2018). Moreover, given that subjective estimation of learning progress was a potential issue, involving participants more directly in the activity selection process, as demonstrated by B. Clément et al., 2024, could be beneficial. Lastly, while individualized training primarily focuses on adjusting task difficulty, other methods such as adaptive storytelling and feedback could also be utilized to enhance motivation (Conati and Manske, 2009).

The training and its assessment could also be improved. For instance, incorporating audio feedback into the MOT task has been shown to positively influence motivation and engagement. Furthermore, to alleviate the monotony associated with repetitive task engagement, the introduction of dual tasks during tracking or the integration of various cognitive training tasks could be a viable approach. The Multiple Identity Task (J. Li et al., 2019b; Oksama and Hyönä, 2004b), the Multiple Object Avoidance Task (Mackenzie et al., 2022) or other variants of the MOT introducing a dual task (F. Joessel, 2022b; Nyquist et al., 2016b), might also serve as effective training tasks when used in conjunction with the ZPDES algorithm, which is able to handle high dimensional activity space. In terms of methodological settings, future studies should include follow up assessments to study potential long term impact. Additionally, ecological transfer could be further assessed with daily life questionnaires (Saba et al., 2021). Finally, considering the results obtained on a healthy young adult population; it would be intriguing to propose such interventions in older populations, as it has been demonstrated that the extent of near transfer varies significantly across age groups (Von Bastian et al., 2013).

5.5 Conclusion

In this study, we compared a multi-armed bandit-based procedure with a traditional staircase procedure to evaluate their efficacy in adjusting the difficulty of cognitive training using the multi-object tracking task. This research, pioneering the simultaneous assessment of five parameters to develop personalized training curricula for each participant, demonstrated that both experimental groups benefited, as evidenced by performance improvements on the training task and cognitive assessments. Furthermore, our procedure, which was guided by learning progress and a model of ZPD, offered a wider variety of tasks, identified multiple areas of progress, and consequently, yielded larger learning gains than the control group at the post-test. This feasibility study lays the foundation for future research involving different tasks or participant demographics.

Appendix

ZPDES - details and hyperparameters

Formalism

As mentionned in the high level description of ZPDES in 5.2; the algorithm learns about the student needs through an interactive learning loop. After proposition of a new activity, the algorithm updates the estimated pedagogical benefit of that activity and update the model of the zone of proximal development. The interactive loop is described in 7. To fit with the classical taxonomy of Intelligent Tutoring Systems, main features of the algorithm are described as our domain, pegagogical and student models in the following sections.

Algorithm 1 Interactive learning loop

Require: : A , a discrete finite set of all possible activities

Require: : R^{ZPD} , a set of ZPD rules

- 1: Initialize G^e a set of experts according to R^{ZPD}
 - 2: **while** learning: **do**
 - 3: $a \leftarrow \text{sample_activity}(G^e)$ ▷ Generate an activity
 - 4: $s \leftarrow \text{answer from learner}$ ▷ Propose it to the participant
 - 5: $G^e \leftarrow \text{update_quality}(G^e, a)$ ▷ Update experts involved in last activity
 - 6: $G^e \leftarrow \text{update_ZPD}(G^e, R^{ZPD})$ ▷ Update ZPD
 - 7: **end while**
-

Domain model: An activity a is a I-dimensional vector where a_i is a parameter. In the MOT training, a_i can be the number of target, the speed, etc.. For each parameter a_i , there are J_i possible values $a_{i,j}$ (continuous parameters are discretized). The cardinal of the set of all activities is therefore equals to $\prod_{j=i}^I J_i$.

Pedagogical model: Add this: As such the expert e_i is a J-dimensional vector that represents the quality of all tracked values. In the high level explanation of the algorithm in ??, we presented a simpler version by restricting to a single pedagogical expert. In practice, each parameter is associated with at least one pedagogical expert e_i . The main objectives of a pedagogical expert are to 1) track the learning progress of each parameter value $a_{i,j}$ and 2) to randomly sample the value with highest learning benefits. At any time t , if a value $a_{i,j}$ is proposed to a participant, the pedagogical expert stores the results of that activity in an history vector $h_{i,j}$ and updates the quality of the associated value with the update rule:

$$e_{i,j}^t \leftarrow \beta e_{i,j}^{t-1} + (1 - \beta) r_{i,j}^t \quad (5.6)$$

The associated reward $r_{i,j}$ used in the update rule can take many forms and is further explained in the section "student model". Then the pedagogical expert updates the quality of all values by adding an exploration term:

$$e'_{i,j} = (1 - \gamma) e_{i,j} + \gamma \xi_u \quad (5.7)$$

Where ξ_u is randomly sampled from a uniform distribution in $[0,1]$. However, in practice, to ensure avoid time consuming hyperparameter tuning (to keep the same order of magnitude) of γ , equation 5.7 is computed as:

$$e'_{i,j} = e_{i,j} + \gamma \sum_j e_{i,j} \quad (5.8)$$

This quality update is further described in 2.

Algorithm 2 Update experts quality estimations

Require: G^e , a graph of pedagogical experts

Require: a , the last activity proposed

Require: S , the result of last activity

```

1: procedure SAMPLE_ACTIVITY( $G^e, a, S$ )
2:   for each parameter value  $a_{i,j}$  in  $a$  do
3:      $e_{i,j}, h_{i,j} \leftarrow \text{FindExpert}(G^e, a_{i,j})$             $\triangleright$  Retrieve quality and history of  $a_{i,j}$ 
4:     Add  $S$  to history  $h_{i,j}$ 
5:      $r_{i,j} \leftarrow \left| \sum_{k=t-d/2}^t \frac{h_{i,j,k}}{D/2} - \sum_{k=t-D}^{t-D/2} \frac{h_{i,j,k}}{D/2} \right|$     $\triangleright$  Compute new reward knowing  $S$ 
6:      $e_{i,j} \leftarrow \beta e_{i,j} + (1 - \beta)r_{i,j}$ 
7:   end for
8: end procedure

```

Finally, a parameter value is sampled by normalizing all qualities according to:

$$p_{i,j} = \frac{e_{i,j}^*}{\sum_j e_{i,j}^*} \quad (5.9)$$

In order to sample an activity a_i , I pedagogical experts can be used to independently draw a new activity vector. However, some pedagogical context require to define dependency between parameters. For instance, we can imagine a meta parameter that defines the type of activity proposed. In the context of a cognitive training, this could be choosing the cognitive task being used such as either a MOT or a gonogo task. This parameters would condition the choice of other parameters as the task are parametrized by very different parameters.

To tackle this issue, a graph of pedagogical experts is defined by gathering experts in groups and by creating a hierarchy between groups. The position of the group in the hierarchy defines a sequence of experts used for sampling. Sampling an activity always starts with experts in the top hierarchical group and then follow the hierarchical structure by sampling subsequent groups of experts. The hierarchical dependency is defined by creating a link between a value $a_{i,j}$ with a subsequent group of experts. Getting back to the previous example, we would have in our graph of experts a top hierarchical group containing the parameter "type of activity" that would take 2 possible values: "MOT or gonogo". Then the MOT value would be associated with a set of parameters (e.g number of targets, speed...) different from the set associated with the "gonogo" value (e.g frequency of stimulus presentation). Moreover, in addition to enable the use of different sets of parameters, this graph permits to create dependencies between parameters. For example, in the MOT training, the top hierarchical group of expert only contains 1 expert tracking the parameter number of targets. Then for each value of the number of target, a group of 4 experts (for the parameters speed, tracking duration, answer duration and spacing between object) is created. This means that it exists several experts of the parameter speed, one for each group, and that they are only consulted after a specific value of number of targets is drawn. The algorithm describing the sampling procedure with the hierarchical structure is described in 3.

Student model: The algorithm assumes very few hypothesis about the student. The first important idea is that the reward associated with the last sampled value $a_{i,j}$ is the

Algorithm 3 Generation of a new activity

Require: G^e , a graph of pedagogical experts

```
1: procedure SAMPLE_ACTIVITY( $G^e$ )
2:   currentGroup  $\leftarrow$  top hierarchical group in  $G^e$ 
3:    $a \leftarrow$  an empty list  $\triangleright$  Initialize a list to store sampled values for one activity
4:   while currentGroup is not empty do
5:     for each expert  $e_i$  in currentGroup do
6:        $p_i \leftarrow$  an empty list  $\triangleright$  Initialize list for  $e_i$ 's probability distribution
7:       for each qualityValue  $e_{i,j}$  in  $e_i$  do
8:          $e_{i,j} \leftarrow e_{i,j} + \gamma \cdot \sum_j e_{i,j}$ 
9:          $ProbaValue_{i,j} \leftarrow \frac{e_{i,j}}{\sum_j e_{i,j}}$ 
10:      Append  $ProbaValue_{i,j}$  to  $p_i$ 
11:    end for
12:     $ParamValue \leftarrow$  sample( $p_i$ )  $\triangleright$  Sample a value according to distribution  $p_i$ 
13:    Append  $ParamValue$  to  $a$ 
14:  end for
15:  if there is a ChildGroup of currentGroup then
16:    CurrentGroup  $\leftarrow$  ChildGroup
17:  else
18:    Break from the loop  $\triangleright$  No more groups to process
19:  end if
20: end while
21: return  $a$   $\triangleright$  Return the list of sampled activities
22: end procedure
```

learning progress of the participant for that particular value:

$$r_{i,j} = \left| \sum_{k=t-D/2}^t \frac{h_{i,j,k}}{D/2} - \sum_{k=t-D}^{t-D/2} \frac{h_{i,j,k}}{D/2} \right| \quad (5.10)$$

The choice of this particular reward is line with the learning progress hypothesis (cf section ref). Taking the absolute learning progress allows to not only targets activities with increasing success rate (meaning learning) but also to track activities with drop of performance (meaning potential fatigue or forget).

In addition to the learning progress hypothesis, we model a zone of proximal development for each participant by restricting and evolving the activity space to a smaller subset of possible activities. In addition to limit the exploration and to avoid combinatorial explosion, this idea allows to induce human knowledge of a prototypical skill development in the activity space and a general sense of difficulty evolution. Initially, pedagogical experts only track a subset of easy parameter values. Untracked values are defined deactivated and can't be sampled. To activate a new value, two possible mechanisms are implemented. The general procedure for ZPD management is described in ??.

The first mechanism for ZPD expansion activates a new value if the average success rate of all open values is above a certain hyperparameter threshold λ_1 . The success rate of a particular value that has been presented $T_{i,j}$ times can be defined as:

$$sr_{i,j} = \begin{cases} 0 & \text{if } m < 2 \\ \frac{1}{m} \sum_{t=T_{i,j}-m}^{T_{i,j}} s_{i,j}^t & \text{else} \end{cases} \quad \text{where } m = \min(T_{i,j}, d) \quad (5.11)$$

It is noteworthy to observe that only a window of minimum 2 and maximum D last activities are considered for success rate calculation. Then, if O_i corresponds to the set of open values for an expert e_i , the mean success rate is computed as:

$$SR(O_i) = \frac{1}{|O_i|} \sum_{j \in O_i} sr_{i,j} \quad (5.12)$$

As observed in eq 5.12, this mechanism is conducted independently of the success of other experts and is called "independant" update. The implemented algorithm is described in 4.

Algorithm 4 Update expert from its own history alone

Require: G^e , a graph of pedagogical experts

```

1: procedure UPDATE_INDEPENDANT_EXPERT( $G^e$ )
2:    $O_i \leftarrow \text{getActivatedValues}(e_i)$ 
3:   openValuesSR  $\leftarrow$  an empty list
4:   for  $j$  in  $O_i$  do
5:     if length of  $h_{i,j} > 2$  then
6:       Add  $\frac{1}{m} \sum_{t=T_{i,j}-m}^{T_{i,j}} h_{i,j,k}$  to openValuesSR
7:     end if
8:   end for
9:   currentExpertSR  $\leftarrow \text{sumOf}(\text{openValuesSR}) / \text{length}(O_i)$ 
10:  if currentExpertSR  $> \lambda_1$  then
11:     $j' \leftarrow \text{getNextValueToOpen}(e_i)$ 
12:     $e_{i,j'} \leftarrow \max(e_i)$   $\triangleright$  Initialize new value with highest quality
13:  end if
14: end procedure

```

The second mechanism for ZPD expansion only concerns experts in top hierarchical groups (especially for values possessing subsequent children expert groups). By defining a set of requirements R_i composed of parameter values in subsequent hierarchical group, a new value is activated if all requirements are fulfilled i.e if all values in R_i have a success rate (calculated with eq 5.11 above a certain hyperparameter threshold λ_2). Different from the previous mechanism, this ZPD update rule is named as a dependant update as it relies on other experts. The implemented algorithm is described in 5.

Finally, the ZPD is also manipulated by deactivating activities with an high success rate. If a parameter value has a success rate above a certain hyperparameter threshold λ_3 , this value is withdrawn from the ZPD and will never be proposed again. This mechanism allows to avoid proposing activities that completely mastered.

Hyperparameters

Several pilot studies were conducted prior to the data presented in this manuscript. The main objective of these different pilots were to validate the hyperparameters presented in Table 5.4.

Algorithm 5 Update expert from history of other experts

Require: e_i , an expert to update

```
1: procedure UPDATE_ZPD( $e_i$ )
2:    $R_i \leftarrow \text{getOtherExpertsValues}(e_i)$ 
3:   updateBool  $\leftarrow$  True
4:   for each  $h'_{i,j}$  in  $R_i$  do
5:     if length of  $h'_{i,j} > 2$  then
6:       currentExpertValueSR  $\leftarrow \frac{1}{m} \sum_{t=T_{i,j}-m}^{T_{i,j}} h_{i,j,k}$ 
7:       if currentExpertValueSR  $< \lambda_2$  then
8:         updateBool  $\leftarrow$  False
9:       end if
10:    else
11:      updateBool  $\leftarrow$  False
12:    end if
13:  end for
14:  if updateBool is True then
15:     $e_{i,j'} \leftarrow \max(e_i)$ 
16:  end if
17: end procedure
```

Table 5.4. All hyperparameters used in our ZPDES version.

Name	Value	Objective	Used in
β	0.2	balance between previous quality estimation and new reward	eq 5.6
γ	0.1	manage exploration/exploitation tradeoff	eq 5.8
D	10	full window size for LP computation	eq 5.10
d	5	window size for SR computation	eq 5.11
λ_1	0.7	threshold on SR for new value activation within an expert	SR comparison
λ_2	0.75	threshold on SR for new value activation across experts	SR comparison
λ_3	0.9	threshold on SR for value deactivation within an expert	SR comparison

Learning trajectories

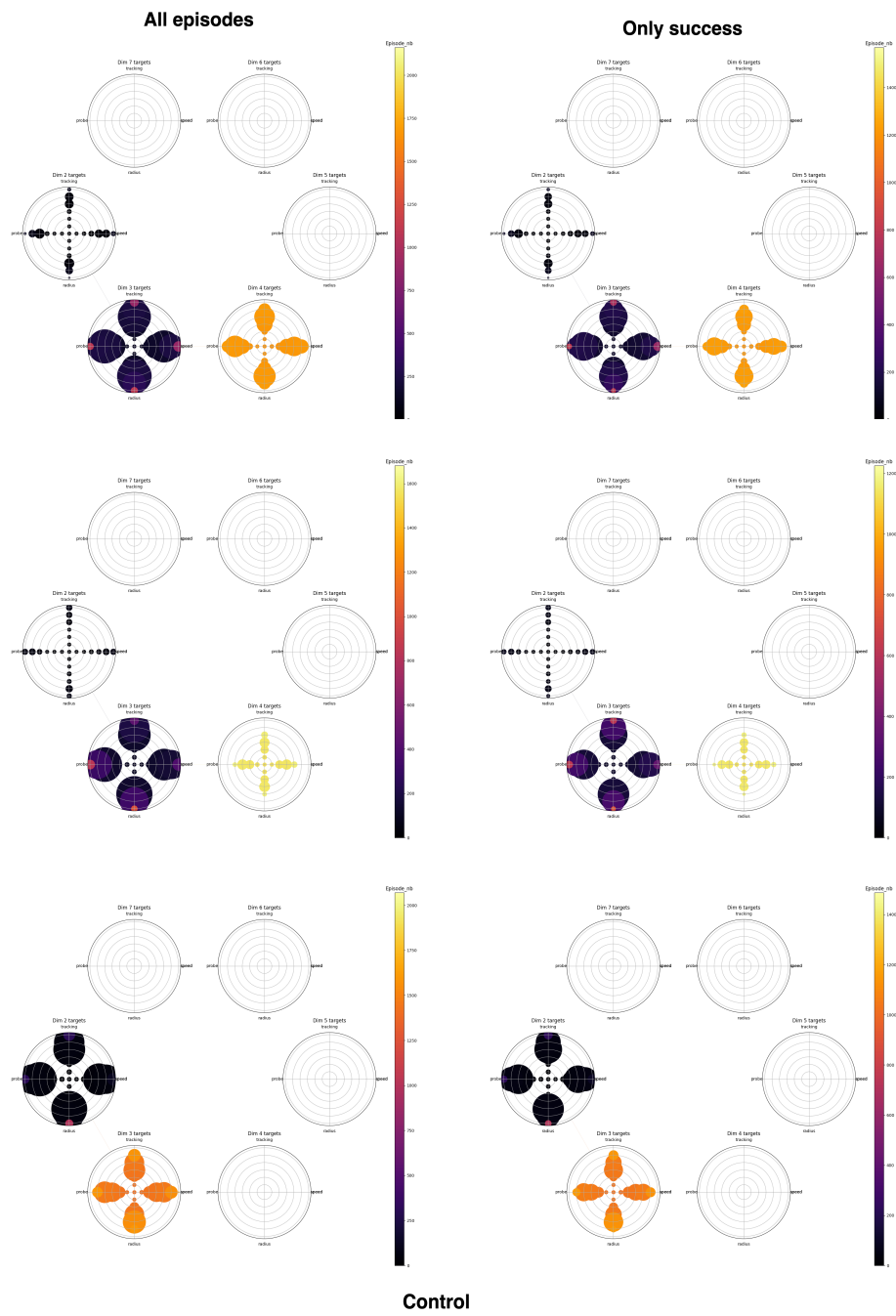


Fig 5.9. Different trajectories from the control group.

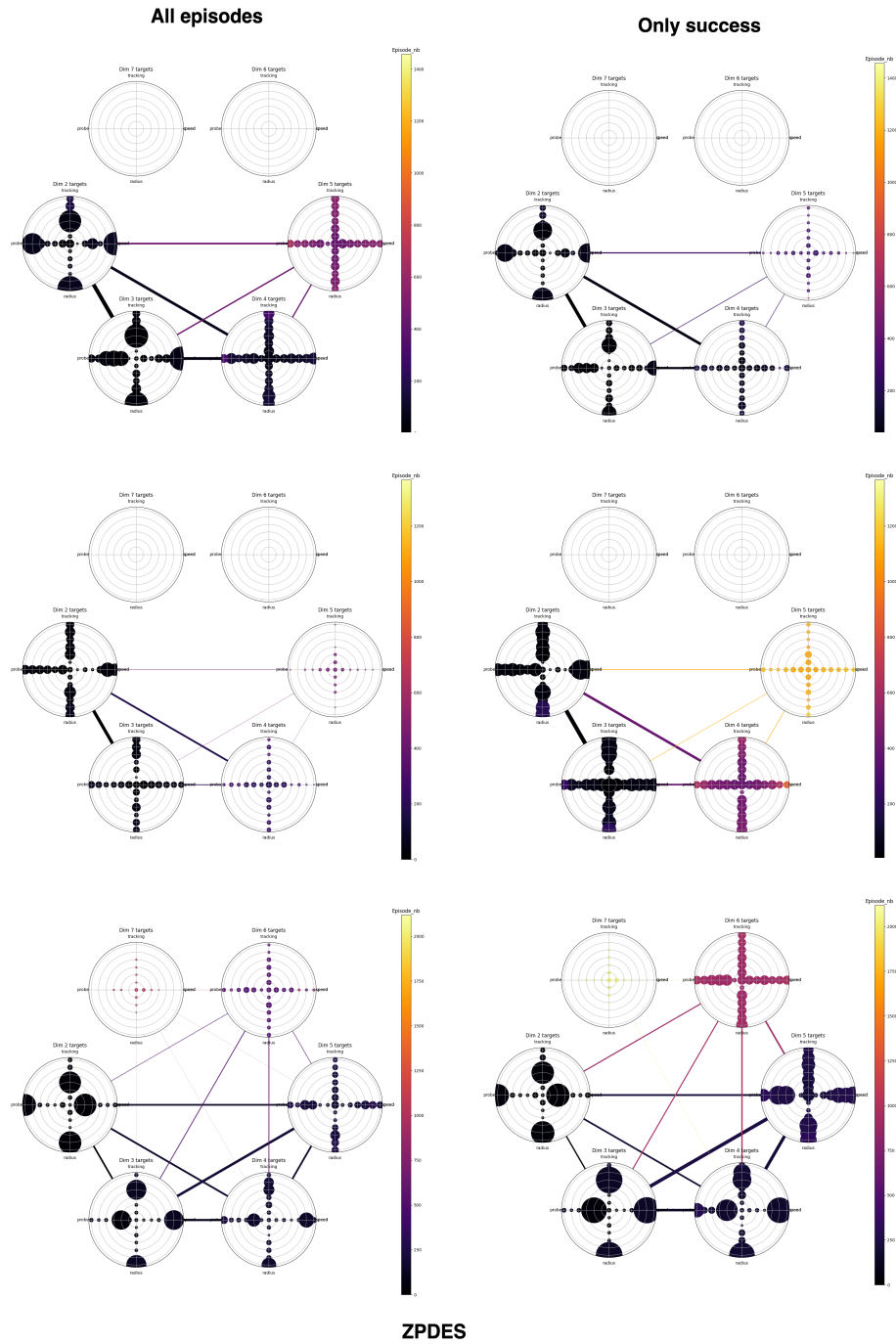


Fig 5.10. Different trajectories from the zpdcs group.

Statistical analysis

To compute the Savage-Dickey density ratio (SDDR) of the different posterior of the cognitive battery analysis, Table 5.5 summarize the different null hypothesis we used:

Table 5.5. Null Hypothesis Models for Different Parameters. N stands for Normal distribution.

Parameter	Null hypothesis model
Accuracy	$N(0, 0.05)$
Reaction Time	$N(0, 20)$
Threshold UFOV	$N(0, 18)$
Switching cost	$N(0, 20)$

Questionnaires

5.5.1 Questionnaires

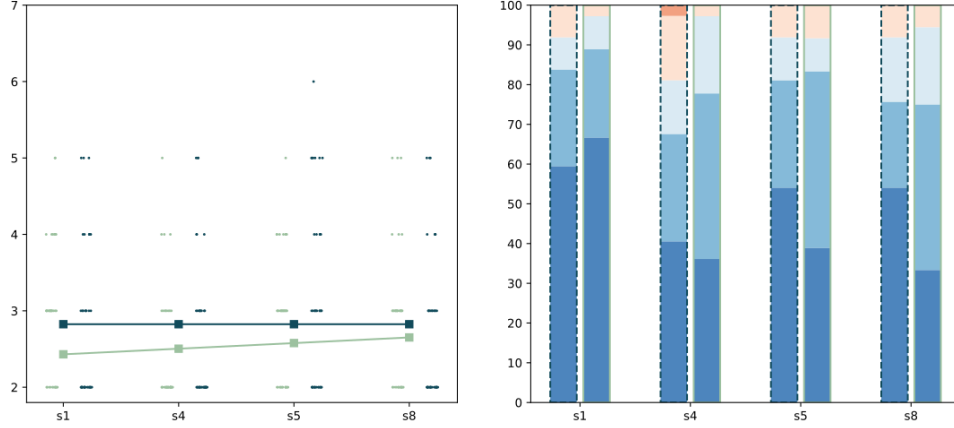
Learning Progress Questionnaire

Possible answers were from 2 to 7.

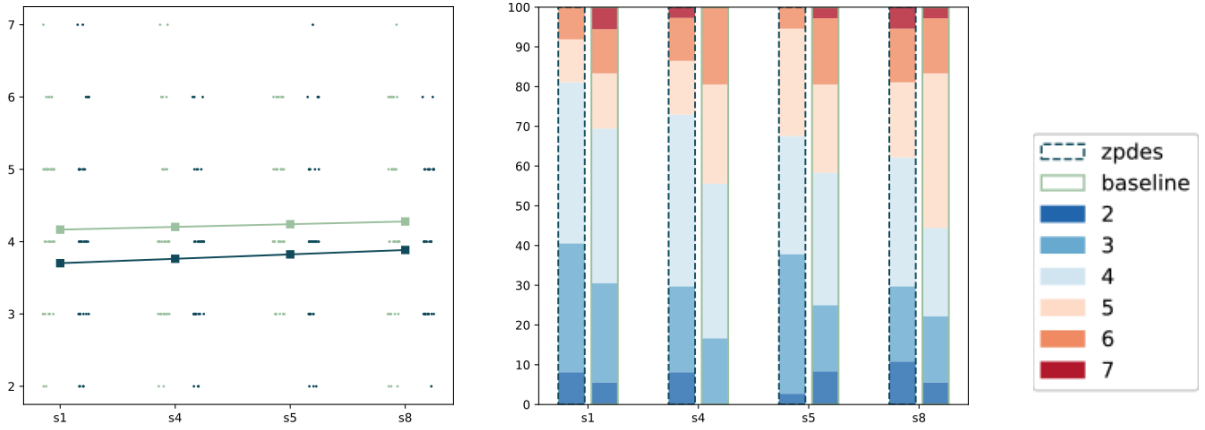
- For the next training activity, how many targets would you like to have?
- How many targets do you think you can track without difficulty?
- How many targets do you think you can track with difficulty?

Learning Progress Questionnaire

How many targets do you think you can track without difficulty?



How many targets do you think you can track with difficulty?



For the next training activity, how many targets would you like to have?

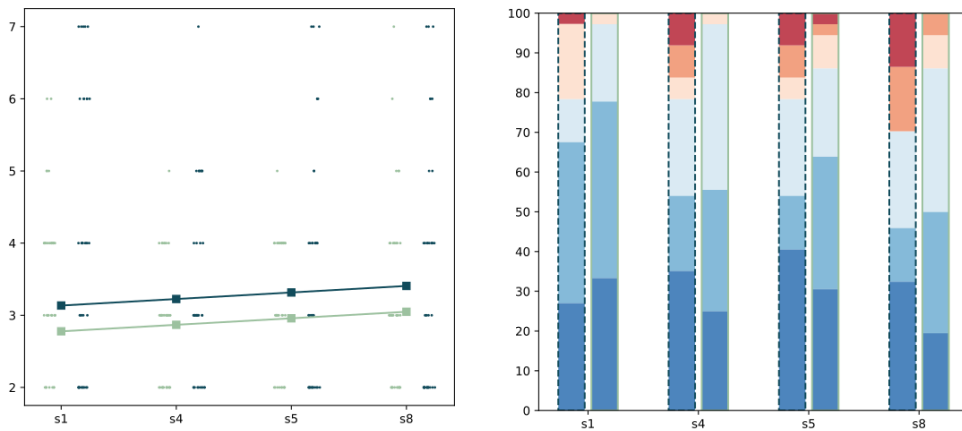


Fig 5.11. LP Questionnaires interaction and distribution.

Nasa-tlx

The question asked is "Reflect on your experience using the game and rate your agreement with the following statements". From "Very Low Very High" on a 20 points scale.

- **Mental Demand:** How mentally demanding was the task?
- **Physical Demand:** How physically demanding was the task?
- **Temporal Demand:** How hurried or rushed was the pace of the task?
- **Performance:** How successful were you in accomplishing what you were asked to do?
- **Effort:** How hard did you have to work to accomplish your level of performance?
- **Frustration:** How insecure, discouraged, irritated, stressed, and annoyed were you?

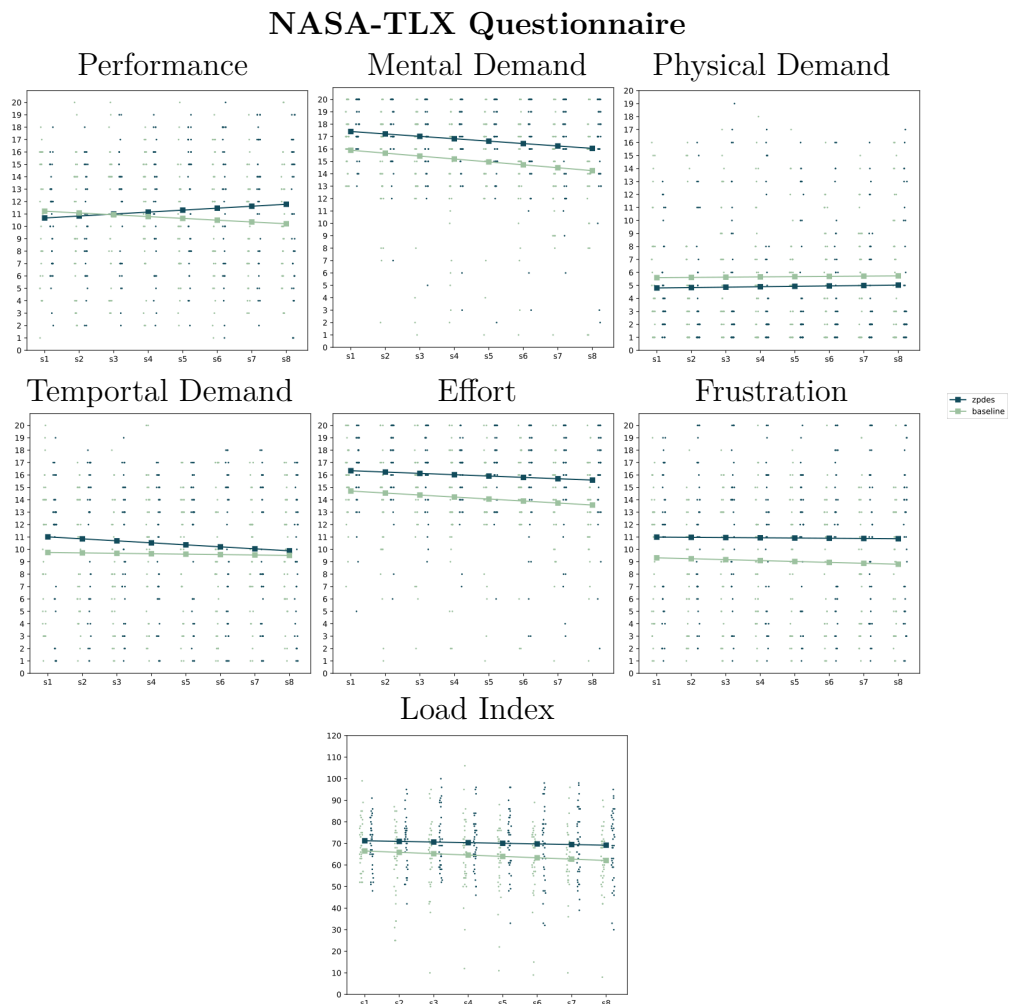


Fig 5.12. NASA-tlx Questionnaires interaction and distribution.

TENS

The question asked is "Reflect on your experience using the game and rate your agreement with the following statements". From "Do not agree Strongly Agree" on a 5 points scale.

- **Competence:** I feel very capable and effective at using the game.
- **Competence:** I feel confident in my ability to use the game.
- **Competence:** Learning how to use the game was difficult. (-)
- **Competence:** I found the interface and controls confusing. (-)
- **Competence:** It wasn't easy to use this game. (-)
- **Autonomy:** The game provides me with useful options and choices
- **Autonomy:** I can get the game to do the things I want it to.
- **Autonomy:** I feel pressured by the game. (-)
- **Autonomy:** The game feels intrusive (-)
- **Autonomy:** The game feels controlling. (-)
- **Relatedness:** The technology helps me to form or sustain relationships that are fulfilling.
- **Relatedness:** The technology helps me to feel part of a larger community.
- **Relatedness:** The technology makes me feel connected to other people.
- **Relatedness:** I don't feel close to other users of the technology. (-)
- **Relatedness:** The technology doesn't support meaningful connections to others. (-)

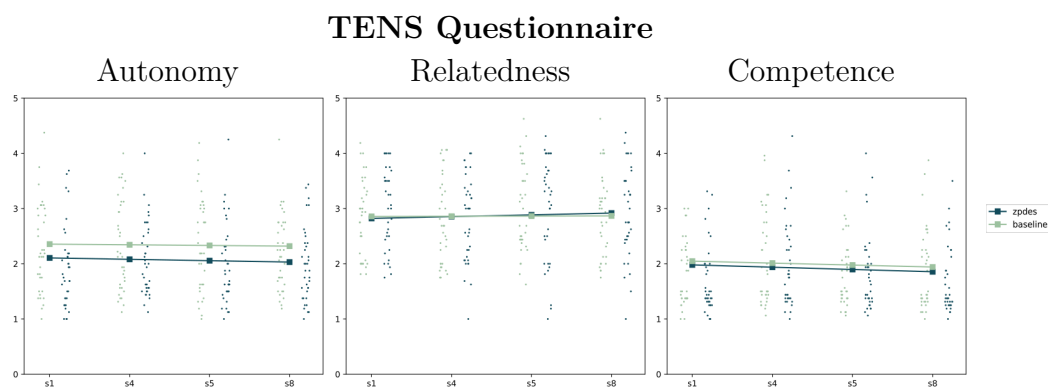


Fig 5.13. TENS Questionnaires interaction and distribution.

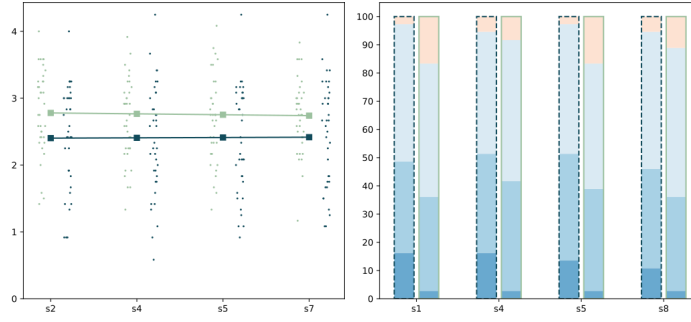
User engagement scale

Question was "Reflect on your experience using the game and rate your agreement with the following statements:". Possible answers were: "Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree".

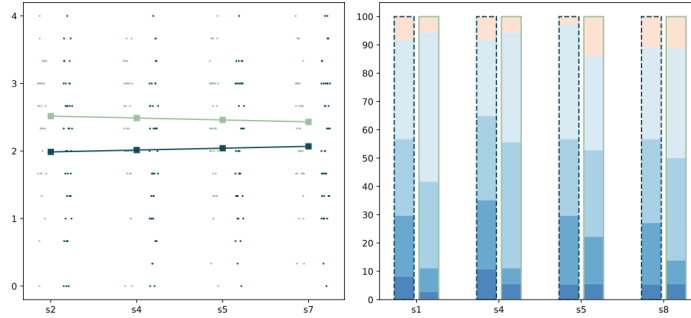
- **Focused Attention:** I lost myself in this experience.
- **Focused Attention:** The time I spent using the game just slipped away.
- **Focused Attention:** I was absorbed in this experience.
- **Perceived Usability:** I felt frustrated while using this game.
- **Perceived Usability:** I found this game confusing to use.
- **Perceived Usability:** Using this game was taxing.
- **Aesthetics:** This game was attractive.
- **Aesthetics:** This game was aesthetically appealing.
- **Aesthetics:** This game appealed to my senses.
- **Reward:** Using the game was worthwhile.
- **Reward:** My experience was rewarding.
- **Reward:** I felt interested in this experience.

UES Questionnaire

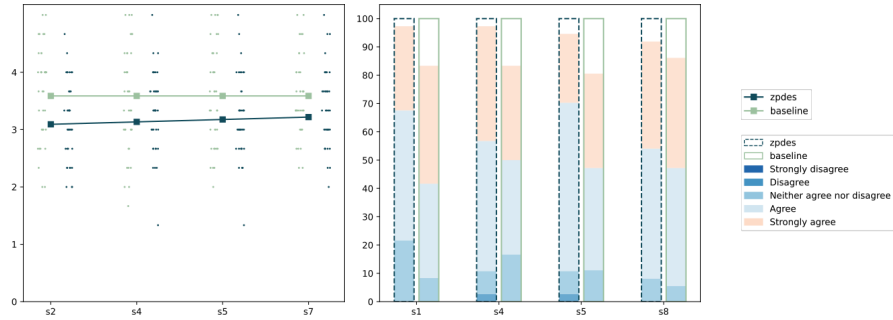
Engagement Score



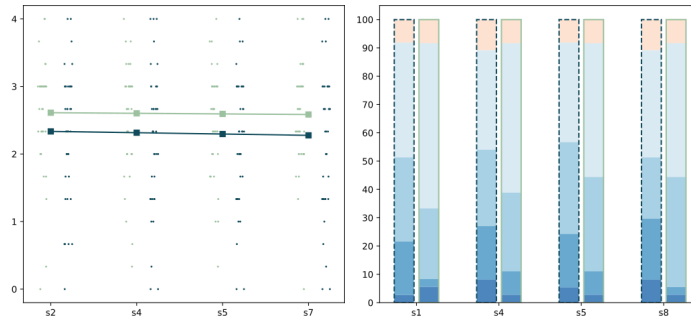
Aesthetics



Perceived Usability



Reward



Focused Attention

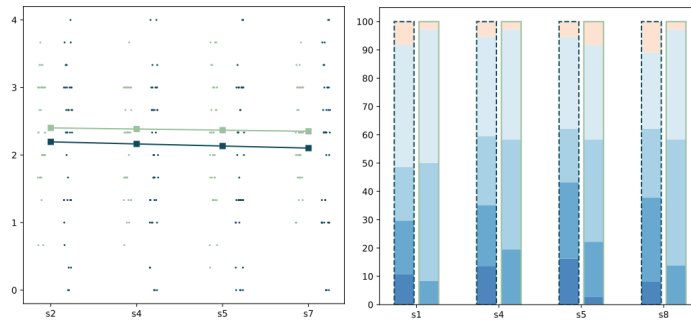


Fig 5.14. UES Questionnaires interaction and distribution.

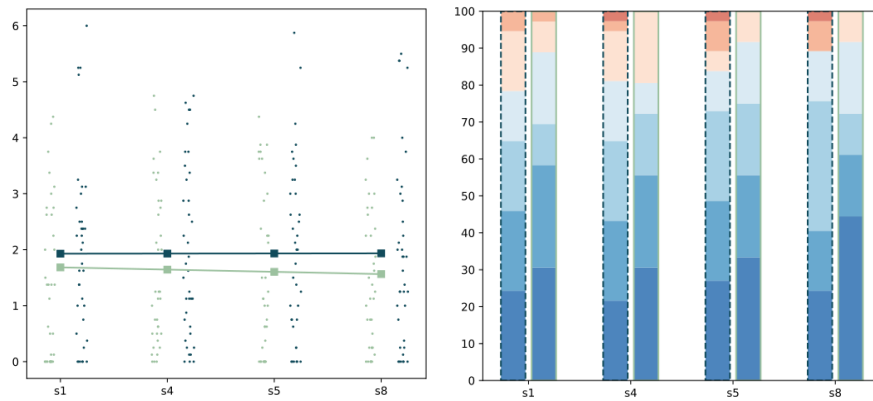
Situational Intrinsic Motivation Scale

The question asked is: Why do you keep practicing the game? Possible answers are form of a likert scale: Not at all Very little A little Somewhat A lot Exactly

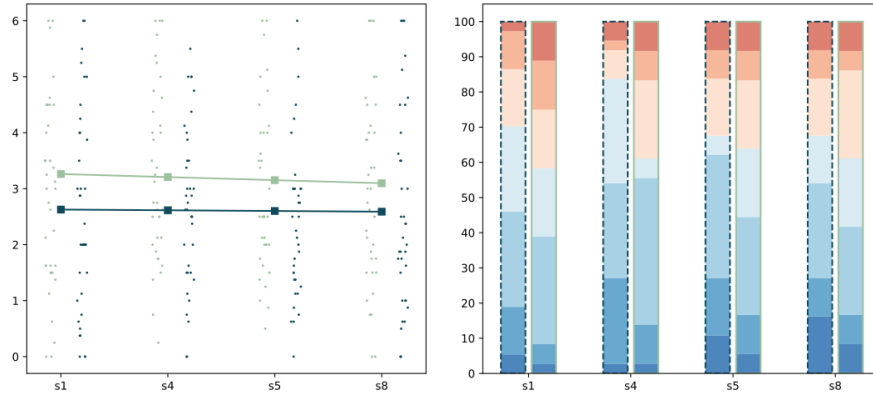
- **Intrinsic motivation:** Because I think that this game is interesting
- **Intrinsic motivation:** Because I think that this activity is pleasant
- **Intrinsic motivation:** Because this game is fun
- **Intrinsic motivation:** Because I feel good when playing this game
- **Identified regulation:** Because I feel like playing this game
- **Identified regulation:** Because I think that this activity is good for me
- **Identified regulation:** It is for my own good
- **Identified regulation:** Because I believe that this game is important for me
- **External regulation:** Because it is something that I have to do
- **External regulation:** Because I don't have any choice
- **External regulation:** Because I am supposed to do it
- **External regulation:** Because I feel that I have to do it
- **Amotivation:** I keep practicing, but I am not sure I should continue
- **Amotivation:** There may be good reasons for practicing this game, but personally I don't see any
- **Amotivation:** I don't know; I don't see what this game brings me
- **Amotivation:** I do this activity but I am not sure if it is worth it

SIMS Questionnaire

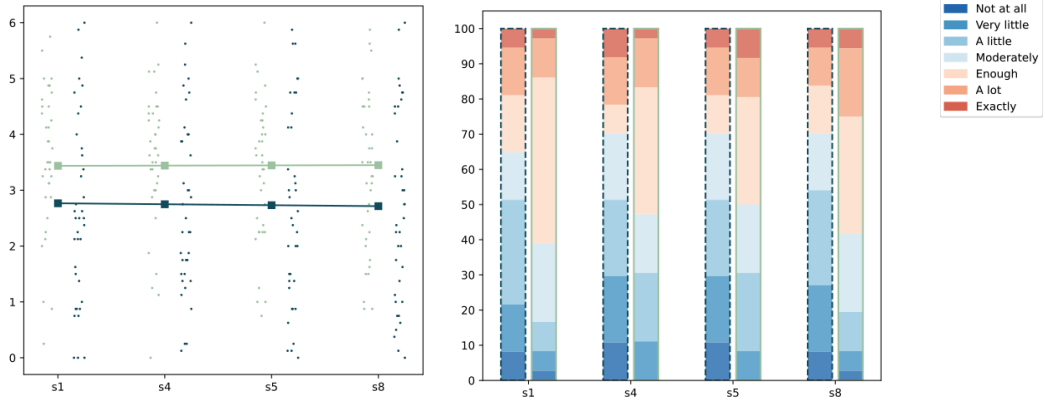
Amotivation



External Regulation



Identified Regulation



Intrinsic Motivation

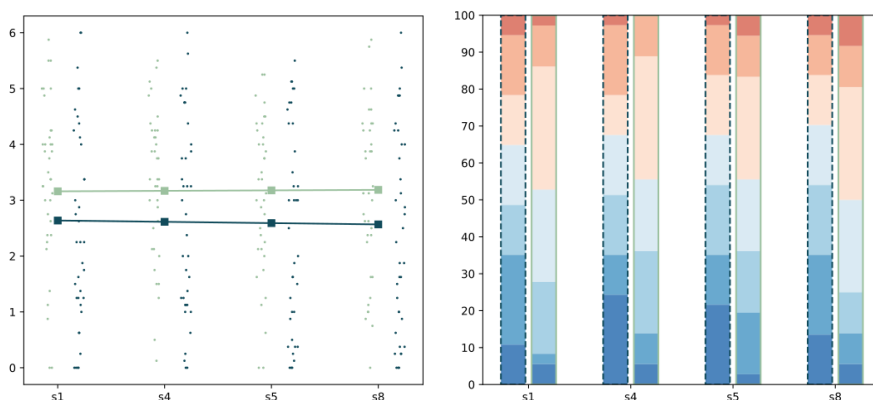


Fig 5.15. SIMS Questionnaires interaction and distribution.

