PhD-FHSE-2023-009
The Faculty of Humanities, Education and Social Sciences

DISSERTATION

Defence held on 30/01/2023 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN PSYCHOLOGIE

by

Morteza ANSARINIA
Born on 10 May 1985 in Sary (Iran)

Towards a Computational Model of General Cognitive Control Using Artificial Intelligence, Experimental Psychology and Cognitive Neuroscience

Dissertation defence committee

Dr Pedro Cardoso-Leite, dissertation supervisor
Professor, Université du Luxembourg

Dr Christine Schiltz, Chairman
Professor, Université du Luxembourg

Dr Jöran Lepsein
Scientific Researcher, Max Planck Institute for Human Cognitive and Brain Sciences

Dr Paul Schrater
Professor, University of Minnesota

Dr Constantin A. Rothkopf
Professor, Technischen Universität Darmstadt

Abstract

Cognitive control is essential to human cognitive functioning as it allows us to adapt and respond to a wide range of situations and environments. The possibility to enhance cognitive control in a way that transfers to real life situations could greatly benefit individuals and society. However, the lack of a formal, quantitative definition of cognitive control has limited progress in developing effective cognitive control training programs. To address this issue, the first part of the thesis focuses on gaining clarity on what cognitive control is and how to measure it. This is accomplished through a large-scale text analysis that integrates cognitive control tasks and related constructs into a cohesive knowledge graph. This knowledge graph provides a more quantitative definition of cognitive control based on previous research, which can be used to guide future research. The second part of the thesis aims at furthering a computational understanding of cognitive control, in particular to study what features of the task (i.e., the environment) and what features of the cognitive system (i.e., the agent) determine cognitive control, its functioning, and generalization. The thesis first presents CogEnv, a virtual cognitive assessment environment where artificial agents (e.g., reinforcement learning agents) can be directly compared to humans in a variety of cognitive tests. It then presents CogPonder, a novel computational method for general cognitive control that is relevant for research on both humans and artificial agents. The proposed framework is a flexible, differentiable end-to-end deep learning model that separates the act of control from the controlled act, and can be trained to perform the same cognitive tests that are used in cognitive psychology to assess humans. Together, the proposed cognitive environment and agent architecture offer unique new opportunities to enable and accelerate the study of human and artificial agents in an interoperable framework.

Research on training cognition with complex tasks, such as video games, may benefit from and contribute to the broad view of cognitive control. The final part of the thesis presents a profile of cognitive control and its generalization based on cognitive training studies, in particular how it may be improved by using action video game training. More specifically,

we contrasted the brain connectivity profiles of people that are either habitual action video game players or do not play video games at all. We focused in particular on brain networks that have been associated with cognitive control. Our results show that cognitive control emerges from a distributed set of brain networks rather than individual specialized brain networks, supporting the view that action video gaming may have a broad, general impact of cognitive control. These results also have practical value for cognitive scientists studying cognitive control, as they imply that action video game training may offer new ways to test cognitive control theories in a causal way.

Taken together, the current work explores a variety of approaches from within cognitive science disciplines to contribute in novel ways to the fascinating and long tradition of research on cognitive control. In the age of ubiquitous computing and large datasets, bridging the gap between behavior, brain, and computation has the potential to fundamentally transform our understanding of the human mind and inspire the development of intelligent artificial agents.

Acknowledgements

First, I would like to thank my advisor, Pedro Cardoso-Leite, for his generosity and kindness, and for creating an environment where I felt safe, valued, and supported to pursue my research interests. Thank you for your tireless skepticism and tough questioning that have improved me in countless ways. I am very grateful for your patience and understating, and very fortunate to have had the opportunity to work with you.

I would like to thank my co-advisor, Jöran Lepsein, and committee members, Florian Waszak, and Christine Schiltz for the comments, discussions, and encouragement in the annual meetings. Paul Schrater in particular has been always helpful in the early stages of my research, and again in the final stages of it. I want to thank him for his scientific wisdom and thoughtful guidance. I would also like to thank Daphne Bavelier and Julia Föcker for the opportunity to learn from them and their generosity in providing valuable data for my research. I am very fortunate to have been inspired by your work and to have received your feedback on my own.

To have been granted the opportunity to pursue PhD studies has indeed been an immense privilege. For that, I would like to thank University of Luxembourg and Max Planck Institute for Human Cognitive and Brain Sciences. I am also grateful for the financial support from Luxembourg National Research Fund. The studies in this thesis were supported by the Luxembourg National Research Fund through the project DIGILEARN. Luxembourg is a wonderful place to live and work, and I am very grateful for the opportunity to have been able to live and work here.

I would also like to thank my friends and family for their love and support, and for being by my side the whole way. I want to extend a special thanks to Saeed Gholami Shahbandi and Shayan Eslami for their genuine and unwavering friendship. Thank you to my parents, my sister, and my bother for their support in everything I have pursued. Thanks also to my friends for contributing to an enjoyable life outside of research, especially Sherwin Ahmadi Fouladi, Mani Monajjemi, Safa Jamali, Soheil Asmari, and Hamid Gholami Shahbandi (and not to forget Marmeladov, Taguchi, and Twigi). And for our many impromptu conversations about cognition and computing, I thank Ali Farjami and Mostafa Salari Rad. As well, I would like to thank members of the xCIT Lab for all the discussions and contributions: Dominic Mussack, Brice Clocher, Aurélien Defossez, Emmanuel Schmück, and Kamelia Jamaati.

Then, I want to thank my partner in science and life (and now wife), Yeganeh Farhzadi, for always believing in me and providing me with love and support, for building and protecting an environment where even silly and awkward ideas are treasured, and for being my best friend. Since 2015, we went through the ups and downs of life together, and I look forward to the rest of it. Thank you Yeganeh, very much. I love you.

As a final note, I would like to thank all the people who have inspired me through their work and lives. For that, it would be remiss of me not to acknowledge all the Iranian people; my many thanks go out to ”woman, life, freedom”.

Dedication

To my beloved Yeganeh

Table of contents

General Introduction **1**

Exploring cognitive control across cognitive sciences disciplines 2

Cognitive Psychology and Neuroscience 3

Artificial intelligence 5

The value and challenges of interdisciplinary research 7

Current research 9

Information sheets 10

1 Linking Theories and Methods in Cognitive Sciences via Joint Embedding of the Scientific Literature: The Example of Cognitive Control **18**

Abstract 18

1.1 Introduction 19

1.2 Methods 21

1.3 Results 24

1.4 Implications 29

2 CogEnv: A Virtual Environment for Contrasting Human and Artificial Agents across Cognitive Tests **33**

Abstract 33

2.1 Introduction 34

2.2 Technical specification 34

2.2.1 Tasks 35
2.2.2 Timing 36
2.2.3 Action space 36
2.2.4 Observation space 36
2.3 Comparing humans and artificial agents 36
2.4 Conclusion 37

3 CogPonder: Towards a Computational Framework of General Cognitive Control **39**
Abstract 39
3.1 Introduction 40
3.2 Desiderata for a general computational cognitive control framework 43
3.2.1 PonderNet 44
3.2.2 TOTE 45
3.3 The CogPonder framework 45
3.4 