Chapter 6

Benefits of Individualized Computerized Cognitive Training on older adults

Aims

After evaluating our procedure on a population of young adults, this chapter aims to apply our approach to a population of Older Adults (OA). While the main design principles remain consistent with the previous chapter, the goal is to determine if the benefits observed in young adults are replicated in a population with different characteristics.

Contents

6.1	Introduction	195
6.2	Design and Methods .	199
6.3	Results	204
6.4	Discussion	212
6.5	Conclusion	216

Abstract

This randomized controlled trial included 50 healthy older adults, divided into AI-based individualized adjustment (ZPDES) and traditional staircase (control) groups. We assessed task performance progression, cognitive transfer across seven tasks, interindividual differences, and subjective experiences via questionnaires. Pre-post comparisons revealed greater training benefits in the MOT task for the ZPDES group. Specific task improvements were noted only in the control group, but no differences were observed at a latent level. Both groups exhibited non-linear intra-training progress: control participants improved initially and then plateaued, while ZPDES participants showed consistent progress throughout the two weeks of training. For both conditions, interindividual differences in prior MOT performance significantly influenced baseline performance in the first training session, but no significant differences were found in performance changes. Differences in training trajectories led to varied subjective experiences: cognitive load decreased more over time for the control group, indicating ZPDES was more demanding. Participants in the ZPDES group reported lower intrinsic and extrinsic motivation but a higher sense of competence. In sum, the ZPDES condition demonstrated greater post-training MOT performance, consistent intra-training progress, and higher competence, despite being more demanding. Consequently, this paper discusses the implications of our approach on training benefits, experience, and engagement, and proposes improvements to the ZPDES algorithm to better address interindividual differences in cognitive aging.

Collaborators

Adolphe Maxime (AM), Pech Marion (PM), Sawayama Masataka (SM), Clement Benjamin (CB), Joessel Freya (JF), Bavelier Daphné (BD), Delmas Alexandra (DA), Maurel Denis (MD), Oudeyer Pierre-Yves (OPY), Sauz on H l ne (SH). ¹ This work is currently *in prep* and represents a preliminary version of a future article.

¹Conceptualization (AM, SM, CB, JF, BD, SH, POY); Data Curation (AM, PM); Formal Analysis (AM, PM); Investigation (AM, PM); Writing - Review & Editing (all).

6.1 Introduction

Thanks to scientific advances in the neurobiology of typical aging, cognitive aging is now understood as a complex phenomenon including declines, maintenance, and even improvement of cognitive functions or skills that are explained through the neuroplasticity operating throughout life (Angel and Isingrini, 2015; Reuter-Lorenz and Park, 2010). As a result of neuroplasticity, aging is crucially characterized by a high degree of intra- and inter-individual variability that are dependent on complex interactions between intrinsic factors (e.g., genetic, physiological, metabolic) and extrinsic factors (e.g., social stimulation, education, life style) when studying cognitive aging (Marioni et al., 2011). A new way to consolidate cognitive skills and maintain cognitive reserve is to develop non-pharmacological interventions that stimulate the brain to compensate for or maintain cognitive functions despite age-related losses and changes (Valenzuela and Sachdev, 2009). Great age-related changes can constitute a risk for the older adults (OA) because they are linked to many disorders, including neurodegenerative diseases, threatening autonomy and quality of life. These interventions are mostly based on cognitive training (von Bastian et al., 2022b) aiming at stimulating one or more cognitive domains that decline with aging, including functions such as decision-making (Fechner et al., 2019; Hartshorne and Germine, 2015), mental flexibility—i.e., the ability to form, change, and update representations processed in working memory (Fechner et al., 2019)—attention (Geerligs et al., 2014), and inhibitory processes related to executive functioning (Collette and Salmon, 2014; Turner and Spreng, 2012). Our present study aims to evaluate attention training, by examining whether AI-automated adaptation of training difficulties to individual abilities improves the benefits of such an intervention compared to a control condition without taking into account initial inter-individual variability on the trained task, and over the course of training.

6.1.1 CT for typical older adults : impact of interindividual differences

Several cognitive interventions or CT for typical OA have been created aiming either domain-specific or multi-domains (I. H. Leung et al., 2015a). Specifically, domain-specific CT focuses on strengthening a particular cognitive domain, such as working memory, throughout the program, whereas multi-domain CT simultaneously addresses multiple cognitive domains through a variety of targeted activities (Gates et al., 2011). The latter approach offers more diverse cognitive challenges, potentially promoting greater brain neuroplasticity and more durable results in the aging population (Ballesteros et al., 2020a; Gates et al., 2011; von Bastian et al., 2022b). Indeed, some of these CT have demonstrated effectiveness in improving various aspects of cognitive functioning, including memory performance, executive functioning, processing speed, attention, and fluid intelligence (I. H. Leung et al., 2015a; Reijnders et al., 2013; von Bastian et al., 2022b). For instance, to enhance attentional flexibility (Belleville et al., 2014), computerized CT involving working memory tasks (e.g., n-back tasks) (Lilienthal et al., 2013) or tasks requiring prioritization and alternation between stimuli (Belleville et al., 2014) has shown positive results among the aging population.

In the CT area, several criteria can be used to assess the efficacy of such interventions. First, improvements in performance on cognitive tasks targeted can be considered, maintenance of improved performance over time, transfer of training effects to different tasks within the same cognitive domain (near transfer) or other domains (far transfer),

and generalization of effects to everyday functioning (Kelly et al., 2014). Nowadays, the challenge of CT lies in achieving far transfer, which aims to trigger improvements in everyday life activities by training specific cognitive processes such as visual attention or working memory (Kelly et al., 2014). Several outcomes have shown near transfer but have not assessed cognitive domains that predict real-world functioning (i.e. ecological measures probing far transfer) in OA even fewer demonstrate generalization to improvements in everyday life activities (Guye, De Simoni, and von Bastian, 2017; Shah et al., 2017a). Thus, the potential transfer of effects to untrained cognitive domains remains unproven. This means no genuine evidence of cognitive training efficacy is today reached.

A key explanation of the challenge of transferring training effects to other tasks is the interindividual differences in CT responsiveness. For this reason, interventions should be tailored to each individual so that the activities remain continuously challenging and not monotonous for maintaining trainee engagement (Reijnders et al., 2013). In this vein, a recent meta-analysis of Z. Li et al., 2024a revealed that engagement, persistence (engagement or persistence rates over 60% are required) and adherence (adherence rates are 80%) are critical determinants influencing the efficacy of cognitive training, especially for memory and visuospatial skills and this, more particularly for healthy OA compared to those with cognitive impairment. This last point raises the question of who benefits from such interventions. It is now well established that some individuals benefit greatly from the intervention (high responders), while others do not (low responders). This phenomenon mirrors the high aging-related heterogeneity, highlighting the limitations of a one-size-fits-all approach for CT design. In a related-way, age-comparative studies have shown larger training effects in younger adults compared to OA (e.g., (Brehmer et al., 2012; Bürki et al., 2014a; Schmiedek et al., 2010; von Bastian et al., 2013)), and in young-old adults compared to old-old adults (Borella et al., 2014; Zinke et al., 2014). These results fit the concept of a magnification effect (Lövdén et al., 2012a), where younger individuals benefit more from cognitive training because they have more cognitive resources available to complete the training tasks successfully (Guye, De Simoni, and von Bastian, 2017). Contrarily, a compensation effect is assumed when children and OA benefited more from training than young adults (Bherer et al., 2008; Karbach and Kray, 2009). This effect can occur because participants with lower prior cognitive functioning have more room for improvement, i.e., a compensation effect (Titz and Karbach, 2014b) as predicted by the Cognitive Compensation theory of aging (Park and Reuter-Lorenz, 2009; Traut et al., 2021b). Yet, recent meta-analyses reflect mixed findings, with some showing age as a moderator of training outcomes (Melby-Lervåg and Hulme, 2013b) and others not (Schwaighofer et al., 2015). General measures of cognitive functioning (such as personality, motivation, cognitive ability) has been identified as factors that likely contribute to the effects of training magnitude and transfer (von Bastian and Oberauer, 2014). Nevertheless, few studies have directly assessed the effect of prior (baseline) cognitive performance on training outcomes (Traut et al., 2021b), with some suggesting that individuals with initially low performance benefited more from training (e.g., (Jaeggi et al., 2008a; Zinke et al., 2014), while others reported the opposite (e.g., (Bürki et al., 2014a). Today, measuring the effectiveness of interventions raises debated issues, partly due to methodological shortcomings in studies on this topic, such as sample size, study design, and cognitive tests used (Simons et al., 2016b; Vater et al., 2021a).

6.1.2 Computerized Cognitive Training for typical older adults : a new promise for healthy aging?

Recently, many computerized CT have been developed both by research and private companies. These interventions and applications can take various forms (such as serious games such as Brainer. (Brainer, s. d.), Neurotracker®. (Vater et al., 2021a), RehaCom®. (RehaCom, s. d.), CogniPlus®.(CogniPlus, s. d.), HappyNeuron Pro®. (Happyneuron, s. d.), the Padua, Rehabilitation Tool (PRT) (Cardullo, 2017), MS Rehab (Gaspari et al., 2020), Cogni-Track (Tacchino et al., 2015) and CogniFit Personal Coach® (CogniFit, s. d.). Among the various levers employed by computerized CT, such as task diversity and game content, attention is recognized as a particularly crucial cognitive function. Attentional processing is ubiquitous and vital for daily activities, and serves as a key indicator of overall cognitive health (Depp et al., 2014b). Consequently, many CT focus specifically on enhancing attention, as it acts as a gateway to other cognitive functions. Impairments in attention can significantly impact everyday activities, such as driving, and then social participation. On this topic, the development of Cogniplus software was based on the Zomerén and Brouwer, 1994 attention model, distinguishing between the intensity and selectivity components of attention and their subcomponents (alertness, selective and divided attention) (Zomerén and Brouwer, 1994). Casutt et al., 2014 found that active training sessions with Cogniplus® improved cognitive and on-road driving performance among OA.

Also focusing on attentional training, Neurotracker offers exercises based on a complex dynamic attentional task called Multi-Object Tracking, which involves selective, sustained and controlled attention. In a typical MOT task, observers are initially presented with several identical objects, a subset of which is then highlighted to indicate their status as targets. Following this, all objects, once again appearing identical, move independently and unpredictably around the display. At the end of the movement phase, observers must identify the original targets (B. Scholl, 2009). Within the MOT task, several parameters can be manipulated such as the number of targets, distractors, the speed of the object's movements, the spatial distance between targets and the tracking duration. These elements make the MOT task a powerful multi-cognitive dimension, as adjusting one of these parameters can affect various cognitive functions (see chapter 3 for a review).

Over the years, the MOT task has garnered attention in cognitive and vision science research, with numerous publications exploring its implications for understanding cognitive functions and object tracking processes (A. O. Holcombe, 2023; Vater et al., 2021a). Its application has extended beyond laboratory settings, demonstrating tangible benefits in daily activities such as driving, gaming, and contributing to cognitive enhancements in executive function and working memory (Vater et al., 2021a; von Bastian and Oberauer, 2014). Neurotracker offers a 3D version of the MOT task, which has been integrated into numerous studies (Vater et al., 2021a) and various contexts (educational, sports, cognitive health). Most of these studies suggest benefits from the practice, with positive results observed in OA (Legault and Faubert, 2012b; Musteata et al., 2019b). However, the results are often inconclusive regarding the evidence of effectiveness of such CT due to the limited standardization of evaluation methods (Simons et al., 2016b). Nevertheless, a common theme is the development of adaptive and tailored solutions to tackle the challenge of the aging population's heterogeneity, with evidence increasingly supporting the individualization of learning paths to address this issue. For instance, in the research field, a randomized double-blind study conducted by Peretz et al., 2011 investigated whether personalized computerized cognitive training provided greater benefits than conventional

computer games for healthy OA. Cognitive performance was assessed at baseline and after 3 months. Both groups improved, but the personalized cognitive training group showed significant improvement in 8 cognitive domains, while the computer games group improved in 6. Personalized training was significantly more effective in improving visuospatial working memory, visuospatial learning, and focused attention. Personalized computerized cognitive training seems to be more effective than computer games in improving cognitive performance in different populations such as younger adults (Plass et al., 2019b) and healthy OA (Bahar-Fuchs et al., 2017; Pedullà et al., 2016a; Peretz et al., 2011).

Regarding the issue of personalization, it is important to consider individualization related to the types of learning paths to assess the effects of such interventions. Most CT programs rely on staircases algorithm to regulate task difficulty (Cornsweet, 1962). This kind of algorithm guides participants towards a predefined optimal difficulty zone by considering parameters that influence task difficulty and the desired success rate can be tailored by specifying the number of successes or failures required to adjust the difficulty level. For instance, numerous studies have employed the 2 up-1 down procedure, i.e., increasing the difficulty after two consecutive successes and decreasing it after two consecutive failures. This approach has limitations such as the linear learning (i.e., the algorithm restricts participants to a fixed trajectory), the challenge of proposing tasks that meet specific participant needs (i.e. tasks being not too easy or impossible to achieve), the need for a variety of tasks (i.e., to promote the learning, motivation and engagement), and the difficulty of addressing these issues in heterogeneous populations (especially in OA).

To address this issue, we have recently conducted a randomized controlled trial with 72 healthy young adults to compare an AI-based individualized adjustment procedure with a traditional staircase method for modulating task difficulty in cognitive training using the MOT task (chapter 5). The study evaluated the impact of these methods on learning trajectories, task performance, and participants' subjective experiences, including cognitive load, engagement, and motivation. The Zone of Proximal Development and Empirical Success (ZPDES) algorithm employs an AI-based method that uses a multi-armed bandit approach informed by Learning Progress and a zone of proximal development (ZPD) model. This algorithm engaged participants with a broader range of tasks, enhancing learning variety to optimize training within the ZPD framework. Performance showed improvements in both the trained task and related tasks for all participants. The AI group demonstrated broader cognitive enhancements not seen in the control group. Subjective assessments indicated a decrease in cognitive load over time, suggesting efficient learning, although the AI-based training was perceived as more demanding, leading to lower engagement and motivation scores. These findings suggest that AI-driven individualized training methods could enhance CT effectiveness and offer significant advantages over traditional methods, though further investigation is needed.

From the overall data, the present study aimed to evaluate the impact of an individualized procedure guided by participant learning progress compared to a classical staircase procedure (one-size-fits-all or linear design) in healthy OA. Specifically, we aimed to investigate if the results of OA are comparable to those of younger adults by replicating the previously employed method. Hence, we investigated three main research questions:

- RQ1- What is the impact of training conditions (ZPDES group [individualized] vs. control group [Staircase]) on observed pre-post assessment ?
- RQ2- What is the impact of training conditions (ZPDES group [individualized] vs. control group [Staircase]) on the MOT intra-training performance? How do individual

differences impact MOT performance evolution, and do training conditions deal with them?

RQ3- What is the impact of training conditions on training experience self-reported in terms of engagement, cognitive load, and motivation?

6.2 Design and Methods

6.2.1 Participants

Participants were recruited through multiple calls to several “Universités du Temps Libre,” French associations that promote cultural and knowledge exchange by organizing workshops and seminars on various topics thanks academic collaborations. To reach our target demographic, we conducted a mailing campaign specifically aimed at individuals over the age of 65, with an explicit goal of enhancing their brain functioning. Following the initial recruitment, we organized an initial meeting either via video call or telephone. During this meeting, we assessed participants to ensure they did not exhibit symptoms of cognitive decline (using the Mini-Mental State Exam (MMSE) score > 27) (Folstein et al., 1975a) or depression (using the 4 items geriatric depression scale (GDS), (J. P. Clément et al., 1997). Only those who met these criteria were included in the study.

6.2.2 Procedure and Materials

The following procedure follows a previous study procedure described in chapter 5. Additionally, all materials used are open-sourced and available at <https://github.com/madolphe/Flowers-OL>.

6.2.3 Schedule

After the initial screening meeting, participants who met the inclusion criteria were invited to a second videoconference. The aim of this session was to guide the participants through the online platform used for the training, assist them in setting up an account, and resolve any technical issues. Once the experimenters ensured that the participants were comfortable with the platform, the remainder of the session consisted of a cognitive evaluation (see section “Cognitive Battery”). Following this first evaluation, the training comprised 16 sessions of approximately 30 minutes each, spanning 8 days (2 sessions per day with a break of 2 days after the first 4 days). During these sessions, participants engaged in the MOT task and completed questionnaires to self-assess various aspects of their experience (see section “Subjective Assessment”). The complete schedule is provided in chapter 5, Figure 5.1.

6.2.4 Pseudo-randomization procedure

After the first session of cognitive evaluation, participants were pseudo-randomly assigned to one of two possible groups: control or intervention (referred to as the ZPDES group in the remainder of the manuscript). Age and attention deficit were controlled to ensure homogeneous groups. As participants began at different initial dates, they were randomly assigned in an online manner, ensuring that age and attention deficit were proportionally balanced between the groups.

6.2.5 Training procedure

The Multi-object tracking task.

The training consisted of practicing the MOT task, a dynamic task where participants are asked to track several targets moving among distractors that share the same visual features. Initially, the targets and distractors were shown moving for a duration of 1000 milliseconds. This was followed by a period of immobility for all objects, also lasting 1000 milliseconds, to ensure the participants identified all the targets to be tracked. After this brief pause, the targets were altered to resemble the distractors in both shape and color. Subsequently, after a variable duration of a few seconds of tracking, all objects ceased movement. Difficulty was adjusted either by a staircase algorithm (control) or a multi-armed bandit algorithm (ZPDES) by manipulating 5 parameters: number of targets, speed, spacing between objects, tracking duration and answer duration.

Additionally, as gamification components have been highlighted as moderators of training effectiveness in both young and older populations (Lumsden et al., 2016b), the training was presented as a series of mini-games embedded in various mini-stories (see chapter 5, for details). Feedback and learning performance were also presented and accessible between trials to help participants better understand their progress (see chapter 5, Figure 5.1).

Difficulty adjustment in the control group

In the control group, the difficulty was manipulated by adjusting five parameters using a 2-up-1-down algorithm (two successes to increase difficulty, one failure to decrease difficulty). While previous studies have used a 3-up-1-down version (F. Joessel, 2022b), we opted for this slightly faster version to avoid large plateaus in progress since we manipulate five parameters concurrently. Additionally, a hierarchical structure was implemented: four parameters (speed, spacing, tracking duration, and answer duration) were managed using a first staircase named S1. To avoid large variations in difficulty, steps in S1 were managed one parameter at a time in a circular manner. Consequently, it required eight consecutive successes for all S1 parameters to advance by one step. Once all parameters reached either their minimum or maximum threshold on S1, the staircase algorithm pertaining to the number of targets, named S2, was updated. Subsequently, all values of S1 were reset to their highest or lowest possible settings. The choice of this control procedure was motivated by its established use as a classical adaptive strategy (F. Joessel, 2022b), and the hierarchical structure allows for better comparability with the ZPDES intervention.

Difficulty adjustment in the ZPDES group

The ZPDES algorithm, initialized and updated independently for each participant, provides a method to balance exploiting known profitable learning activities and exploring new activities that might be beneficial. According to the learning progress hypothesis (Oudeyer et al., 2016b), the feeling of LP leads to higher intrinsic motivation levels, which in turn fosters more learning progress, creating a positive feedback loop. As such, the ZPDES algorithm selects the next activity randomly, giving preference to activities with higher learning progress. To ensure efficient sampling, this mechanism is coupled with a model of the ZPD, which describes the most pedagogically optimal activities for learning that are neither too easy nor too difficult. In practice, the ZPD is defined by an initial subset of interesting activities that evolves over time according to rules that allow for the addition

or deletion of activities within the zone. In sum, the ZPDES algorithm aims to propose activities with the highest learning progress by randomly exploring activities within a model of ZPD. This strategy avoids creating a single route in the activity space and allows for the creation of individualized trajectories for each participant.

6.2.6 Assessments

Cognitive Battery

Seven cognitive tasks from Adolphe et al., 2022a were proposed. To detect near to far transfer effects, the selection of tasks was designed to evaluate tasks that showed a correlation in performance and tapped into similar cognitive dimensions as the MOT (near effect), as well as different control tasks (far effects). We employed a comprehensive set of tests to broadly assess cognitive functions, as follows: measures of selective and divided attention (UFOV (Ball et al., 1988; L. Hoffman et al., 2005)), Loadblindness (Eayrs and Lavie, 2018a), Enumeration (Eayrs and Lavie, 2018a), attentional control (task switching (Monsell et al., 2003b)), Go/NoGo (T. Mani et al., 2005), working memory (Corsi (Berch et al., 1998b)), and episodic memory (memorability (Isola et al., 2011a)). The evaluation was planned to last 1 hour and 30 minutes. All tasks are freely accessible for demonstration: (https://flowers-ol.bordeaux.inria.fr/flowers_demo) and open-source software (<https://github.com/flowersteam/cognitive-testbattery>).

For each task, we used several difficulty conditions and randomly shuffled the order of presentation for each participant. The UFOV task was the only task using a staircase method following (Yung et al., 2015b). Performance metrics were calculated for each difficulty condition and then averaged over all trials. Depending on the task, metrics included accuracy (MOT, Loadblindness, Enumeration, Go/NoGo, working memory, memorability), reaction times (memorability), switching costs (task switching), and the last staircase threshold (UFOV).

Intra-training evaluation

Several hidden evaluations throughout the intervention were used to assess and compare performance between groups. Four evaluation sessions were conducted at the beginning of sessions 1, 4, 5, and 9 (chapter 5). While participants were unaware of these evaluations, they provided reliable and comparable performance assessments between groups throughout the intervention on the trained tasks. The evaluation set consisted of 12 activities, representing all possible combinations of the number of targets (2, 4, 6), speed (low or high), and spacing (low or high), each randomly presented four times.

Subjective Assessment

To investigate the training experience, several questionnaires were administered throughout the intervention (Figure ??). First, some single-use questionnaires were employed to better understand participants' profiles, such as demographic questionnaires, video game questionnaires², and self-assessments of attention deficits³. Additionally, other standardized questionnaires were administered multiple times during the training to assess the impact of the intervention on cognitive load (NASA-TL, (Hart, 2006), engagement (with the

²<https://www.unige.ch/fapse/brainlearning/vgq>

³<https://add.org/wp-content/uploads/2015/03/adhd-questionnaire-ASRS111.pdf>

User Engagement Scale - UES, (O’Brien et al., 2018)), self-determination feelings (with the Technology-based Experience of Need Satisfaction–Interface, TENS, (Peters et al., 2018), and self-perceived intrinsic motivations (Situational motivation scale, SIMS, (Guay, Vallerand, and Blanchard, 2000)). A non-standardized, homemade questionnaire was also used to collect participants’ self-rating of their progress (i.e., Feeling of learning/progress), asking questions about the difficulty levels they were able to handle or would like to train on (Appendix chapter 5). Complementary, we asked for free comments from participants as feedback from the training.

6.2.7 Analyses

Framework

The data analysis was organized to address our three research questions. First, to evaluate the impact of the intervention on cognitive performance, we examined performance evolution through intra-training evaluations and the cognitive battery. Second, to explore the impact of individual variability, we used latent growth models, focusing on individual trajectories and the influence of individual factors. Finally, to assess subjective performance evolution, we analyzed questionnaire responses and conducted a qualitative analysis of the collected verbal feedback.

Q1.1. MOT Cognitive battery

To investigate the efficiency of AI-based training pre-post intervention differences are studied on MOT task. Bayesian mixed models were used to assess potential main effects of time, group, or an interaction between group and time on average performance measured in each session of MOT evaluation. Estimated values (mean and High Density Interval) for main and interaction effects were reported. Bayes Factor (BF) were used to assess whether the parameters were credibly non-null. To estimate BF, we used the Savage-Dickey density ratio (SDDR) as described in (Wagenmakers et al., 2010). SDDR is an efficient way to compute BF when two nested models are being compared. For example, to estimate the credibility of a model using a main effect of time, we compare a model with the main effect of time (alternative model) to a model without it (null model). If the resulting BF $H1 = 5$, it means the data are five times more likely under the alternative model. Note that BF $H0 = 1/5$, so in this scenario, the data are 0.2 times more likely under the null hypothesis. BF values between 1 and 3 correspond to ambiguous evidence, between 3 and 10 to substantial evidence, between 10 and 30 to strong evidence, between 30 and 100 to very strong evidence, and above 100 to decisive evidence. Similar analyses have been performed for the six others tests of cognitive battery (see Appendix 6.5).

Q1.2. Cognitive battery

A Latent Factor Analysis (PCA, details in chapter 5) is used to investigate the global impact of condition training on the seven tasks probing attention, short-term memory, working memory and long-term memory.

Q2.1. Intra training evaluation

Intra training evaluation is performed with descriptive data as well as Latent Growth Curve Models (LGCM). Latent score models are powerful tools for studying performance changes

over time with multiple measurements, helping identify who benefits from cognitive training. To estimate training trajectories, we used LGCM on our data. LGCM employs structural equation modeling (SEM) to examine interindividual differences in intraindividual changes over time, addressing common research issues like missing data, non-normal data, and non-linear changes (P. Curran et al., 2010).

LGCM account for measurement error and provide separate estimates for baseline cognitive performance (intercept) and changes in training performance (slope), allowing us to understand how baseline performance relates to changes: a positive relationship indicates magnification effects, while a negative relationship indicates compensation effects. LGCM uses two latent factors: baseline performance and change in performance, predicted by repeated observed measures (sessions). Loadings between the baseline and observed variables are set to 1, and loadings between change and observed variables start at 0 and can either reflect the time between sessions or be freely estimated. Error variances and covariances between latent factors or manifest variables can also be included.

Following Guye, De Simoni, and von Bastian, 2017, we fitted LGCM to investigate performance changes in the MOT task during training sessions. For each group, we tested three models: no growth, linear, and non-linear. The no-growth model includes only baseline performance. The linear model constrains the change loadings to the temporal spacing between observations. The non-linear model allows for the free estimation of change loadings.

Model fits were assessed and compared using standard metrics: the chi-square statistic (χ^2), with good fits indicated by values between 0 and 2 times the number of degrees of freedom (DF); the standardized root-mean-square residual (SRMR), with good fits indicated by values smaller than 0.08; and the comparative fit index (CFI), with good fits indicated by values greater than 0.95. For each group, the best-fitting model was retained, and changes over time were investigated by examining the mean baseline performance (μ_i), mean change (μ_s), and the type of evolution (if nonlinear, λ_3 , λ_4). Additionally, the variances of initial performance, slope, and the covariance between the two latent variables were reported. These estimates allow us to assess the heterogeneity within each group and detect magnification effects (significantly negative covariance between latents) or compensation effects (significantly positive covariance between latents).

For the variances of latent factors and the covariance between latent factors, we reported the p-value associated with the Wald test (assessing the null hypothesis that the estimated parameter is zero) and the BF associated with both the null and alternative hypotheses. We used the Bayesian Information Criterion (BIC) to estimate BF (Wagenmakers, 2007):

$$BF_{H1} = \exp(0.5 \times (BIC_2 - BIC_1))$$

where BIC2 represents the BIC for the null model (without the predictor) and BIC1 represents the BIC for the alternative model (free estimation of the predictor of interest). BF evaluates the evidence in favor of the alternative hypothesis over the null hypothesis. For example, to assess the credibility of the hypothesis that the covariance between the initial and change parameters is non-zero, we compared a model with the covariance constrained to 0 (null model) with a model freely estimating this parameter.

We also investigated the impact of individual differences on performance evolution by adding six predictors to the best-fitting model: pre-test cognitive battery performance, education level, engagement, cognitive load, and self-determination index. These predictors were regressed on both the baseline and change factors, with estimated values, p-values, and BF reported for each predictor. For the LGCM the analyses were conducted in R

(version 3.2.3; R Core Team 2015) using the “lavaan” package (version 0.5.23; Rosseel, 2012). We use Python and R for all others analyses.

Q3.1. Subjective Assessment

To investigate the impact of training on subjective experience, several questionnaires were administered throughout the intervention (Figure 5.1). First, learning progress is evaluated by asking three questions about target tracking in training activities (see Appendix chapter 5), focusing on the number of targets participants desire versus those they estimate they can track easily or with difficulty. This approach helps calculate the ratio between feasible and challenging targets. The Nasa-TLX (Hart, 2006) uses a ‘load index’ to represent overall cognitive load. The Situational Intrinsic Motivation Scale (SIMS, Guay, Vallerand, and Blanchard, 2000) measures motivations in specific situations, encompassing intrinsic motivation and identified regulation. The User Engagement Scale (UES, O’Brien et al., 2018) assesses engagement through four dimensions: focused attention (the sensation of attentional capture), intrinsic reward (a sense of satisfaction from learning and success), perceived usability, and the visual appeal of the software device. Lastly, the TENS (Peters et al., 2018) analyzes how technology interfaces meet essential user needs, including autonomy, competence, and social relatedness according to the Self-determination theory (Ryan and Deci, 2017).

Q3.2. Qualitative analysis

Following the training phase, participants are invited to provide open-ended feedback on their training experience, perceptions, and potential areas for improvement.

6.3 Results

6.3.1 Descriptive statistics

There were 66 participants who initiated the experiment, 2 participants dropped out immediately after the initial platform orientation session (1 participant for the control group and 1 for the ZPDES group). Among the 64 participants who finished the first session corresponding to the pre cognitive battery evaluation, 14 participants discontinued their participation during the training, with 6 belonging to the ZPDES condition (attrition rate of 9%) and 8 to the control condition (Attrition rate of 12%). In total, 50 participants completed the entire experiment, with 24 in the control condition and 26 in the ZPDES condition (see Figure 6.1). Thus, the participation rate was 84%. This rate suggests good participant engagement according to the meta-analysis by Z. Li et al., 2024a.

6.3.2 Q1. What is the impact of training conditions (ZPDES group [individualized] vs. control group [Staircase]) on observed pre-post assessment?

Q1.1. Pre-and post-measures of MOT from cognitive battery

Evolution from pre to post-test was observed for both groups for moteval (ZPDES group strong evidence: mean=0.08, HDI=[0.01, 0.14], BF=7.87; control group weak evidence: mean=0.04, HDI=[-0.01, 0.10], BF=1.88)

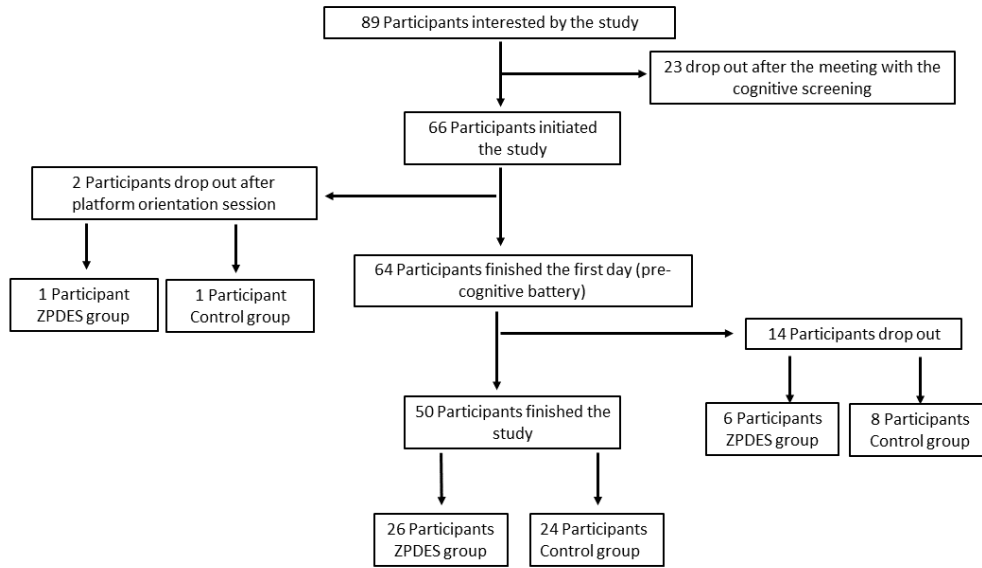


Fig 6.1. Flowchart of the study

In addition to detection of credible evolution within groups, we compare the magnitude of change between groups. For metrics where change was detected for both groups, only the moteval task revealed evidence in favor of greater magnitude for the ZPDES group (mean between-group difference in change=0.03, HDI=[-0.05, 0.12]) (Figure. 6.2). All other results for the 6 other tests of cognitive battery are available in the Appendix 6.5. They are not presented as the main results in order to focus on the overall effects from the latent factor analysis.

Q1.2. Multi domain task: a focus on a latent factor analysis related to the pre-post cognitive performance

Following Kaiser Criterion, 7 principal components were kept for latent factor analysis explaining 73.61% of variance (Figure 6.3). For all principal components (PC), no difference was found between groups at pre-test with all BF being smaller than 0.7 (Appendix Table 6.3). Regarding pre-post evolution, very strong evidence in favor of evolution from pre-to post-test was observed on PC1 with BF01=404.07 and BF01=2175 for ZPDES and control respectively. All other PCs did not show any credible effect.

6.3.3 Q2. What is the impact of training conditions (ZPDES group [individualized] vs. control group [Staircase]) on the MOT intra-training performance? How do individual differences impact MOT performance changes, and do training conditions influence them?

Q2.1. Impact of training condition in terms of path difference

The descriptive data of intra-training performance showed that the ZPDES condition resulted in various training paths among the OA while the control condition produced a

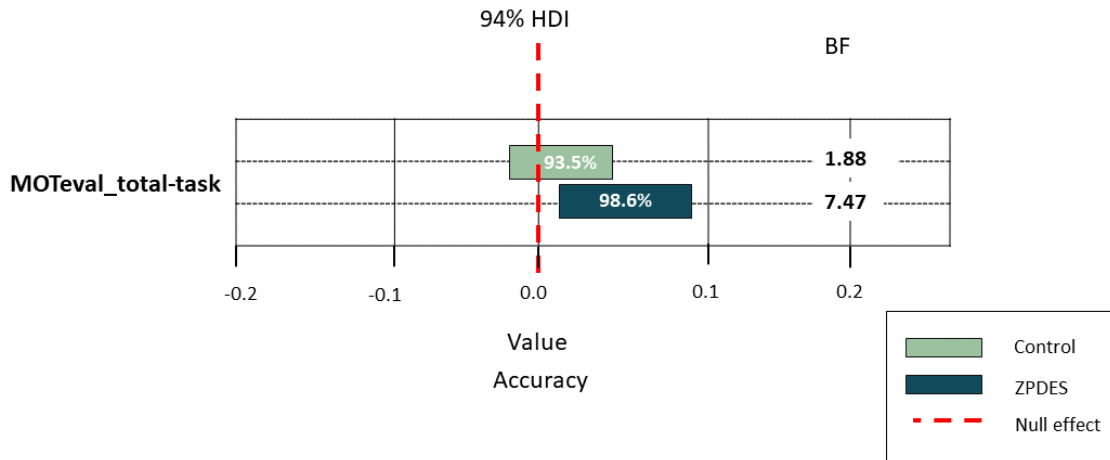


Fig 6.2. Pre-post evolution for MOT task accuracy (left) metric. Horizontal bars represent the 94% HDI of the posterior distribution for each parameter. Inside these bars, the probability of the parameter being either greater than zero ($P(\text{param} > 0)$) or less than zero ($P(\text{param} < 0)$) is shown, depending on which probability is higher.

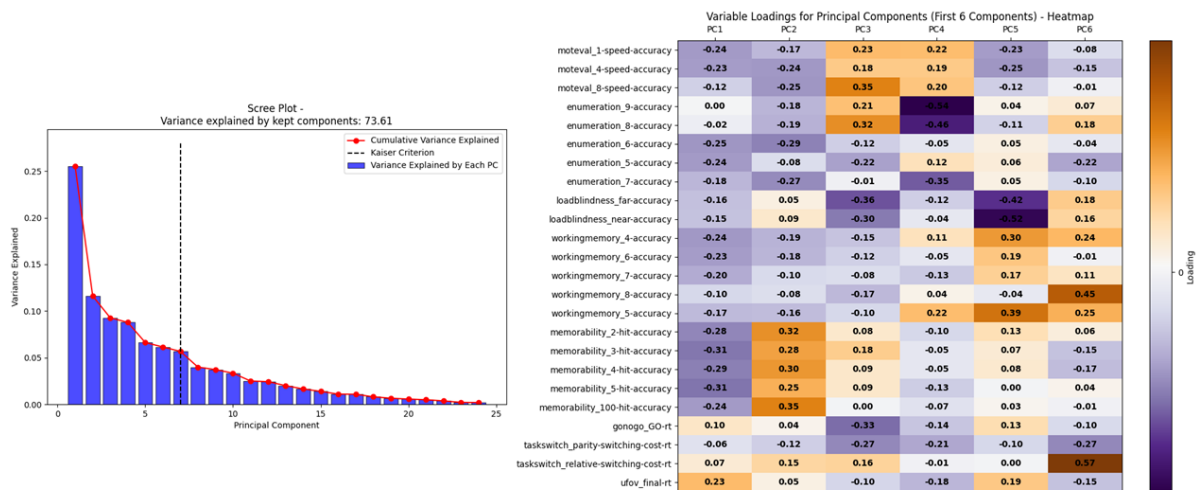


Fig 6.3. Scree plot of the PCA (left) and Matrix loadings of the PCA (right).

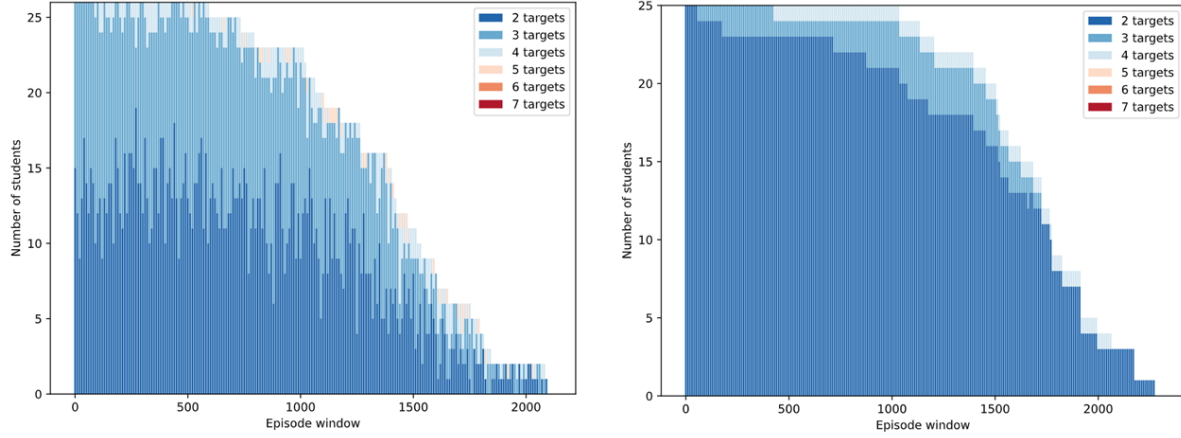


Fig 6.4. Proportion of the population on each different number of targets (2 to 7 targets) through time.

unique and uniform path for all participants (Figures 6.4, 6.5).

Related to the training performances, linear mixed models fitted on the average performance of each evaluation session revealed a main effect of group, with the ZPDES group showing overall lower performance (from the beginning) (mean=-0.04, HDI=[-0.1, 0.00], BF=2.17). Both groups improved through training, as indicated by a main effect of time (mean=0.03, HDI=[0.02, 0.04], BF » 1000). No credible group*time interaction was detected (mean=0.01, HDI=[-0.01, 0.02], BF=0.18).

Q2.2. How does inter-individual differences impact the MOT intra-training performance? Do training conditions (ZPDES vs. control) affect these changes?

Latent Growth Curve models The comparisons of 3 models (no growth, linear growth, non linear growth) first allowed us to investigate the type of performance evolution. Models performances and comparisons are summarized in Table 6.1. For the ZPDES group, the non-linear model fitted best the data with $\chi^2(3) = 2.58$, $p=0.46$, SRMR=0.05 and CFI=1. For the control group, the non-linear model fitted best the data with $\chi^2(3) = 5.95$, $p=0.114$, SRMR=0.13 and CFI=0.95.

Estimates and heterogeneity All following results are summarize in Figure 6.7. For the ZPDES group, results indicate that individuals started on session 1 with a mean performance score of $\mu_i=2.48$ and significantly increased their performance by $\mu_s=0.28$ leading to training performance across the 4 sessions of 2.48 (session 1), 2.76 (session 2), 2.76 (session 3, estimated coefficient $\lambda_3=1.01$) and 2.82 (session 4, estimated coefficient $\lambda_4=1.22$). We found a significant negative association between intercept and slope but with weak evidence in favor the alternative hypothesis ($\sigma_{i,s}=-0.07$, SE=0.03, $p=0.044$, $BF_{H1}=1.23$). However there was substantial evidence for individual differences in the variance of baseline performance ($\sigma_i^2=0.18$, SE=0.06, $p=0.002$, $BF_{H1}=3.95$) and

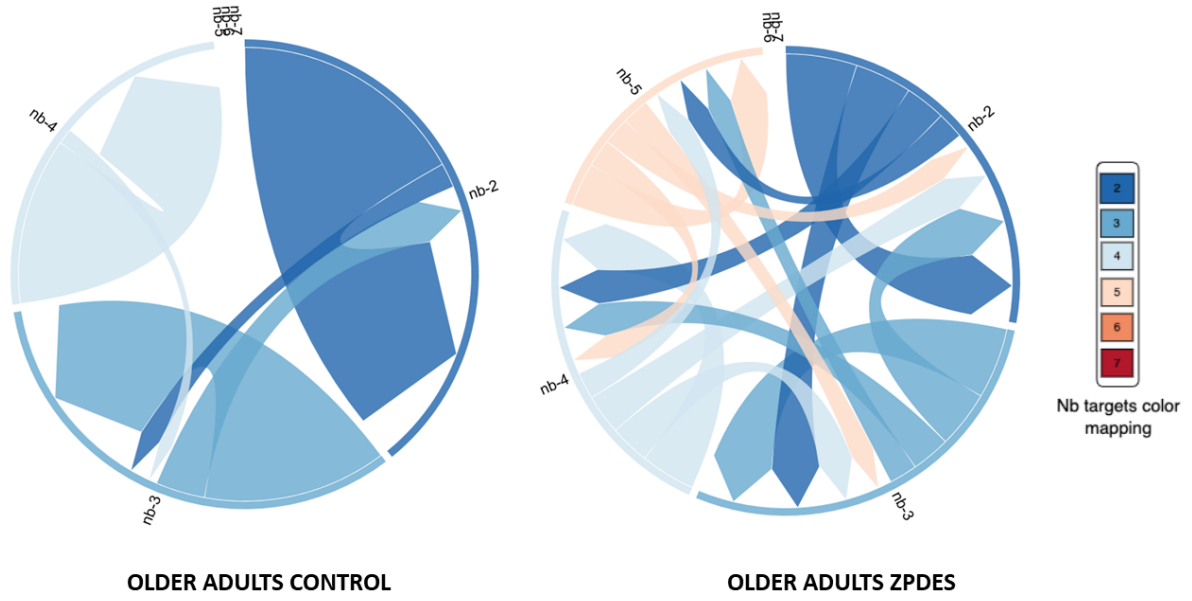


Fig 6.5. Number of transitions between activities in each group for the two studied populations.

significant but ambiguous evidence in the variance of change ($\sigma_s^2=0.08$, $SE=0.04$, $p=0.031$, $BF_{H1}=1.20$). For the control group, results indicate that individuals started on session 1 with a mean performance score of $\mu_i=2.65$ and significantly increased their performance by $\mu_s=0.32$ leading to training performance across the 4 sessions of 2.65 (session 1), 2.97 (session 2), 2.94 (session 3, estimated coefficient $\lambda_3=0.92$) and 2.90 (session 4, estimated coefficient $\lambda_4=0.81$). We found no significant association between intercept and slope ($\sigma_{i,s}=-0.06$, $SE=0.06$, $p=0.268$, $BF_{H1}=1.17$). However there was substantial evidence for individual differences in the variance of baseline performance ($\sigma_i^2=0.11$, $SE=0.04$, $p=0.006$, $BF_{H1}=2.21$) but ambiguous evidence in change ($\sigma_s^2=0.09$, $SE=0.1$, $p=0.398$, $BF_{H1}=1.00$). In sum, the two conditions exhibited a trend toward different performance trajectories: the ZPDES group showed a consistent, monotonic increase in performance over the two weeks, whereas the control group began to plateau at the start of the second week. To illustrate this, Figure 6.6 shows the changes in performance scores across the sessions.

Impact of individual differences We investigated the impact of individual differences to both the initial performance and change through training. Pre-test performance had a significant impact on baseline performance for both the ZPDES group and the control group. Engagement and SDI had a significant impact on baseline performance in the control group. Engagement was the only significant predictor for change in the ZPDES group. No predictor was found significant for change in the control group. Overall, regarding the impact of adding these variables to the model, no credible impact can be reported for any of the predictors ($BF < 3$). All findings are summarized in Table 6.2.

Table 6.1. Models fit indices and comparisons. Significant χ^2 and $\Delta\chi^2$ are in bold (p<0.05).

Model	χ^2	df	BIC	Comparison	$\Delta\chi^2$	Δdf
	ZPDES					
No growth	40.69	8	69.30	—	—	—
Linear Growth	14.63	5	52.89	No Growth vs. Linear	26.06	3
Non linear Growth	2.58	3	47.28	Linear vs Non Linear	12.05	2
	Control					
No growth	37.33	8	29.62	—	—	—
Linear Growth	16.49	5	40.93	No Growth vs. Linear	20.84	3
Non Linear Growth	5.95	3	25.43	Linear vs Non Linear	10.54	2

Table 6.2. Impact of individual differences variables on baseline and change performance. Notes: b are estimated parameters, and p are their respective p-values (assessing null). Predictors are regressed on both change and baseline, therefore model comparison between null (no predictor) and alternative (free estimation of association between individual differences variables and both slopes and initial performance) are provided in the same column.

Individual Differences	Impact on baseline				Impact on change				Model comparison			
	ZPDES		Control		ZPDES		Control		ZPDES		Control	
	b	p	b	p	b	p	b	p	BF_{H1}	BF_{H0}	BF_{H1}	BF_{H0}
Pre-test	0.5	0.00	0.41	0.04	0.03	0.90	-0.20	0.49	1.33	0.75	0.80	1.25
Education level	-0.12	0.58	0.31	0.11	-0.1	0.68	-0.22	0.29	0.77	1.29	0.72	1.39
Engagement	-0.22	0.30	-0.42	0.03	0.9	0.04	0.49	0.19	0.96	1.04	0.75	1.32
Cognitive Load	0.15	0.46	-0.26	0.27	-0.08	0.73	0.01	0.98	0.74	1.21	0.71	1.41
SDI	-0.1	0.65	-0.48	0.01	0.38	0.15	0.66	0.07	0.82	1.34	0.77	1.30

6.3.4 Q3. What is the impact of training conditions on trainee experience in terms of engagement, reported cognitive load and motivation ?

NASA-tlx The load index suggests a significant main effect of time with a slight decrease over time (HDI=[-1.180, -0.20] ; BF=8.77). Most components of the NASA-TLX, displayed a slight but consistent decrease over time, indicating a general reduction in cognitive load (Mental Demand: HDI=[-0.747,-0.439], BF=8.06e+139; Temporal Demand: HDI=[-0.588, -0.229], BF=2918.582; Effort: HDI= [-0.455, -0.151], BF=76.143). A Significant group effect was found for performance dimension, where the ZPDES group showed lowest ratings compared to the control group (HDI=[-2.563, -0.318], BF= 10.652). Furthermore, a significant interaction between the conditions over the time was observed for effort perception (HDI=[0.170,0.579], BF= 35.206), mental demand (HDI=[0.214, 0.630], BF=6.85e+01) and load index (HDI=[0.154, 1.445], BF=4.32).

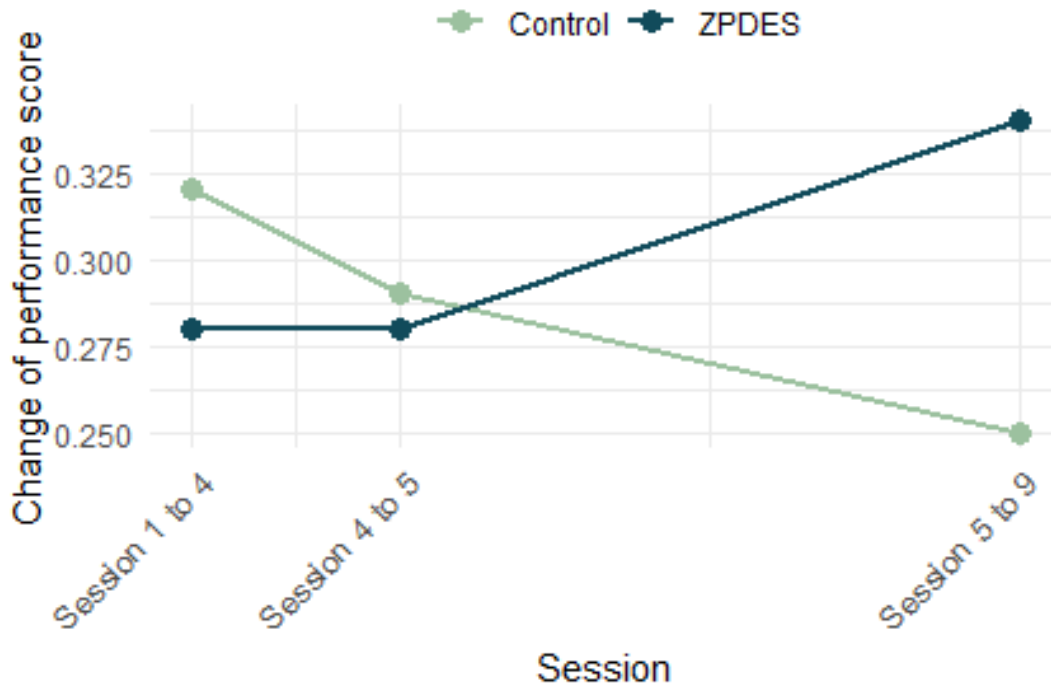


Fig 6.6. Change of performance score across the sessions

SIMS Regarding the type of motivation throughout intervention there were a main effect of the time for the SDI ($HDI=[-1.940,-1.100]$, $BF=3.574e+108$), intrinsic motivation ($HDI=[-0.260,-0.120]$, $BF=3.88e+12$) and identified regulation ($HDI=[-0.285,-0.133]$, $BF=6.29e+23$). There was a group effect, ZPDES group scoring lower External Regulation ($HDI=[-1.841-0.534]$, $BF=78.506$) and for the SDI ($HDI=[-3.0, 0.21]$, $BF=3.53$). No significant interaction was found between the two groups over time.

UES No significant effect of the time and of training condition was observed for the total score of the trainee engagement scale. By contrast, one group effect for the perceived usability of software device ($HDI=[-1.004, -0.166]$, $BF=6.846$), the ZPDES group scoring lower than the control group. No interaction effects were reported for the engagement scores.

TENS A group effect was found for the competence where the ZPDES group has higher level ($HDI=[0.162,0.963]$, $BF=7.324$) than control group. No significant time effects and interaction between the group over the time were noted.

Feeling of Learning Progress There is no significant effect of time, group or time*group for the feeling of learning progress. All subjective dimensions are detailed in Appendix ??.

Thematic analysis : feedback of user experience According to content analysis, four main themes emerged from the participants' feedback. The first theme highlighted difficulties with the cognitive test battery, particularly in understanding certain exercises. One participant expressed that they found the tests "*destabilizing*." The second theme focused on perceptions of attention training. A participant remarked, "*I found it very*

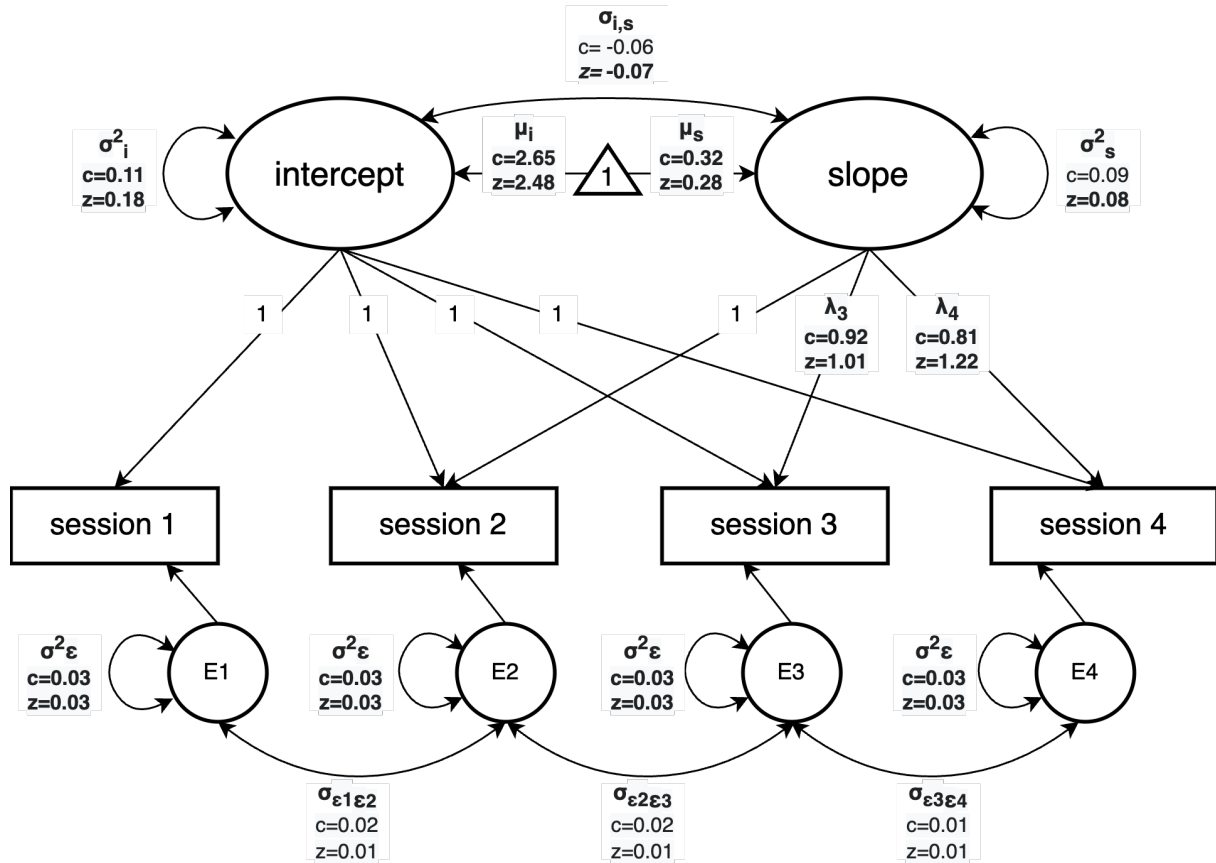


Fig 6.7. on linear latent growth model of performance in intra-training evaluations for both groups. Estimated coefficients are provided for both control (c) and ZPDES (z). Non-null significant ($p < 0.05$) values are displayed in bold.

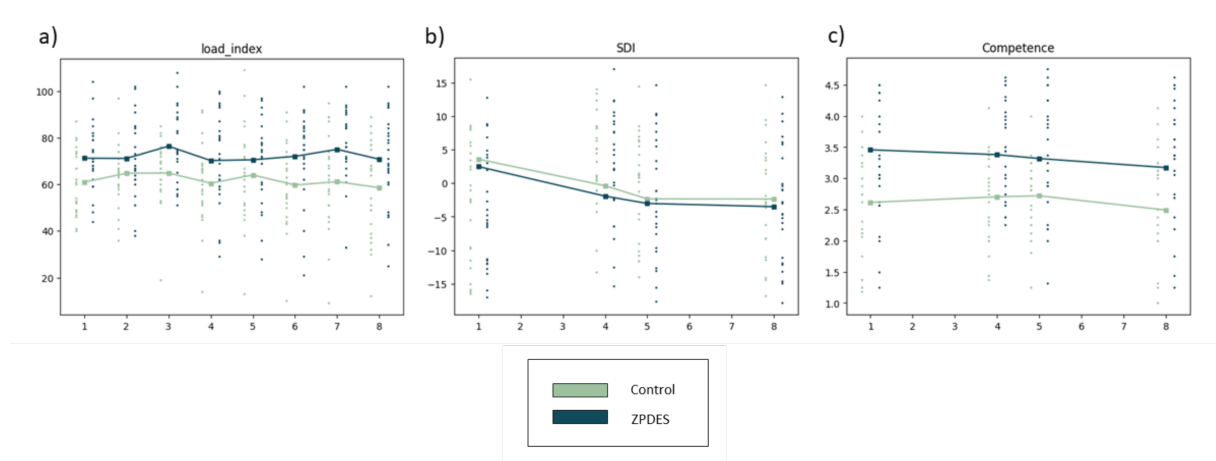


Fig 6.8. Subjective questionnaires with a) load Index related to the NASA-TLX; b) SDI related to the SIMS; d) feeling of competence related to the TENS.

interesting despite being repetitive.” Another participant conveyed frustration with the repetitive nature of the attention task, stating, “*Very interesting but frustrating.*” (as probed by the subscale “perceived usability” from Engagement scale). The third theme encompassed feedback related to the overall trainee experience during the intervention. One example was a participant who said, “*I found the experience interesting and the environment friendly. Thank you for allowing me to participate.*” Finally, a fourth theme

emerged concerning participants' difficulties with computer use. An example of this feedback included: *"Some technical problems due to my screen's incorrect configuration (my error), but overall it went well"*. Thematic differences between the two groups seem to emerge. Participants from the ZPDES group expressed more difficulties in performing the cognitive battery exercises and greater frustrations with the tasks. However, they also provided more positive feedback regarding their overall training experience. It is noteworthy that participants in both groups reported experiencing performance anxiety. One participant stated, *"At the beginning of the experiment, I was confident and relaxed, but as I made more mistakes, I became a bit stressed."*

6.4 Discussion

Based on previous results in young adults, our aim was to study the impact of an individualized procedure, guided by participant learning progress, compared to a classical staircase procedure, on a sample of healthy OA.

6.4.1 Q1. Impact of training (ZPDES vs. Control) conditions on cognitive gains (pre- vs. post-measures)

The first objective was to evaluate how different training conditions (individualized vs. staircase) affect cognitive enhancement, specifically through performance in the MOT task. Both groups demonstrated improvements from pre- to post-test in MOT evaluation. Importantly, the ZPDES group exhibited a greater increase in MOT evaluation. This indicates that both groups enhanced their divided dynamic attention in the MOT task, with the ZPDES group achieving slightly higher gains, aligning with their greater learning gains over eight sessions. This supports that ZPDES worked well for enhanced specific attentional processes involved in MOT tasks. In terms of other extra-MOT outcomes, both training conditions led to similar enhancements, with far-reaching effects on working memory (enumeration and Corsi tasks) and memory (immediate and delayed memorability tests). These improvements were evidenced by gains in both latent factors in PCA and observed variables (see Appendix 6.5 for detailed results for each task of the cognitive battery). However, no improvement was noted in other components, indicating variability in developmental trajectories. The near and far effects obtained for both conditions fitted with previous studies using MOT or related-tasks as Neurotracker where far effects, although real, fall short of the near effects observed (Vater et al., 2021a).

Taken together, these results indicated that ZPDES is enough to promote near effects in OA, but not for eliciting far effects. In other words, compared to previous findings in young participants, OA's ZPDES progress was primarily observed in the specifically trained task, while younger individuals exhibited both near and far effects, including improvements in working memory and various memory dimensions (chapter 5). Therefore, the results in OA are mixed in succeeding increased near effects while failing to increase the far effects.

6.4.2 Q2. Impact of Training Conditions (ZPDES vs. Control) on MOT training Performance

The descriptive analysis of intra-training performance revealed that the ZPDES condition actually led to varied training paths among OA, whereas the control condition resulted

in a uniform training path, as expected. Additionally, ZPDES condition led to higher levels of MOT task difficulty for some OAs than the control condition, revealing that ZPDES is more cognitively challenging and demanding through the exercise adaptation to individual success and capabilities. Taken together, this means that ZPDES behaved as we had expected and already observed in the young adults. Moreover, the bayesian analysis of MOT intra-training performance according to 2 factors-design (Time * group) revealed that both groups demonstrated similar intra-training improvements cross time (time effect) although OAs in the ZPDES group had lower prior performance compared to the control group. Thus, within intra-training performance, no magnification (i.e control group with higher performance change) or compensation (i.e ZPDES group with higher performance change) effect was actually observed across the two groups of OAs.

Regarding the progress trajectory, the intra-training data described a nonlinear progression for both groups across training sessions, but with different trajectories in terms of magnitude of progress from one session to another. As noted in the introduction, the literature acknowledges the heterogeneity of training effects due to age, along with magnification and compensation effects (Guye, De Simoni, and von Bastian, 2017). Our study found nonlinear intra-training progress for both groups, with differing trajectories. In the control condition, participants showed improvement in the first week, which then plateaued throughout the second week. Conversely, participants in the ZPDES group exhibited a monotonic increase in progress over both weeks. Furthermore, a closed eye put on interindividual differences revealed that baseline performance in the first training session for both conditions (ZPDES and control) was influenced by prior MOT performance (pre-test), confirming initial interindividual differences where the participants with the best initial performance were those who performed best at the start of training, regardless of the personalization conditions provided by our two training conditions. Such result supported the cognitive heterogeneity amongst OAs (Guye, De Simoni, and von Bastian, 2017).

Interestingly, engagement was the only factor in the ZPDES condition to show significant inter-individual differences for performance changes (albeit baseline and change models were not significant). This suggests that the ZPDES tutoring strategy, involving random exploration guided by LP, induces a closed link between learning gains and subject's engagement, which is not observed in the control condition. It is important to bear in mind that our measure of engagement is the TENS scale, which comprises 4 dimensions, two of which (Focused attention and Intrinsic reward) directly relate to the motivation and volition involved in self-regulated learning (K. McGrew, 2022). Thus, during training, in the ZPDES condition, and only in this condition, a synergy is created between changes in performance and engagement, as predicted in the Learning Progress hypothesis (Oudeyer et al., 2016b) on which the design of the ZPDES algorithm is based. Consequently, it is possible that some participants might engage more significantly in the training, resulting in substantial change, while others may not, leading to little or no change, as recently pinpointed by the meta-analysis of Z. Li et al., 2024a. In the same vein, another argument to this assumption is the links demonstrated between the perceived need to train one's brain and the power of engagement on training effects (Traut et al., 2021b). Individuals who feel this need are likely to be more motivated to continue the training over the long term compared to those who do not feel the need, with the results of greater training benefits. It is also noteworthy for both training conditions that factors such as education level, and cognitive load, known to impact training in the literature, did not emerge as individual source of differences in our study for any of the groups, replicating similar analyses by Guye, De Simoni, and von Bastian, 2017. These findings should be interpreted

with caution, as the analysis is based on small sample sizes, potentially resulting in low statistical power.

In summary, the ZPDES approach seems to yield no quantitative over-time change in intra-training performance, but to produce qualitative changes with more incremental progress across time compared to those in control condition. The trainee's engagement could play a greater role in qualitative training-related performance changes in ZPDES condition compared to control condition. These latter results deserve deep examination and could be insighted from the analysis of subjective data from the trainees' experience over time.

6.4.3 Q.3. The impact of cognitive load on performance, engagement, competence and motivation during trained tasks.

Cognitive load across time

Several trends were highlighted, including an overall reduction in cognitive load (Nasa-TLX). More specifically, the ZPDES group showed higher levels in dimensions such as mental demand, temporal demand, and effort, which have been identified as proxies in typical learning situations (Efklides, 2006a). The ZPDES group scored lower on the performance dimension than the control group. The participants seem to experience more challenging learning paths, indicating that this type of training can result in perceived more effortful trajectories for the ZPDES group. The lower performance in the ZPDES group may also be related to the variety of difficulty levels presented in the MOT training, which can give participants the impression of increased task difficulty, even if the task is not actually harder. We can ask whether the variability of activities might impact metacognitive predictions and, more specifically, if it led participants to encounter more situations/activities where they experienced difficulties compared to the control group (Ten et al., 2021a). In this way, variability might help individuals gain a better understanding of their objective performance because they were exposed to a greater number of activities and faced more challenges, leading to more frequent failures. In sum, the results appear relatively similar to those of the young adults (chapter 5).

Engagement across time

Paradoxically, participants' engagement in the ZPDES condition remained unchanged despite their increased cognitive load. Such a result in OAs differed from observation in younger individuals, where the trainee engagement decreased with the increasing of cognitive load in ZPDES condition, as predicted by the economic metacognitive view of self-regulated learning by Feldon et al., 2019. As previously assumed, it may be that, as people age, other factors influence the maintenance of engagement, including metacognitive considerations or beliefs about the merits of attentional training for one's own cognitive health (Traut et al., 2021b), as promoted in the recent health policies regarding active aging (WHO, 2014).

Motivation and Self-Determination feelings across time.

Our results suggest trends related to motivation and the feeling of competence. Specifically, there was a decrease in self-determination index and intrinsic motivation (SIMS) for both groups, but a higher level of feeling of self-determined competence among the ZPDES group

(TENS). Furthermore, the ZPDES group experienced a significant decrease in external regulation. Despite the effortful nature of the training task, participants in the ZPDES group could perceive themselves as more competent with a decrease of external regulation, but the demanding nature of ZPDES training may have led to a decrease in intrinsic motivation. These results can be contextualized with metacognitive defaults associated to aging that impact how OA perceive and regulate their cognitive functioning, especially in cognitively demanding situations or tasks where OAs exhibits metacognitive failures for accurately monitoring and assessing their memory, as well as learning performance (e.g., Hertzog et al., 2008). Systematic reviews and meta-analyses have highlighted the benefits of addressing metacognitive dimensions during training, leading to more adaptive cognitive functioning (Sella et al., 2023). In our case, It is difficult to argue that this empowered feeling of competence is linked to performance improvements that are better perceived in the ZPDES condition than in the control condition. Indeed, the subjective measures of learning progress during training did not change as a function of time or training condition, as testimony of aging-related metacognitive defaults, especially in the context of small training-related performance changes. Beside of this, studies have shown that OA who follow a training program at home can increase their self-efficacy (Bailey et al., 2009; West et al., 2008) as they feel more competent through the use of technological tools. In a similar way, ZPDES can lead to a virtuous cycle, where challenging activities make individuals feel more empowered in their MOT- related competency as training progresses.

Overall, the studied aspects of learning experience (cognitive load, motivation, engagement) revealed significant complexity in metacognitive feelings elicited by a CCT, warranting further exploration as it is likely essential for sustained engagement in training (Nahum-Shani et al., 2022).

6.4.4 Lessons learned from the participants' feedback

. Among the participants contacted (a posteriori) to ask for their opinions on their experiences and who agreed to participate, it appeared that the thematic analysis corroborates with the previous findings according to the subjective questionnaires. Several participants met difficulties with the cognitive test battery that was perceived somewhat complex and sometimes without meaningfulness for everyday life. Emotions such as cognitive frustration and anxiety were reported, as measured by the Nasa-TLX (Hart, 2006). These emotions are often exacerbated among OA when using digital tools (Czaja et al., 2019; Quillion-Dupré et al., 2016). Despite these emotional challenges, most participants expressed satisfaction with the training experience in both conditions, which is encouraging for future experiments with CCT.

6.4.5 How can we increase, boost objectify transfer effects?

In light of the differences in progress trajectory into intra-training performance, proposing additional training sessions could elucidate the extent of changes observed in the ZPDES condition compared to the control group. More sessions with ZPDES may yield more incremental progress, and in turn, more opportunities of far and ecological transfers. This strategy may enhance transfer, but also reinforce both LP and metacognition (i.e., awareness of one's learning) with, at the end, perhaps improved feelings of learning and then improved intrinsic motivation, or at least maintained as engagement for a better synergy between immediate motivational states and self-regulated volitional strategies (K. McGrew, 2022). Adding sessions for MOT-based training is also demonstrated as beneficial (Hung

and Seitz, 2014b). Complementary, incorporating breaks for incubation/consolidation of learning, akin to distributed versus massed learning, may be advantageous (see the review, (Son and Simon, 2012)), especially for older adults (Kornell et al., 2010).

To further objectify transfer effects, functional measures such as neurofeedback, specifically EEG, can be employed to mark cerebral changes induced by CT (Park and Reuter-Lorenz, 2009). This method applied to ZPDES condition of MOT-based CCT could enhance perceptual-cognitive processing and could provide objective neurofunctional support of improvements in processing, such as increased processing speed or better management of attentional resources. EEG data analysis could reveal specific brain activation patterns during training, highlighting either compensatory processes (activation of new neural circuits for task performance) or optimization processes, such as improving cognitive performance through more intense activation of task-relevant areas (Salthouse, 1994, see the recent review by (Baykara et al., 2021)). The brain connectome approach can provide additional valuable biomarkers for facilitating learners’ learning processes (see (Taya et al., 2015) for review). Additionally, providing higher quality feedback on participants’ progress, specifically more explicit information about the functional dimensions they are improving, could encourage participants in their learning progress and metacognition. For example, personalized messages such as “You have improved in processing speed” could motivate participants. This strategy might help observe effects on the feeling of progress (balance between easy and hard zones) that did not emerge in our study.

6.4.6 Limitations and perspectives

Despite several strengths of the present study, there are some limitations. In addition to the small sample size discussed in the previous paragraph, it is noteworthy that while averaging across several training sessions improved the robustness of our performance indicators, it inevitably introduced some shortcomings. Although our group sizes were considerably larger than the median group size in the cognitive training literature ($n = 22$; Lampit et al., 2014), they are still fairly small when using SEM. In the presence of small sample sizes, the likelihood of type 2 errors increases (Guye, De Simoni, and von Bastian, 2017). To overcome this limitation, we also evaluated the evidence for and against the existence of links between the individual differences variables and changes in training performance using BFs, as they vary less when power is low (Dienes, 2014). The size of the BFs indicates that our sample sizes were not sufficient to provide strong conclusive evidence. Also, the recruitment bias should be noted, as the participants who underwent attention training were retired individuals enrolled in the Université du Temps Libre (lecture-based courses). These participants are likely aware of the benefits of continuous cognitive stimulation throughout life. This awareness likely contributed to the fact that, despite the perceived time-consuming nature of the MOT training, participants completed the program, resulting in a very low dropout rate. However, we did not evaluate the long-term effects of the training to observe ecological transfer effects in daily life activities. Nonetheless, some qualitative feedback indicated that a few participants felt improvements in attentional activities, such as practicing music.

6.5 Conclusion

Taken together, the ZPDES method represents a new avenue for enhancing training efficacy, providing more tailored individualization for a population with highly varied

needs. Future studies based on similar AI-based training should focus on fine-tuning the ZPDES algorithm, adjusting the training dosage, and improving intra-training assessments using metacognitive scales. This will help better understand the relationships between learning progress and intrinsic motivation among the aging population, as emphasized by recent investigations (e.g., for review, Sakaki et al., 2018). To our knowledge, this is the first time a truly individualized approach based on the LP algorithm has been integrated into a CT study focusing on older adults. Finally, we aim to offer a turnkey training solution for both patients and practitioners.

Appendix

Initial differences and evolution for other tasks

Initial differences were observed with the control group showing likely higher performance at pre-test for the loadblindness (mean difference=-0.1, HDI=[-0.2, 0.0], BF=5.88). Ambiguous evidence of group difference in favor the control participants was observed for three task as followed : moteval (mean difference=-0.05, HDI=[-0.10, 0.01], BF=2.32), enumeration (mean difference=-0.05, HDI=[-0.11, 0.01], BF=2.58), and for reaction times (RT) metric: gonogo (mean difference=-23.40, HDI=[-52.5, 4.37], BF=2.37).

Evolution from pre to post-test was observed for both groups for moteval (ZPDES group strong evidence: mean=0.08, HDI=[0.01, 0.14], BF=7.87; control group weak evidence: mean=0.04, HDI=[-0.01, 0.10], BF=1.88) and memorability (strong evidence for both groups: ZPDES, mean=0.08, HDI=[-0.00, 0.16], BF=4.21; control, mean=0.08, HDI=[0.02, 0.14], BF=12.36).

Evolution from pre to post-test was only observed for the control group with strong evidence for gonogo RT (mean=-36.49, HDI=[-65.58, -8.45], BF=12.97).

Latent Factor Analysis

All results for each principal component of the pre-post difference in the latent space:

Table 6.3. Pre-test Difference and Post-Pre Evolution for Each Component

Component	Group	Pre-test Difference				Post-Pre Evolution			
		d	BF01	t-stats	p-value	d	BF01	t-stats	p-value
PC1	ZPDES	0.44	0.77	1.57	0.12	0.45	404.01	4.79	6.310e-05
	Control					0.65	2175.91	5.62	9.976e-06
PC2	ZPDES	0.47	0.89	-1.68	0.09	0.13	0.28	0.82	0.41
	Control					0.19	0.42	1.24	0.22
PC3	ZPDES	0.11	0.30	0.40	0.68	0.25	0.70	-1.67	0.10
	Control					0.10	0.24	-0.52	0.60
PC4	ZPDES	0.17	0.33	0.61	0.54	0.30	0.56	-1.5	0.14
	Control					0.40	0.48	-1.35	0.18
PC5	ZPDES	0.35	0.53	-1.24	0.22	0.06	0.22	-0.41	0.68
	Control					0.01	0.21	0.08	0.93
PC6	ZPDES	0.07	0.29	-0.25	0.79	0.18	0.28	-0.81	0.42
	Control					0.17	0.31	0.94	0.35
PC7	ZPDES	0.09	0.295	0.32	0.74	0.13	0.23	-0.55	0.58
	Control					0.18	0.24	0.46	0.57
Global Score	ZPDES	0.77	5.96	-2.78	0.007	0.36	9.42	3.12	0.004
	Control					0.61	319.3	4.76	8.316e-05

Questionnaires

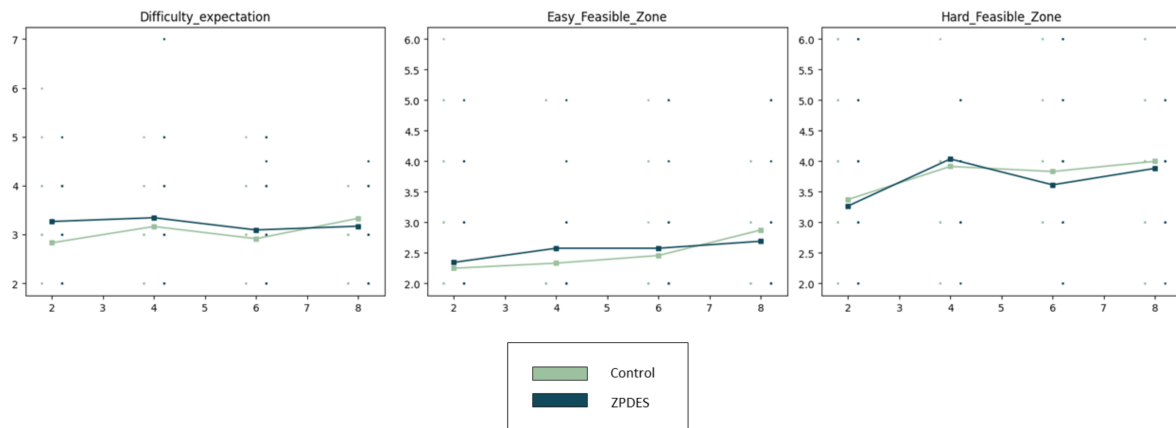


Fig 6.9. Feeling of learning across the training.

Part IV

Discussions and conclusion

Chapter 7

Conclusion and discussion

Aims

After reviewing the literature and evaluating a new individualization procedure in MOT-based training, it is time to conclude this work by reflecting on the various observations we collected.

Contents

7.1 Main Takeaway of the Results from Chapters 6 and 7	222
7.2 General summary of the work	226
7.3 Looking at the LP hypothesis	231
7.4 Perspectives of the work	234
7.5 Limits of the work . . .	240
7.6 Conclusions	243

Abstract

First, a review of the main results collected in Chapters 5 and 6 is presented. Four takeaway posters summarize the primary directions of our analysis: training trajectories, intra-training performance evolution, pre-post cognitive evolution, and questionnaire responses. Next, all results are discussed to provide a comprehensive overview of the key insights derived from the collected data. This leads to a discussion of more fundamental questions regarding the nature of the LP hypothesis and the integration of other components such as cognitive load, metacognition, and personality traits into the model. With these theoretical perspectives, practical recommendations are proposed, including several enhancements to address the limitations of the ZPDES algorithm. Finally, taking a broader view, we outline future projects that stem from this PhD thesis.

Collaborators

Adolphe Maxime, Pech Marion, Oudeyer Pierre-Yves, Delmas Alexandra, Maurel Denis, Sauz on H l ne ¹

¹Following Contributor Roles Taxonomy, all collaborators contributed for "Writing - Review & Editing".

7.1 Main Takeaway of the Results from Chapters 6 and 7

7.1.1 Training Performance

Summary of the analysis aims and main metrics

The primary dimension of our analysis investigates the distinct training trajectories proposed to participants in each group (ZPDES or control). We utilized exploratory visualizations, as shown in Figures 7.1 and 7.2, and computed differences between groups using heuristics that characterize several aspects such as success rates, variability, or trajectory coherence.

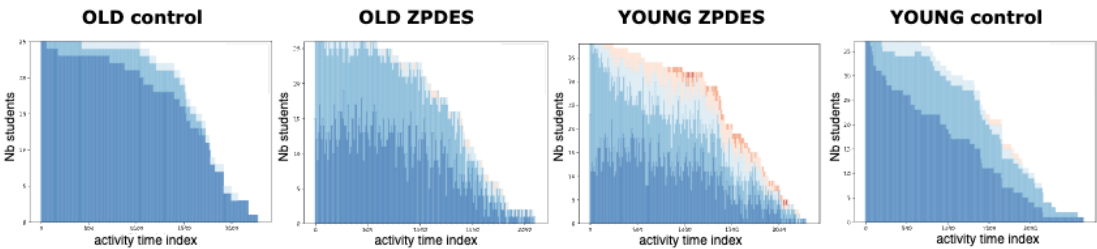


Fig 7.1. Proportion of the population on each different number of targets (2 to 7 targets) through time. More details on this figures can be found in chapter 6.

Types of activity proposed through time:

- Training trajectories exhibit greater variability in the ZPDES groups compared to the control group;
- Both young and older adults in the ZPDES groups are presented with a higher number of targets compared to their respective control groups.

Other general observations on trajectories:

- ZPDES participants showed a generally lower success rate than the control groups;
- ZPDES participants succeeded with a higher number of targets (max=7 for young adults and 5 for older adults) while simultaneously failing at a low number of targets but at very high values of speed, tracking duration, or low spacing between stimuli;
- There was no significant difference between groups in terms of total training duration.

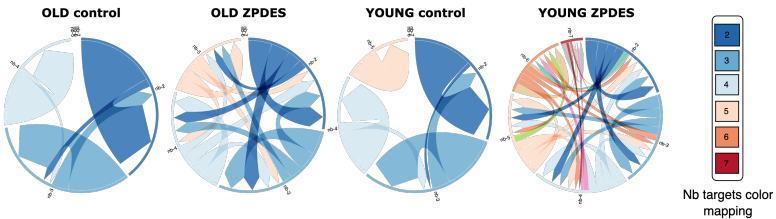


Fig 7.2. Number of transitions between activities in each group for the two studied populations. More details on this figures can be found in chapter 6

Diversity of transition through time:

- Transitions between activities are more diverse in the ZPDES group for both older and younger adults;
- The increased variability of activities in the ZPDES group is also evidenced by the higher number of different transitions.

7.1.2 Intra-training evaluation

Summary of the analysis aims and main metrics

The secondary dimension of our analysis examines the evolution of performance using an evaluation of the MOT task, which was not disclosed to participants. This evaluation was conducted four times during training (sessions 1, 4, 5, and 9). We computed the mean success rate across groups and sessions (Figure 7.3) and fitted a LGCM (Figure 7.4) to identify the type of learning curve, as well as to investigate the impact of individual differences, including potential magnification or compensation effects.

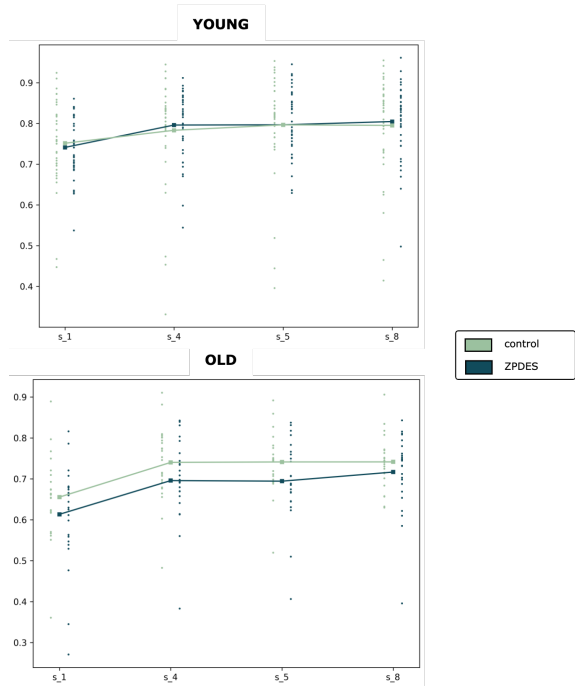


Fig 7.3. Descriptive graph of the success rates for both groups in both studies across the four sessions of intra-training evaluation. Scatter plots represent individual performances.

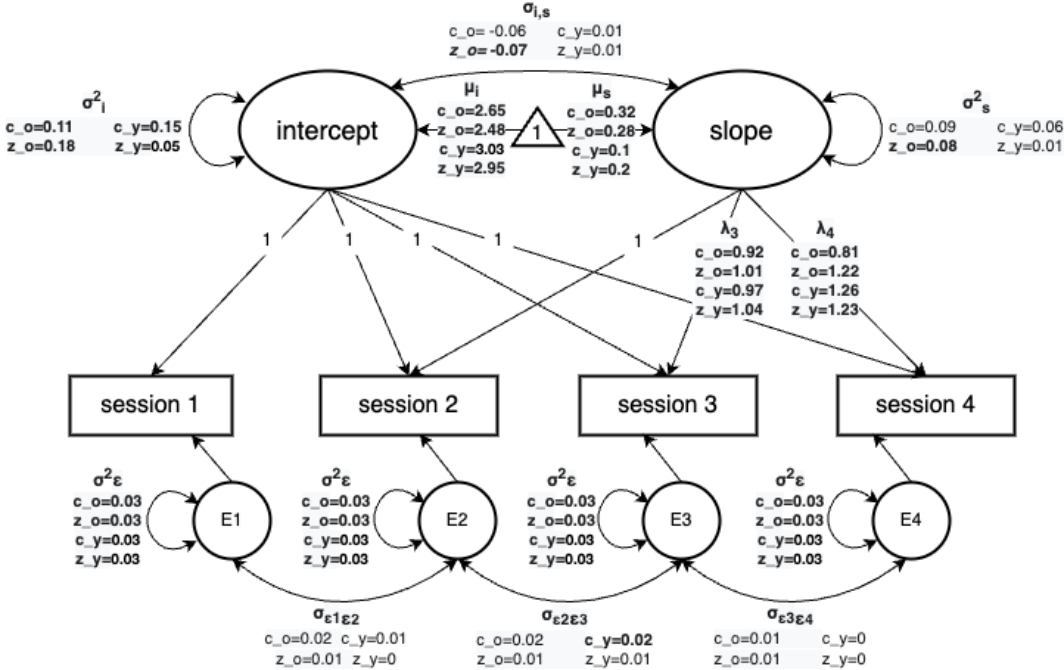


Fig 7.4. Latent Growth Curve Model parameter estimates. Variables from the older adult studies are suffixed with “_o,” while variables from the young adult studies are suffixed with “_y.” Significant non-zero values are bolded. More details on this figure can be found in Chapter 6.

Intra-training performance evolution:

- A general increase in performance is observed for all groups across all studies, though this improvement is not linear over time;
- For older adults in the control group, performance appears to plateau during week 2, while it continues to improve for the ZPDES group;
- A credible, though weak, covariance has been identified in the older adult population, with negative signs indicating a potential compensation effect. This was not observed in the younger population.

7.1.3 Pre-post cognitive evolution

Summary of the analysis aims and main metrics

The third dimension of our analysis evaluates performance evolution using a cognitive battery of seven tasks introduced in Chapter 3, administered before and after training. For each task, we computed the posterior distribution of performance at pre-test, post-test, and the change from pre- to post-test. From these posterior distributions (Figures 7.5 and 7.6), we derived credible intervals to detect significant performance evolution within groups and differences between groups.

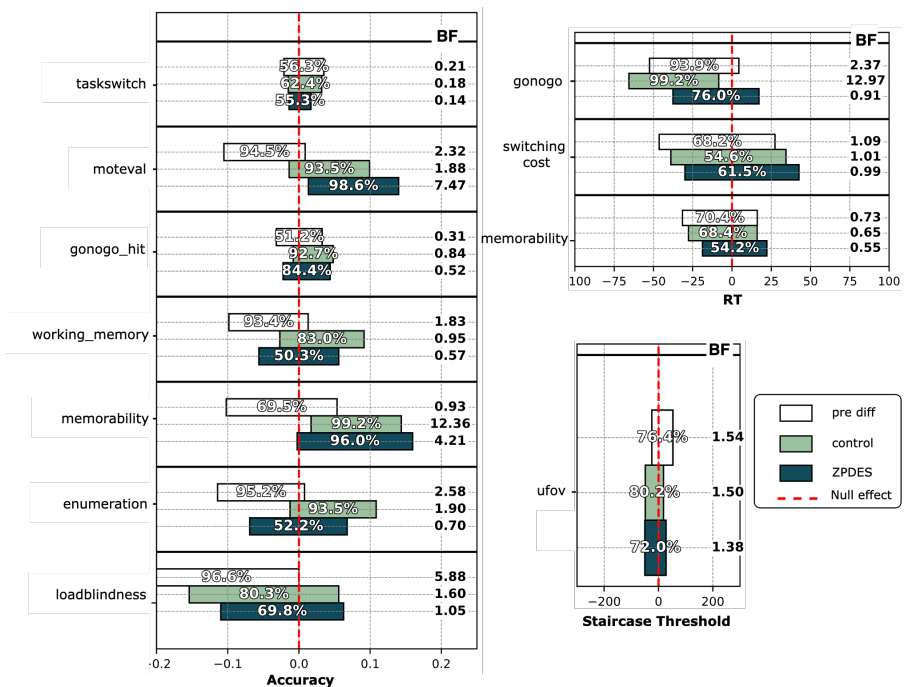
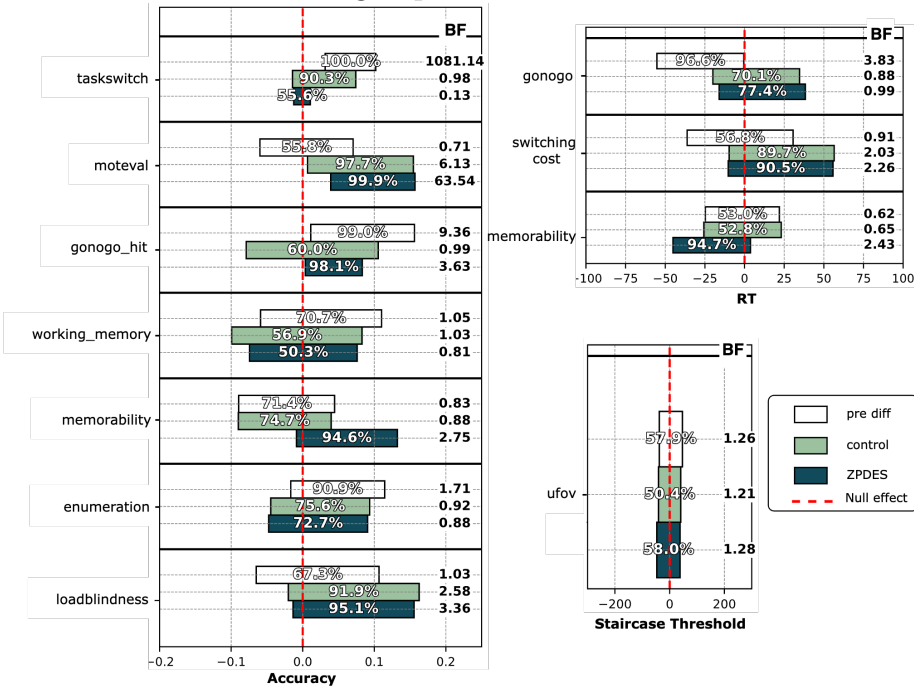


Fig 7.5. Forest plots of posterior distribution of the pre-test performance (white bars) and post-pre evolution (colored bars) for the young population. More details can be found in chapter 5.

Fig 7.6. Forest plots of the posterior distribution of pre-test performance (white bars) and post-pre evolution (colored bars) for the older population. More details can be found in Chapter 6.

Pre-post evolution main outcomes:

- Initial differences were noted at pre-test: young adults in the ZPDES group scored higher on 3 metrics compared to the control group, while older adults in the ZPDES group scored lower on 4 metrics compared to the control group;
- Training improved performance in all groups across all studies for the MOT task;
- In the young sample, the ZPDES group showed broader performance improvement (6 metrics vs. 3 in the control group);
- In the older sample, the control group showed broader performance improvement (4 metrics vs. 2 in the ZPDES group).

7.1.4 Questionnaires

Summary of the analysis aims and main metrics

The fourth dimension of our analysis evaluates self-reported subjective dimensions such as cognitive load, motivation, engagement, and feeling of competence. Questionnaires were administered throughout the week to track changes over the course of the training. We used mixed models to investigate within-group evolution throughout the intervention and differences between groups.

Outcomes from the different questionnaires (rounded numbers in the main text correspond to specific rows in Figure 7.7):.

- ① Cognitive load decreased over time for both groups in both young and older populations, particularly in terms of mental demand and effort;
- ② The use of ZPDES algortihm impacted mental demand and effort. In the young population, there was a main effect of time, while in the older population, there was an interaction effect (group × time). This suggests that for participants using ZPDES, the intervention was generally more demanding or effortful;
- ③ For the performance component, ZPDES participants reported lower levels overall. While the impact of time was absent for older adults, young adults in the ZPDES group reported a stronger increase in performance over time compared to the control group;
- ④ Intrinsic motivation was lower for ZPDES groups in both young and older populations. In the young adult study, ZPDES participants showed lower levels on several components linked to intrinsic motivation (SDI, engagement, perceived usability, focused attention, reward);
- ⑤ Intrinsic motivation decreased over time only in the older adults study;
- ⑥ Competence was higher in the older adults ZPDES group compared to the control group. No group effect was observed in the young study on this dimension. However, the young ZPDES group reported lower levels in the “hard feasible zone,” a proxy for competence.

Questionnaire	Component		OLD	YOUNG	
Nasa-TLX	Load Index	1	-	- +	2
	Mental Demand		-	- +	
	Effort		-	- +	
	Frustration				
	Physical demand			+	
	Temporal Demand		-		
	Performance	3		- - + + +	
SIMS	SDI	4	- -	-	5
TENS	Autonomy			-	
	Relatedness				
UES	Competence	6	+	-	
	Engagement			-	
	Perceived Usability		-	-	
	Focused Attention			-	
LP	Reward			-	
	Diff expectation				
	Easy feasible zone				
	Hard feasible zone	7		-	
Credible effect of group Credible effect of time Credible effect of interaction + Positive coefficient - Negative coefficient					

Fig 7.7. Table summarizing the questionnaire results. Bayes Factors were used to assess the null hypothesis for each parameter estimate of the linear mixed model (see Chapter 5 for details). A credible non-null estimate is defined by a criterion of BF > 1.5.

7.2 General summary of the work

The two experimental studies presented in this dissertation provide an initial understanding of the impact of new individualized procedures on CT based on the MOT task. Previous sections (sections 7.1.2, 7.1.1, 7.1.3, 7.1.4) offer a preliminary overview of the main outcomes from the two experimental chapters. This section aims to summarize and discuss the work more comprehensively. As outlined in chapters 5 and 6, the discussion can be categorized into three main dimensions: (1) analysis of training trajectories and the evolution of task performance, (2) analysis of cognitive performance changes within the pre-post cognitive battery, and (3) the impact on self-reported subjective measures.

7.2.1 Training trajectories

The ZPDES algorithm is distinguished by its ability to create unique, tailored paths for each participant. This “branched-path” design, which contrasts with the linear design of the staircase algorithm (used as a control procedure), results in diverse training trajectories. This diversity is demonstrated by the variety of activities proposed, the proportion of participants engaged in each activity type (categorized by the number of targets), and the number of transitions between activity types.

Furthermore, the ZPDES algorithm generally proposed activities with a higher number of targets compared to the control procedure. This behavior, governed by the rules managing ZPD evolution, contrasts with the control group patterns. The staircase algorithm, by design, can only adjust along a single predefined degree of freedom, leading to significant performance plateaus in the control group. Fine-grained analysis of individual trajectories revealed that participants stuck at certain levels of number of targets progressed once they overcame these steps, indicating that the difficulty did not always increase as expected in the staircase algorithm.

Additionally, our observations replicated previous findings (Alvarez and Franconeri, 2007b) showing that ZPDES participants could reach and succeed at high target levels (up to seven targets) when other parameters were set to easier values. This was particularly evident in our sample of older adults, some of whom achieved the five-target level, which was unattainable for the control group.

This difference in activity proposition warrants further discussion. In chapter 3, we demonstrated that the MOT task’s performance could be unpredictable when multiple parameters are manipulated simultaneously. This complexity was also noted by F. Joessel, 2022a, whose experiments used a staircase procedure to manage the difficulty of a dual MOT task. Their procedure did not show as much learning as other conditions that manipulated fewer parameters.

Moreover, the systematic review in chapter 2 identified several strategies for implementing a branched-path design, primarily relying on reinforcement learning algorithms. The algorithms included in the review managed a very small set of parameters to maintain sample efficiency. For instance, interventions using tabular methods such as Q-learning (Zini et al., 2022) exemplify this approach. By employing a hybrid method that combines expert rules (the model of ZPD) and a hierarchy of multi-armed bandit algorithms, ZPDES represents a first step toward managing larger activity spaces.

Observations of activity space exploration alone are insufficient to conclude the approach’s effectiveness. Our initial attempt to investigate the average success rate of participants during training revealed an interesting trend: the average success rate of ZPDES groups in both samples was lower than that of the control group (approximately

50% versus 66%). However, this lower success rate must be interpreted cautiously, as the trajectories differed between groups and the ZPDES algorithm focuses on learning progress rather than success rate.

Most interventions assessing MOT training impact use intra-training performance as a comparative measure (e.g., Romeas et al., 2016c, Vater et al., 2021d). In our experimental design, direct comparisons are not possible. The staircase algorithm provides an estimate of the maximum difficulty a participant can reach (highest step achieved), but this is not the case for the ZPDES group. To address this issue, previous research (Clement et al., 2015) suggested mapping the activity space to a difficulty space where activities are ordered by difficulty levels. However, our context does not allow for easy prediction of such mappings (e.g., is high-speed/large-spacing more challenging than low-speed/small-spacing?).

Thus, following practices in educational contexts (e.g., Rosenbaum, 1994) and machine learning benchmarks (e.g., Bellemare et al., 2013; Portelas, Colas, Hofmann, and Oudeyer, 2020), we designed a common, blind evaluation for participants. This evaluation consists of 12 activities of varying difficulties (small to large number of targets, low to high speed, small to large spacing) to provide a reference frame for evaluating and comparing mastery between groups.

Is ZPDES a better procedure to enhance intra-training MOT performance?

Results from the intra-training evaluations indicated that all groups improved during training. However, no significant differences in progress were observed between the groups in either study. Weak evidence from LGCM analysis suggested that ZPDES participants, both young and old, continued to improve during the second week of training, unlike the control group. However, this should be considered a trend rather than a definitive conclusion (small sample sizes). Although no notable differences were detected in this evaluation, the methodology holds promise for identifying dynamics within learning trajectories (Byrne and Crombie, 2003). To our knowledge, this approach is not widely used but should be considered by future researchers, particularly in pilot studies, to gain clearer insights into the impact of training.

In practice, the absence of performance differences — despite differences in training trajectories — might be attributed to several features of the evaluation. First, as the evaluation consists of 12 different activities, the general performance score might obscure progress in specific activity types. Investigations for each activity type revealed some progress, but the limited number of trials per activity type (four each) may result in noisy estimates of participants’ actual mastery levels. Additionally, visual parameters varied between participants during the same evaluation session due to the gamified intervention’s randomized visual skins. This variation might have influenced some evaluations, introducing further noise into the measures. These limitations should be considered, especially given the different training trajectories and performance evolution observed during the pre-post evaluation discussed in the next section.

7.2.2 Cognitive Performance

The MOT evaluation task

Previous studies using MOT as a training task have consistently observed performance increases in the trained task (Harenberg et al., 2021; Legault and Faubert, 2012b; Moen et al., 2018b; Musteata et al., 2019b; Parsons and Faubert, 2021; Parsons et al., 2016b;

Tullo, Faubert, and Bertone, 2018b). In our studies (chapters 5, 6), the MOT task used for evaluation within the cognitive battery differed slightly from the training procedure. For example, it included occlusions, and objects followed linear trajectories, whereas training trajectories deviated randomly. The performance increase on this version, which differs slightly from the MOT version used in training, is noteworthy. Previous research Strong and Alvarez, 2017b has shown that variations in trajectory type or retinotopic locations during evaluation typically result in lower performance improvements compared to evaluations conducted under identical settings to the training.

Moreover, for both studies, the ZPDES group showed a slightly stronger performance increase compared to the control groups, as indicated by the HDI of accuracy post-pre evolution. While this observation requires replication and larger sample sizes for validation, it suggests a potentially more suitable curriculum and greater task progress for the ZPDES group. Although all participants exhibited similar progress patterns on the MOT evaluation task, notable differences in other tasks within the battery highlight the nuanced effects of the interventions and the differences between the studied populations.

Transfer effects and impacts of MOT parameters

Investigating other tasks within the cognitive battery revealed differences between groups and studies. For the young sample, both groups improved in the loadblindness and task-switch tasks, whereas no improvements were observed for these tasks in the older sample. This difference is notable, as these tasks require attention abilities (e.g., dividing resources for the dual task in loadblindness or sustained attention for task-switching to ensure objective reconfiguration), cognitive flexibility, and high processing speed. This disparity in training benefits illustrates the complexity of assessing and understanding transfer effects. One possible explanation is the magnification effect, where individuals who already possess some required skills show greater improvements from training. Young adults, who typically have more efficient attention abilities, cognitive flexibility, and processing speeds, are more likely to benefit from such training. In contrast, older adults may lack these foundational skills to the same extent, limiting their capacity to improve through similar training programs (Traut et al., 2021a).

Following an opposite pattern, a difference was found in the memorability accuracy metric, which improved for both groups in the older sample but only for the ZPDES group in the young sample. In Chapter 3, we demonstrated that memorability is correlated with tasks requiring attention, as attention is critical for sustaining focus through blocks and accurately capturing the intrinsic features of images for recall upon second presentation. This suggests a compensation effect, where participants with lower initial performance benefit more from training (Traut et al., 2021a). While young participants in the control group may have begun with higher attention abilities, the potential improvement was too small to show a significant change. However, adding to this complex picture, young adults using ZPDES did benefit from the training. This could indicate that the ZPDES training focuses on specific activities relying on different cognitive components thereby enhancing their performance.

A difference was observed, with performance enhancements noted only for the ZPDES group in the young sample and, conversely, only for the control group in the older sample. Specifically, the older control group showed performance improvements in the enumeration and Go/No-Go (reaction times) tasks. These enhancements may be attributed to improvements in fast and automatic processes. This was not observed in the older ZPDES group, possibly due to the proposed staircase trajectory.

In the control group, plateauing on specific tasks characterized by high speeds, low spacing, long tracking duration, but a low number of targets, could have allowed participants to fine-tune certain automatic processes, increasing efficiency in quick decision-making situations such as enumeration or Go/No-Go tasks. In contrast, ZPDES participants encountered higher variability and might have spent less time fine-tuning these particular processes.

Additionally, the hypothesis of enhancing automatic processes — referred to as “skills” within the Prism model (N. A. Taatgen, 2013) — could be further explored through cognitive modeling. As demonstrated with the Attentional Blindness task in Hoekstra et al., 2022, computational assessments of the operators and skills used to perform a task can clarify the characteristics of the trained components. This would help delineate the concepts and constructs under discussion. Consequently, this idea presents new perspectives and necessitates further research to fully explain our collected data.

While these discussions are interesting, they rely on numerous hypotheses and unobserved concepts without a clear description of the “process” and “cognitive components” (Traut et al., 2021a). Factually, in the study on young adults, ZPDES participants progressed more, while in the older study, the control group showed greater improvement. This observation was further confirmed using PCA to assess performance evolution within each Principal Component, viewed as cognitive constructs. For instance, in the young population, the ZPDES group exhibited broader enhancement across three Principal Components compared to the control group, which improved in only one of the three.

By observing the different impacts of the ZPDES algorithm in young and old samples, it is worth questioning the quality and effectiveness of the individualization provided by ZPDES. We hypothesized that a larger training impact indicates a more adapted training path, suggesting ZPDES’s efficiency in proposing a tailored curriculum. However, it is possible that the greater learning effect in the young ZPDES group was primarily due to the larger variability in the learning trajectory (Raviv et al., 2022). Conversely, this variability might have been counterproductive for the older sample (Lampit et al., 2014d).

The next sections will explore these questions using other strategies. The first strategy involves directly examining the impact of individual differences on learning outcomes, and the second involves studying self-reported measures.

7.2.3 Is ZPDES really providing an individualized path?

To investigate whether ZPDES provides a trajectory tailored to participants’ needs, we can hypothesize that if the trajectories were adapted, initial individual differences should not predict changes in performance. In practice, this statement holds if we consider the objective of individualization as bringing all participants to a minimum amount of performance enhancement, meaning progress is observed for all participants regardless of initial differences. In contrast, other conceptions of individualization might aim to maximize each participant’s potential progress, where initial differences would still predict progress, but all participants would improve to their highest potential. Even though ZPDES assumptions do not imply one or the other conception, we employed a LGCM, a classical strategy for studying performance change when multiple measures are available (Karbach et al., 2017b, Guye, De Simoni, and von Bastian, 2017) to test this hypothesis. In our context, performance was measured by the results of the MOT evaluation through the intervention (four sessions). We observed that the covariance between initial performance and progress (i.e., change over time) was weakly credible only for the older population in both ZPDES and control groups, but not for any group in the young sample study.

These observations indicate larger variability in the older adult groups and suggest a potential magnification effect (of the MOT performance), where covariance was negatively correlated with initial performance, a phenomenon observed in several previous studies (Lövdén et al., 2012b).

Replicating a strategy introduced in Guye, De Simoni, and von Bastian, 2017, we employed an additional method: besides a baseline model fitting only on two latent factors (intercept and slope), we added different covariates such age, motivation or engagement and compared model fits with and without it, investigating estimated coefficient values. Some observations aligned with prior assumptions; for instance, pre-test performance in the cognitive battery significantly predicted baseline performance across all groups (i.e. high performers at the cognitive battery show high performance in the first session of the MOT evaluation). However, the addition of single covariates did not significantly enhance the models, suggesting that individual differences, as considered with our strategy, did not causally impact performance changes. This finding raises questions about the initial inquiry posited at the beginning of the manuscript, where we stated that individualization was needed because initial differences between participants were an important cause of small effects or lack of effects in CT studies.

However, our LGCM analysis should be interpreted with caution. First, some individual difference metrics are aggregate scores that mask the inherent complexity of variables such as pre-test performance and subjective metrics. Second, the small sample size in our study may increase the risk of type II errors. Various rules of thumb for determining the minimal sample size required for robust conclusions include having at least 100 to 200 participants (Boomsma, 1982) or 5 to 10 observations per estimated parameter (Bollen, 2014). In our context, these rules can lead to different guidelines, ranging from 100-200 participants to 70-140 participants, depending on the rule applied. Therefore, power calculations should be tailored to our specific model and require a detailed analysis, which we plan to conduct in future work. As suggested by Wolf et al., 2013, we will use Monte Carlo simulations to gain a clearer understanding of the reliability of our conclusions.

7.2.4 Questionnaires

Although some questionnaires can be criticized for their lack of internal validity and reliability (Taherdoost, 2016), they provide insightful information on the impact on trainee experiences and their related metacognitive states. As illustrated in 7.1.4, both groups in both studies showed a significant decrease in load index. This decrease aligns with the objective performance increases observed in training trajectories and the cognitive battery. Additionally, consistent with objective observations such as lower success rates and higher variability, the ZPDES group in both young and older adults showed a superior cognitive load, indicated by higher senses of effort, mental demand, and physical demand. Notably, this cognitive load decreased less over time for the older ZPDES group revealing the greater cognitive demand of ZPDES condition.

On the motivation and engagement scales, the studies showed slight differences. In the young sample, the ZPDES group, in addition to experiencing higher cognitive load, exhibited generally lower intrinsic motivation (as indicated by SDI or autonomy) and engagement (as reflected in all components of the User Engagement Scale (UES)). In the older sample, the ZPDES group also showed lower intrinsic motivation (SDI or autonomy), but engagement was less impacted (albeit affecting the Perceived Usability component) compared to the young sample. These different results in young and older adults might provide evidence of the evolving role and impact of intrinsic motivation and engagement

with age. As discussed in section 7.3.2, it is possible that beliefs or expectations played a role in the engagement of older participants. For instance, despite the training not being inherently motivating, older adults may have engaged more because they believed the training was beneficial for their health or cognition. In contrast, this was likely less the case for the younger participants.

Feeling of performance (NASA-tlx) and competence (TENS) showed varied behaviors between the studies. In the young adult sample, the feeling of performance increased over the week for the ZPDES group, while it decreased for the control group. This change was not observed in the older sample, where the sense of performance remained stationary. However, the older ZPDES group reported a general lower level of performance than the control group. This feeling of lower performance is consistent with the lower average success rate in training and the differences observed in the pre-post battery. Interestingly, competence was reported higher in the older ZPDES group compared to the control, whereas no difference between groups was found in the young sample.

Given these diverse results in both objective and subjective performances across studies, the next sections will discuss these findings in the context of the learning progress hypothesis, the limitations of our work, and the perspectives they offer.

7.3 Looking at the LP hypothesis

7.3.1 Cognitive Load and Metacognition

The LP hypothesis suggests that curiosity is driven by the need to engage in learnable/empowering activities and that the experience of making learning progress is intrinsically rewarding, thereby reinforcing curiosity and intrinsic motivation to continue the task (Oudeyer et al., 2016b). In our intervention, the ZPDES algorithm operationalizes this idea by employing a multi-armed bandit algorithm guided by an estimation of LP. Our results indicate that the LP hypothesis is influenced by certain factors that are not directly accounted for by our individualized procedure.

The first factor is cognitive load. Our results indicate that the load, effort, and demands (both physical and temporal) were higher in the ZPDES groups. Concurrently, we observed lower motivation and engagement in these groups. Given the lower success rates and higher variability in proposed activities, it is unsurprising that the ZPDES paths were reported as more demanding and effortful. According to the LP hypothesis, ZPDES participants, experiencing higher objective LP, should be more motivated. This pattern suggests that cognitive load potentially moderates the actual LP experience, indirectly affecting the positive feedback loop between progress experience and intrinsic motivation. In cognitive load theory (Sweller, 2011), motivation has traditionally been viewed as a precursor to learning, necessary for engaging in effortful training (Salomon, 1983). However, recent research (see Poupard et al., 2022; Feldon et al., 2019 for reviews) has suggested that levels of cognitive load can impact motivation. This implies that the relationship between motivation and cognitive load may be bidirectional: mental effort is invested when there is sufficient motivational belief, but inadequate results can decrease motivation. We attempted to represent this perspective in Figure 7.8. The arrow from intrinsic motivation to objective learning progress suggests that optimal cognitive load is required to experience learning progress (the historical view of cognitive load as in Salomon, 1983). Indirectly, the arrow from objective learning progress to motivation (via the feedback loop) suggests that cognitive load also moderates motivation levels (the more recent view of the relationship

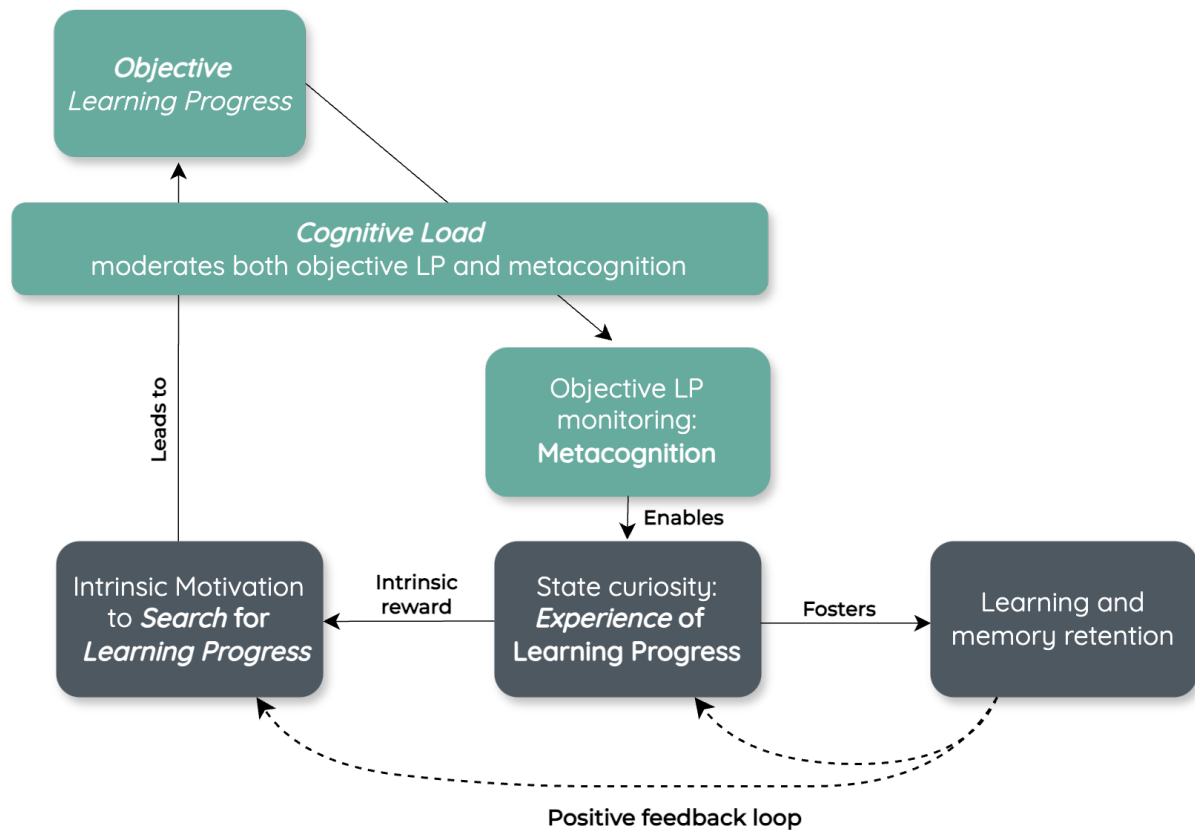


Fig 7.8. A new perspective on the LP hypothesis: Integrating cognitive load and metacognition

between cognitive load and motivation as in Feldon et al., 2019).

The second factor is metacognition. Figure 7.8 introduces another distinction in the model: separating LP into an “objective” component (actual performance evolution in a task) and an “experienced” LP moderated by metacognitive processes. Our results suggest that the higher variability induced by the individualized procedure may have made it more difficult for participants to monitor and estimate their own LP. Practically, we acknowledge that the collected data and experimental design were not sufficient to fully develop a new model, and these components were likely already included in the existing model. For instance, metacognition is often assessed by examining the association between participants’ accuracy in a task and their confidence levels, which are typically evaluated through self-reports (Fleming and Lau, 2014). This method was not utilized in our study. However, Figure 7.8 provides new objectives for operationalizing the concepts into an algorithm. As such, section 7.4 will discuss how these findings might offer new perspectives on enhancing ZPDES or other individualized procedures.

7.3.2 Inter-individual differences

While we have previously explored the connections between cognitive performance, cognitive load, and metacognition, the role of affect and personality traits remains underexamined. This dimension, though not addressed in our work thus far, holds significant potential for further investigation. Indeed, Sakaki et al., 2018 found that curiosity, despite declining with age, predicts healthy aging and correlates with better learning abilities and cognitive functioning in older adults. For instance, curiosity trait encourages individuals to seek

out novel experiences, triggering the production of dopamine and norepinephrine. These neurotransmitters enhance learning by modulating hippocampal activity, which is crucial as the hippocampus is particularly vulnerable to age-related deterioration. In this regard, Sakaki et al., 2018 highlights that such stimulation could prevent the impact of aging on memory. Consequently, in our CT scenario, this dimension could be explored; for instance, curious older adults might exemplify magnification effects. It is thus important to discuss the interplay between cognition, conation (motivation and volition), and affect (emotions, temperament and personality) in our context. This triptych, introduced as the classic trilogy-of-the-mind model by E. R. Hilgard, 1980, can be further examined within a learning context using the Cognitive-Affective-Motivation Model of Learning (CAMML) model introduced by K. S. McGrew, 2022.

The CAMML model integrates several existing frameworks. Firstly, cognition is described according to the CHC taxonomy (see chapter 1, section 1.1.1 for an overview). Secondly, conation, defined as “*the proactive (as opposed to habitual) part of motivation that connects knowledge, affect, drives, desires, and instincts to behavior*” (from *APA dictionary of Psychology*), encompasses both motivation and volition constructs. Motivational constructs can be seen as the drivers of behavior, providing the spark and fuel that propel one forward. In the model, these constructs include achievement orientations (e.g., “Do I want to do this activity? Why? What are my goals?”) and self-beliefs (e.g., “Can I be successful? Am I capable of doing this activity?”). Volition, on the other hand, is the post-decisional, action or performance phase of motivated learning, referred to in contemporary literature as self-regulated learning or action control (K. S. McGrew, 2022). It consists of strategies or mechanisms to self-regulate learning (e.g., “What do I need to do to succeed? How am I doing?”). In sum, conation comprises motivational constructs that initiate the commitment pathway to learning, followed by volition (or self-regulated learning strategies) that direct the course of action to achieve the goal. Finally, affect is described through the Big Five personality traits (Roccas et al., 2002), enhanced with socio-emotional constructs, which assess individuals based on five major dimensions of personality: openness, conscientiousness, extraversion, agreeableness, and neuroticism.

Figure 7.9, taken from K. S. McGrew, 2022, illustrates how these three components connect. There is a first mediating link between affective constructs and conative constructs. This link can be further refined by considering that construct such as open-mindedness (related to curiosity to learn, creativity...) mainly impacts the motivation part of conation (i.e the drivers of motivated learning) whereas conscientiousness (i.e determination, focus, persistence...) mainly impacts the volition part of conation (related to self-regulated learning). As such, the different affective constructs can also be seen as mediating different temporal dynamics of a self-regulated learning. The second link, illustrated in Figure 7.9, involves conation and cognition. Within the cognitive component, following Cattell’s Investment theory (Cattell, 1987), the investment in different cognitive abilities is mediated by motivational constructs, as depicted in the “Personal Investment” section of Figure 7.9.

This model provides insights into the results we collected. For instance, an interesting pattern emerged in the ZPDES groups: young adults exhibited higher cognitive load, lower motivation, and engagement but a stronger sense of performance compared to the control group. In contrast, older adults showed similar trends in motivation and cognitive load; however, they maintained stable engagement (which decreased for young adults) and reported higher competence (which remained stable for young adults) than their control group.

This pattern could be explained by the different impacts of motivational and volitional constructs between age groups. In the young sample, the larger objective impact of ZPDES

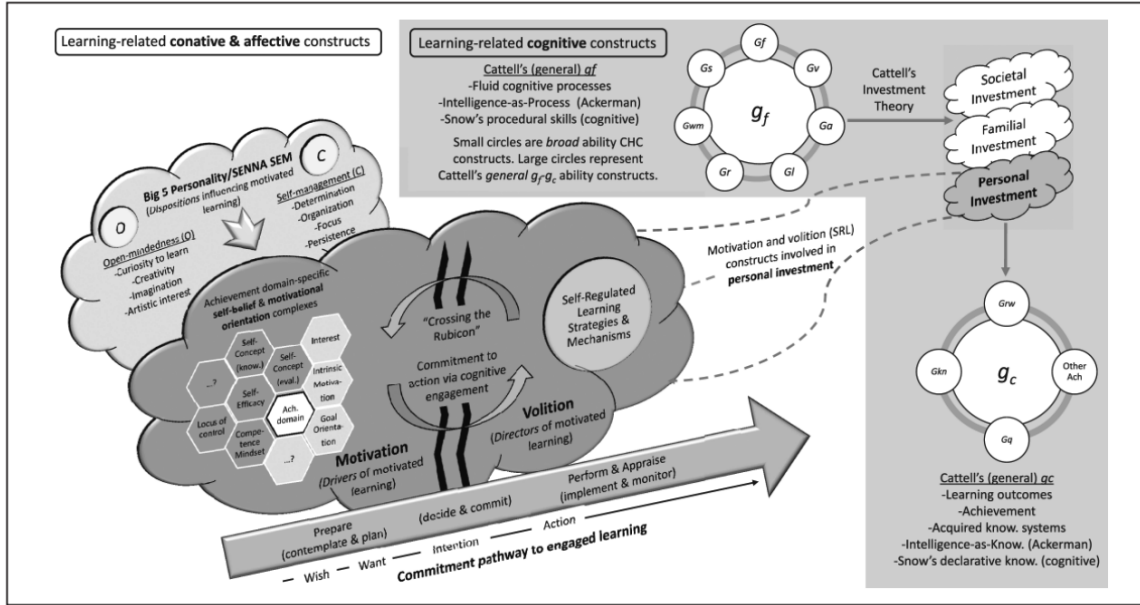


Fig 7.9. The CAMML model, taken from K. S. McGrew, 2022

might be linked to the higher sense of performance and less related to engagement or self-beliefs. Conversely, for older adults in ZPDES, motivational constructs were less important. Previous work has shown that open-mindedness, which mediates motivation in the CAMML model, decreases with age (Sakaki et al., 2018). Consequently, in the older adults group, volition and regulation processes may have played a more significant role in maintaining engagement. This aligns with our LGCM results, where engagement was a weak but significant predictor of change in the ZPDES group. It can then be posited that this causally impacted the lower learning benefit in the older adult ZPDES group. This would be in line with Z. Li et al., 2024b who highlighted in their meta-analysis that engagement is a critical determinant of training efficacy in older adults.

These discussions are theoretical, and no strong evidence allows us to conclude definitively on these observations. However, this preliminary insight suggests that future research could further assess these findings by evaluating affective traits, expectancies, and training efficacy-related beliefs before the intervention. Additionally, as highlighted in the previous section, this discussion underscores how the LP hypothesis is a foundational step toward understanding intrinsically motivated learning and exploration, while also revealing many other dimensions that could be incorporated into the model.

7.4 Perspectives of the work

7.4.1 ZPDES algorithm

Estimation of student's level

An important inquiry for future research on ZPDES is how learner mastery levels are estimated and considered. As discussed in the previous section, high cognitive load negatively impacts motivation and engagement. Therefore, in addition to targeting progressive activities, the algorithm should also consider actual mastery levels or performance in the online generation of the curriculum. In our intervention, mastery level is only indirectly

considered. The multi-armed bandit algorithm sampling allows for the automatic detection of several learning situations related to mastery levels, such as unfeasible activities, mastered choices (i.e., null or low learning progress), or drops in performance due to forgetting or fatigue (i.e., non-null learning progress). However, this mechanism alone does not provide a clear boundary between mastered and unmastered activities and does not directly predict participants’ actual performance. Therefore, we employed a second mechanism—a model of ZPD—to eliminate activities that reach a certain success rate threshold, defining them as “mastered.” Even with the deletion of mastered activities, we could consider performance directly as an additional objective in the multi-armed bandit setting. This idea was tested in Ten, 2022; Ten et al., 2021b by fitting several computational models on a free-choice exploration task. The authors showed that the best-fitting model was a bivariate one defined by both a learning progress component and a Percent Correct component. This suggests that in exploratory tasks, participants are sensitive to both learning progress and their actual performance.

As such, adding a second component to the objective of ZPDES exploration, such as ensuring an overall minimal success rate, might be an avenue to explore. This idea connects the LP-based procedure with other adaptive learning approaches that set a criterion for minimal success rate, also known as the 85% optimal rule for learning. This 85% objective—tightly related to the concepts of optimal difficulty and flow (i.e. the “Goldilocks zone” (Kidd et al., 2012)) — has been demonstrated to be the optimal training accuracy in binary classification tasks with gradient-based learning algorithms and other neural network-based strategies (Wilson et al., 2019).

This objective could be added as a general component describing a sense of overall performance (i.e., not computed independently for each activity but over all activities proposed within a block of several trials) that ensures a minimal level of performance or self-efficacy, defined as an individual’s belief in their ability to successfully execute tasks. In other words, in this proposition, the LP component—computed for each activity—would ensure micro-adaptation (driven by high LP activities), while the performance component would ensure longer-term and macro-adaptation (typically at the scale of a block of activities).

Additionally, the model of ZPD is useful for restricting the exploration of the activity space, but deletion in ZPD can be sample inefficient. Activities nearing their threshold are proposed less frequently because their learning progress decreases, reducing their probability of being selected. A classical approach to estimate learner mastery is Bayesian Knowledge Tracing (BKT) (Pelánek, 2017b), a hidden Markov model that models a binary state of mastery (often denoted as θ) for an activity. In BKT, an activity can initially be mastered with probability P_i and, if not already mastered, can transition to a mastered state with probability P_l . In an unmastered state, the learner has a probability P_g to correctly answer, whereas in a mastered state, the learner has a probability P_s to fail the activity. After the learner practices the activity, Bayes’ rule is applied to update the state of mastery for that activity (Corbett and Anderson, 1994). Similarly, Shabana et al., 2022 proposed using ZPDES with a change point detection algorithm to establish clear boundaries between mastered and unmastered states. This approach can potentially improve the efficiency of mastery estimation in the ZPDES framework.

An additional perspective to explore in ZPD management is the integration of expert knowledge into the temporal dynamics of the generated sequence. Participants may struggle to accurately assess their mastery levels when faced with highly variable sequences of difficult and easy activities. Currently, ZPDES does not impose constraints on the pedagogical sequence, except for enforcing the same number of targets for three consecutive

trials to reduce variability. However, adding other mechanisms to overly stabilize the sequence might negate the benefits of variability. Thus, mechanisms to better control the presentation dynamics of already mastered activities could be beneficial. Presenting these activities periodically can reinforce participants’ mastery and LP self-assessment while providing less intensive tasks, which has been shown to be advantageous for learning (Hung and Seitz, 2014a). Currently, the algorithm randomly selects activities with low learning progress (including mastered ones) without controlled timing. To address this, Mu et al., 2018 proposed a forgetting mechanism that suggests activities outside the ZPD when their memory trace falls below a certain threshold.

Efficient description of the activity space

Staircase strategies are limited by the number of parameters managed concurrently. The ZPDES approach, with its two-level hierarchical graph structure, allows for more flexible exploration of the activity space. However, this representation has limitations because it assumes that parameters in the lower hierarchical levels (speed, spacing, tracking duration, and probe duration) are independent. This independence assumption is problematic as it introduces noise into the learning progress estimations computed by each bandit associated with the different parameters.

An interesting perspective to tackle this issue comes from Automatic Curriculum Learning strategies (Bengio et al., 2009; Portelas, Colas, Weng, et al., 2020, that is to say machine learning methods used to automatically find learning trajectories for artificial agents by challenging them with tasks adapted to their capacities. Specifically, Portelas, Colas, Hofmann, and Oudeyer, 2020 proposed ALP-GMM (Absolute Learning Progress with Gaussian Mixture Model) an algorithm able to select high learning progress tasks in continuous multidimensional activity spaces. After an initialization phase of random sampling uniformly across the activity space, a gaussian mixture model is fitted on the observed proposed tasks concatenated to their respective ALP measure. In their scenario, ALP is computed as the reward difference between the last sampled activity and the closest (in space) previously sampled activity. GMM allows to detect different niches of progress: fitted Gaussians are randomly chosen following a bandit scheme where mean absolute progress is used as the utility of each arm. An activity is finally sampled with the chosen Gaussian. After a certain period of time, GMM are fitted again to adjust with agents non stationary mastery level. While this algorithm showed to be efficient with Deep-Reinforcement Learning agents, several aspects would need to be resolved to be used within human time scales. However, easy adjustments could be thought such as discretization of activity spaces (several parameters are discrete by nature or assumptions of continuity can be hold knowing limits of human perceptions) or even the use of a model of ZPD (as done in ZPDES) to restrict exploration to only a subsets of interesting activities.

7.4.2 Perspectives on future individualized training

Metacognitive consideration

Our results highlight a situation where cognitive load was high and intrinsic motivation and engagement were low, despite high objective LP—contrasting with predictions of the LP hypothesis. We discussed various aspects that might moderate these outcomes, focusing on the impact of metacognition as a crucial element to monitor or feel objective LP. Thus,

it is worth asking whether tailoring learning should occur at both the “cognitive” level and the “metacognitive” level for a given learning context. Several interventions have previously shown that metacognition is trainable Abdelghani, Law, et al., 2023; Callender et al., 2016; de Boer et al., 2018. For instance, in Abdelghani, Law, et al., 2023, the authors proposed a series of 8 sessions to 8-10-year-old children. These sessions included animated videos presenting declarative knowledge about curiosity and metacognitive skills, as well as practice sessions to apply these skills during a reading-comprehension task. The results showed a positive impact on children’s metacognitive efficiency and their ability to express curiosity through question-asking behaviors.

Additionally, as metacognition enhancement has been shown to improve task performance (Callender et al., 2016), it could be an additional pedagogical objective of training interventions. To implement this idea, several strategies would need to be developed to estimate metacognitive efficiency (Fleming and Lau, 2014) before, during, and after a given learning activity (Ozturk, 2015). For instance, B. Clément et al., 2024 proposed enhancing the existing interaction loop (i.e activity proposition - answer - ZPDES update - activity proposition) with a direct query to the learner to collect a declarative statement of their needs. This strategy could be extended to include measures of confidence as proxies for metacognitive sensitivity (Fleming and Lau, 2014) and to use this signal in the sampling decision.

To construct a curriculum that also trains metacognitive processes, scaffolding and individualization should be considered. Indeed, it is not always evident a priori if a specific activity will be useful for developing metacognition. For instance, an unfeasible activity that is far above the student’s current level may be useless for mastery. However, encountering such activities can enhance the student’s ability to detect unfeasible tasks, which is a valuable metacognitive skill. Therefore, when developing metacognitive skills training, an optimal zone for metacognitive development should be considered. Previous works have proposed that optimal activities for metacognitive development overlap with the ZPD (Baker, 1994; Metcalfe, 2009). This idea is exemplified in 7.10 where the ZPD overlaps with a zone where metacognitive development is also optimal. However, recent work by (Abdelghani, Law, et al., 2023), has shown that certain metacognitive training can lead to progress in metacognitive efficiency without impacting learning performance. This raises the question of the actual overlap between the ZPD and an optimal metacognitive developmental zone, as proposed in Figure 7.10. More research is needed to understand how to integrate components of metacognitive training into classical training. In this vein, some previous ITS have attempted to incorporate these dimensions to adapt training (V. A. Aleven and Koedinger, 2002; Mathan and Koedinger, 2005). These procedures often rely heavily on language, as they require learners to self-explain their progress and experience. As such, language-based strategies enabled by Generative Artificial Intelligence (GAI) offer new avenues in this direction. The next section will discuss these possibilities.

Generative AI context

Recent advances in deep learning that can convincingly generate text (OpenAI et al., 2024; Zhao et al., 2023), images (Betker et al., n.d.; Rombach et al., 2022), or videos (Ho et al., 2022; Singer et al., 2022), known as Generative Artificial Intelligence (GAI) tools, are promising for the future of educational technologies (Chan, 2023). Particularly, the use of Large Language Models (Large Language Model (LLM))—autoregressive models trained to predict the next word in a sequence given the previous words (Brown et al., 2020; Vaswani et al., 2023)—has gained significant interest for their capacity to provide

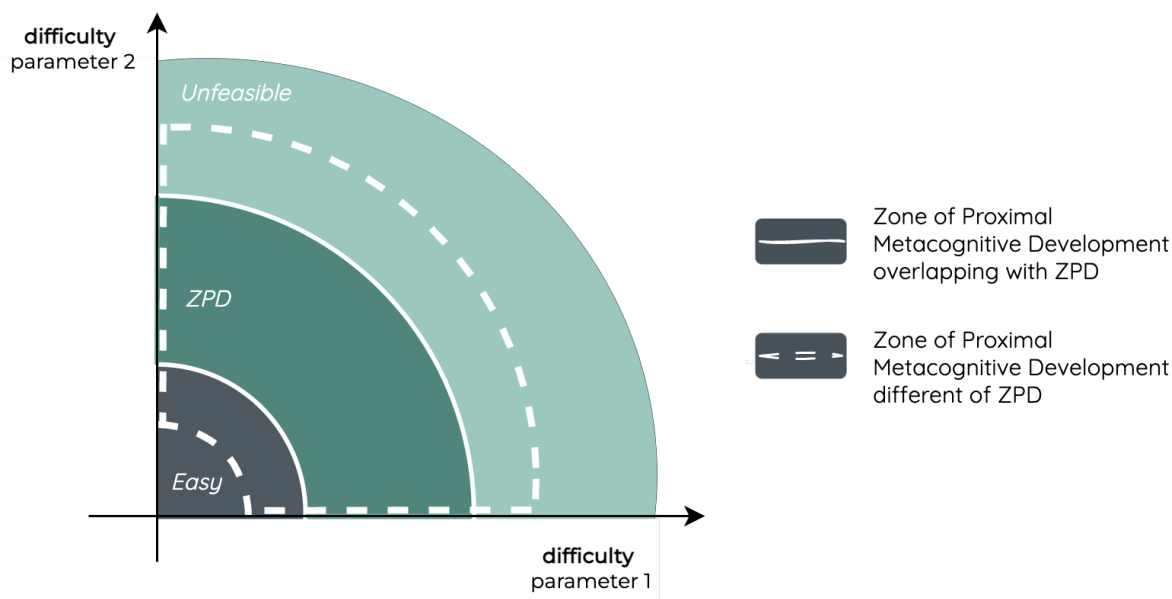


Fig 7.10. An illustration of two possible hypotheses: a ZPD matching an optimal metacognitive development zone (solid white borders) and an overlapping but larger metacognitive development zone (dashed white borders).

learners with natural language interfaces (Andreu and Palmeira, 2024; Baskara et al., 2023; Firaina and Sulisworo, 2023; Kiryakova and Angelova, 2023; T. Wang et al., 2023). For instance, EdTech companies are already selling LLM-based products such as Khanmigo, an AI virtual teacher at Khan Academy (<https://www.khanmigo.ai/>), and a RolePlay feature in Duolingo (<https://blog.duolingo.com/duolingo-max/>), which engages learners in everyday situations to practice real-world conversations.

In early ITS, the development ratio, that is, the number of hours required to author one hour of instructional content, was estimated to vary between 200 and 300 hours for one intervention (Schmucker et al., 2023). Therefore, tools that help teachers build curriculum and exercises can be incredibly beneficial. For instance, Hwang et al., n.d. propose using GPT-3.5 to generate multiple-choice questions and a strategy to validate the generated material following Bloom’s taxonomy. Similarly, H. A. Nguyen et al., 2022 provide heuristics and methods for human validation, while Pourcel et al., 2024 offer an LLM-based strategy to generate diverse and qualitative programming puzzles in Python.

In addition to curriculum building, tools that help teachers test pedagogical approaches or new methods are also promising. For example, Markel et al., 2023 propose generating several GPT-simulated students to train teaching assistants, allowing them to practice without the pressure of affecting real students. Another strategy facilitated by LLM is the learning-by-teaching framework, where learners act as teachers. Previous systems relied on predefined text sequences to facilitate dialogs between the learner and the system (E. Law et al., 2020). Enhancing this framework, Schmucker et al., 2023 provide both an LLM-based teacher and student, asking the learner to teach the virtual student while receiving hints or feedback from a virtual teacher.

Moreover, tools for automatically correcting exercises are also in development (e.g Creutz, 2024, S. Moore et al., 2022). For instance, Lee et al., 2022 propose an interface that supports children and parents in rewriting stories together with the help of AI techniques. Furthermore, automatic feedback on tasks with plausible human-like interaction is another valuable application. Roest et al., 2023 use an LLM (GPT-3.5 turbo) to generate next-step

hints for an introductory Python programming course. The generated feedback was often perceived as personalized, useful, and clear by both students and experts. However, some feedback contained misleading information, highlighting the challenge of controlling the accuracy of information provided by these stochastic models.

Finally, predicting student performance from background distal factors, such as academic trajectory or socioeconomic status, and proximal cognitive and non-cognitive features, like student motivation and engagement, is another potential use of GAI. Hayat and Hasan, 2023 propose using a transformer-based architecture (FLAN-T5) for this purpose.

In sum, these new tools are part of major changes in today's education. A case study involving 20 stakeholders in higher education Zhou et al., 2024 already showed that participants valued how LLMs, particularly ChatGPT, provide "autonomy and control" over other AI tools and appreciated the experience being "personalized" and "general-purpose". However, they also expressed ethical concerns regarding "inaccuracies of responses, hallucinations, bias, or privacy".

Moreover, while these tools offer opportunities, they raise questions about their impact on student learning dynamics. For instance, T. Wang et al., 2023 interviewed computer science instructors who expressed concern that first-year students often trust and use LLM-generated code with little consideration. Instructors worried that "students with underdeveloped mental models may experience shallow learning when using AI assistants." Therefore, it remains to be determined how Generative AI can provide a personalized, interactive, and empowering educational environment while avoiding passiveness, loss of curiosity, overestimation of one's competencies, and lack of critical thinking Abdelghani, Wang, et al., 2023.

7.4.3 Enhancing the psychologist' toolbox with AI

```
I am a researcher specialized in cognitive science with a
passion for AI, psychology, mathematics, and educational sciences.
My role is to revise the PhD manuscript of one of my students,
ensuring that the English is corrected and the text is formal.
First, I will give a mark to the corresponding paragraph (from
1: very bad to 5: very good), and I will also comment with
1 sentence on the most important thing to change to make it
perfect. Then, while I will stay close to the original input,
I will make necessary corrections to the English. I won't
hesitate to rewrite some parts to ensure the tone fits with
top scientific journals like Nature or Science (and to overcome
the issue I raised in my comment). The text should be always
clear for readers and I will make sure transition are always
very efficient: the reading should always feel very natural. If
I spot that a paragraph misses transition or is not logical, I
will provide a comment after generating the rephrased paragraph.
I will avoid using words such as 'to delve', 'unravel', and
'intricates'. When providing LaTeX references such as \cite{} or
\ac{}, the output will remain unchanged.
```

Pre-prompt used to make manuscript correction with GPT-4.

GAI has and will continue to have a broader impact beyond the educational setting. Indeed, LLMs provide tools to enhance researchers' work, such as scientific writing. For

instance, the prompt included above was used to ensure correct and clear English while writing this dissertation. Using the ‘myGPT’ feature offered by OpenAI², this prompt allowed for configuring an agent to provide feedback and reformulations of the manuscript content. In addition to this use, LLMs were employed as experts to explore diverse ideas or statistical methods.

Many common routines will be profoundly transformed by the use of GAI. For instance, in chapter 6, we experimented with using GPT-4 as an additional annotator in a qualitative analysis of verbatims collected post-training. This idea has been previously tested (Xiao et al., 2023) and we are currently developing a tutorial to demonstrate how researchers conducting qualitative analyses could benefit from these techniques. As shown in R. E. Wang et al., 2023, such tools could help accelerate repetitive tasks when datasets scale. Similarly, these tools can be leveraged for systematic reviews or meta-analyses. For instance, LLMs can assist in the inclusion/exclusion phase, with humans judging a subset of the collected articles and using inter-rater reliability metrics to assess the approach.

Image generation capabilities will also provide additional tools for researchers. In this work, we used image generators to create several illustrations to gamify our intervention. Additionally, we used models to create various illustrations included in this dissertation.

While these strategies will empower researchers and accelerate repetitive and time-consuming processes, some aspects of the scientific process will face new challenges. For example, during the submission of our systematic review to a journal (cf chapter 2), we encountered LLM-generated reviews. Although we advocate for the common use of GPT for reformulation, all authors agreed on the lack of value in the generated content. Therefore, it is crucial to develop new gatekeeping mechanisms.

7.5 Limits of the work

7.5.1 Ecological validity

Our study faced several limitations regarding ecological validity. Firstly, the relatively small sample size constrained the broad applicability of our findings. Additionally, we could have investigated the reasons behind participant dropouts to gain more comprehensive insights into the study’s dynamics (for instance following the method proposed in F. Joessel, 2022a).

The computerized and online nature of our intervention raised questions about controlling screen parameters. Some calibrations, such as those detailed in Adolphe et al., 2022a, were not ensured.

We also faced issues with participant pseudo-randomization. Pseudo-randomization led to unexpected pre-test differences in the older population, likely due to between-group differences in the study’s duration. This factor should have been accounted for to enhance the study’s validity.

For a more robust qualitative analysis, incorporating open-ended questions more systematically could have been beneficial. Indeed, this method allows for gathering finer-grained feedback on what participants engaged with and could help identify ways to enhance the interventions or better detect the main barriers to our intervention. As discussed in previous sections, leveraging AI to process these responses would have provided deeper insights into participant experiences and perceptions.

²<https://help.openai.com/en/articles/8554397-creating-a-gpt>

The cognitive battery used for assessment posed subjectivity challenges, as we primarily selected activities correlating with the training task. This raises questions about the notion of generalization. Furthermore, the lack of follow-up and absence of real-world assessments, such as daily life questionnaires, limited our ability to evaluate the long-term impacts and practical applicability of the training. Additionally, we did not pre-register our study, which is considered a good practice to enhance transparency and reproducibility in research (P. Simmons et al., 2021). Pre-registration could have provided a clearer framework for our hypotheses, methodology, and analysis plan, reducing potential biases such as selective reporting and data-driven hypothesis adjustments.

In future studies, addressing these limitations by expanding sample sizes, improving randomization techniques, incorporating more comprehensive assessment tools, both qualitative and quantitative, and adopting pre-registration practices will be crucial for enhancing ecological validity and the overall robustness of our research.

7.5.2 Bayesian vs Frequentist view

A significant aspect of our study was the use of Bayesian inference as the main framework for statistical analysis. Unlike the frequentist approach, which relies on long-run frequencies and p-values, Bayesian inference provides a probabilistic framework that incorporates prior knowledge and updates beliefs with new data. Historically, the adoption of Bayesian methods was limited due to the computational intensity of running Markov Chain Monte Carlo (MCMC) simulations (Kruschke, 2010). However, with modern advancements in computing power, these techniques have become more accessible and feasible for extensive analysis.

Despite these advancements, Bayesian methods still face several challenges. One major issue is the general lack of knowledge and understanding around these methodologies. Additionally, there is a notable absence of standardized practices. While some journals are beginning to accept Bayesian analyses (e.g. Kruschke, 2021c), the appropriate methods and standards for conducting such analyses are not always clear. For example, calculating a Bayes factor can be complex and varies significantly depending on the method used, such as methods relying on the difference in Bayesian Information Criterion (BIC) or the Savage-Dickey density ratio. These approaches require a strong foundation in probability and mathematics, posing a barrier to their widespread adoption.

Looking ahead, we anticipate that the increased availability of high-quality learning materials and user-friendly statistical software will facilitate broader adoption and understanding of Bayesian methods. Resources such as JASP and various R libraries are making these techniques more accessible to researchers. As the field continues to evolve, the integration of Bayesian inference into mainstream research practices will likely become more standardized and widely accepted.

7.5.3 Future directions

Starting scientific project

The material developed in Chapter 4 will be re-used for cognitive assessment in a new cohort led by researchers from Bordeaux Population Health. This cohort will include 2,500 children from CP-CE1-CE2 and 2,500 children from 6ème and 5ème grades, aiming to understand the causes and mechanisms leading to the increasing prevalence of myopia and its possible links with the increased use of digital screens. Several longitudinal measures will be

conducted, such as ophthalmological exams, measurements of optical correction (if wearing glasses), visual acuity without correction, and biometric assessments (including axial length measurement). Additionally, demographic questionnaires will gather information on socio-demographic characteristics, places of residence, lifestyle, screen habits, academic achievement, and neurodevelopment. In this context, the cognitive battery will be adapted to meet the needs of these children and serve as a neuropsychological assessment.

Another direction for future research involves applying our design to a population with cerebral small vessel disease (cSVD). In France, cSVD affects over 4 million people over the age of 60 and is strongly associated with dementia, often manifesting as decreased attention and working memory. Currently, there is no effective treatment for these cognitive deficits. However, previous studies, such as (Pantoni et al., 2017), have shown that attention training can benefit attention and working memory and may increase activity in brain circuits involved in cognitive processes. This project is currently led within the new Vascular Brain Health Institute, created in 2024 at Bordeaux. It gathers three research units : (1) Bordeaux Population Health Lab (Inserm and Univ. of Bordeaux), (2) Institute of Neurodegenerative diseases (Neuromodulation and Neuroprosthetics Team) and (3) Inria centre of University of Bordeaux (Flowers and Titan teams). This new research program aims to assess the efficacy of non-invasive brain stimulation (NIBS) with and without MOT-based cognitive training for this population. It is a significant advancement for the ideas presented in this dissertation as it involves studying a clinical population, incorporating additional measures such as EEG recordings (i.e., neurofunctional changes as guide for assessing CT effectiveness), and exploring the potential synergy between individualized training and neuromodulation techniques like transcranial alternating current stimulation.

Using the Work in an Industrial Context: Onepoint

Onepoint, by assembling a dedicated R&D team, aims to propose evidence-based solutions and offer high-standard expertise and counsel to its customers. In addition to its Corporate Social Responsibility research projects, the company has focused on developing enhanced learning programs in various domains such as software development, cloud computing, data management, low code, and even soft skills development (e.g public speaking, stress management...) ³. As part of these efforts, a platform featuring various serious games ⁴ has been developed. Initially, our research project, which focuses on individualized learning procedures, was seen as a potential enhancement to existing training programs. However, serious games typically follow a specific narrative path in their learning material, which has limited the integration of our approach. Despite this, the expertise gained in ITS and educational technologies has facilitated valuable discussions with potential customers, particularly during bid responses and dissemination talks.

Moreover, this work has laid the foundation for future projects inspired by the developed ideas. For instance, an upcoming project will explore the benefits of using LLM to automatically generate flashcards for a spaced repetition study technique (Lu et al., 2021), aimed at helping learners efficiently memorize information. Key research questions include: Can LLM generate diverse and high-quality flashcards? How can instructional design theories be used to control the generation of these flashcards? How can spaced-repetition instructional design be enhanced by considering learner performance signals?

In addition to these scientific inquiries, several industrial perspectives could be leveraged by the company. On the learning side, these strategies could enhance the current training

³<https://www.groupeonepoint.com/fr/ecole-onepoint/>

⁴<https://www.gamabu.com/en/>

offerings by proposing new evidence-based methods to customers. This is particularly relevant as many calls for tenders demand micro-learning strategies, where learners engage in brief, focused sessions (maximum 5 minutes). On the AI side, the methodologies used to assess and develop the tool could be extended to other contexts, such as the automatic generation of presentation slides, or applied to various domains, given the extensive range of training offerings.

7.6 Conclusions

In this dissertation, we embarked on a comprehensive exploration of individualized cognitive training enhanced with AI techniques. Initially, we conducted a literature review on existing individualized cognitive training methods, highlighting the integration of AI to enhance personalization. Subsequently, we focused on the MOT task literature, specifically examining parameter manipulation to understand how different factors can be adjusted for effective training. These first two chapters provided a foundation for understanding the current landscape of personalized CT and identified the key parameters to manipulate in the MOT task. In Chapter 3, we developed and validated a cognitive test battery, which was subsequently used in the following two studies.

We then conducted two studies with similar experimental designs: a pre-post cognitive battery assessment and an eight-hour training program over two weeks. The first study involved young adults, while the second focused on older adults. The main findings revealed general improvements in both groups on the trained tasks and some near transfer effects that varied between the young and old samples. Additionally, we observed that motivation and engagement were lower for participants using the individualized procedure.

These results were further discussed in the discussion section, identifying several areas for improvement in such procedures. This interdisciplinary work underscores how the interconnected fields of AI and cognitive psychology can mutually benefit each other. Our research highlights the importance of considering individual differences in future studies, emphasizing that personalized approaches can enhance the effectiveness of cognitive training. This dissertation serves as a step forward in integrating AI techniques with cognitive psychology, aiming to improve personalized cognitive training methodologies for diverse populations.

By bridging these fields, we have opened new avenues for research and application, striving to create more adaptive and effective training programs. Ultimately, this work not only advances our understanding of cognitive training but also sets a foundation for future innovations that honor the uniqueness of each learner.

Bibliography

- Abdelghani, R., Law, E., Desvaux, C., Oudeyer, P.-Y., & Sauzéon, H. (2023). Interactive environments for training children's curiosity through the practice of metacognitive skills : A pilot study. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 495–501. <https://doi.org/10.1145/3585088.3593880>
- Abdelghani, R., Wang, Y.-H., Yuan, X., Wang, T., Lucas, P., Sauzéon, H., & Oudeyer, P.-Y. (2023). GPT-3-Driven Pedagogical Agents to Train Children's Curious Question-Asking Skills. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-023-00340-7>
- Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., Osthege, M., Vieira, R., Wiecki, T., & Zinkov, R. (2023). PyMC: A modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, 9, e1516. <https://doi.org/10.7717/peerj-cs.1516>
- Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of Working Memory: Differences in Definition, Degree of Modularity, Role of Attention, and Purpose. *Language, Speech, and Hearing Services in Schools*, 49(3), 340–355. https://doi.org/10.1044/2018_LSHSS-17-0114
- Adolphe, M., Pech, M., Sawayama, M., Maurel, D., Delmas, A., Oudeyer, P.-Y., & Sauzéon, H. (2023, December 26). Exploring the Potential of Artificial Intelligence in Individualized Cognitive Training: A Systematic Review. <https://doi.org/10.31234/osf.io/2wg59>
- Adolphe, M., Sawayama, M., Maurel, D., Delmas, A., Oudeyer, P.-Y., & Sauzéon, H. (2022a). An Open-Source Cognitive Test Battery to Assess Human Attention and Memory. *Frontiers in Psychology*, 13, 880375. <https://doi.org/10.3389/fpsyg.2022.880375>
- Adolphe, M., Sawayama, M., Maurel, D., Delmas, A., Oudeyer, P.-Y., & Sauzéon, H. (2022b). An Open-Source Cognitive Test Battery to Assess Human Attention and Memory. *Frontiers in Psychology*, 13, 880375. <https://doi.org/10.3389/fpsyg.2022.880375>
- Adolphe, M., Sawayama, M., Maurel, D., Delmas, A., Oudeyer, P.-Y., & Sauzéon, H. (2022c). An adaptive curriculum design for cognitive training of attention.
- Aiken, C. (1896). *Methods of Mind-training: Concentrated Attention and Memory*. Harper & Brothers Publishers.
- Aleven, V., Mclaughlin, E., Glenn, A., & Koedinger, K. (2016). *Instruction Based on Adaptive Learning Technologies*. Retrieved June 8, 2024, from <https://www.semanticscholar.org/paper/Instruction-Based-on-Adaptive-Learning-Technologies-Aleven-Mclaughlin/f72d4d1a67079d3d64d481d37646b16c1aa2b136>
- Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, 26(2), 147–179. https://doi.org/10.1207/s15516709cog2602_1
- Allaire, J. C., Bäckman, L., Balota, D. A., Bavelier, D., Bjork, R., Bower, G. H., & Zelinski, E. (2014). A consensus on the brain training industry from the scientific community. [Max planck institute for human development and stanford center on longevity.]. <https://longevity.stanford.edu/a-consensus-on-the-brain-training-industry-from-the-scientific-community-2/>
- Allen, J. (2006). Object Tracking Using Camshift Algorithm and Multiple Quantized Feature Spaces.

- Allen, R., McGeorge, P., Pearson, D. G., & Milne, A. (2006). Multiple-target tracking: A role for working memory? *Quarterly journal of experimental psychology*, 59(6), 1101–1116.
- Alloway, T. P., Bibile, V., & Lau, G. (2013). Computerized working memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior*, 29(3), 632–638. <https://doi.org/10.1016/j.chb.2012.10.023>
- Alnawmasi, M. M., & Khuu, S. K. (2022). Deficits in multiple object-tracking and visual attention following mild traumatic brain injury [Number: 1 Publisher: Nature Publishing Group]. *Scientific Reports*, 12(1), 13727. <https://doi.org/10.1038/s41598-022-18163-2>
- Alvarez, G. A., & Cavanagh, P. (2005). Independent Resources for Attentional Tracking in the Left and Right Visual Hemifields. *Psychological Science*, 16(8), 637–643. <https://doi.org/10.1111/j.1467-9280.2005.01587.x>
- Alvarez, G. A., & Franconeri, S. L. (2007a). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, 7(13), 14. <https://doi.org/10.1167/7.13.14>
- Alvarez, G. A., & Franconeri, S. L. (2007b). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, 7(13), 14. <https://doi.org/10.1167/7.13.14>
- Anderson, B. A. (2018). Controlled information processing, automaticity, and the burden of proof. *Psychonomic Bulletin & Review*, 25(5), 1814–1823. <https://doi.org/10.3758/s13423-017-1412-7>
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, Ray. (1995a). Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences*, 4(2), 167–207. https://doi.org/10.1207/s15327809jls0402_2
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, Ray. (1995b). Cognitive Tutors: Lessons Learned. *Journal of the Learning Sciences*, 4(2), 167–207. https://doi.org/10.1207/s15327809jls0402_2
- Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A Theory of Higher Level Cognition and Its Relation to Visual Attention. *Human-Computer Interaction*, 12(4), 439–462. https://doi.org/10.1207/s15327051hci1204_5
- Andreu, J. M. P., & Palmeira, A. L. (2024). Quick review of pedagogical experiences using GPT-3 in education. *Journal of Technology and Science Education*, 14(2), 633–647. <https://doi.org/10.3926/jotse.2111>
- Andrews, M., & Baguley, T. (2013). Prior approval: The growth of bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, 66(1), 1–7.
- Angel, L., & Isingrini, M. (2015). Le vieillissement neurocognitif : entre pertes et compensation. *L'Année psychologique*, 115(2), 289–324. <https://doi.org/10.3917/anpsy.152.0289>
- Annabi, L., & Nguyen, S. M. (2023). Prerequisite Structure Discovery in Intelligent Tutoring Systems. *2023 IEEE International Conference on Development and Learning (ICDL)*, 176–181. <https://doi.org/10.1109/ICDL55364.2023.10364416>
- Archipub (**typedirector**). (2021, January 2). *Publicité chocolat Lanvin avec Dali - 1968* [Video]. Retrieved July 10, 2024, from https://www.youtube.com/watch?v=Impcn_2WcB8
- Armitage, J., & Eerola, T. (2020). Reaction time data in music cognition: Comparison of pilot data from lab, crowdsourced, and convenience web samples. *Frontiers in psychology*, 10, 2883.

- Au, J., Gibson, B. C., Bunarjo, K., Buschkuehl, M., & Jaeggi, S. M. (2020). Quantifying the Difference Between Active and Passive Control Groups in Cognitive Interventions Using Two Meta-analytical Approaches. *Journal of Cognitive Enhancement*, 4(2), 192–210. <https://doi.org/10.1007/s41465-020-00164-6>
- Azeiteiro, F., & Lopes, M. (2019). Between Clones and Snow-Flakes: Personalization in Intelligent Tutoring Systems. In P. Moura Oliveira, P. Novais, & L. P. Reis (Eds.), *Progress in Artificial Intelligence* (pp. 15–26). Springer International Publishing. https://doi.org/10.1007/978-3-030-30241-2_2
- Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated Learning with MetaTutor: Advancing the Science of Learning with MetaCognitive Tools. In M. S. Khine & I. M. Saleh (Eds.), *New Science of Learning: Cognition, Computers and Collaboration in Education* (pp. 225–247). Springer. https://doi.org/10.1007/978-1-4419-5716-0_11
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/s1364-6613\(00\)01538-2](https://doi.org/10.1016/s1364-6613(00)01538-2)
- Baddeley, A. D., & Hitch, G. (1974, January 1). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (pp. 47–89, Vol. 8). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., Siebourg-Polster, J., Steiert, B., & Zhang, J. D. (2020). An Introduction to Machine Learning [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpt.1796>]. *Clinical Pharmacology & Therapeutics*, 107(4), 871–885. <https://doi.org/10.1002/cpt.1796>
- Bahar-Fuchs, A., Webb, S., Bartsch, L., Clare, L., Rebok, G., Cherbuin, N., & Anstey, K. J. (2017). Tailored and Adaptive Computerized Cognitive Training in Older Adults at Risk for Dementia: A Randomized Controlled Trial. *Journal of Alzheimer's disease: JAD*, 60(3), 889–911. <https://doi.org/10.3233/JAD-170404>
- Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking [Publisher: Routledge eprint: <https://doi.org/10.1080/13506280344000158>]. *Visual Cognition*, 10(8), 949–963. <https://doi.org/10.1080/13506280344000158>
- Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Metacognitive Training at Home: Does It Improve Older Adults' Learning? *Gerontology*, 56(4), 414–420. <https://doi.org/10.1159/000266030>
- Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific reports*, 8(1), 1–11.
- Baker, L. (1994, January 1). Fostering Metacognitive Development. In H. W. Reese (Ed.), *Advances in Child Development and Behavior* (pp. 201–239, Vol. 25). JAI. [https://doi.org/10.1016/S0065-2407\(08\)60053-1](https://doi.org/10.1016/S0065-2407(08)60053-1)
- Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, 344(6182), 424–427.
- Ball, K. K., Beard, B. L., Roenker, D. L., Miller, R. L., & Griggs, D. S. (1988). Age and visual search: Expanding the useful field of view. *JOSA A*, 5(12), 2210–2219. <https://doi.org/10.1364/JOSAA.5.002210>
- Ballesteros, S., Rieker, J. A., Mayas, J., Prieto, A., Toril, P., Jiménez, M. P., & Reales, J. M. (2020a). Effects of multidomain versus single-domain training on executive control and memory in older adults: Study protocol for a randomized controlled trial. *Trials*, 21. <https://doi.org/10.1186/s13063-020-04293-3>
- Ballesteros, S., Rieker, J. A., Mayas, J., Prieto, A., Toril, P., Jiménez, M. P., & Reales, J. M. (2020b). Effects of multidomain versus single-domain training on executive

- control and memory in older adults: Study protocol for a randomized controlled trial [Publisher: BMC]. *Trials*, 21. <https://doi.org/10.1186/s13063-020-04293-3>
- Barkley, R. A. (1991). The ecological validity of laboratory and analogue assessment methods of ADHD symptoms. *Journal of Abnormal Child Psychology*, 19(2), 149–178. <https://doi.org/10.1007/BF00909976>
- Barnett, S. M., & Ceci, S. J. (2002a). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637. <https://doi.org/10.1037/0033-2909.128.4.612>
- Barnett, S. M., & Ceci, S. J. (2002b). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, 128(4), 612.
- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(3), 570–585. <https://doi.org/10.1037/0278-7393.33.3.570>
- Bartolomé, A., Castañeda, L., & Adell, J. (2018). Personalisation in educational technology: The absence of underlying pedagogies [Publisher: SpringerOpen]. *International Journal of Educational Technology in Higher Education*, 15(1), 1–17.
- Baskara, F. X., Puri, A., & Wardhani, A. (2023). ChatGPT and the Pedagogical Challenge: Unveiling the Impact on Early-Career Academics in Higher Education. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 5, 311–322. <https://doi.org/10.23917/ijolae.v5i3.22966>
- Bavelier, D., Achtman, R., Mani, M., & Föcker, J. (2012). Neural bases of selective attention in action video game players. *Vision Research*, 61, 132–143. <https://doi.org/10.1016/j.visres.2011.08.007>
- Bavelier, D., & Green, C. S. (2019). Enhancing Attentional Control: Lessons from Action Video Games. *Neuron*, 104(1), 147–163. <https://doi.org/10.1016/j.neuron.2019.09.031>
- Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, 35, 391–416. <https://doi.org/10.1146/annurev-neuro-060909-152832>
- Baykara, E., Könen, T., Unger, K., & Karbach, J. (2021). MRI Predictors of Cognitive Training Outcomes. *Journal of Cognitive Enhancement*, 5. <https://doi.org/10.1007/s41465-020-00188-y>
- Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018a). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, 144(1), 77–110. <https://doi.org/10.1037/bul0000130>
- Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018b). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, 144(1), 77–110. <https://doi.org/10.1037/bul0000130>
- Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018c). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, 144(1), 77–110. <https://doi.org/10.1037/bul0000130>
- Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2023a). Effects of action video game play on cognitive skills: A meta-analysis. Retrieved June 11, 2024, from https://assets.pubpub.org/fn3lcckh/tmb_tmb0000102-51686605934050.pdf
- Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2023b). Effects of action video game play on cognitive skills: A meta-analysis. Retrieved

- July 14, 2024, from https://assets.pubpub.org/fn3lcckh/tmb_tmb0000102-51686605934050.pdf
- Bejjanki, V. R., Zhang, R., Li, R., Pouget, A., Green, C. S., Lu, Z.-L., & Bavelier, D. (2014). Action video game play facilitates the development of better perceptual templates. *Proceedings of the National Academy of Sciences of the United States of America*, 111(47), 16961–16966. <https://doi.org/10.1073/pnas.1417056111>
- Belchior, P., Yam, A., Thomas, K. R., Bavelier, D., Ball, K. K., Mann, W. C., & Marsiske, M. (2019). Computer and Videogame Interventions for Older Adults’ Cognitive and Everyday Functioning. *Games for Health Journal*, 8(2), 129–143. <https://doi.org/10.1089/g4h.2017.0092>
- Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research*, 47, 253–279. <https://doi.org/10.1613/jair.3912>
- Belleville, S., Mellah, S., Boysson, C. de, Demonet, J.-F., & Bier, B. (2014). The Pattern and Loci of Training-Induced Brain Changes in Healthy Older Adults Are Predicted by the Nature of the Intervention. *PLOS ONE*, 9(8), e102710. <https://doi.org/10.1371/journal.pone.0102710>
- Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 41–48. <https://doi.org/10.1145/1553374.1553380>
- Benoit, J. J., Roudaia, E., Johnson, T., Love, T., & Faubert, J. (2020). The neuropsychological profile of professional action video game players. *PeerJ*, 8, e10211.
- Berch, D. B., Krikorian, R., & Huha, E. M. (1998a). The corsi block-tapping task: Methodological and theoretical considerations. *Brain and cognition*, 38(3), 317–338.
- Berch, D. B., Krikorian, R., & Huha, E. M. (1998b). The Corsi Block-Tapping Task: Methodological and Theoretical Considerations. *Brain and Cognition*, 38(3), 317–338. <https://doi.org/10.1006/brcg.1998.1039>
- Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. McGraw-Hill Book Company. <https://doi.org/10.1037/11164-000>
- Bertoni, S., Franceschini, S., Ronconi, L., Gori, S., & Facoetti, A. (2019a). Is excessive visual crowding causally linked to developmental dyslexia? *Neuropsychologia*, 130. <https://doi.org/10.1016/j.neuropsychologia.2019.04.018>
- Bertoni, S., Franceschini, S., Ronconi, L., Gori, S., & Facoetti, A. (2019b). Is excessive visual crowding causally linked to developmental dyslexia? *Neuropsychologia*, 130, 107–117. <https://doi.org/10.1016/j.neuropsychologia.2019.04.018>
- Bertrand, S., Favier, P.-A., & André, J.-M. (2021). Pragmatic Software Maintainability Management Using a Multi-agent System Working in Collaboration with the Development Team. In S. Rodríguez González, A. González-Briones, A. Gola, G. Katranas, M. Ricca, R. Loukanova, & J. Prieto (Eds.), *Distributed Computing and Artificial Intelligence, Special Sessions, 17th International Conference* (pp. 201–204, Vol. 1242). Springer International Publishing. https://doi.org/10.1007/978-3-030-53829-3_21
- Betker, J., Goh, G., Jing, L., TimBrooks, †., Wang, J., Li, L., LongOuyang, †., JuntangZhuang, †., JoyceLee, †., YufeiGuo, †., WesamManassra, †., PrafullaDhariwal, †., CaseyChu, †., YunxinJiao, †., & Ramesh, A. (n.d.). Improving Image Generation with Better Captions. Retrieved July 1, 2024, from <https://www.semanticscholar.org/paper/Improving-Image-Generation-with-Better-Captions-Betker-Goh/cfee1826dd4743eab44c6e27a0cc5970effa4d80>

- Bettencourt, K. C., & Somers, D. C. (2009a). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of vision*, 9(7), 9–9.
- Bettencourt, K. C., & Somers, D. C. (2009b). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, 9(7), 9. <https://doi.org/10.1167/9.7.9>
- Bettencourt, K. C., & Somers, D. C. (2009c). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, 9(7), 9. <https://doi.org/10.1167/9.7.9>
- Bherer, L., Kramer, A. F., Peterson, M. S., Colcombe, S., Erickson, K., & Becic, E. (2008). Transfer effects in task-set cost and dual-task cost after dual-task training in older and younger adults: Further evidence for cognitive plasticity in attentional control in late adulthood. *Experimental Aging Research*, 34(3), 188–219. <https://doi.org/10.1080/03610730802070068>
- Binder, J. C., Martin, M., Zöllig, J., Röcke, C., Mérillat, S., Eschen, A., Jäncke, L., & Shing, Y. L. (2016). Multi-domain training enhances attentional control. *Psychology and Aging*, 31(4), 390–408. <https://doi.org/10.1037/pag0000081>
- Blacker, K. J., Hamilton, J., Roush, G., Pettijohn, K. A., & Biggs, A. T. (2019). Cognitive Training for Military Application: A Review of the Literature and Practical Guide. *Journal of Cognitive Enhancement*, 3(1), 30–51. <https://doi.org/10.1007/s41465-018-0076-1>
- Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476), 307–310.
- BLOOM, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4–16. <https://doi.org/10.3102/0013189X013006004>
- Blumberg, F. (2014). *Learning by Playing: Video Gaming in Education*. Oxford University Press.
- Bokura, H., Yamaguchi, S., & Kobayashi, S. (2001). Electrophysiological correlates for response inhibition in a go/nogo task. *Clinical neurophysiology*, 112(12), 2224–2232.
- Bollen, K. A. (2014, August 28). *Structural Equations with Latent Variables*. John Wiley & Sons.
- Book, S., Jank, M., Pendergrass, A., & Graessel, E. (2022). Individualised computerised cognitive training for community-dwelling people with mild cognitive impairment: Study protocol of a completely virtual, randomised, controlled trial. *Trials*, 23(1), 371. <https://doi.org/10.1186/s13063-022-06152-9>
- Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. *Part I*, 1.
- Boot, W. R. (2015). Video games as tools to achieve insight into cognitive processes. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00003>
- Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, 129(3), 387–398. <https://doi.org/10.1016/j.actpsy.2008.09.005>
- Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects [Publisher: Sage Publications Sage CA: Los Angeles, CA]. *Perspectives on psychological science*, 8(4), 445–454.

- Borella, E., Carretti, B., Cantarella, A., Riboldi, F., Zavagnin, M., & De Beni, R. (2014). Benefits of training visuospatial working memory in young-old and old-old. *Developmental Psychology*, 50(3), 714–727. <https://doi.org/10.1037/a0034293>
- Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, 104(2), 198–230. <https://doi.org/10.1016/j.cognition.2006.05.009>
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241), 177–178. <https://doi.org/10.1038/226177a0>
- Bowers, A., Anastasio, J., Howe, P., O'Connor, M., Hollis, A., Kapust, L., Bronstad, M., & Horowitz, T. (2011). Dynamic Attention as a Predictor of Driving Performance in Clinical Populations: Preliminary Results. *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : driving assessment 2011*, 307–313. <https://doi.org/10.17077/drivingassessment.1413>
- Bowers, A., Anastasio, R. J., Howe, P., connor, M., Hollis, A., Kapust, L., Bronstad, M., & Horowitz, T. (2011). Dynamic Attention as a Predictor of Driving Performance in Clinical Populations: Preliminary Results. <https://doi.org/10.17077/drivingassessment.1413>
- Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the Trail Making Test. *Nature Protocols*, 1(5), 2277–2281. <https://doi.org/10.1038/nprot.2006.390>
- Brainer. (n.d.). Brainer | Riabilitazione Cognitiva. Retrieved June 26, 2023, from <https://www.brainer.it/>
- Bredemeier, K., & Simons, D. J. (2012). Working memory and inattention blindness. *Psychonomic Bulletin & Review*, 19(2), 239–244.
- Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, 6, 63. <https://doi.org/10.3389/fnhum.2012.00063>
- Bronner, G. (2021, January 6). *Apocalypse cognitive*. Humensis.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020, July 22). *Language Models are Few-Shot Learners*. arXiv: 2005.14165 [cs]. <https://doi.org/10.48550/arXiv.2005.14165>
- Bunderson, C. V., & Martinez, M. (2000). Building interactive World Wide Web (Web) learning environments to match and support individual learning differences [Publisher: Association for the Advancement of Computing in Education (AACE)]. *Journal of Interactive Learning Research*, 11(2), 163–195.
- Bürki, C. N., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014a). Individual differences in cognitive plasticity: An investigation of training curves in younger and older adults. *Psychological Research*, 78(6), 821–835. <https://doi.org/10.1007/s00426-014-0559-3>
- Bürki, C. N., Ludwig, C., Chicherio, C., & de Ribaupierre, A. (2014b). Individual differences in cognitive plasticity: An investigation of training curves in younger and older adults. *Psychological Research*, 78(6), 821–835. <https://doi.org/10.1007/s00426-014-0559-3>
- Butler, M., McCreedy, E., Nelson, V. A., Desai, P., Ratner, E., Fink, H. A., Hemmy, L. S., McCarten, J. R., Barclay, T. R., Brasure, M., Davila, H., & Kane, R. L. (2018). Does Cognitive Training Prevent Cognitive Decline?: A Systematic Review. *Annals of Internal Medicine*, 168(1), 63–68. <https://doi.org/10.7326/M17-1531>
- Bylinskii, Z., Goetschalckx, L., Newman, A., & Oliva, A. (2021). Memorability: An image-computable measure of information utility. *arXiv preprint arXiv:2104.00805*.

- Bylinskii, Z., Goetschalckx, L., Newman, A., & Oliva, A. (2022). Memorability: An Image-Computable Measure of Information Utility. In B. Ionescu, W. A. Bainbridge, & N. Murray (Eds.), *Human Perception of Visual Information: Psychological and Computational Perspectives* (pp. 207–239). Springer International Publishing. https://doi.org/10.1007/978-3-030-81465-6_8
- Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015a). Intrinsic and extrinsic effects on image memorability. *Vision research*, 116, 165–178.
- Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015b). Intrinsic and extrinsic effects on image memorability. *Vision Research*, 116, 165–178. <https://doi.org/10.1016/j.visres.2015.03.005>
- Byrne, B. M., & Crombie, G. (2003). Modeling and Testing Change: An Introduction to the Latent Growth Curve Model. *Understanding Statistics*, 2(3), 177–203. https://doi.org/10.1207/S15328031US0203_02
- Callender, A. A., Franco-Watkins, A. M., & Roberts, A. S. (2016). Improving metacognition in the classroom through instruction, training, and feedback. *Metacognition and Learning*, 11(2), 215–235. <https://doi.org/10.1007/s11409-015-9142-6>
- Cardullo, S. (2017). New frontiers in neuropsychology. The Padua Rehabilitation Tool: A new software for rehabilitation using touch-screen technology [Publisher: Università degli studi di Padova].
- Carter, E., & Blank, G. D. (2013). An Intelligent Tutoring System to Teach Debugging [Book Title: Artificial Intelligence in Education Series Title: Lecture Notes in Computer Science]. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39112-5_134
[TLDR] A web-based ITS is discussed to teach introductory level Computer Science students debugging skills, using and teaching case-based reasoning.
- Casutt, G., Theill, N., Martin, M., Keller, M., & Jäncke, L. (2014). The drive-wise project: Driving simulator training increases real driving performance in healthy older drivers. *Frontiers in Aging Neuroscience*, 6, 85. <https://doi.org/10.3389/fnagi.2014.00085>
- Cattell, R. B. (1987, July 1). *Intelligence: Its Structure, Growth and Action*. Elsevier.
- Cavanagh, P., & Alvarez, G. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9(7), 349–354. <https://doi.org/10.1016/j.tics.2005.05.009>
- Cavanagh, P., & Alvarez, G. A. (2005a). Tracking multiple targets with multifocal attention. *Trends in cognitive sciences*, 9(7), 349–354.
- Cavanagh, P., & Alvarez, G. A. (2005b). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9(7), 349–354. <https://doi.org/10.1016/j.tics.2005.05.009>
- Cecala, A. L. (2016). Psychophysics: A Practical Introduction (2nd Ed.) *Journal of Undergraduate Neuroscience Education*, 15(1), R21. Retrieved August 22, 2023, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5105982/>
- Chabay, R. W., & Sherwood, B. A. (1992). A practical guide for the creation of educational software. In *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches* (pp. 151–186). Lawrence Erlbaum Associates, Inc.

- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1), 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Chang, M., D’Aniello, G., Gaeta, M., Orciuoli, F., Sampson, D., & Simonelli, C. (2020). Building ontology-driven tutoring models for intelligent tutoring systems using data mining [Publisher: IEEE]. *IEEE Access*, 8, 48151–48162.
- Chein, J., & Schneider, W. (2012). The Brain’s Learning and Control Architecture. *Current Directions in Psychological Science - CURR DIRECTIONS PSYCHOL SCI*, 21, 78–84. <https://doi.org/10.1177/0963721411434977>
- Choi, Y., Qi, F., Gordon, J., & Schweighofer, N. (2008). Performance-Based Adaptive Schedules Enhance Motor Learning. *Journal of Motor Behavior*, 40(4), 273–280. <https://doi.org/10.3200/JMBR.40.4.273-280>
- Cicerone, K. D., Goldin, Y., Ganci, K., Rosenbaum, A., Wethe, J. V., Langenbahn, D. M., Malec, J. F., Bergquist, T. F., Kingsley, K., Nagele, D., Trexler, L., Fraas, M., Bogdanova, Y., & Harley, J. P. (2019). Evidence-Based Cognitive Rehabilitation: Systematic Review of the Literature From 2009 Through 2014. *Archives of Physical Medicine and Rehabilitation*, 100(8), 1515–1533. <https://doi.org/10.1016/j.apmr.2019.02.011>
- Clare, L., & Woods, B. (2003). Cognitive rehabilitation and cognitive training for early-stage Alzheimer’s disease and vascular dementia [Publisher: John Wiley & Sons, Ltd]. *Cochrane Database of Systematic Reviews*, (4). <https://doi.org/10.1002/14651858.CD003260>
- Clement, B. (2018, December 12). *Adaptive Personalization of Pedagogical Sequences using Machine Learning* [These de doctorat]. Bordeaux. Retrieved April 9, 2024, from <https://theses.fr/2018BORD0373>
- Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2013a). Multi-armed bandits for intelligent tutoring systems. *arXiv preprint arXiv:1310.3174*.
- Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2013b). Multi-armed bandits for intelligent tutoring systems. *arXiv preprint arXiv:1310.3174*.
- Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2015, June 19). *Multi-Armed Bandits for Intelligent Tutoring Systems*. arXiv: 1310.3174 [cs]. <https://doi.org/10.48550/arXiv.1310.3174>
- Clément, B., Sauzéon, H., Roy, D., & Oudeyer, P.-Y. (2024, January 16). *Improved Performances and Motivation in Intelligent Tutoring Systems: Combining Machine Learning and Learner Choice*. arXiv: 2402.01669 [cs]. Retrieved April 8, 2024, from <http://arxiv.org/abs/2402.01669>
- Clément, J. P., Nassif, R. F., Léger, J. M., & Marchan, F. (1997). [Development and contribution to the validation of a brief French version of the Yesavage Geriatric Depression Scale]. *L’Encephale*, 23(2), 91–99.
- CogniFit. (n.d.). CogniFit : Entraînement cérébral - La santé mentale simplifiée. Retrieved June 26, 2023, from <https://www.cognifit.com/fr>
- CogniPlus. (n.d.). CogniPlus - entraînement cognitif de SCHUHFRIED. Retrieved June 26, 2023, from <https://www.schuhfried.com/fr/cogniplus/>
- Collette, F., & Salmon, É. (2014). Fonctionnement exécutif et réseaux cérébraux. *Revue de neuropsychologie*, 6(4), 256–266. <https://doi.org/10.1684/nrp.2014.0321>
- Conati, C., Gertner, A., & VanLehn, K. (2002). Using Bayesian Networks to Manage Uncertainty in Student Modeling. *User Modeling and User-Adapted Interaction*, 12(4), 371–417. <https://doi.org/10.1023/A:1021258506583>

- Conati, C., & Manske, M. (2009). Evaluating Adaptive Feedback in an Educational Computer Game. In Z. Ruttkay, M. Kipp, A. Nijholt, & H. H. Villhjálmsson (Eds.), *Intelligent Virtual Agents* (pp. 146–158). Springer. https://doi.org/10.1007/978-3-642-04380-2_18
- Connors, C. K., Staff, M., Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2000). Connors' continuous performance test ii (cpt ii v. 5). *Multi-Health Syst Inc*, 29, 175–96.
- Cooke, J. R. H., Horst, A. C. t., Beers, R. J. v., & Medendorp, W. P. (2017). Effect of depth information on multiple-object tracking in three dimensions: A probabilistic perspective [Publisher: Public Library of Science]. *PLOS Computational Biology*, 13(7), e1005554. <https://doi.org/10.1371/journal.pcbi.1005554>
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278. <https://doi.org/10.1007/BF01099821>
- Corbetta, M., & Shulman, G. (1998). Human cortical mechanisms of visual attention during orienting and search | Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. Retrieved April 18, 2024, from <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1998.0289>
- Corbin-Berrigan, L.-A., Kowalski, K., Faubert, J., Christie, B., & Gagnon, I. (2018). Three-dimensional multiple object tracking in the pediatric population: The NeuroTracker and its promising role in the management of mild traumatic brain injury. *Neuroreport*, 29(7), 559–563. <https://doi.org/10.1097/WNR.0000000000000988>
- Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88(4), 715–730. <https://doi.org/10.1037/0022-0663.88.4.715>
- Cornsweet, T. N. (1962). The Staircase-Method in Psychophysics. *The American Journal of Psychology*, 75(3), 485–491. <https://doi.org/10.2307/1419876>
- 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. <https://doi.org/10.1017/S0140525X01003922>
- Coyle, H., Traynor, V., & Solowij, N. (2015). Computerized and Virtual Reality Cognitive Training for Individuals at High Risk of Cognitive Decline: Systematic Review of the Literature. *The American Journal of Geriatric Psychiatry*, 23(4), 335–359. <https://doi.org/10.1016/j.jagp.2014.04.009>
- Creutz, M. (2024). Correcting Challenging Finnish Learner Texts With Claude, GPT-3.5 and GPT-4 Large Language Models, 1–10. Retrieved June 25, 2024, from <https://aclanthology.org/2024.wnut-1.1>
- Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety*. Jossey-bass.
- Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. 80(5), 2657–2670.
- Curran, P., Obeidat, K., & Losardo, D. (2010). Twelve Frequently Asked Questions About Growth Curve Modeling. *Journal of cognition and development : official journal of the Cognitive Development Society*, 11, 121–136. <https://doi.org/10.1080/15248371003699969>
- Curran, T., Hills, A., Patterson, M. B., & Strauss, M. E. (2001). Effects of aging on visuospatial attention: An erp study. *Neuropsychologia*, 39(3), 288–301.

- Czaja, S. J., Boot, W. R., Charness, N., & Rogers, W. A. (2019). *Designing for older adults: Principles and creative human factors approaches, 3rd ed.* CRC Press/Routledge/Taylor & Francis Group. <https://doi.org/10.1201/b22189>
- da Silva, T. B. L., Bratkauskas, J. S., Barbosa, M. E. d. C., da Silva, G. A., Zumkeller, M. G., de Moraes, L. C., Lessa, P. P., Cardoso, N. P., Ordonez, T. N., & Brucki, S. M. D. (2022). Long-term studies in cognitive training for older adults: A systematic review. *Dementia & Neuropsychologia*, 16(2), 135–152. <https://doi.org/10.1590/1980-5764-DN-2021-0064>
- De Simoni, C., & von Bastian, C. C. (2018). Working memory updating and binding training: Bayesian evidence supporting the absence of transfer [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: General*, 147(6), 829–858. <https://doi.org/10.1037/xge0000453>
- de Boer, H., Donker, A. S., Kostons, D. D. N. M., & van der Werf, G. P. C. (2018). Long-Term Effects of Metacognitive Strategy Instruction on Student Academic Performance: A Meta-Analysis. *Educational Research Review*, 24, 98–115. <https://doi.org/10.1016/j.edurev.2018.03.002>
- Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668, discussion 692–700. <https://doi.org/10.1037/0033-2909.125.6.627>
- Deci, E. L. (1975). *Intrinsic Motivation*. Springer US. <https://doi.org/10.1007/978-1-4613-4446-9>
- Dehaene, S., & Changeux, J.-P. (1991). The Wisconsin Card Sorting Test: Theoretical Analysis and Modeling in a Neuronal Network. *Cerebral Cortex*, 1(1), 62–79. <https://doi.org/10.1093/cercor/1.1.62>
- de Leeuw, J. R., & Motz, B. A. (2016). Psychophysics in a web browser? comparing response times collected with javascript and psychophysics toolbox in a visual search task. *Behavior Research Methods*, 48(1), 1–12.
- Delmas, A., Clement, B., Oudeyer, P.-Y., & Sauzéon, H. (2018). Fostering Health Education With a Serious Game in Children With Asthma: Pilot Studies for Assessing Learning Efficacy and Automatized Learning Personalization. *Frontiers in Education*, 3. <https://doi.org/10.3389/educ.2018.00099>
- Delvenne, J.-F. (2005). The capacity of visual short-term memory within and between hemifields. *Cognition*, 96(3), B79–88. <https://doi.org/10.1016/j.cognition.2004.12.007>
- Depp, C. A., Harmell, A. L., & Jeste, D. (2014a). Strategies for Successful Aging: A Research Update. *Current psychiatry reports*, 16(10), 476. <https://doi.org/10.1007/s11920-014-0476-6>
- Depp, C. A., Harmell, A. L., & Jeste, D. (2014b). Strategies for Successful Aging: A Research Update. *Current psychiatry reports*, 16(10), 476. <https://doi.org/10.1007/s11920-014-0476-6>
- Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12), 1827–1837. [https://doi.org/10.1016/0042-6989\(95\)00294-4](https://doi.org/10.1016/0042-6989(95)00294-4)
- Deunk, M., Doolaard, S., Smale-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive effects of differentiation practices*. RUG/GION.
- Deunk, M. I., Doolaard, S., Smalle-Jacobse, A., & Bosker, R. J. (2015). *Differentiation within and across classrooms: A systematic review of studies into the cognitive*

- effects of differentiation practices*. GION onderwijs/onderzoek, Rijksuniversiteit Groningen.
- Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Diard, J., Lobier, M., & Valdois, S. (2010). Bayesian Modeling of Human Performance in a Visual Processing Training Software. <https://doi.org/10.1201/EBK1439835111-50>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00781>
- Dixon, W. J., & Mood, A. M. (1948). A Method for Obtaining and Analyzing Sensitivity Data [Publisher: Taylor & Francis _eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1948.10483254>]. *Journal of the American Statistical Association*, 43(241), 109–126. <https://doi.org/10.1080/01621459.1948.10483254>
- Domingues-Montanari, S. (2017). Clinical and psychological effects of excessive screen time on children. *Journal of Paediatrics and Child Health*, 53(4), 333–338. <https://doi.org/10.1111/jpc.13462>
- Doroudi, S., Aleven, V., & Brunskill, E. (2019). Where's the Reward? *International Journal of Artificial Intelligence in Education*, 29(4), 568–620. <https://doi.org/10.1007/s40593-019-00187-x>
- Dünser, A., & Mancero, G. (2009). The Use of Depth in Change Detection and Multiple Object Tracking. *The Ergonomics Open Journal*, 2(1). Retrieved April 18, 2024, from <https://benthamopen.com/ABSTRACT/TOERGJ-2-142>
- Dye, M. W., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision research*, 50(4), 452–459.
- Eayrs, J., & Lavie, N. (2018a). Establishing individual differences in perceptual capacity. *Journal of Experimental Psychology: Human Perception and Performance*, 44(8), 1240–1257. <https://doi.org/10.1037/xhp0000530>
- Eayrs, J., & Lavie, N. (2018b). Establishing individual differences in perceptual capacity. *Journal of Experimental Psychology: Human Perception and Performance*, 44(8), 1240.
- Eayrs, J., & Lavie, N. (2018c). Establishing individual differences in perceptual capacity: Journal of Experimental Psychology: Human Perception and Performance [Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, 44(8), 1240–1257. <https://doi.org/10.1037/xhp0000530>
- Edwards, J. D., Wadley, V. G., Vance, D. E., Wood, K., Roenker, D. L., & Ball, K. K. (2005). The impact of speed of processing training on cognitive and everyday performance [Publisher: Routledge _eprint: <https://doi.org/10.1080/13607860412331336788>]. *Ageing & Mental Health*, 9(3), 262–271. <https://doi.org/10.1080/13607860412331336788>
- Edwards, J. D., Fausto, B. A., Tetlow, A. M., Corona, R. T., & Valdés, E. G. (2018). Systematic review and meta-analyses of useful field of view cognitive training. *Neuroscience & Biobehavioral Reviews*, 84, 72–91.
- Efklides, A. (2006a). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3–14. <https://doi.org/10.1016/j.edurev.2005.11.001>
- Efklides, A. (2006b). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1, 3–14. <https://doi.org/10.1016/j.edurev.2005.11.001>

- El-Sabagh, H. A. (2021). Adaptive e-learning environment based on learning styles and its impact on development students' engagement. *International Journal of Educational Technology in Higher Education*, 18(1), 53. <https://doi.org/10.1186/s41239-021-00289-4>
- Erica. (n.d.). Erica. Retrieved June 26, 2023, from <https://erica.giuntipsy.com/>
- Ericson, J., & Klingberg, T. (2023). A dual-process model for cognitive training. *npj Science of Learning*, 8(1), 1–8. <https://doi.org/10.1038/s41539-023-00161-2>
- Ericson, J. M., & Beck, M. R. (2013). Changing target trajectories influences tracking performance. *Psychonomic Bulletin & Review*, 20(5), 951–956. <https://doi.org/10.3758/s13423-013-0424-1>
- Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a Memory Skill. *Science*, 208(4448), 1181–1182. <https://doi.org/10.1126/science.7375930>
- Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: Some results and methodological considerations [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology*, 79(2, Pt.2), 1–16. <https://doi.org/10.1037/h0026873>
- Eun, S.-J., Kim, E. J., & Kim, J. Y. (2022). Development and Evaluation of an Artificial Intelligence–Based Cognitive Exercise Game: A Pilot Study [Publisher: Hindawi]. *Journal of Environmental and Public Health*, 2022, e4403976. <https://doi.org/10.1155/2022/4403976>
- Facoetti, A., & Molteni, M. (2001). The gradient of visual attention in developmental dyslexia. *Neuropsychologia*, 39(4), 352–357. [https://doi.org/10.1016/S0028-3932\(00\)00138-X](https://doi.org/10.1016/S0028-3932(00)00138-X)
- Faria, A. L., Paulino, T., & Badia, S. B. i. (2019). Comparing adaptive cognitive training in virtual reality and paper-pencil in a sample of stroke patients [ISSN: 2331-9569]. *2019 International Conference on Virtual Rehabilitation (ICVR)*, 1–7. <https://doi.org/10.1109/ICVR46560.2019.8994746>
- Faubert, J. (2013a). Professional athletes have extraordinary skills for rapidly learning complex and neutral dynamic visual scenes. *Scientific reports*, 3(1), 1–3.
- Faubert, J. (2013b). Professional athletes have extraordinary skills for rapidly learning complex and neutral dynamic visual scenes [Publisher: Nature Publishing Group]. *Scientific Reports*, 3(1), 1154. <https://doi.org/10.1038/srep01154>
- Faubert, J., & Sidebottom, L. (2012). Perceptual-cognitive training of athletes [Place: US Publisher: Human Kinetics]. *Journal of Clinical Sport Psychology*, 6(1), 85–102.
- Fechner, H. B., Pachur, T., & Schooler, L. J. (2019). How does aging impact decision making? The contribution of cognitive decline and strategic compensation revealed in a cognitive architecture. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(9), 1634–1663. <https://doi.org/10.1037/xlm0000661>
- Fehd, H., & Seiffert, A. (2008). Attention to the center of the target array during multiple object tracking. *Journal of Vision*, 8(6), 504. <https://doi.org/10.1167/8.6.504>
- Fehd, H. M. (2010). Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, 10(4), 1–13. <https://doi.org/10.1167/10.4.19>
- Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive Load as Motivational Cost. *Educational Psychology Review*, 31(2), 319–337. <https://doi.org/10.1007/s10648-019-09464-6>
- Feldon, D. F., Franco, J., Chao, J., Peugh, J., & Maahs-Fladung, C. (2018). Self-efficacy change associated with a cognitive load-based intervention in an undergraduate biology course. *Learning and Instruction*, 56, 64–72. <https://doi.org/10.1016/j.learninstruc.2018.04.007>

- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2018). Studies of expertise from psychological perspectives: Historical foundations and recurrent themes. In *The Cambridge handbook of expertise and expert performance*, 2nd ed (pp. 59–83). Cambridge University Press. <https://doi.org/10.1017/9781316480748.006>
- Fermé, E., Bermúdez i Badia, S., Sirsat, M., & Almeida, Y. (2020). Ai-rehab: A framework for ai driven neurorehabilitation training-the profiling challenge. *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies, Valletta, Malta*, 845–853.
- Firaina, R., & Sulisworo, D. (2023). Exploring the Usage of ChatGPT in Higher Education: Frequency and Impact on Productivity. *Buletin Edukasi Indonesia*, 2(01), 39–46. <https://doi.org/10.56741/bei.v2i01.310>
- Flanagan, D. P., & McDonough, E. M. (2018, September 17). *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. Guilford Publications.
- Fleddermann, M.-T., Heppe, H., & Zentgraf, K. (2019). Off-Court Generic Perceptual-Cognitive Training in Elite Volleyball Athletes: Task-Specific Effects and Levels of Transfer. *Frontiers in Psychology*, 10. Retrieved March 4, 2024, from <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2019.01599>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00443>
- Föcker, J., Cole, D., Beer, A. L., & Bavelier, D. (2018). Neural bases of enhanced attentional control: Lessons from action video game players. *Brain and Behavior*, 8(7), e01019. <https://doi.org/10.1002/brb3.1019>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975a). “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3), 189–198. Retrieved June 19, 2024, from https://www.academia.edu/download/48794085/0022-3956_2875_2990026-620160913-28497-y2c5d5.pdf
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975b). “mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. 12(3), 189–198.
- Franceschini, S., & Bertoni, S. (2019). Improving action video games abilities increases the phonological decoding speed and phonological short-term memory in children with developmental dyslexia. *Neuropsychologia*, 130, 100–106. <https://doi.org/10.1016/j.neuropsychologia.2018.10.023>
- Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action Video Games Make Dyslexic Children Read Better. *Current Biology*, 23(6), 462–466. <https://doi.org/10.1016/j.cub.2013.01.044>
- Franceschini, S., Trevisan, P., Ronconi, L., Bertoni, S., Colmar, S., Double, K., Facoetti, A., & Gori, S. (2017). Action video games improve reading abilities and visual-to-auditory attentional shifting in English-speaking children with dyslexia. *Scientific Reports*, 7(1), 5863. <https://doi.org/10.1038/s41598-017-05826-8>
- Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychological Science*, 21(7), 920–925. <https://doi.org/10.1177/0956797610373935>
- Franconeri, S., Jonathan, S., & Scimeca, J. (2010). Tracking Multiple Objects Is Limited Only by Object Spacing, Not by Speed, Time, or Capacity. *Psychological Science*, 21(7), 920–925. <https://doi.org/10.1177/0956797610373935>
- Franiatte, N., Boissin, E., Delmas, A., & De Neys, W. (2024). Boosting debiasing: Impact of repeated training on reasoning. *Learning and Instruction*, 89, 101845.

Retrieved July 10, 2024, from <https://www.sciencedirect.com/science/article/pii/S0959475223001147>

- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415. <https://doi.org/10.1073/pnas.1319030111>
- Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of experimental psychology: General*, *137*(2), 201.
- Friesen, N., & Feenberg, A. (2007). 'Ed Tech in Reverse': Information technologies and the cognitive revolution. *Educational Philosophy and Theory*, *39*, 720–736. <https://doi.org/10.1111/j.1469-5812.2007.00314.x>
- Front Matter. (2015, January 1). In J. K. Kruschke (Ed.), *Doing Bayesian Data Analysis (Second Edition)* (pp. i–ii). Academic Press. <https://doi.org/10.1016/B978-0-12-405888-0.09999-2>
- Fu, L., Kessels, R. P. C., & Maes, J. H. R. (2020). The effect of cognitive training in older adults: Be aware of CRUNCH [Publisher: Routledge _eprint: <https://doi.org/10.1080/13825585.2019.1708251>]. *Aging, Neuropsychology, and Cognition*, *27*(6), 949–962. <https://doi.org/10.1080/13825585.2019.1708251>
- Fulvio, J., Green, C. S., & Schrater, P. (2013). Specificity in learning: Blame the paradigm. *Journal of Vision*, *13*(9), 246. <https://doi.org/10.1167/13.9.246>
- Furley, P., & Wood, G. (2016). Working memory, attentional control, and expertise in sports: A review of current literature and directions for future research [Place: Netherlands Publisher: Elsevier Science]. *Journal of Applied Research in Memory and Cognition*, *5*(4), 415–425. <https://doi.org/10.1016/j.jarmac.2016.05.001>
- García-Rudolph, A., & Gibert, K. (2014). A data mining approach to identify cognitive NeuroRehabilitation Range in Traumatic Brain Injury patients [Publisher: Elsevier]. *Expert systems with applications*, *41*(11), 5238–5251.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. Basic Books.
- Gaspari, M., Zini, F., & Stecchi, S. (2020). Enhancing cognitive rehabilitation in multiple sclerosis with a disease-specific tool [Publisher: Taylor & Francis]. *Disability and Rehabilitation: Assistive Technology*, 1–14.
- Gates, N. J., Sachdev, P. S., Fiatarone Singh, M. A., & Valenzuela, M. (2011). Cognitive and memory training in adults at risk of dementia: A systematic review. *BMC geriatrics*, *11*, 55. <https://doi.org/10.1186/1471-2318-11-55>
- Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language*, *105*, 19–42. <https://doi.org/10.1016/j.jml.2018.10.003>
- Geerligs, L., Saliassi, E., Maurits, N. M., Renken, R. J., & Lorist, M. M. (2014). Brain mechanisms underlying the effects of aging on different aspects of selective attention. *NeuroImage*, *91*, 52–62. <https://doi.org/10.1016/j.neuroimage.2014.01.029>
- Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*(1), 1–40. <https://doi.org/10.1006/cogp.1996.0011>
- Gobin, C. M., Banks, J. B., Fins, A. I., & Tartar, J. L. (2015). Poor sleep quality is associated with a negative cognitive bias and decreased sustained attention. *Journal of Sleep Research*, *24*(5), 535–542. <https://doi.org/10.1111/jsr.12302>

- González, V. M., & Mark, G. (2004). "Constant, constant, multi-tasking craziness": Managing multiple working spheres. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 113–120. <https://doi.org/10.1145/985692.985707>
- Gori, S., & Facoetti, A. (2015). How the visual aspects can be crucial in reading acquisition? The intriguing case of crowding and developmental dyslexia. *Journal of Vision*, 15(1), 15.1.8. <https://doi.org/10.1167/15.1.8>
- Gori, S., Seitz, A. R., Ronconi, L., Franceschini, S., & Facoetti, A. (2016). Multiple Causal Links Between Magnocellular-Dorsal Pathway Deficit and Developmental Dyslexia. *Cerebral Cortex (New York, N.Y.: 1991)*, 26(11), 4356–4369. <https://doi.org/10.1093/cercor/bhv206>
- Gozli, D. G., Bavelier, D., & Pratt, J. (2014). The effect of action video game playing on sensorimotor learning: Evidence from a movement tracking task. *Human Movement Science*, 38, 152–162. <https://doi.org/10.1016/j.humov.2014.09.004>
- Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research*, 1(1), 35–51. [https://doi.org/10.1016/S1389-0417\(99\)00005-4](https://doi.org/10.1016/S1389-0417(99)00005-4)
- Graves, F. P. (1915). *A Student's History of Education*. Macmillan.
- Gredler, M. E. (2012). Understanding Vygotsky for the Classroom: Is It Too Late? *Educational Psychology Review*, 24(1), 113–131. <https://doi.org/10.1007/s10648-011-9183-6>
- Green, C. S., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging*, 23(4), 692–701. <https://doi.org/10.1037/a0014345>
- Green, C. S., & Bavelier, D. (2003a). Action video game modifies visual selective attention. *Nature*, 423(6939), 534–537.
- Green, C. S., & Bavelier, D. (2003b). Action video game modifies visual selective attention. *Nature*, 423(6939), 534–537. <https://doi.org/10.1038/nature01647>
- Green, C. S., & Bavelier, D. (2006a). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of experimental psychology: Human perception and performance*, 32(6), 1465.
- Green, C. S., & Bavelier, D. (2006b). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, 101(1), 217–245.
- Green, C. S., & Bavelier, D. (2006c). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology. Human Perception and Performance*, 32(6), 1465–1478. <https://doi.org/10.1037/0096-1523.32.6.1465>
- Green, C. S., Bavelier, D., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., & Witt, C. M. (2019). Improving methodological standards in behavioral interventions for cognitive enhancement [Publisher: Springer Science and Business Media LLC]. *Journal of Cognitive Enhancement*, 3(1), 2–29.
- Green, C. S., Pouget, A., & Bavelier, D. (2010). Improved Probabilistic Inference as a General Learning Mechanism with Action Video Games. *Current Biology*, 20(17), 1573–1579. <https://doi.org/10.1016/j.cub.2010.07.040>
- Green, C., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, 101(1), 217–245. <https://doi.org/10.1016/j.cognition.2005.10.004>
- Greenwood, P. M., & Parasuraman, R. (2016). The mechanisms of far transfer from cognitive training: Review and hypothesis. *Neuropsychology*, 30(6), 742–755. <https://doi.org/10.1037/neu0000235>

- Greenwood, P. M., Parasuraman, R., & Haxby, J. V. (1993). Changes in visuospatial attention over the adult lifespan. *Neuropsychologia*, 31(5), 471–485.
- Gronwall, D. M. A. (1977). Paced auditory serial-addition task: A measure of recovery from concussion [Publisher: SAGE Publications Sage CA: Los Angeles, CA]. *Perceptual and motor skills*, 44(2), 367–373.
- Guay, F., Vallerand, R., & Blanchard, C. (2000). On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24, 175–213. <https://doi.org/10.1023/A:1005614228250>
- Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, 24(3), 175–213. <https://doi.org/10.1023/A:1005614228250>
- Gurnsey, R., Roddy, G., & Chanab, W. (2011). Crowding is size and eccentricity dependent. *Journal of Vision*, 11(7), 15. <https://doi.org/10.1167/11.7.15>
- Guye, S., De Simoni, C., & von Bastian, C. (2017). Do Individual Differences Predict Change in Cognitive Training Performance? A Latent Growth Curve Modeling Approach. *Journal of Cognitive Enhancement*, 1, 374–393. <https://doi.org/10.1007/s41465-017-0049-9>
- Guye, S., De Simoni, C., & von Bastian, C. C. (2017). Do Individual Differences Predict Change in Cognitive Training Performance? A Latent Growth Curve Modeling Approach. *Journal of Cognitive Enhancement*, 1(4), 374–393. <https://doi.org/10.1007/s41465-017-0049-9>
- Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer [Place: US Publisher: American Psychological Association]. *Psychology and Aging*, 32(8), 732–746. <https://doi.org/10.1037/pag0000206>
- Hallock, H., Collins, D., Lampit, A., Deol, K., Fleming, J., & Valenzuela, M. (2016a). Cognitive Training for Post-Acute Traumatic Brain Injury: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience*, 10. Retrieved July 27, 2023, from <https://www.frontiersin.org/articles/10.3389/fnhum.2016.00537>
- Hallock, H., Collins, D., Lampit, A., Deol, K., Fleming, J., & Valenzuela, M. (2016b). Cognitive Training for Post-Acute Traumatic Brain Injury: A Systematic Review and Meta-Analysis. *Frontiers in Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00537>
- Happyneuron. (n.d.). Happyneuron @ la maison : Envoi d’activités de rééducation aux patient. Retrieved June 26, 2023, from <https://www.happyneuronpro.com/orthophonie/alamaison/therapeute/>
- Harbour, R., & Miller, J. (2001). A new system for grading recommendations in evidence based guidelines. *BMJ (Clinical research ed.)*, 323(7308), 334–336. <https://doi.org/10.1136/bmj.323.7308.334>
- Hardy, J. L., Nelson, R. A., Thomason, M. E., Sternberg, D. A., Katovich, K., Farzin, F., & Scanlon, M. (2015). Enhancing Cognitive Abilities with Comprehensive Training: A Large, Online, Randomized, Active-Controlled Trial. *PloS One*, 10(9), e0134467. <https://doi.org/10.1371/journal.pone.0134467>
- Hardy, S., Reuter, C., Göbel, S., Steinmetz, R., Baller, G., Kalbe, E., Moussaoui, A. E., Abels, S., Dienst, S., Dornhöfer, M., & Fathi, M. (2015). NeuroCare—Personalization and Adaptation of Digital Training Programs for Mild Cognitive Impairments. In R. Wichert & H. Klausning (Eds.), *Ambient Assisted Living: 7. AAL-Kongress 2014 Berlin, Germany, January 21-22, 2014* (pp. 53–63). Springer International Publishing. https://doi.org/10.1007/978-3-319-11866-6_4

- Harenberg, S., St. Onge, J., Robinson, J., Eguakun, O., Feinstein, A., Dorsch, K., Kakar, R. S., Abdulhakim, R., Rehman, Z., Shawush, M., & Pillay, V. (2021). Effectiveness of Three-Dimensional Multiple-Object Tracking in Patients with Multiple Sclerosis. *International Journal of MS Care*, 23(4), 143–149. <https://doi.org/10.7224/1537-2073.2020-007>
- Harris, D. J., Wilson, M. R., Crowe, E. M., & Vine, S. J. (2020). Examining the roles of working memory and visual attention in multiple object tracking expertise. *Cognitive Processing*, 21(2), 209–222. <https://doi.org/10.1007/s10339-020-00954-y>
- Harris, D. J., Wilson, M. R., Smith, S. J. R., Meder, N., & Vine, S. J. (2020). Testing the Effects of 3D Multiple Object Tracking Training on Near, Mid and Far Transfer. *Frontiers in Psychology*, 11. Retrieved March 7, 2024, from <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.00196>
- Harris, D. J., Wilson, M. R., Smith, S. J., Meder, N., & Vine, S. J. (2020). Testing the effects of 3d multiple object tracking training on near, mid and far transfer. *Frontiers in Psychology*, 11, 196.
- Harris, D. J., Wilson, M. R., & Vine, S. J. (2018a). A Systematic Review of Commercial Cognitive Training Devices: Implications for Use in Sport. *Frontiers in Psychology*, 9. Retrieved June 23, 2023, from <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00709>
- Harris, D. J., Wilson, M. R., & Vine, S. J. (2018b). A Systematic Review of Commercial Cognitive Training Devices: Implications for Use in Sport. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00709>
- Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–908. <https://doi.org/10.1177/154193120605000909>
- Hartshorne, J. K., & Germine, L. T. (2015). When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. *Psychological Science*, 26(4), 433–443. <https://doi.org/10.1177/0956797614567339>
- Harvey, P. D., McGurk, S. R., Mahncke, H., & Wykes, T. (2018). Controversies in computerized cognitive training [Publisher: Elsevier]. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(11), 907–915.
- Hayat, A., & Hasan, M. R. (2023). The Power of Personalization and Contextualization: Early Student Performance Forecasting with Language Models. *The 2023 NeurIPS (Neural Information Processing Systems) Workshop on Generative AI for Education (GAIED)*. Retrieved June 25, 2024, from <https://par.nsf.gov/biblio/10499296-power-personalization-contextualization-early-student-performance-forecasting-language-models>
- He, S., Cavanagh, P., & Intriligator, J. (1997). Attentional resolution. *Trends in Cognitive Sciences*, 1(3), 115–121. [https://doi.org/10.1016/S1364-6613\(97\)89058-4](https://doi.org/10.1016/S1364-6613(97)89058-4)
- Hertzog, C., Kramer, A. F., Wilson, R. S., & Lindenberger, U. (2008). Enrichment Effects on Adult Cognitive Development: Can the Functional Capacity of Older Adults Be Preserved and Enhanced? *Psychological Science in the Public Interest*, 9(1), 1–65. <https://doi.org/10.1111/j.1539-6053.2009.01034.x>
- Hilbig, B. E. (2016). Reaction time effects in lab-versus web-based research: Experimental evidence. *Behavior Research Methods*, 48(4), 1718–1724.
- Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. *Journal of the History of the Behavioral Sciences*, 16(2), 107–117. [https://doi.org/10.1002/1520-6696\(198004\)16:2<107::AID-JHBS2300160202>3.0.CO;2-Y](https://doi.org/10.1002/1520-6696(198004)16:2<107::AID-JHBS2300160202>3.0.CO;2-Y)

- Hilgard, J., Sala, G., Boot, W. R., & Simons, D. J. (2019). Overestimation of Action-Game Training Effects: Publication Bias and Salami Slicing (R. Zwaan & R. Zwaan, Eds.). *Collabra: Psychology*, 5(1), 30. <https://doi.org/10.1525/collabra.231>
- Hill, N. T., Mowszowski, L., Naismith, S. L., Chadwick, V. L., Valenzuela, M., & Lampit, A. (2017a). Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis. *American Journal of Psychiatry*, 174(4), 329–340. <https://doi.org/10.1176/appi.ajp.2016.16030360>
- Hill, N. T., Mowszowski, L., Naismith, S. L., Chadwick, V. L., Valenzuela, M., & Lampit, A. (2017b). Computerized Cognitive Training in Older Adults With Mild Cognitive Impairment or Dementia: A Systematic Review and Meta-Analysis [Publisher: American Psychiatric Publishing]. *American Journal of Psychiatry*, 174(4), 329–340. <https://doi.org/10.1176/appi.ajp.2016.16030360>
- Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., Kingma, D. P., Poole, B., Norouzi, M., Fleet, D. J., & Salimans, T. (2022, October 5). *Imagen Video: High Definition Video Generation with Diffusion Models*. arXiv: 2210.02303 [cs]. <https://doi.org/10.48550/arXiv.2210.02303>
- Hoekstra, C., Martens, S., & Taatgen, N. A. (2022). Testing the skill-based approach: Consolidation strategy impacts attentional blink performance. *PLOS ONE*, 17(1), e0262350. <https://doi.org/10.1371/journal.pone.0262350>
- Hoffman, L., McDowd, J. M., Atchley, P., & Dubinsky, R. (2005). The role of visual attention in predicting driving impairment in older adults. *Psychology and Aging*, 20(4), 610–622. <https://doi.org/10.1037/0882-7974.20.4.610>
- Hoffman, M. D., & Gelman, A. (2011, November 17). *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo*. arXiv: 1111.4246 [cs, stat]. <https://doi.org/10.48550/arXiv.1111.4246>
- Holcombe, A. (2023). *Attending to moving objects*. Cambridge University Press. Retrieved June 27, 2024, from <https://www.cambridge.org/core/elements/attending-to-moving-objects/0914ABF2EF7D03676124F7250874071A>
- Holcombe, A. O. (2023). *From Attention to Distraction*. Retrieved May 30, 2024, from <https://psyc2016.whatanimalssee.com/>
- Holcombe, A. O., & Chen, W.-Y. (2012). Exhausting attentional tracking resources with a single fast-moving object. *Cognition*, 123(2), 218–228. <https://doi.org/10.1016/j.cognition.2011.10.003>
- Holcombe, A. O., & Chen, W.-Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to <3 Hz when tracking three. *Journal of Vision*, 13(1), 12. <https://doi.org/10.1167/13.1.12>
- Horowitz, T. S., & Cohen, M. A. (2010). Direction information in multiple object tracking is limited by a graded resource. *Attention, Perception, & Psychophysics*, 72(7), 1765–1775. <https://doi.org/10.3758/APP.72.7.1765>
- Hosokawa, K., Maruya, K., Nishida, S., Takahashi, M., & Nakadomari, S. (2019). Gamified vision test system for daily self-check. *2019 IEEE Games, Entertainment, Media Conference (GEM)*, 1–8.
- Huang, L., Mo, L., & Li, Y. (2012a). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, 38(2), 414–428. <https://doi.org/10.1037/a0026314>
- Huang, L., Mo, L., & Li, Y. (2012b). Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. *Journal of Experimental*

- Psychology. Human Perception and Performance*, 38(2), 414–428. <https://doi.org/10.1037/a0026314>
- Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. Appleton-Century.
- Hung, S.-C., & Seitz, A. R. (2014a). Prolonged Training at Threshold Promotes Robust Retinotopic Specificity in Perceptual Learning. *Journal of Neuroscience*, 34(25), 8423–8431. <https://doi.org/10.1523/JNEUROSCI.0745-14.2014>
- Hung, S.-C., & Seitz, A. R. (2014b). Prolonged training at threshold promotes robust retinotopic specificity in perceptual learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(25), 8423–8431. <https://doi.org/10.1523/JNEUROSCI.0745-14.2014>
- Hwang, K., Challagundla, S., Alomair, M., Chen, L., & Choa, F. (n.d.). Towards AI-Assisted Multiple Choice Question Generation and Quality Evaluation at Scale: Aligning with Bloom’s Taxonomy. Retrieved June 25, 2024, from <https://www.semanticscholar.org/paper/Towards-AI-Assisted-Multiple-Choice-Question-and-at-Hwang-Challagundla/8c20862d07859a7707bb2f6fa2d8dc9b3aba19b7>
- IJsselstein, W., de Kort, Y., & Poels, K. (2013). *The Game Experience Questionnaire*. Technische Universiteit Eindhoven.
- Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011a). Understanding the Intrinsic Memorability of Images. *Advances in Neural Information Processing Systems*, 24. Retrieved April 13, 2024, from <https://proceedings.neurips.cc/paper/2011/hash/286674e3082feb7e5afb92777e48821f-Abstract.html>
- Isola, P., Parikh, D., Torralba, A., & Oliva, A. (2011b). *Understanding the intrinsic memorability of images* (tech. rep.). MASSACHUSETTS INST OF TECH CAMBRIDGE.
- Iterbeke, K., De Witte, K., & Schelfhout, W. (2021). The effects of computer-assisted adaptive instruction and elaborated feedback on learning outcomes. A randomized control trial [Publisher: Elsevier]. *Computers in Human Behavior*, 120, 106666.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008a). Improving fluid intelligence with training on working memory [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 105(19), 6829–6833. <https://doi.org/10.1073/pnas.0801268105>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008b). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105(19), 6829–6833. <https://doi.org/10.1073/pnas.0801268105>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 108(25), 10081–10086. <https://doi.org/10.1073/pnas.1103228108>
- Jaeggi, S. M., Buschkuhl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory*, 18(4), 394–412. <https://doi.org/10.1080/09658211003702171>
- Jaegle, A., Mehrpour, V., Mohsenzadeh, Y., Meyer, T., Oliva, A., & Rust, N. (2019). Population response magnitude variation in inferotemporal cortex predicts image memorability. *Elife*, 8, e47596.
- Jeffress, L. A. (**typedactor**). (1951). *Cerebral mechanisms in behavior; the Hixon Symposium*. Wiley.
- Jeffreys, H. (1998, August 6). *The Theory of Probability*. OUP Oxford.

- Jeong, S.-H., & Hwang, Y. (2012). Does multitasking increase or decrease persuasion? Effects of multitasking on comprehension and counterarguing. *Journal of Communication*, 62(4), 571–587. <https://doi.org/10.1111/j.1460-2466.2012.01659.x>
- Jeunet, C., Tonin, L., Albert, L., Chavarriaga, R., Bideau, B., Argelaguet, F., Millán, J. D. R., Lécuyer, A., & Kulpa, R. (2020). Uncovering EEG Correlates of Covert Attention in Soccer Goalkeepers: Towards Innovative Sport Training Procedures. *Scientific Reports*, 10(1), 1705. <https://doi.org/10.1038/s41598-020-58533-2>
- Joessel, A. (2022). Development of a video game to investigate the AVG features promoting attentional control. *Published online*. Retrieved April 18, 2024, from <https://access.archive-ouverte.unige.ch/access/metadata/c32af7d3-7b6c-460b-9a4b-f992a70243f4/download>
- Joessel, F. (2022a, October 31). *Development of a Video Game to Investigate the AVG Features Promoting Attentional Control*. <https://doi.org/10.13097/archive-ouverte/unige:164866>
- Joessel, F. (2022b, October 31). *Development of a Video Game to Investigate the AVG Features Promoting Attentional Control*. <https://doi.org/10.13097/archive-ouverte/unige:164866>
- Joessel, F., Pichon, S., & Bavelier, D. (2023). A video-game-based method to induce states of high and low flow. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-023-02251-w>
- Johann, V. E., & Karbach, J. (2020). Effects of game-based and standard executive control training on cognitive and academic abilities in elementary school children. *Developmental Science*, 23(4), e12866. <https://doi.org/10.1111/desc.12866>
- Johansson, H., Folkerts, A.-K., Hammarström, I., Kalbe, E., & Leavy, B. (2023). Effects of motor-cognitive training on dual-task performance in people with Parkinson's disease: A systematic review and meta-analysis. *Journal of Neurology*, 270(6), 2890–2907. <https://doi.org/10.1007/s00415-023-11610-8>
- Jonge, M., Tabbers, H., Pecher, D., & Zeelenberg, R. (2012). The Effect of Study Time Distribution on Learning and Retention: A Goldilocks Principle for Presentation Rate. *Journal of experimental psychology. Learning, memory, and cognition*, 38, 405–12. <https://doi.org/10.1037/a0025897>
- Jonsdottir, S., Bouma, A., Sergeant, J. A., & Scherder, E. J. (2006). Relationships between neuropsychological measures of executive function and behavioral measures of ADHD symptoms and comorbid behavior. *Archives of Clinical Neuropsychology*, 21(5), 383–394. <https://doi.org/10.1016/j.acn.2006.05.003>
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992a). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219. [https://doi.org/10.1016/0010-0285\(92\)90007-O](https://doi.org/10.1016/0010-0285(92)90007-O)
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992b). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, 24(2), 175–219. [https://doi.org/10.1016/0010-0285\(92\)90007-O](https://doi.org/10.1016/0010-0285(92)90007-O)
- Kallio, E.-L., Öhman, H., Kautiainen, H., Hietanen, M., & Pitkälä, K. (2017). Cognitive Training Interventions for Patients with Alzheimer's Disease: A Systematic Review [Publisher: IOS Press]. *Journal of Alzheimer's Disease*, 56(4), 1349–1372. <https://doi.org/10.3233/JAD-160810>
- Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology Research and Development*, 53(3), 83–93. <https://doi.org/10.1007/BF02504800>

- Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-y., & Camerer, C. F. (2009). The Wick in the Candle of Learning: Epistemic Curiosity Activates Reward Circuitry and Enhances Memory. *Psychological Science*, 20(8), 963–973. <https://doi.org/10.1111/j.1467-9280.2009.02402.x>
- Karbach, J., Könen, T., & Spengler, M. (2017a). Who Benefits the Most? Individual Differences in the Transfer of Executive Control Training Across the Lifespan. *Journal of Cognitive Enhancement*, 1(4), 394–405. <https://doi.org/10.1007/s41465-017-0054-z>
- Karbach, J., Könen, T., & Spengler, M. (2017b). Who Benefits the Most? Individual Differences in the Transfer of Executive Control Training Across the Lifespan. *Journal of Cognitive Enhancement*, 1(4), 394–405. <https://doi.org/10.1007/s41465-017-0054-z>
- Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, 12(6), 978–990. <https://doi.org/10.1111/j.1467-7687.2009.00846.x>
- Karbach, J., & Kray, J. (2021). Executive Function Training. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 199–212). Springer International Publishing. https://doi.org/10.1007/978-3-030-39292-5_14
- Karbach, J., Strobach, T., & Schubert, T. (2015). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence*, 21(3), 285–301. <https://doi.org/10.1080/09297049.2014.899336>
- Karbach, J., & Unger, K. (2014). Executive control training from middle childhood to adolescence [Publisher: Frontiers]. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00390>
- Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E., & Green, C. S. (2017). Perceptual Learning Generalization from Sequential Perceptual Training as a Change in Learning Rate. *Current Biology*, 27(6), 840–846. <https://doi.org/10.1016/j.cub.2017.01.046>
- Katz, B., Jones, M. R., Shah, P., Buschkuehl, M., & Jaeggi, S. M. (2016). Individual Differences and Motivational Effects. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 157–166). Springer International Publishing. https://doi.org/10.1007/978-3-319-42662-4_15
- Katz, B., Jones, M. R., Shah, P., Buschkuehl, M., & Jaeggi, S. M. (2021). Individual differences in cognitive training research. In *Cognitive training: An overview of features and applications, 2nd ed* (pp. 107–123). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-39292-5_8
- Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., & Brennan, S. (2014). The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: A systematic review and meta-analysis. *Ageing Research Reviews*, 15, 28–43. <https://doi.org/10.1016/j.arr.2014.02.004>
- Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to Learn Causal Models. *Cognitive Science*, 34(7), 1185–1243. <https://doi.org/10.1111/j.1551-6709.2010.01128.x>
- Khadimallah, R., Abdelkefi, M., & Kallel, I. (2020). Emotion Regulation in Intelligent Tutoring Systems: A Systematic Literature Review [ISSN: 2470-6698]. *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 363–370. <https://doi.org/10.1109/TALE48869.2020.9368372>

- Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015). Understanding and predicting image memorability at a large scale. *Proceedings of the IEEE international conference on computer vision*, 2390–2398.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLOS ONE*, 7(5), e36399. <https://doi.org/10.1371/journal.pone.0036399>
- Kim, J. J., Kim, Y.-J., Lee, H.-M., Lee, S.-H., & Chung, S.-T. (2018). Personalized Recommendation System for Efficient Integrated Cognitive Rehabilitation Training Based on Bigdata. In C. Stephanidis (Ed.), *HCI International 2018 – Posters' Extended Abstracts* (pp. 32–39). Springer International Publishing. https://doi.org/10.1007/978-3-319-92279-9_4
- Kiryakova, G., & Angelova, N. (2023). ChatGPT—A Challenging Tool for the University Professors in Their Teaching Practice. *Education Sciences*, 13(10), 1056. <https://doi.org/10.3390/educsci13101056>
- Kitakoshi, D., Hanada, R., Iwata, K., & Suzuki, M. (2015). Cognitive training system for dementia prevention using memory game based on the concept of human-agent interaction [Publisher: Fuji Technology Press Ltd.]. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 19(6), 727–737.
- Kitakoshi, D., Suzuki, K., & Suzuki, M. (2020a). A Study on Coordination of Exercise Difficulty in Cognitive Training System for Older Adults. *2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)*, 1–6.
- Kitakoshi, D., Suzuki, K., & Suzuki, M. (2020b). A Study on Coordination of Exercise Difficulty in Cognitive Training System for Older Adults. *2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS)*, 1–6.
- Klein, E. (2018). *Matie?Re a? Contredire: Essai de Philo-Physique*. E?ditions de l'Observatoire.
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317–324. <https://doi.org/10.1016/j.tics.2010.05.002>
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., Gillberg, C. G., Forssberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD—a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44(2), 177–186. <https://doi.org/10.1097/00004583-200502000-00010>
- Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24(6), 781–791. <https://doi.org/10.1076/jcen.24.6.781.8395>
- Koedinger, K. R., & Corbett, A. (2006). *Cognitive tutors: Technology bringing learning sciences to the classroom*. na.
- Komarudin, K., Mulyana, M., Berliana, B., & Purnamasari, I. (2021). NeuroTracker Three-Dimensional Multiple Object Tracking (3D-MOT): A Tool to Improve Concentration and Game Performance among Basketball Athletes. *Annals of Applied Sport Science*, 9(1), 0–0. <https://doi.org/10.29252/aassjournal.946>
- Könen, T., & Karbach, J. (2015). The benefits of looking at intraindividual dynamics in cognitive training data. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00615>

- Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, 25(2), 498–503. <https://doi.org/10.1037/a0017807>
- Kovbasiuk, A., Lewandowska, P., Brzezicka, A., & Kowalczyk-Grębska, N. (2022). Neuroanatomical predictors of complex skill acquisition during video game training. *Frontiers in Neuroscience*, 16. <https://doi.org/10.3389/fnins.2022.834954>
- Kraus, T. H., & Breznitz, Z. (2009). Can the error detection mechanism benefit from training the working memory? a comparison between dyslexics and controls—an erp study. *PloS one*, 4(9), e7141.
- Kruschke, J. K. (2010). Bayesian data analysis. *WIREs Cognitive Science*, 1(5), 658–676. <https://doi.org/10.1002/wcs.72>
- Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology. General*, 142(2), 573–603. <https://doi.org/10.1037/a0029146>
- Kruschke, J. K. (2021a). Bayesian analysis reporting guidelines. 5(10), 1282–1291.
- Kruschke, J. K. (2021b). Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour*, 5(10), 1282–1291. <https://doi.org/10.1038/s41562-021-01177-7>
- Kruschke, J. K. (2021c). Bayesian Analysis Reporting Guidelines. *Nature Human Behaviour*, 5(10), 1282–1291. <https://doi.org/10.1038/s41562-021-01177-7>
- Kueider, A. M., Parisi, J. M., Gross, A. L., & Rebok, G. W. (2012a). Computerized cognitive training with older adults: A systematic review. *PloS one*, 7(7), e40588.
- Kueider, A. M., Parisi, J. M., Gross, A. L., & Rebok, G. W. (2012b). Computerized Cognitive Training with Older Adults: A Systematic Review. *PLOS ONE*, 7(7), e40588. <https://doi.org/10.1371/journal.pone.0040588>
- LaBar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M.-M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: A functional mri comparison within subjects. *Neuroimage*, 10(6), 695–704.
- Lampit, A., Hallock, H., & Valenzuela, M. (2014a). Computerized cognitive training in cognitively healthy older adults: A systematic review and meta-analysis of effect modifiers [Publisher: Public Library of Science]. *PLoS medicine*, 11(11), e1001756.
- Lampit, A., Hallock, H., & Valenzuela, M. (2014b). Computerized cognitive training in cognitively healthy older adults: A systematic review and meta-analysis of effect modifiers. *PLoS medicine*, 11(11), e1001756.
- Lampit, A., Hallock, H., & Valenzuela, M. (2014c). Computerized cognitive training in cognitively healthy older adults: A systematic review and meta-analysis of effect modifiers. *PLoS medicine*, 11(11), e1001756. <https://doi.org/10.1371/journal.pmed.1001756>
- Lampit, A., Hallock, H., & Valenzuela, M. (2014d). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. *PLOS Medicine*, 11(11), e1001756. <https://doi.org/10.1371/journal.pmed.1001756>
- Lampit, A., Heine, J., Finke, C., Barnett, M. H., Valenzuela, M., Wolf, A., Leung, I. H. K., & Hill, N. T. M. (2019). Computerized Cognitive Training in Multiple Sclerosis: A Systematic Review and Meta-analysis. *Neurorehabilitation and Neural Repair*, 33(9), 695–706. <https://doi.org/10.1177/1545968319860490>
- Lapierre, M. D., Cropper, S. J., & Howe, P. D. L. (2017). Shared processing in multiple object tracking and visual working memory in the absence of response order and task order confounds [Publisher: Public Library of Science]. *PLOS ONE*, 12(4), e0175736. <https://doi.org/10.1371/journal.pone.0175736>

- Lapierre, M. D., Cropper, S. J., & Howe, P. D. (2017). Shared processing in multiple object tracking and visual working memory in the absence of response order and task order confounds. *PloS one*, *12*(4), e0175736.
- Larrabee, G. J. (2015). The multiple validities of neuropsychological assessment. *American Psychologist*, *70*(8), 779.
- Law, E., Ravari, P., Chhibber, N., Kulic, D., Lin, S., Pantasdo, K., Ceha, J., Suh, S., & Dillen, N. (2020). Curiosity Notebook: A Platform for Learning by Teaching Conversational Agents, 1–9. <https://doi.org/10.1145/3334480.3382783>
- Law, E. L.-C., Kickmeier, M., Albert, D., & Holzinger, A. (2008). Challenges in the Development and Evaluation of Immersive Digital Educational Games. In *HCI and Usability for Education and Work, 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008* (pp. 19–30, Vol. 5298). Springer. https://doi.org/10.1007/978-3-540-89350-9_2
- Law, E. L.-C., Kickmeier-Rust, M. D., Albert, D., & Holzinger, A. (2008). Challenges in the Development and Evaluation of Immersive Digital Educational Games. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (pp. 19–30). Springer. https://doi.org/10.1007/978-3-540-89350-9_2
- Learning “Useless” Things in School is (Usually) NOT Useless. (2020, February 7). The Learning Scientists. Retrieved July 2, 2024, from <https://www.learningscientists.org/blog/2020/2/7-1>
- Lee, Y., Kim, T. S., Chang, M., & Kim, J. (2022, May). Interactive Children’s Story Rewriting Through Parent-Children Interaction. In T.-H. ’ Huang, V. Raheja, D. Kang, J. J. Y. Chung, D. Gissin, M. Lee, & K. I. Gero (Eds.), *Proceedings of the First Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2022)* (pp. 62–71). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.in2writing-1.9>
- Legault, I., Allard, R., & Faubert, J. (2013a). Healthy Older Observers Show Equivalent Perceptual-Cognitive Training Benefits to Young Adults for Multiple Object Tracking. *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00323>
- Legault, I., Allard, R., & Faubert, J. (2013b). Healthy older observers show equivalent perceptual-cognitive training benefits to young adults for multiple object tracking. *Frontiers in psychology*, *4*, 323.
- Legault, I., & Faubert, J. (2012a). Perceptual-cognitive training improves biological motion perception: Evidence for transferability of training in healthy aging. *23*(8), 469–473.
- Legault, I., & Faubert, J. (2012b). Perceptual-cognitive training improves biological motion perception: Evidence for transferability of training in healthy aging. *Neuroreport*, *23*, 469–73. <https://doi.org/10.1097/WNR.0b013e328353e48a>
- Legault, I., & Faubert, J. (2012c). Perceptual-cognitive training improves biological motion perception: Evidence for transferability of training in healthy aging. *Neuroreport*, *23*(8), 469–473. <https://doi.org/10.1097/WNR.0b013e328353e48a>
- Lemberger, P., & Saillenfest, A. (2024, April 29). *Explaining Text Classifiers with Counterfactual Representations*. arXiv: 2402.00711 [cs]. <https://doi.org/10.48550/arXiv.2402.00711>
- Leshner, E. L., & Berryhill, J. S. (1994). Validation of the geriatric depression scale-short form among inpatients [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/1097-4679%28199403%2950%3A2%3C256%3A%3AAID-JCLP2270500218%3E3.0.CO%3B2-E>]. *Journal of Clinical Psychology*, *50*(2),

- 256–260. [https://doi.org/10.1002/1097-4679\(199403\)50:2<256::AID-JCLP2270500218>3.0.CO;2-E](https://doi.org/10.1002/1097-4679(199403)50:2<256::AID-JCLP2270500218>3.0.CO;2-E)
- Leung, I. H., Walton, C. C., Hallock, H., Lewis, S. J., Valenzuela, M., & Lampit, A. (2015a). Cognitive training in Parkinson disease: A systematic review and meta-analysis [Publisher: AAN Enterprises]. *Neurology*, *85*(21), 1843–1851.
- Leung, I. H., Walton, C. C., Hallock, H., Lewis, S. J., Valenzuela, M., & Lampit, A. (2015b). Cognitive training in Parkinson disease. *Neurology*, *85*(21), 1843–1851. <https://doi.org/10.1212/WNL.0000000000002145>
- Leung, N. T. Y., Tam, H. M. K., Chu, L. W., Kwok, T. C. Y., Chan, F., Lam, L. C. W., Woo, J., & Lee, T. M. C. (2015). Neural plastic effects of cognitive training on aging brain. *Neural Plasticity*, *2015*. <https://doi.org/10.1155/2015/535618>
- Li, J., Oksama, L., & Hyönä, J. (2019a). Model of Multiple Identity Tracking (MOMIT) 2.0: Resolving the serial vs. parallel controversy in tracking. *Cognition*, *182*, 260–274. <https://doi.org/10.1016/j.cognition.2018.10.016>
- Li, J., Oksama, L., & Hyönä, J. (2019b). Model of Multiple Identity Tracking (MOMIT) 2.0: Resolving the serial vs. parallel controversy in tracking. *Cognition*, *182*, 260–274. <https://doi.org/10.1016/j.cognition.2018.10.016>
- Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for participants' viewing distance in large-scale, psychophysical online experiments using a virtual chinrest. *Scientific reports*, *10*(1), 1–11.
- Li, Z., He, H., Chen, Y., & Guan, Q. (2024a). Effects of engagement, persistence and adherence on cognitive training outcomes in older adults with and without cognitive impairment: A systematic review and meta-analysis of randomised controlled trials. *Age and Ageing*, *53*(1), afad247. <https://doi.org/10.1093/ageing/afad247>
- Li, Z., He, H., Chen, Y., & Guan, Q. (2024b). Effects of engagement, persistence and adherence on cognitive training outcomes in older adults with and without cognitive impairment: A systematic review and meta-analysis of randomised controlled trials. *Age and Ageing*, *53*(1), afad247. <https://doi.org/10.1093/ageing/afad247>
- Liang, S., Guo, Y., Cheng, S., Wu, S., Wang, X., Wang, X., Lu, D., & Liu, X. (2022). Experimental Studies of the Attention Processing Model in Multiple Object Tracking Task. *Brain Sciences*, *12*(12), 1686. <https://doi.org/10.3390/brainsci12121686>
- Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, *20*(1), 135–141. <https://doi.org/10.3758/s13423-012-0335-6>
- Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-Object Tracking Is Based on Scene, Not Retinal, Coordinates. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 235–247. <https://doi.org/10.1037/0096-1523.31.2.235>
- Liu, T., Chen, W., Xuan, Y., & Fu, X. (2009). The Effect of Object Features on Multiple Object Tracking and Identification. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 206–212). Springer. https://doi.org/10.1007/978-3-642-02728-4_22
- Lochner, M. J., & Trick, L. M. (2014). Multiple-object tracking while driving: The multiple-vehicle tracking task. *Attention, Perception, & Psychophysics*, *76*(8), 2326–2345. <https://doi.org/10.3758/s13414-014-0694-3>
- Loewenstein, G. (1994). The Psychology of Curiosity: A Review and Reinterpretation. *Psychological Bulletin*, *116*, 75–98. <https://doi.org/10.1037/0033-2909.116.1.75>
- Logie, R., Camos, V., & Cowan, N. (2021, January 5). *Working Memory: State of the Science*. Oxford University Press.

- Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010a). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, 136(4), 659–676. <https://doi.org/10.1037/a0020080>
- Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010b). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, 136(4), 659–676. <https://doi.org/10.1037/a0020080>
- Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012a). Training-induced compensation versus magnification of individual differences in memory performance [Publisher: Frontiers Media SA]. *Frontiers in human neuroscience*, 6, 141.
- Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012b). Training-induced compensation versus magnification of individual differences in memory performance. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00141>
- Lu, M., Farhat, J. H., & Beck Dallaghan, G. L. (2021). Enhanced Learning and Retention of Medical Knowledge Using the Mobile Flash card Application Anki. *Medical Science Educator*, 31(6), 1975–1981. <https://doi.org/10.1007/s40670-021-01386-9>
- Lucia, S., Bianco, V., Boccacci, L., & Di Russo, F. (2021). Effects of a Cognitive-Motor Training on Anticipatory Brain Functions and Sport Performance in Semi-Elite Basketball Players. *Brain Sciences*, 12(1), 68. <https://doi.org/10.3390/brainsci12010068>
- Luck, S. J., & Vogel, E. K. (1997a). The capacity of visual working memory for features and conjunctions [Publisher: Nature Publishing Group UK London]. *Nature*, 390(6657), 279–281. Retrieved April 18, 2024, from <https://www.nature.com/articles/36846>
- Luck, S. J., & Vogel, E. K. (1997b). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279–281.
- Ludt, R., & Goodrich, G. L. (2002). Change in Visual Perceptual Detection Distances for Low Vision Travelers as a Result of Dynamic Visual Assessment and Training [Publisher: SAGE Publications Inc]. *Journal of Visual Impairment & Blindness*, 96(1), 7–21. <https://doi.org/10.1177/0145482X0209600103>
- Luis-Ruiz, S., Caldú, X., Sánchez-Castañeda, C., Pueyo, R., Garolera, M., & Jurado, M. Á. (2020). Is cognitive training an effective tool for improving cognitive function and real-life behaviour in healthy children and adolescents? A systematic review. *Neuroscience & Biobehavioral Reviews*, 116, 268–282. <https://doi.org/10.1016/j.neubiorev.2020.06.019>
- Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016a). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR serious games*, 4(2), e5888.
- Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016b). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy [Company: JMIR Serious Games Distributor: JMIR Serious Games Institution: JMIR Serious Games Label: JMIR Serious Games Publisher: JMIR Publications Inc., Toronto, Canada]. *JMIR Serious Games*, 4(2), e5888. <https://doi.org/10.2196/games.5888>
- Ma, W. J., & Huang, W. (2009). No capacity limit in attentional tracking: Evidence for probabilistic inference under a resource constraint. *Journal of Vision*, 9(11), 3. <https://doi.org/10.1167/9.11.3>
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, 17(3), 347–356. <https://doi.org/10.1038/nn.3655>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis [Place: US Publisher: American Psychological

- Association]. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Macdonald, J. S., & Lavie, N. (2008). Load induced blindness. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1078.
- Mackenzie, A. K., Vernon, M. L., Cox, P. R., Crundall, D., Daly, R. C., Guest, D., Muhl-Richardson, A., & Howard, C. J. (2022). The Multiple Object Avoidance (MOA) task measures attention for action: Evidence from driving and sport. *Behavior Research Methods*, 54(3), 1508–1529. <https://doi.org/10.3758/s13428-021-01679-2>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, 98(1), 185.
- Madigan, S., McArthur, B. A., Anhorn, C., Eirich, R., & Christakis, D. A. (2020). Associations Between Screen Use and Child Language Skills: A Systematic Review and Meta-analysis. *JAMA pediatrics*, 174(7), 665–675. <https://doi.org/10.1001/jamapediatrics.2020.0327>
- Makowski, D., Ben-Shachar, M. S., Chen, S., & Lüdtke, D. (2019). Indices of effect existence and significance in the bayesian framework. *Frontiers in psychology*, 10, 2767.
- Mampadi, F., Chen, S., Ghinea, G., & Chen, M.-P. (2011). Design of adaptive hypermedia learning systems: A cognitive style approach. *Computers & Education*, 56, 1003–1011. <https://doi.org/10.1016/j.compedu.2010.11.018>
- Mancas, M., & Le Meur, O. (2013). Memorability of natural scenes: The role of attention. *2013 IEEE International Conference on Image Processing*, 196–200.
- Mani, T., Bedwell, J., & Miller, L. (2005). Age-related decrements in performance on a brief continuous performance test. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists*, 20, 575–86. <https://doi.org/10.1016/j.acn.2004.12.008>
- Mani, T. M., Bedwell, J. S., & Miller, L. S. (2005). Age-related decrements in performance on a brief continuous performance test. *Archives of clinical neuropsychology*, 20(5), 575–586.
- Mann, D., Williams, A., Ward, P., & Janelle, C. (2007). Perceptual-Cognitive Expertise in Sport: A Meta-Analysis. *Journal of sport & exercise psychology*, 29, 457–78. <https://doi.org/10.1123/jsep.29.4.457>
- Marinelli, M., Sunyer, J., Alvarez-Pedrerol, M., Iñiguez, C., Torrent, M., Vioque, J., Turner, M. C., & Julvez, J. (2014). Hours of television viewing and sleep duration in children: A multicenter birth cohort study. *JAMA pediatrics*, 168(5), 458–464. <https://doi.org/10.1001/jamapediatrics.2013.3861>
- Marioni, R., Hout, A. van den, Valenzuela, M., Brayne, C., Matthews, F., & Ageing, F. M. C. a. S. (2011). O3-4.2 Cognitive reserve and cognitive decline: Are individual sub-components of reserve driving the associations? *Journal of Epidemiology & Community Health*, 65, A36–A36. <https://doi.org/10.1136/jech.2011.142976a.98>
- Mark, G., Iqbal, S. T., Czerwinski, M., Johns, P., & Sano, A. (2016). Neurotics Can't Focus: An *in situ* Study of Online Multitasking in the Workplace. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1739–1744. <https://doi.org/10.1145/2858036.2858202>

- Markel, J. M., Opferman, S. G., Landay, J. A., & Piech, C. (2023). GPTEach: Interactive TA Training with GPT-based Students. *Proceedings of the Tenth ACM Conference on Learning @ Scale*, 226–236. <https://doi.org/10.1145/3573051.3593393>
- Marr, D. (2010, July 9). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. MIT Press.
- Massart, J., & Chaumon, M.-E. B. (2021). Usage du réseau social numérique d'entreprise en contexte de télétravail et covid-19: Implications pour la qualité de vie au travail. *Les incidences psycho-sociales et socio-organisationnelles de la crise sanitaire covid sur le travail et la santé des salariés*. Retrieved July 10, 2024, from <https://hal.science/hal-03838629/>
- Massart, J., Delmas, A.-A. D.-M., Constantin, A., & Chaumon, M.-E. B. (2024). Développement d'un serious game sur la qualité de vie et des conditions de travail: Conception participative et tests-utilisateurs. *IHM'24-35e Conférence Internationale Francophone Sur l'Interaction Humain-Machine*. Retrieved July 10, 2024, from <https://hal.science/hal-04493636/>
- Mathan, S., & Koedinger, K. (2005). Fostering the Intelligent Novice: Learning From Errors With Metacognitive Tutoring. *Educational Psychologist - EDUC PSYCHOL*, 40, 257–265. https://doi.org/10.1207/s15326985ep4004_7
- Maunsell, J. H. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science*, 1, 373–391.
- Mayer, R. (2017). Using multimedia for e-learning [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12197>]. *Journal of Computer Assisted Learning*, 33(5), 403–423. <https://doi.org/10.1111/jcal.12197>
- Mayer, R. E., Parong, J., & Bainbridge, K. (2019). Young adults learning executive function skills by playing focused video games. *Cognitive Development*, 49, 43–50. <https://doi.org/10.1016/j.cogdev.2018.11.002>
- Mazon, C., Clément, B., Roy, D., Oudeyer, P.-Y., & Sauzéon, H. (2023). Pilot study of an intervention based on an intelligent tutoring system (ITS) for instructing mathematical skills of students with ASD and/or ID. *Education and Information Technologies*, 28(8), 9325–9354. <https://doi.org/10.1007/s10639-022-11129-x>
- McDowd, J. M. (2007). An overview of attention: Behavior and brain. *Journal of neurologic physical therapy: JNPT*, 31(3), 98–103. <https://doi.org/10.1097/NPT.0b013e31814d7874>
- McGrew, K. S. (2022). The Cognitive-Affective-Motivation Model of Learning (CAMML): Standing on the shoulders of giants. *Canadian Journal of School Psychology*, 37(1), 117–134. <https://doi.org/10.1177/08295735211054270>
- McGrew, K. (2022). *The Cognitive-Affective-Motivation Model of Learning (CAMML): Standing on the Shoulders of Giants - Kevin S. McGrew, 2022*. Retrieved July 19, 2024, from <https://journals.sagepub.com/doi/full/10.1177/08295735211054270>
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, memory, and cognition*, 22(6), 1423.
- Melby-Lervåg, M., & Hulme, C. (2013a). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49(2), 270–291. <https://doi.org/10.1037/a0028228>
- Melby-Lervåg, M., & Hulme, C. (2013b). Is working memory training effective? A meta-analytic review [Place: US Publisher: American Psychological Association]. *Developmental Psychology*, 49(2), 270–291. <https://doi.org/10.1037/a0028228>
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of “Far

- Transfer”: Evidence From a Meta-Analytic Review. *Perspectives on Psychological Science*, 11(4), 512–534. <https://doi.org/10.1177/1745691616635612>
- Metcalfe, J. (2009). Metacognitive Judgments and Control of Study. *Current directions in psychological science*, 18(3), 159–163. <https://doi.org/10.1111/j.1467-8721.2009.01628.x>
- Metcalfe, J., Schwartz, B. L., & Eich, T. S. (2020). Epistemic curiosity and the region of proximal learning. *Current Opinion in Behavioral Sciences*, 35, 40–47. <https://doi.org/10.1016/j.cobeha.2020.06.007>
- Meyer, S. (2019a). Conception et évaluation d’Evasion, un logiciel éducatif d’entraînement des capacités d’attention visuelle impliquées en lecture.
- Meyer, S. (2019b, January 22). *Conception et évaluation d’Evasion, un logiciel éducatif d’entraînement des capacités d’attention visuelle impliquées en lecture* [Doctoral dissertation, Université Grenoble Alpes]. Retrieved February 13, 2024, from <https://theses.hal.science/tel-02402422>
- Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017a). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, 79(5), 1255–1274. <https://doi.org/10.3758/s13414-017-1338-1>
- Meyerhoff, H. S., Papenmeier, F., & Huff, M. (2017b). Studying visual attention using the multiple object tracking paradigm: A tutorial review. *Attention, Perception, & Psychophysics*, 79(5), 1255–1274. <https://doi.org/10.3758/s13414-017-1338-1>
- Michaels, J., Chaumillon, R., Mejia-Romero, S., Bernardin, D., & Faubert, J. (2022). Three-dimensional multiple object tracking improves young adult cognitive abilities associated with driving: Evidence for transfer to the useful field of view. *NeuroReport*, 33(12), 504. <https://doi.org/10.1097/WNR.0000000000001807>
- Mielke, M. M., Machulda, M. M., Hagen, C. E., Edwards, K. K., Roberts, R. O., Pankratz, V. S., Knopman, D. S., Jack Jr, C. R., & Petersen, R. C. (2015). Performance of the cogstate computerized battery in the mayo clinic study on aging. *Alzheimer’s & Dementia*, 11(11), 1367–1376.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>
- Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3), 141–144. [https://doi.org/10.1016/S1364-6613\(03\)00029-9](https://doi.org/10.1016/S1364-6613(03)00029-9)
- Milne, S., Cook, J., Shiu, E., & McFadyen, A. (1997). Adapting to learner attributes: Experiments using an adaptive tutoring system [Publisher: Taylor & Francis]. *Educational psychology*, 17(1-2), 141–155.
- MILNER, B. R. E. N. D. A. (1963). Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes. *Archives of Neurology*, 9(1), 90–100. <https://doi.org/10.1001/archneur.1963.00460070100010>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000a). 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.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000b). 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. <https://doi.org/10.1006/cogp.1999.0734>

- Moen, F., Hrozanova, M., & Stiles, T. (2018a). The effects of perceptual-cognitive training with Neurotracker on executive brain functions among elite athletes. *Cogent Psychology*, 5. <https://doi.org/10.1080/23311908.2018.1544105>
- Moen, F., Hrozanova, M., & Stiles, T. (2018b). The effects of perceptual-cognitive training with Neurotracker on executive brain functions among elite athletes (A. K. Travlos, Ed.). *Cogent Psychology*, 5(1), 1544105. <https://doi.org/10.1080/23311908.2018.1544105>
- Mohr, D. C., Schueller, S. M., Riley, W. T., Brown, C. H., Cuijpers, P., Duan, N., Kwasny, M. J., Stiles-Shields, C., & Cheung, K. (2015). Trials of Intervention Principles: Evaluation Methods for Evolving Behavioral Intervention Technologies [Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada]. *Journal of Medical Internet Research*, 17(7), e4391. <https://doi.org/10.2196/jmir.4391>
- Mohsenzadeh, Y., Mullin, C., Oliva, A., & Pantazis, D. (2019). The perceptual neural trace of memorable unseen scenes. *Scientific reports*, 9(1), 1–10.
- Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, 7(3), 134–140.
- Monsell, S., Sumner, P., & Waters, H. (2003a). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & cognition*, 31(3), 327–342.
- Monsell, S., Sumner, P., & Waters, H. (2003b). Task-set reconfiguration with predictable and unpredictable task switches. *Memory & Cognition*, 31(3), 327–342. <https://doi.org/10.3758/BF03194391>
- Monterrat, B., Yessad, A., Bouchet, F., Lavoué, E., & Luengo, V. (2017). MAGAM : Un modèle générique pour l’adaptation multi-aspects dans les EIAH. *Environnements Informatiques Pour l’Apprentissage Humain*, 29–40. Retrieved June 8, 2024, from <https://hal.science/hal-01517137>
- Moore, J. (2011). Behaviorism. *The Psychological Record*, 61(3), 449–463. <https://doi.org/10.1007/BF03395771>
- Moore, S., Nguyen, H. A., Bier, N., Domadia, T., & Stamper, J. (2022). Assessing the Quality of Student-Generated Short Answer Questions Using GPT-3. *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption: 17th European Conference on Technology Enhanced Learning, EC-TEL 2022, Toulouse, France, September 12–16, 2022, Proceedings*, 243–257. https://doi.org/10.1007/978-3-031-16290-9_18
- Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods [Publisher: Taylor & Francis]. *Interactive Learning Environments*, 29(1), 142–163.
- Mu, T., Goel, K., & Brunskill, E. (2017). Program2Tutor: Combining automatic curriculum generation with multi-armed bandits for intelligent tutoring systems. *Conference on Neural Information Processing Systems*. Retrieved April 10, 2024, from http://www.teaching-machines.cc/nips2017/papers/nips17-teaching_paper-12.pdf
- Mu, T., Wang, S., Andersen, E., & Brunskill, E. (2018). Combining adaptivity with progression ordering for intelligent tutoring systems. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, 1–4. <https://doi.org/10.1145/3231644.3231672>
- Mu, T., Wang, S., Andersen, E., & Brunskill, E. (2021). Automatic Adaptive Sequencing in a Webgame. In A. I. Cristea & C. Troussas (Eds.), *Intelligent Tutoring Systems*

- (pp. 430–438, Vol. 12677). Springer International Publishing. https://doi.org/10.1007/978-3-030-80421-3_47
- Muir, S. W., Gopaul, K., & Montero Odasso, M. M. (2012). The role of cognitive impairment in fall risk among older adults: A systematic review and meta-analysis. *Age and Ageing*, 41(3), 299–308. <https://doi.org/10.1093/ageing/afs012>
- Müller, H., & Krummenacher, J. (2006). Visual search and selective attention. *Visual Cognition - VIS COGN*, 14, 389–410. <https://doi.org/10.1080/13506280500527676>
- Musteata, S., Yoshida, K., Baranzini, D., Spaner, C., Taneja, C., Abutalebi, J., Christie, B., Brian, R., & Christie. (2019a). Perceptual-Cognitive Training Can Improve Cognition in Older Adults with Subjective Cognitive Decline, 1–15.
- Musteata, S., Yoshida, K., Baranzini, D., Spaner, C., Taneja, C., Abutalebi, J., Christie, B., Brian, R., & Christie. (2019b). Perceptual-Cognitive Training Can Improve Cognition in Older Adults with Subjective Cognitive Decline, 1–15.
- Nahum-Shani, I., Shaw, S. D., Carpenter, S. M., Murphy, S. A., & Yoon, C. (2022). Engagement in digital interventions. *American Psychologist*, 77(7), 836–852. <https://doi.org/10.1037/amp0000983>
- Nakai, T., & Nishimoto, S. (2020). Quantitative models reveal the organization of diverse cognitive functions in the brain. *Nature communications*, 11(1), 1–12.
- Nash, K., Schiller, B., Gianotti, L. R., Baumgartner, T., & Knoch, D. (2013). Electrophysiological indices of response inhibition in a go/nogo task predict self-control in a social context. *PLoS One*, 8(11), e79462.
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. 53(4), 695–699.
- Neri, F., Smeralda, C. L., Momi, D., Sprugnoli, G., Menardi, A., Ferrone, S., Rossi, S., Rossi, A., Di Lorenzo, G., & Santarnecchi, E. (2021). Personalized Adaptive Training Improves Performance at a Professional First-Person Shooter Action Videogame. *Frontiers in Psychology*, 12. Retrieved June 26, 2023, from <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.598410>
- NeuroTracker. (n.d.). NeuroTracker | #1 Cognitive Training System in the World. Retrieved June 26, 2023, from <https://www.neurotrackerx.com/>
- Nguyen, H. A., Bhat, S., Moore, S., Bier, N., & Stamper, J. (2022). Towards Generalized Methods for Automatic Question Generation in Educational Domains. In I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, & T. Farrell (Eds.), *Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption* (pp. 272–284). Springer International Publishing. https://doi.org/10.1007/978-3-031-16290-9_20
- Nguyen, L., Murphy, K., & Andrews, G. (2019). Immediate and long-term efficacy of executive functions cognitive training in older adults: A systematic review and meta-analysis. [Publisher: American Psychological Association]. *Psychological bulletin*, 145(7), 698.
- Nguyen, L. H., & Holmes, S. (2019). Ten quick tips for effective dimensionality reduction. *PLoS computational biology*, 15(6), e1006907.
- Nichols, D., & Walker, D. (2005). Infants' and Toddlers' Television Viewing and Language Outcomes. *American Behavioral Scientist*, 48. <https://doi.org/10.1177/0002764204271505>
- Nigg, J. T., Willcutt, E. G., Doyle, A. E., & Sonuga-Barke, E. J. S. (2005). Causal Heterogeneity in Attention-Deficit/Hyperactivity Disorder: Do We Need Neuropsychological

- chologically Impaired Subtypes? *Biological Psychiatry*, 57(11), 1224–1230. <https://doi.org/10.1016/j.biopsych.2004.08.025>
- Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, 24(4), 427–469. <https://doi.org/10.1007/s40593-014-0029-5>
- Nyquist, J. B., Lappin, J. S., Zhang, R., & Tadin, D. (2016a). Perceptual training yields rapid improvements in visually impaired youth. *Scientific Reports*, 6(1), 37431. <https://doi.org/10.1038/srep37431>
- Nyquist, J. B., Lappin, J. S., Zhang, R., & Tadin, D. (2016b). Perceptual training yields rapid improvements in visually impaired youth. *Scientific Reports*, 6(1), 37431. <https://doi.org/10.1038/srep37431>
- 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. <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- Obukhova, L. F., & Korepanova, I. A. (2009). The Zone of Proximal Development: A Spatiotemporal Model. *Journal of Russian & East European Psychology*, 47(6), 25–47. <https://doi.org/10.2753/RPO1061-0405470602>
- Oksama, L., & Hyönä, J. (2004a). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach [Publisher: Routledge _eprint: <https://doi.org/10.1080/13506280344000473>]. *Visual Cognition*, 11(5), 631–671. <https://doi.org/10.1080/13506280344000473>
- Oksama, L., & Hyönä, J. (2004b). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, 11(5), 631–671. <https://doi.org/10.1080/13506280344000473>
- Oksama, L., & Hyönä, J. (2004c). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition*, 11(5), 631–671. <https://doi.org/10.1080/13506280344000473>
- Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology*, 56(4), 237–283. <https://doi.org/10.1016/j.cogpsych.2007.03.001>
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024, March 4). *GPT-4 Technical Report*. arXiv: 2303.08774 [cs]. <https://doi.org/10.48550/arXiv.2303.08774>
- Oswald, T. K., Rumbold, A. R., Kedzior, S. G. E., & Moore, V. M. (2020). Psychological impacts of “screen time” and “green time” for children and adolescents: A systematic scoping review. *PLOS ONE*, 15(9), e0237725. <https://doi.org/10.1371/journal.pone.0237725>
- Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic Motivation Systems for Autonomous Mental Development [Conference Name: IEEE Transactions on Evolutionary Computation]. *IEEE Transactions on Evolutionary Computation*, 11(2), 265–286. <https://doi.org/10.1109/TEVC.2006.890271>

- Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, 1. <https://doi.org/10.3389/neuro.12.006.2007>
- Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016a). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Progress in brain research*, 229, 257–284. Retrieved June 11, 2024, from <https://www.sciencedirect.com/science/article/pii/S0079612316300589>
- Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016b). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Progress in brain research*, 229, 257–284. Retrieved April 21, 2024, from <https://www.sciencedirect.com/science/article/pii/S0079612316300589>
- Owsley, C. (2016). Vision and aging. *Annual review of vision science*, 2, 255–271.
- Ozturk, N. (2015). A short review of research on metacognition training with elementary students.
- P. Simmons, J., D. Nelson, L., & Simonsohn, U. (2021). Pre-registration: Why and How. *Journal of Consumer Psychology*, 31(1), 151–162. <https://doi.org/10.1002/jcpy.1208>
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture. *Instructional Science*, 32(1/2), 1–8. Retrieved May 1, 2024, from <https://www.jstor.org/stable/41953634>
- Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. *Journal of Educational Psychology*, 86, 122–133. <https://doi.org/10.1037/0022-0663.86.1.122>
- Pagani, L. S., Fitzpatrick, C., & Barnett, T. A. (2013). Early childhood television viewing and kindergarten entry readiness. *Pediatric Research*, 74(3), 350–355. <https://doi.org/10.1038/pr.2013.105>
- Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, 592(7855), 601–605.
- Pantoni, L., Poggesi, A., Diciotti, S., Valenti, R., Orsolini, S., Della Rocca, E., Inzitari, D., Mascalchi, M., & Salvadori, E. (2017). Effect of Attention Training in Mild Cognitive Impairment Patients with Subcortical Vascular Changes: The RehAtt Study. *Journal of Alzheimer's disease: JAD*, 60(2), 615–624. <https://doi.org/10.3233/JAD-170428>
- Park, D. C., & Reuter-Lorenz, P. (2009). The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annual Review of Psychology*, 60, 173–196. <https://doi.org/10.1146/annurev.psych.59.103006.093656>
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744. <https://doi.org/10.1038/89532>
- Parsons, B., & Faubert, J. (2021). Enhancing learning in a perceptual-cognitive training paradigm using EEG-neurofeedback. *Scientific Reports*, 11(1), 4061. <https://doi.org/10.1038/s41598-021-83456-x>
- Parsons, B., Magill, T., Boucher, A., Zhang, M., Zogbo, K., Bérubé, S., Scheffer, O., Beauregard, M., & Faubert, J. (2016a). Enhancing Cognitive Function Using Perceptual-Cognitive Training. *Clinical EEG and neuroscience*, 47(1), 37–47. <https://doi.org/10.1177/1550059414563746>
- Parsons, B., Magill, T., Boucher, A., Zhang, M., Zogbo, K., Bérubé, S., Scheffer, O., Beauregard, M., & Faubert, J. (2016b). Enhancing Cognitive Function Using Perceptual-Cognitive Training. *Clinical EEG and Neuroscience*, 47(1), 37–47. <https://doi.org/10.1177/1550059414563746>

- Pech, M. (2023, March 17). *Détection de situations à risques en vie quotidienne chez des personnes âgées fragiles au moyen de méthodes d'intelligence artificielle : Expérimentation d'un dispositif en population générale* [These de doctorat]. Bordeaux. Retrieved June 11, 2024, from <https://theses.fr/2023BORD0057>
- Pedullà, L., Brichetto, G., Tacchino, A., Vassallo, C., Zaratin, P., Battaglia, M. A., Bonzano, L., & Bove, M. (2016a). Adaptive vs. non-adaptive cognitive training by means of a personalized App: A randomized trial in people with multiple sclerosis [Publisher: BioMed Central]. *Journal of neuroengineering and rehabilitation*, 13(1), 1–10.
- Pedullà, L., Brichetto, G., Tacchino, A., Vassallo, C., Zaratin, P., Battaglia, M. A., Bonzano, L., & Bove, M. (2016b). Adaptive vs. non-adaptive cognitive training by means of a personalized App: A randomized trial in people with multiple sclerosis. *Journal of NeuroEngineering and Rehabilitation*, 13(1), 88. <https://doi.org/10.1186/s12984-016-0193-y>
- Pelánek, R. (2017a). Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27(3), 313–350. <https://doi.org/10.1007/s11257-017-9193-2>
- Pelánek, R. (2017b). Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27(3), 313–350. <https://doi.org/10.1007/s11257-017-9193-2>
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4(12), 1136–1169. <https://doi.org/10.1167/4.12.12>
- Peng, P., & Miller, A. C. (2016a). Does attention training work? A selective meta-analysis to explore the effects of attention training and moderators. *Learning and Individual Differences*, 45, 77–87. <https://doi.org/10.1016/j.lindif.2015.11.012>
- Peng, P., & Miller, A. C. (2016b). Does attention training work? A selective meta-analysis to explore the effects of attention training and moderators. *Learning and Individual Differences*, 45, 77–87. <https://doi.org/10.1016/j.lindif.2015.11.012>
- Peretz, C., Korczyn, A. D., Shatil, E., Aharonson, V., Birnboim, S., & Giladi, N. (2011). Computer-based, personalized cognitive training versus classical computer games: A randomized double-blind prospective trial of cognitive stimulation. *Neuroepidemiology*, 36(2), 91–99. <https://doi.org/10.1159/000323950>
- Pergher, V., Alizadeh Shalchy, M., Pahor, A., Van Hulle, M., Jaeggi, S., & Seitz, A. (2020). Divergent Research Methods Limit Understanding of Working Memory Training. *Journal of Cognitive Enhancement*, 4. <https://doi.org/10.1007/s41465-019-00134-7>
- Peters, D., Calvo, R. A., & Ryan, R. M. (2018). Designing for Motivation, Engagement and Wellbeing in Digital Experience. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.00797>
- Pillay, N. (2020). The impact of genetic programming in education. *Genetic Programming and Evolvable Machines*, 21(1-2), 87–97. <https://doi.org/10.1007/s10710-019-09362-4>
- [TLDR] An overview of the impact that genetic programming has had in teaching and learning is provided, the use of genetic programming in intelligent tutoring systems, predicting student performance and designing learning environments is examined, and a critical analysis of Genetic programming in education is provided.
- Plass, J. L., Homer, B. D., Pawar, S., Brenner, C., & MacNamara, A. P. (2019a). The effect of adaptive difficulty adjustment on the effectiveness of a game to develop executive function skills for learners of different ages. *Cognitive Development*, 49, 56–67. <https://doi.org/10.1016/j.cogdev.2018.11.006>

- Plass, J. L., Homer, B. D., Pawar, S., Brenner, C., & MacNamara, A. P. (2019b). The effect of adaptive difficulty adjustment on the effectiveness of a game to develop executive function skills for learners of different ages. *Cognitive Development*, 49, 56–67. <https://doi.org/10.1016/j.cogdev.2018.11.006>
- Pliakos, K., Joo, S.-H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education*, 137, 91–103. <https://doi.org/10.1016/j.compedu.2019.04.009>
- Portelas, R., Colas, C., Hofmann, K., & Oudeyer, P.-Y. (2020). Teacher algorithms for curriculum learning of Deep RL in continuously parameterized environments. *Proceedings of the Conference on Robot Learning*, 835–853. Retrieved June 24, 2024, from <https://proceedings.mlr.press/v100/portelas20a.html>
- Portelas, R., Colas, C., Weng, L., Hofmann, K., & Oudeyer, P.-Y. (2020, May 28). *Automatic Curriculum Learning For Deep RL: A Short Survey*. arXiv: 2003.04664 [cs, stat]. <https://doi.org/10.48550/arXiv.2003.04664>
- Pothier, K., Benguigui, N., Kulpa, R., & Chavoix, C. (2015). Multiple Object Tracking While Walking: Similarities and Differences Between Young, Young-Old, and Old-Old Adults. *The Journals of Gerontology: Series B*, 70(6), 840–849. <https://doi.org/10.1093/geronb/gbu047>
- Poupard, M., Larrue, F., Sauzéon, H., & Tricot, A. (2022, December). *A systematic review of immersive technologies for education: Effects of cognitive load and curiosity state on learning performance*. Retrieved May 2, 2024, from <https://hal.science/hal-03906797>
- Pourcel, J., Colas, C., Molinaro, G., Oudeyer, P.-Y., & Teodorescu, L. (2024, May 29). *ACES: Generating Diverse Programming Puzzles with Autotelic Generative Models*. arXiv: 2310.10692 [cs]. <https://doi.org/10.48550/arXiv.2310.10692>
- Preiss, M., Shatil, E., Cermakova, R., Cimermannova, D., & Flesher, I. (2013). Personalized cognitive training in unipolar and bipolar disorder: A study of cognitive functioning. *Frontiers in human neuroscience*, 7, 108.
- Pylyshyn, Z. W. (2001a). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1-2), 127–158. [https://doi.org/10.1016/s0010-0277\(00\)00156-6](https://doi.org/10.1016/s0010-0277(00)00156-6)
- Pylyshyn, Z. W. (2001b). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80(1-2), 127–158. [https://doi.org/10.1016/s0010-0277\(00\)00156-6](https://doi.org/10.1016/s0010-0277(00)00156-6)
- Pylyshyn, Z. (1994). Some primitive mechanisms of spatial attention. *Cognition*, 50(1), 363–384. [https://doi.org/10.1016/0010-0277\(94\)90036-1](https://doi.org/10.1016/0010-0277(94)90036-1)
- Pylyshyn, Z. W., & Storm, R. W. (1988a). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3), 179–197. <https://doi.org/10.1163/156856888X00122>
- Pylyshyn, Z. W., & Storm, R. W. (1988b). Tracking multiple independent targets: Evidence for a parallel tracking mechanism [Publisher: Brill]. *Spatial vision*, 3(3), 179–197. Retrieved June 25, 2024, from https://brill.com/downloadpdf/view/journals/sv/3/3/article-p179_3.pdf
- Pylyshyn, Z. W., & Storm, R. W. (1988c). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3), 179–197.
- Quaglini, S., Panzarasa, S., Giorgiani, T., Zucchella, C., Bartolo, M., Sinforiani, E., & Sandrini, G. (2009). Ontology-Based Personalization and Modulation of Computerized Cognitive Exercises. In C. Combi, Y. Shahar, & A. Abu-Hanna (Eds.), *Artificial Intelligence in Medicine* (pp. 240–244). Springer. https://doi.org/10.1007/978-3-642-02976-9_34

- Quillion-Dupré, L., Monfort, E., & Rialle, V. (2016). Mieux comprendre l'usage et la transmission des technologies d'information et de communication aux personnes âgées. *NPG Neurologie-Psychiatrie-Gériatrie*, 16(96), 305–312. Retrieved June 12, 2024, from <https://www.sciencedirect.com/science/article/pii/S1627483016300976>
- Radesky, J. S., & Christakis, D. A. (2016). Increased Screen Time: Implications for Early Childhood Development and Behavior. *Pediatric Clinics of North America*, 63(5), 827–839. <https://doi.org/10.1016/j.pcl.2016.06.006>
- Rathnayaka, M., Watawala, W., Manamendra, M. G., Silva, S., Kasthurirathna, D., & Jayalath, T. (2021). Cognitive Rehabilitation based Personalized Solution for Dementia Patients using Reinforcement Learning. *2021 IEEE International Systems Conference (SysCon)*, 1–6.
- Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, 26(6), 462–483. <https://doi.org/10.1016/j.tics.2022.03.007>
- Ray, R. D., & Belden, N. (2007). Teaching college level content and reading comprehension skills simultaneously via an artificially intelligent adaptive computerized instructional system [Publisher: Springer]. *The Psychological Record*, 57, 201–218.
- RehaCom. (n.d.). RehaCom – Cognitive Therapy | HASOMED GmbH. Retrieved June 26, 2023, from <https://hasomed.de/en/products/rehacom/>
- Reidy, L., Chan, D., Nduka, C., & Gunes, H. (2020). Facial electromyography-based adaptive virtual reality gaming for cognitive training. *Proceedings of the 2020 International Conference on Multimodal Interaction*, 174–183.
- Reijnders, J., van Heugten, C., & van Boxtel, M. (2013). Cognitive interventions in healthy older adults and people with mild cognitive impairment: A systematic review [Publisher: Elsevier]. *Ageing research reviews*, 12(1), 263–275.
- Reuter-Lorenz, P. A., & Park, D. C. (2010). Human Neuroscience and the Aging Mind: A New Look at Old Problems. *The Journals of Gerontology: Series B*, 65B(4), 405–415. <https://doi.org/10.1093/geronb/gbq035>
- Richert, R., Robb, M., Fender, J., & Wartella, E. (2010). Word learning from baby videos. *Archives of Pediatrics and Adolescent Medicine*, 164(5), 432–437. <https://doi.org/10.1001/archpediatrics.2010.24>
- Ritter, F. E., Tehranchi, F., & Oury, J. D. (2019). ACT-R: A cognitive architecture for modeling cognition. *WIREs Cognitive Science*, 10(3), e1488. <https://doi.org/10.1002/wcs.1488>
- Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five Personality Factors and Personal Values. *Personality and Social Psychology Bulletin*, 28(6), 789–801. <https://doi.org/10.1177/0146167202289008>
- Roest, L., Keuning, H., & Jeuring, J. (2023, December 3). *Next-Step Hint Generation for Introductory Programming Using Large Language Models*. arXiv: 2312.10055 [cs]. <https://doi.org/10.48550/arXiv.2312.10055>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis With Latent Diffusion Models, 10684–10695. Retrieved July 1, 2024, from https://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html
- Romeas, T., Guldner, A., & Faubert, J. (2016a). 3D-Multiple Object Tracking training task improves passing decision-making accuracy in soccer players. *Psychology of Sport and Exercise*, 22, 1–9. Retrieved June 11, 2024, from <https://www.sciencedirect.com/science/article/pii/S1469029215000631>

- Romeas, T., Guldner, A., & Faubert, J. (2016b). 3D-Multiple Object Tracking training task improves passing decision-making accuracy in soccer players. *Psychology of Sport and Exercise*, 22, 1–9. <https://doi.org/10.1016/j.psychsport.2015.06.002>
- Romeas, T., Guldner, A., & Faubert, J. (2016c). 3D-Multiple Object Tracking training task improves passing decision-making accuracy in soccer players. *Psychology of Sport and Exercise*, 22, 1–9. <https://doi.org/10.1016/j.psychsport.2015.06.002>
- Ropelato, S., Zünd, F., Menozzi, M., Sumner, R., & Magnenat, S. (2017). Adaptive Tutoring on a Virtual Reality Driving Simulator.
- Rosenbaum, J. (1994). Assessment: An Overview. In *Assessing Communication Education*. Routledge.
- Rosseel, Y. (2012). **Lavaan** : An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryan, R. M., & Deci, E. L. (2017, February). *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness* [Google-Books-ID: Bc_DDAAAQBAJ]. Guilford Publications.
- Saba, M., Rwabihama, J.-P., Bouvard, É., Mettling, P., Sztulman, É., Lemarié, N., Piolino, P., & Blanchet, S. (2021, August 31). *The Effect of Attention Process Training (APT-II) On Cognitive and Daily Life Functioning in Patients With a Mild Cognitive Impairment: A Randomized Controlled Trial*. <https://doi.org/10.21203/rs.3.rs-861371/v1>
- Sakaki, M., Yagi, A., & Murayama, K. (2018). Curiosity in old age: A possible key to achieving adaptive aging. *Neuroscience & Biobehavioral Reviews*, 88, 106–116. <https://doi.org/10.1016/j.neubiorev.2018.03.007>
- Sala, G. (2017, August). *Once Upon a Time There Was Far Transfer* [Doctoral dissertation]. Retrieved August 23, 2023, from <https://www.semanticscholar.org/paper/Once-Upon-a-Time-There-Was-Far-Transfer-Sala/40ddc1c44dd875206e6959a8e17664da1a11cc56>
- Sala, G., & Gobet, F. (2017a). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, 53(4), 671–685. <https://doi.org/10.1037/dev0000265>
- Sala, G., & Gobet, F. (2017b). Does Far Transfer Exist? Negative Evidence From Chess, Music, and Working Memory Training. *Current Directions in Psychological Science*, 26(6), 515–520. <https://doi.org/10.1177/0963721417712760>
- Salomon, G. (1983). The differential investment of mental effort in learning from different sources. *Educational Psychologist*, 18(1), 42–50. <https://doi.org/10.1080/00461528309529260>
- Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, 8(4), 535.
- Sampayo-Vargas, S., Cope, C., He, Z., & Byrne, G. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Education*, 69, 452–462. <https://doi.org/10.1016/j.compedu.2013.07.004>
- Sandeep, S., Shelton, C. R., Pahor, A., Jaeggi, S. M., & Seitz, A. R. (2020a). Application of machine learning models for tracking participant skills in cognitive training. *Frontiers in Psychology*, 11, 1532.

- Sandeep, S., Shelton, C. R., Pahor, A., Jaeggi, S. M., & Seitz, A. R. (2020b). Application of machine learning models for tracking participant skills in cognitive training [Publisher: Frontiers Media SA]. *Frontiers in Psychology*, 11, 1532.
- Sansone, C. (1986). A question of competence: The effects of competence and task feedback on intrinsic interest. *Journal of Personality and Social Psychology*, 51(5), 918–931. <https://doi.org/10.1037/0022-3514.51.5.918>
- Santos, C. F. G. D., & Papa, J. P. (2022). Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks. *ACM Computing Surveys*, 54, 213:1–213:25. <https://doi.org/10.1145/3510413>
- Santos, R. M. S., Mendes, C. G., Marques Miranda, D., & Romano-Silva, M. A. (2022). The Association between Screen Time and Attention in Children: A Systematic Review. *Developmental Neuropsychology*, 47(4), 175–192. <https://doi.org/10.1080/87565641.2022.2064863>
- Sasaki, K., & Yamada, Y. (2019). Crowdsourcing visual perception experiments: A case of contrast threshold. *PeerJ*, 7, e8339.
- Sawayama, M., Dobashi, Y., Okabe, M., Hosokawa, K., Koumura, T., Saarela, T., Olkkonen, M., & Nishida, S. (2022). Visual discrimination of optical material properties: A large-scale study. *Journal of Vision*, 22(?), ?–?
- Scharfen, H.-E., & Memmert, D. (2021a). Cognitive training in elite soccer players: Evidence of narrow, but not broad transfer to visual and executive function. *German Journal of Exercise and Sport Research*, 51(2), 135–145. <https://doi.org/10.1007/s12662-020-00699-y>
- Scharfen, H.-E., & Memmert, D. (2021b). Cognitive training in elite soccer players: Evidence of narrow, but not broad transfer to visual and executive function. *German Journal of Exercise and Sport Research*, 51(2), 135–145. <https://doi.org/10.1007/s12662-020-00699-y>
- Schmidhuber, J. (1991). Curious model-building control systems. *[Proceedings] 1991 IEEE International Joint Conference on Neural Networks*, 1458–1463 vol.2. <https://doi.org/10.1109/IJCNN.1991.170605>
- Schmiedek, F. (2016a). Methods and Designs. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 9–18). Springer International Publishing. https://doi.org/10.1007/978-3-319-42662-4_2
- Schmiedek, F. (2016b). Methods and Designs. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 9–18). Springer International Publishing. https://doi.org/10.1007/978-3-319-42662-4_2
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. *Frontiers in Aging Neuroscience*, 2, 27. <https://doi.org/10.3389/fnagi.2010.00027>
- Schmucker, R., Xia, M., Azaria, A., & Mitchell, T. (2023, November 14). *Ruffle&Riley: Towards the Automated Induction of Conversational Tutoring Systems*. arXiv: 2310.01420 [cs]. <https://doi.org/10.48550/arXiv.2310.01420>
- Schneider, W., & McGrew, K. (2012, January 1). The Cattell-Horn-Carroll model of intelligence. In *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144).
- Schneider, W. J., & McGrew, K. S. (2013). Individual differences in the ability to process information. In *The handbook of educational theories* (pp. 767–782). IAP Information Age Publishing.

- Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, 80(1-2), 159–177. [https://doi.org/10.1016/s0010-0277\(00\)00157-8](https://doi.org/10.1016/s0010-0277(00)00157-8)
- Scholl, B. (2009). What Have We Learned about Attention from Multiple-Object Tracking (and Vice Versa)? *Computation, Cognition, and Pylyshyn*.
- Scholl, B. J. (2009). What have we learned about attention from multiple-object tracking (and vice versa)? Retrieved April 18, 2024, from https://direct.mit.edu/books/edited-volume/chapter-pdf/2282825/9780262255196_cab.pdf
- Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, 50(2), 138–166. <https://doi.org/10.1080/00461520.2015.1036274>
- Segijn, C. (2016, December 29). Second Screen Advertising: A Typology of Multiscreening. <https://doi.org/10.1108/978-1-78560-313-620151003>
- Segijn, C. M., Voorveld, H. A. M., Vandeberg, L., & Smit, E. G. (2017). The Battle of the Screens: Unraveling Attention Allocation and Memory Effects When Multiscreening. *Human Communication Research*, 43(2), 295–314. <https://doi.org/10.1111/hcre.12106>
- Seitz, A. R. (2018). A New Framework of Design and Continuous Evaluation to Improve Brain Training. *Journal of Cognitive Enhancement*, 2(1), 78–87. <https://doi.org/10.1007/s41465-017-0058-8>
- Sekuler, R., McLaughlin, C., & Yotsumoto, Y. (2008). Age-related changes in attentional tracking of multiple moving objects. *Perception*, 37(6), 867–876.
- Self, J. (1998). The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education*, 10, 350. Retrieved June 11, 2024, from <https://telearn.hal.science/hal-00197346>
- Sella, E., Carbone, E., Vincenzi, M., Toffalini, E., & Borella, E. (2023). Efficacy of memory training interventions targeting metacognition for older adults: A systematic review and meta-analysis. *Aging & Mental Health*, 27(4), 674–694. <https://doi.org/10.1080/13607863.2022.2122931>
- Shabana, K. M., Lakshminarayanan, C., & Anil, J. K. (2022). CurriculumTutor: An Adaptive Algorithm for Mastering a Curriculum. In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (pp. 319–331). Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_26
- Shah, T. M., Weinborn, M., Verdile, G., Sohrabi, H. R., & Martins, R. N. (2017a). Enhancing Cognitive Functioning in Healthy Older Adults: A Systematic Review of the Clinical Significance of Commercially Available Computerized Cognitive Training in Preventing Cognitive Decline. *Neuropsychology Review*, 27(1), 62–80. <https://doi.org/10.1007/s11065-016-9338-9>
- Shah, T. M., Weinborn, M., Verdile, G., Sohrabi, H. R., & Martins, R. N. (2017b). Enhancing Cognitive Functioning in Healthy Older Adults: A Systematic Review of the Clinical Significance of Commercially Available Computerized Cognitive Training in Preventing Cognitive Decline. *Neuropsychology Review*, 27(1), 62–80. <https://doi.org/10.1007/s11065-016-9338-9>
- Shaked, D., Faulkner, L. M. D., Tolle, K., Wendell, C. R., Waldstein, S. R., & Spencer, R. J. (2020). Reliability and validity of the Conners' Continuous Performance Test. *Applied Neuropsychology: Adult*, 27(5), 478–487. <https://doi.org/10.1080/23279095.2019.1570199>

- Shanon, B. (2018). On the place of representations in cognition. In *Thinking* (pp. 33–49). Routledge. Retrieved June 25, 2024, from <https://www.taylorfrancis.com/chapters/edit/10.4324/9781315802015-4/place-representations-cognition-benny-shanon>
- Shao, Y.-k., Mang, J., Li, P.-l., Wang, J., Deng, T., & Xu, Z.-x. (2015). Computer-Based Cognitive Programs for Improvement of Memory, Processing Speed and Executive Function during Age-Related Cognitive Decline: A Meta-Analysis. *PLOS ONE*, 10(6), e0130831. <https://doi.org/10.1371/journal.pone.0130831>
- Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory*, 5(2), 179–187. <https://doi.org/10.1037/0278-7393.5.2.179>
- Shen, X., & Xu, C. (2021). Research on children’s cognitive development for learning disabilities using recommendation method [Publisher: Wiley Online Library]. *Concurrency and Computation: Practice and Experience*, 33(9), e6097.
- Silva, T., Bratkauskas, J., Barbosa, M., Silva, G., Zumkeller, M., Moraes, L., Lessa, P., Cardoso, N., Ordonez, T., & Brucki, S. (2022). Long-term studies in cognitive training for older adults: A systematic review. *Dementia & Neuropsychologia*, 16. <https://doi.org/10.1590/1980-5764-dn-2021-0064>
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016a). Do “Brain-Training” Programs Work? *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 17(3), 103–186. <https://doi.org/10.1177/1529100616661983>
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016b). Do “Brain-Training” Programs Work? [Publisher: SAGE Publications Inc]. *Psychological Science in the Public Interest*, 17(3), 103–186. <https://doi.org/10.1177/1529100616661983>
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016c). Do “Brain-Training” Programs Work? [Publisher: SAGE Publications Inc]. *Psychological Science in the Public Interest*, 17(3), 103–186. <https://doi.org/10.1177/1529100616661983>
- Singer, U., Polyak, A., Hayes, T., Yin, X., An, J., Zhang, S., Hu, Q., Yang, H., Ashual, O., Gafni, O., Parikh, D., Gupta, S., & Taigman, Y. (2022, September 29). *Make-A-Video: Text-to-Video Generation without Text-Video Data*. arXiv: 2209.14792 [cs]. <https://doi.org/10.48550/arXiv.2209.14792>
- Singh, A., Chakraborty, S., He, Z., Tian, S., Zhang, S., Lustria, M. L. A., Charness, N., Roque, N. A., Harrell, E. R., & Boot, W. R. (2022). Deep learning-based predictions of older adults’ adherence to cognitive training to support training efficacy. *Frontiers in Psychology*, 13. Retrieved June 30, 2023, from <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.980778>
- Singley, M. K., & Anderson, J. R. (1989). *The Transfer of Cognitive Skill*. Harvard University Press.
- Skinner, B. F. (1961). Teaching Machines. *Scientific American*, 205(5), 90–106. Retrieved April 9, 2024, from <https://www.jstor.org/stable/24937132>
- Skogsberg, K., Grabowecky, M., Wilt, J., Revelle, W., Iordanescu, L., & Suzuki, S. (2015a). A relational structure of voluntary visual-attention abilities. *Journal of Experimental Psychology. Human Perception and Performance*, 41(3), 761–789. <https://doi.org/10.1037/a0039000>
- Skogsberg, K., Grabowecky, M., Wilt, J., Revelle, W., Iordanescu, L., & Suzuki, S. (2015b). A relational structure of voluntary visual-attention abilities: Journal of Experimental Psychology: Human Perception and Performance [Publisher: American

- Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, 41(3), 761–789. <https://doi.org/10.1037/a0039000>
- Solana, J., Cáceres, C., García-Molina, A., Chausa, P., Opisso, E., Roig-Rovira, T., Menasalvas, E., Tormos-Muñoz, J. M., & Gómez, E. J. (2014). Intelligent Therapy Assistant (ITA) for cognitive rehabilitation in patients with acquired brain injury [Publisher: Springer]. *BMC medical informatics and decision making*, 14, 1–13.
- Son, L. K., & Simon, D. A. (2012). Distributed Learning: Data, Metacognition, and Educational Implications. *Educational Psychology Review*, 24(3), 379–399. <https://doi.org/10.1007/s10648-012-9206-y>
- Soofi, A. A., & Ahmed, M. U. (2019). A systematic review of domains, techniques, delivery modes and validation methods for intelligent tutoring systems [Publisher: Science and Information (SAI) Organization Limited]. *International Journal of Advanced Computer Science and Applications*, 10(3).
- Soreq, E., Violante, I. R., Daws, R. E., & Hampshire, A. (2021). Neuroimaging evidence for a network sampling theory of individual differences in human intelligence test performance. *Nature communications*, 12(1), 1–13.
- Sowden, P. T., Rose, D., & Davies, I. R. L. (2002). Perceptual learning of luminance contrast detection: Specific for spatial frequency and retinal location but not orientation. *Vision Research*, 42(10), 1249–1258. [https://doi.org/10.1016/s0042-6989\(02\)00019-6](https://doi.org/10.1016/s0042-6989(02)00019-6)
- Stahl, A. E., & Feigenson, L. (2015). Cognitive development. Observing the unexpected enhances infants’ learning and exploration. *Science (New York, N. Y.)*, 348(6230), 91–94. <https://doi.org/10.1126/science.aaa3799>
- Star Arts (**typedirector**). (2016). *Entretien exclusif avec Salvador Dali* [Video]. Retrieved July 10, 2024, from <https://www.youtube.com/watch?v=3LWbfkAURvY>
- Steyvers, M., & Schafer, R. J. (2020a). Inferring latent learning factors in large-scale cognitive training data [Publisher: Nature Publishing Group UK London]. *Nature Human Behaviour*, 4(11), 1145–1155.
- Steyvers, M., & Schafer, R. J. (2020b). Inferring latent learning factors in large-scale cognitive training data. *Nature Human Behaviour*, 4(11), 1145–1155.
- Strobach, T., & Karbach, J. (2016). Introduction. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 1–5). Springer International Publishing. https://doi.org/10.1007/978-3-319-42662-4_1
- Strobach, T., & Karbach, J. (Eds.). (2021). *Cognitive Training: An Overview of Features and Applications*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-39292-5>
- Strong, R. W., & Alvarez, G. A. (2017a). Training enhances attentional expertise, but not attentional capacity: Evidence from content-specific training benefits. *Journal of Vision*, 17(4), 4. <https://doi.org/10.1167/17.4.4>
- Strong, R. W., & Alvarez, G. A. (2017b). Training enhances attentional expertise, but not attentional capacity: Evidence from content-specific training benefits. *Journal of Vision*, 17(4), 4. <https://doi.org/10.1167/17.4.4>
- Stroop, J. R. (1935). Studies of Interference in Serial Verbal Reactions. *Journal of Experimental Psychology*, 18(6), 643. <https://doi.org/10.1037/h0054651>
- Stubbert, E. (2016). *Tracking multiple objects in space: Similarities between children with attention deficit hyperactivity disorder and typically developing children*. McGill University (Canada). Retrieved April 18, 2024, from <https://search.proquest.com/openview/16cb3439fd3ac1a3be6b0f9fe793adeb/1?pq-origsite=gscholar&cbl=18750&diss=y>

- Stubbert, E., Tullo, D., Faubert, J., Bertone, A., & Burack, J. A. (2023). Biological motion and multiple object tracking performance develop similarly from childhood through early adolescence. *Cognitive Development*, 67, 101360. <https://doi.org/10.1016/j.cogdev.2023.101360>
- Subramanian, A., Chitlangia, S., & Baths, V. (2022). Reinforcement learning and its connections with neuroscience and psychology. *Neural Networks*, 145, 271–287. <https://doi.org/10.1016/j.neunet.2021.10.003>
- Suganuma, M., & Yokosawa, K. (2006). Grouping and Trajectory Storage in Multiple Object Tracking: Impairments Due to Common Item Motions. *Perception*, 35(4), 483–495. <https://doi.org/10.1068/p5487>
- Sun, S., Joy, M., & Griffiths, N. (2007). The use of learning objects and learning styles in a multi-agent education system [Publisher: Association for the Advancement of Computing in Education (AACE)]. *Journal of Interactive Learning Research*, 18(3), 381–398.
- Sungeelee, V., Lorette, A., Sigaud, O., & Caramiaux, B. (2024). Interactive curriculum learning increases and homogenizes motor smoothness. *Scientific Reports*, 14(1), 2843. <https://doi.org/10.1038/s41598-024-53253-3>
- Sweller, J. (2011, January 1). CHAPTER TWO - Cognitive Load Theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (pp. 37–76, Vol. 55). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Taatgen, N. (2013). The Nature and Transfer of Cognitive Skills. *Psychological review*, 120, 439–471. <https://doi.org/10.1037/a0033138>
- Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological review*, 120(3), 439. Retrieved July 1, 2024, from <https://psycnet.apa.org/record/2013-19654-001>
- Taatgen, N. A. (2021). Theoretical Models of Training and Transfer Effects. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 41–54). Springer International Publishing. https://doi.org/10.1007/978-3-030-39292-5_4
- Tacchino, A., Pedullà, L., Bonzano, L., Vassallo, C., Battaglia, M. A., Mancardi, G., Bove, M., & Brichetto, G. (2015). A New App for At-Home Cognitive Training: Description and Pilot Testing on Patients with Multiple Sclerosis [Company: JMIR mHealth and uHealth Distributor: JMIR mHealth and uHealth Institution: JMIR mHealth and uHealth Label: JMIR mHealth and uHealth Publisher: JMIR Publications Inc., Toronto, Canada]. *JMIR mHealth and uHealth*, 3(3), e4269. <https://doi.org/10.2196/mhealth.4269>
- Tadin, D., Nyquist, J. B., Lusk, K. E., Corn, A. L., & Lappin, J. S. (2012). Peripheral Vision of Youths with Low Vision: Motion Perception, Crowding, and Visual Search. *Investigative Ophthalmology & Visual Science*, 53(9), 5860–5868. <https://doi.org/10.1167/iovs.12-10350>
- Taherdoost, H. (2016). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management (IJARM)*, 5. Retrieved July 1, 2024, from <https://hal.science/hal-02546799>

- Taya, F., Sun, Y., Babiloni, F., Thakor, N., & Bezerianos, A. (2015). Brain enhancement through cognitive training: A new insight from brain connectome. *Frontiers in Systems Neuroscience*, 9. <https://doi.org/10.3389/fnsys.2015.00044>
- Taylor, V. A., Daneault, V., Grant, J., Scavone, G., Breton, E., Roffe-Vidal, S., Courtemanche, J., Lavarenne, A. S., Marrelec, G., Benali, H., & Beauregard, M. (2013). Impact of meditation training on the default mode network during a restful state. *Social Cognitive and Affective Neuroscience*, 8(1), 4–14. <https://doi.org/10.1093/scan/nsr087>
- Ten, A. (2022). *The Role of Progress-Based Intrinsic Motivation in Learning: Evidence from Human Behavior and Future Directions* [Doctoral dissertation, Université de Bordeaux]. Retrieved June 24, 2024, from <https://theses.hal.science/tel-03675261/>
- Ten, A., Kaushik, P., Oudeyer, P.-Y., & Gottlieb, J. (2021a). Humans monitor learning progress in curiosity-driven exploration. *Nature Communications*, 12(1), 5972. <https://doi.org/10.1038/s41467-021-26196-w>
- Ten, A., Kaushik, P., Oudeyer, P.-Y., & Gottlieb, J. (2021b). Humans monitor learning progress in curiosity-driven exploration. *Nature Communications*, 12(1), 5972. <https://doi.org/10.1038/s41467-021-26196-w>
- Thompson, T. W., Gabrieli, J. D. E., & Alvarez, G. A. (2010). Adaptive Training in Multiple Object Tracking Expands Attentional Capacity. *Journal of Vision*, 10(7), 308. <https://doi.org/10.1167/10.7.308>
- Thorndike, E. (1912). *Education. A first book*. Macmillan.
- Titz, C., & Karbach, J. (2014a). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research*, 78(6), 852–868. <https://doi.org/10.1007/s00426-013-0537-1>
- Titz, C., & Karbach, J. (2014b). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research*, 78(6), 852–868. <https://doi.org/10.1007/s00426-013-0537-1>
- Tombaugh, T. N. (2006). A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology*, 21(1), 53–76. <https://doi.org/10.1016/j.acn.2005.07.006>
- Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A Comprehensive Review [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1532-5415.1992.tb01992.x>]. *Journal of the American Geriatrics Society*, 40(9), 922–935. <https://doi.org/10.1111/j.1532-5415.1992.tb01992.x>
- Traut, H. J., Guild, R. M., & Munakata, Y. (2021a). Why Does Cognitive Training Yield Inconsistent Benefits? A Meta-Analysis of Individual Differences in Baseline Cognitive Abilities and Training Outcomes. *Frontiers in Psychology*, 12, 662139. <https://doi.org/10.3389/fpsyg.2021.662139>
- Traut, H. J., Guild, R. M., & Munakata, Y. (2021b). Why Does Cognitive Training Yield Inconsistent Benefits? A Meta-Analysis of Individual Differences in Baseline Cognitive Abilities and Training Outcomes. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.662139>
- Treutwein, B. (1995). Adaptive psychophysical procedures [Publisher: Elsevier]. *Vision research*, 35(17), 2503–2522.
- Treviño, M., Zhu, X., Lu, Y., Scheuer, L., Passell, E., Huang, G., Germine, L., & Horowitz, T. (2021). How do we measure attention? Using factor analysis to establish construct validity of neuropsychological tests. *Cognitive Research: Principles and Implications*, 6, 51. <https://doi.org/10.1186/s41235-021-00313-1>

- Treviño, M., Zhu, X., Lu, Y. Y., Scheuer, L. S., Passell, E., Huang, G. C., Germine, L. T., & Horowitz, T. S. (2021). How do we measure attention? Using factor analysis to establish construct validity of neuropsychological tests. *Cognitive Research: Principles and Implications*, 6, 51. <https://doi.org/10.1186/s41235-021-00313-1>
- Trick, L. M., Jaspers-Fayer, F., & Sethi, N. (2005). Multiple-object tracking in children: The “Catch the Spies” task. *Cognitive Development*, 20(3), 373–387. <https://doi.org/10.1016/j.cogdev.2005.05.009>
- Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 331.
- Tricot, A. (1998). Charge cognitive et apprentissage. Une présentation des travaux de John Sweller. *Revue de Psychologie de L'Education*, 1.
- Tsiakas, K., Abujelala, M., & Makedon, F. (2018). Task Engagement as Personalization Feedback for Socially-Assistive Robots and Cognitive Training [Number: 2 Publisher: Multidisciplinary Digital Publishing Institute]. *Technologies*, 6(2), 49. <https://doi.org/10.3390/technologies6020049>
- Tullo, D., Bertone, A., Guy, J., & Faubert, J. (2018). Training with a three-dimensional multiple object-tracking (3D-MOT) paradigm improves attention in students with a neurodevelopmental condition: A randomized controlled trial [Publisher: Wiley-Blackwell]. *Developmental Science*, 21(6), N.PAG–N.PAG. <https://doi.org/10.1111/desc.12670>
- Tullo, D., Faubert, J., & Bertone, A. (2018a). The characterization of attention resource capacity and its relationship with fluid reasoning intelligence: A multiple object tracking study. *Intelligence*, 69, 158–168. <https://doi.org/10.1016/j.intell.2018.06.001>
- Tullo, D., Faubert, J., & Bertone, A. (2018b). Examining the benefits of training attention with Multiple Object-Tracking for individuals diagnosed with a neurodevelopmental condition: A cross-over, cognitive training study. *Journal of Vision*, 18(10), 1021. <https://doi.org/10.1167/18.10.1021>
- Tullo, D., Guy, J., Faubert, J., & Bertone, A. (2018). Training with a three-dimensional multiple object-tracking (3D-MOT) paradigm improves attention in students with a neurodevelopmental condition: A randomized controlled trial. *Developmental Science*, 21(6), e12670. <https://doi.org/10.1111/desc.12670>
- Turnbull, A., Seitz, A., Tadin, D., & Lin, F. V. (2022). Unifying framework for cognitive training interventions in brain aging [Publisher: Elsevier]. *Ageing Research Reviews*, 101724.
- Turner, G. R., & Spreng, R. N. (2012). Executive functions and neurocognitive aging: Dissociable patterns of brain activity. *Neurobiology of Aging*, 33(4), 826.e1–13. <https://doi.org/10.1016/j.neurobiolaging.2011.06.005>
- Twarog, J. P., Politis, M. D., Woods, E. L., Boles, M. K., & Daniel, L. M. (2015). Daily television viewing time and associated risk of obesity among U.S. preschool aged children: An analysis of NHANES 2009-2012. *Obesity Research & Clinical Practice*, 9(6), 636–638. <https://doi.org/10.1016/j.orcp.2015.09.004>
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498–505. <https://doi.org/10.3758/BF03192720>
- Valenzuela, M., & Sachdev, P. (2009). Can cognitive exercise prevent the onset of dementia? Systematic review of randomized clinical trials with longitudinal follow-up. *The American Journal of Geriatric Psychiatry: Official Journal of the American*

- Association for Geriatric Psychiatry*, 17(3), 179–187. <https://doi.org/10.1097/JGP.0b013e3181953b57>
- Vallat, R. (2018). Pingouin: Statistics in python. *J. Open Source Softw.*, 3(31), 1026.
- van der Donk, M. L., van Viersen, S., Hiemstra-Beernink, A.-C., Tjeenk-Kalff, A. C., van der Leij, A., & Lindauer, R. J. (2017). Individual Differences in Training Gains and Transfer Measures: An Investigation of Training Curves in Children with Attention-Deficit/Hyperactivity Disorder. *Applied Cognitive Psychology*, 31(3), 302–314. <https://doi.org/10.1002/acp.3327>
- van der Velde, M., Sense, F., Borst, J., & van Rijn, H. (2021). Alleviating the Cold Start Problem in Adaptive Learning using Data-Driven Difficulty Estimates. *Computational Brain & Behavior*, 4(2), 231–249. <https://doi.org/10.1007/s42113-021-00101-6>
- van de Ven, R. M., Murre, J. M. J., Veltman, D. J., & Schmand, B. A. (2016). Computer-Based Cognitive Training for Executive Functions after Stroke: A Systematic Review. *Frontiers in Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00150>
- Vandewaetere, M., Desmet, P., & Clarebout, G. (2011a). The contribution of learner characteristics in the development of computer-based adaptive learning environments [Publisher: Elsevier]. *Computers in Human Behavior*, 27(1), 118–130.
- Vandewaetere, M., Desmet, P., & Clarebout, G. (2011b). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27(1), 118–130. <https://doi.org/10.1016/j.chb.2010.07.038>
- Vandewaetere, M., Desmet, P., & Clarebout, G. (2011c). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27, 118–130. <https://doi.org/10.1016/j.chb.2010.07.038>
- VanLEHN, K. U. R. T. (2011a). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- VanLEHN, K. U. R. T. (2011b). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- vanMarle, K., & Scholl, B. J. (2003). Attentive Tracking of Objects Versus Substances [Publisher: SAGE Publications Inc]. *Psychological Science*, 14(5), 498–504. <https://doi.org/10.1111/1467-9280.03451>
- Vartanian, O., Coady, L., & Blackler, K. (2016). 3D Multiple Object Tracking Boosts Working Memory Span: Implications for Cognitive Training in Military Populations [Publisher: Routledge _eprint: <https://doi.org/10.1037/mil0000125>]. *Military Psychology*, 28(5), 353–360. <https://doi.org/10.1037/mil0000125>
- Vassoyan, J., Vie, J.-J., & Lemberger, P. (2023, May 10). *Towards Scalable Adaptive Learning with Graph Neural Networks and Reinforcement Learning*. arXiv: 2305.06398 [cs]. <https://doi.org/10.48550/arXiv.2305.06398>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023, August 1). *Attention Is All You Need*. arXiv: 1706.03762 [cs]. <https://doi.org/10.48550/arXiv.1706.03762>
- Vater, C., Gray, R., & Holcombe, A. O. (2021a). A critical systematic review of the Neurotracker perceptual-cognitive training tool [Publisher: Springer]. *Psychonomic bulletin & review*, 1–26.

- Vater, C., Gray, R., & Holcombe, A. O. (2021b). A critical systematic review of the Neurotracker perceptual-cognitive training tool [Publisher: Springer]. *Psychonomic bulletin & review*, 1–26.
- Vater, C., Gray, R., & Holcombe, A. O. (2021c). A critical systematic review of the neurotracker perceptual-cognitive training tool. *28*(5), 1458–1483.
- Vater, C., Gray, R., & Holcombe, A. O. (2021d). A critical systematic review of the Neurotracker perceptual-cognitive training tool. *Psychonomic Bulletin & Review*, *28*(5), 1458–1483. <https://doi.org/10.3758/s13423-021-01892-2>
- Vermeent, S., Dotsch, R., Schmand, B., Klamming, L., Miller, J. B., & van Elswijk, G. (2020). Evidence of validity for a newly developed digital cognitive test battery. *Frontiers in psychology*, *11*, 770.
- Vermeir, J. F., White, M. J., Johnson, D., Crombez, G., & Ryckeghem, D. M. L. V. (2020). The Effects of Gamification on Computerized Cognitive Training: Systematic Review and Meta-Analysis [Company: JMIR Serious Games Distributor: JMIR Serious Games Institution: JMIR Serious Games Label: JMIR Serious Games Publisher: JMIR Publications Inc., Toronto, Canada]. *JMIR Serious Games*, *8*(3), e18644. <https://doi.org/10.2196/18644>
- Verniani, A., Galvin, E., Tredinnick, S., Putman, E., Vance, E. A., Clark, T. K., & Anderson, A. P. (2024). Features of adaptive training algorithms for improved complex skill acquisition. *Frontiers in Virtual Reality*, *5*. Retrieved March 4, 2024, from <https://www.frontiersin.org/articles/10.3389/frvir.2024.1322656>
- Viswanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking [Place: United Kingdom Publisher: Pion]. *Perception*, *31*(12), 1415–1437. <https://doi.org/10.1068/p3432>
- Vladisaukas, M., Belloli, L. M. L., Fernández Slezak, D., & Goldin, A. P. (2022). A Machine Learning Approach to Personalize Computerized Cognitive Training Interventions. *Frontiers in Artificial Intelligence*, *5*, 788605. <https://doi.org/10.3389/frai.2022.788605>
- Võ, M. L.-H., Bylinskii, Z., & Oliva, A. (2017). Image memorability in the eye of the beholder: Tracking the decay of visual scene representations. *bioRxiv*, 141044.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92–114. <https://doi.org/10.1037/0096-1523.27.1.92>
- Von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, *41*(4), 611–624. <https://doi.org/10.3758/s13421-012-0280-7>
- Von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity [Publisher: Elsevier]. *Journal of Memory and Language*, *69*(1), 36–58. Retrieved April 18, 2024, from <https://www.sciencedirect.com/science/article/pii/S0749596X13000120>
- von Bastian, C. C., Belleville, S., Udale, R. C., Reinhartz, A., Essounni, M., & Strobach, T. (2022a). Mechanisms underlying training-induced cognitive change. *Nature Reviews Psychology*, *1*(1), 30–41. <https://doi.org/10.1038/s44159-021-00001-3>
- von Bastian, C. C., Belleville, S., Udale, R. C., Reinhartz, A., Essounni, M., & Strobach, T. (2022b). Mechanisms underlying training-induced cognitive change [Number: 1 Publisher: Nature Publishing Group]. *Nature Reviews Psychology*, *1*(1), 30–41. <https://doi.org/10.1038/s44159-021-00001-3>

- von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological Research*, 80(2), 181–194. <https://doi.org/10.1007/s00426-015-0655-z>
- von Bastian, C. C., Langer, N., Jäncke, L., & Oberauer, K. (2013). Effects of working memory training in young and old adults. *Memory & Cognition*, 41(4), 611–624. <https://doi.org/10.3758/s13421-012-0280-7>
- von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A review. *Psychological Research*, 78(6), 803–820. <https://doi.org/10.1007/s00426-013-0524-6>
- Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes ‘expert’ in the cognitive laboratory? A meta-analytic review of cognition and sport expertise [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/acp.1588>]. *Applied Cognitive Psychology*, 24(6), 812–826. <https://doi.org/10.1002/acp.1588>
- Vul, E., Alvarez, G., Tenenbaum, J., & Black, M. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in Neural Information Processing Systems*, 22. Retrieved July 14, 2024, from https://proceedings.neurips.cc/paper_files/paper/2009/hash/d79aac075930c83c2f1e369a511148fe-Abstract.html
- Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. A. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. *Advances in neural information processing systems*.
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes* (M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman, **typeredactors**). Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Vygotsky, L. S., & Cole, M. (1978a). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.
- Vygotsky, L. S., & Cole, M. (1978b). *Mind in society: Development of higher psychological processes*. Harvard university press.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of values. *Psychonomic Bulletin & Review*, 14(5), 779–804. <https://doi.org/10.3758/BF03194105>
- Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology*, 60(3), 158–189. <https://doi.org/10.1016/j.cogpsych.2009.12.001>
- Wang, R. E., Wirawarn, P., Goodman, N., & Demszky, D. (2023, June 15). *SIGHT: A Large Annotated Dataset on Student Insights Gathered from Higher Education Transcripts*. arXiv: 2306.09343 [cs]. <https://doi.org/10.48550/arXiv.2306.09343>
- Wang, T., Vargas-Díaz, D., Brown, C., & Chen, Y. (2023). Exploring the Role of AI Assistants in Computer Science Education: Methods, Implications, and Instructor Perspectives. *2023 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 92–102. <https://doi.org/10.1109/VL-HCC57772.2023.00018>
- Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, 17(3), 10. <https://doi.org/10.1167/17.3.10>
- Watters, A. (2023, February 7). *Teaching Machines: The History of Personalized Learning*. MIT Press.
- Webb, S. L., Loh, V., Lampit, A., Bateman, J. E., & Birney, D. P. (2018a). Meta-analysis of the effects of computerized cognitive training on executive functions: A cross-disciplinary taxonomy for classifying outcome cognitive factors [Publisher: Springer]. *Neuropsychology review*, 28, 232–250.

- Webb, S. L., Loh, V., Lampit, A., Bateman, J. E., & Birney, D. P. (2018b). Meta-Analysis of the Effects of Computerized Cognitive Training on Executive Functions: A Cross-Disciplinary Taxonomy for Classifying Outcome Cognitive Factors. *Neuropsychology Review*, 28(2), 232–250. <https://doi.org/10.1007/s11065-018-9374-8>
- Webb, S. L., Loh, V., Lampit, A., Bateman, J. E., & Birney, D. P. (2018c). Meta-Analysis of the Effects of Computerized Cognitive Training on Executive Functions: A Cross-Disciplinary Taxonomy for Classifying Outcome Cognitive Factors. *Neuropsychology Review*, 28(2), 232–250. <https://doi.org/10.1007/s11065-018-9374-8>
- Webb, S. L., Loh, V., Lampit, A., Bateman, J. E., & Birney, D. P. (2018d). Meta-Analysis of the Effects of Computerized Cognitive Training on Executive Functions: A Cross-Disciplinary Taxonomy for Classifying Outcome Cognitive Factors. *Neuropsychology Review*, 28(2), 232–250. <https://doi.org/10.1007/s11065-018-9374-8>
- West, R. L., Bagwell *, D. K., & Dark-Freudeman, A. (2008). Self-Efficacy and Memory Aging: The Impact of a Memory Intervention Based on Self-Efficacy. *Aging, Neuropsychology, and Cognition*, 15(3), 302–329. <https://doi.org/10.1080/13825580701440510>
- Whipple, G. M. (1910). The effect of practice upon the range of visual attention and of visual apprehension. *Journal of Educational Psychology*, 1(5), 249–262. <https://doi.org/10.1037/h0075300>
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66(5), 297–333. <https://doi.org/10.1037/h0040934>
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15(4), 160–168. <https://doi.org/10.1016/j.tics.2011.02.005>
- WHO. (2014). *Active ageing : A policy framework*. Retrieved July 19, 2024, from <https://iris.who.int/handle/10665/67215>
- Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, 3, 159–177. <https://doi.org/10.1080/14639220210123806>
- Wierzbicki, M., Rupaszewski, K., & Styrkowiec, P. (2023). Comparing highly trained handball players' and non-athletes' performance in a multi-object tracking task. *The Journal of General Psychology*, 0(0), 1–13. <https://doi.org/10.1080/00221309.2023.2241950>
- Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., Morris, J. N., Rebok, G. W., Unverzagt, F. W., Stoddard, A. M., Wright, E., & ACTIVE Study Group, f. t. (2006). Long-term Effects of Cognitive Training on Everyday Functional Outcomes in Older Adults. *JAMA*, 296(23), 2805–2814. <https://doi.org/10.1001/jama.296.23.2805>
- Wilms, I. (2011). Using artificial intelligence to control and adapt level of difficulty in computer based, cognitive therapy—an explorative study. *J. Cyberther. Rehabil*, 4(3), 387–396.
- Wilson, R. C., Shenhav, A., Straccia, M., & Cohen, J. D. (2019). The Eighty Five Percent Rule for optimal learning. *Nature Communications*, 10(1), 4646. <https://doi.org/10.1038/s41467-019-12552-4>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. *Educational and psychological measurement*, 76(6), 913–934. <https://doi.org/10.1177/0013164413495237>
- Wollesen, B., Janssen, T. I., Müller, H., & Voelcker-Rehage, C. (2022). Effects of cognitive-motor dual task training on cognitive and physical performance in healthy children

- and adolescents: A scoping review. *Acta Psychologica*, 224, 103498. <https://doi.org/10.1016/j.actpsy.2022.103498>
- Woodworth, R. S., & Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological Review*, 8(3), 247–261. <https://doi.org/10.1037/h0074898>
- Woutersen, K., Guadron, L., van den Berg, A. V., Boonstra, F. N., Theelen, T., & Goossens, J. (2017). A meta-analysis of perceptual and cognitive functions involved in useful-field-of-view test performance. *Journal of Vision*, 17(14), 11. <https://doi.org/10.1167/17.14.11>
- Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P.-Y. (2023). Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. *28th International Conference on Intelligent User Interfaces*, 75–78. <https://doi.org/10.1145/3581754.3584136>
- Xu, Q., Subbaraju, V., Cheong, C. H., Wang, A., Kang, K., Bashir, M., Dong, Y., Li, L., & Lim, J.-H. (2018). Personalized serious games for cognitive intervention with lifelog visual analytics. *Proceedings of the 26th ACM international conference on Multimedia*, 328–336.
- Yantis, S. (1992a). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24(3), 295–340. [https://doi.org/10.1016/0010-0285\(92\)90010-y](https://doi.org/10.1016/0010-0285(92)90010-y)
- Yantis, S. (1992b). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology*, 24(3), 295–340. [https://doi.org/10.1016/0010-0285\(92\)90010-Y](https://doi.org/10.1016/0010-0285(92)90010-Y)
- Yazdani, M., & Lawler, R. W. (1986). Artificial intelligence and education: An overview. *Instructional Science*, 14(3-4), 197–206. <https://doi.org/10.1007/BF00051820>
- Yung, A., Cardoso-Leite, P., Dale, G., Bavelier, D., & Green, S. (2015a). Methods to Test Visual Attention Online. *Journal of Visualized Experiments*, in press. <https://doi.org/10.3791/52470>
- Yung, A., Cardoso-Leite, P., Dale, G., Bavelier, D., & Green, S. (2015b). Methods to Test Visual Attention Online. *Journal of Visualized Experiments*, in press. <https://doi.org/10.3791/52470>
- Zebda, E., Manca, M., & Paternò, F. (2022). Document - Towards adaptation of humanoid robot behaviour in serious game scenarios using reinforcement learning. Retrieved June 30, 2023, from https://openportal.isti.cnr.it/doc?id=people_____::361e469effa025f8b2481ded331245b0
- Zentgraf, K., Heppe, H., & Fleddermann, M.-T. (2017a). Training in interactive sports. *German Journal of Exercise and Sport Research*, 47(1), 2–14. <https://doi.org/10.1007/s12662-017-0441-8>
- Zentgraf, K., Heppe, H., & Fleddermann, M.-T. (2017b). Training in interactive sports. *German Journal of Exercise and Sport Research*, 47(1), 2–14. <https://doi.org/10.1007/s12662-017-0441-8>
- Zhang, H., Huntley, J., Bhome, R., Holmes, B., Cahill, J., Gould, R. L., Wang, H., Yu, X., & Howard, R. (2019). Effect of computerised cognitive training on cognitive outcomes in mild cognitive impairment: A systematic review and meta-analysis [Publisher: British Medical Journal Publishing Group Section: Mental health]. *BMJ Open*, 9(8), e027062. <https://doi.org/10.1136/bmjopen-2018-027062>
- Zhang, R.-Y., Chopin, A., Shibata, K., Lu, Z.-L., Jaeggi, S. M., Buschkuhl, M., Green, C. S., & Bavelier, D. (2021a). Action video game play facilitates "learning to learn". *Communications Biology*, 4(1), 1154. <https://doi.org/10.1038/s42003-021-02652-7>

- Zhang, R.-Y., Chopin, A., Shibata, K., Lu, Z.-L., Jaeggi, S. M., Buschkuhl, M., Green, C. S., & Bavelier, D. (2021b). Action video game play facilitates “learning to learn”. *Communications Biology*, 4(1), 1–10. <https://doi.org/10.1038/s42003-021-02652-7>
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., ... Wen, J.-R. (2023, November 24). *A Survey of Large Language Models*. arXiv: 2303.18223 [cs]. <https://doi.org/10.48550/arXiv.2303.18223>
- Zhong, S.-h., Ma, Z., Wilson, C., Liu, Y., & Flombaum, J. I. (2014). Why do people appear not to extrapolate trajectories during multiple object tracking? a computational investigation. *Journal of Vision*, 14(12), 12–12.
- Zhou, K. Z., Kilhoffer, Z., Sanfilippo, M. R., Underwood, T., Gumusel, E., Wei, M., Choudhry, A., & Xiong, J. (2024, January 22). “The teachers are confused as well”: *A Multiple-Stakeholder Ethics Discussion on Large Language Models in Computing Education*. arXiv: 2401.12453 [cs]. <https://doi.org/10.48550/arXiv.2401.12453>
- Zini, F., Le Piane, F., & Gaspari, M. (2022). Adaptive Cognitive Training with Reinforcement Learning [Publisher: ACM New York, NY]. *ACM Transactions on Interactive Intelligent Systems (TiIS)*, 12(1), 1–29.
- Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*, 50(1), 304–315. <https://doi.org/10.1037/a0032982>
- Zomeran, A. H., & Brouwer, W. H. (1994). *Clinical Neuropsychology of Attention*. Oxford University Press.
- Zorzi, M., Barbiero, C., Facoetti, A., Lonciari, I., Carrozzi, M., Montico, M., Bravar, L., George, F., Pech-Georgel, C., & Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, 109(28), 11455–11459. <https://doi.org/10.1073/pnas.1205566109>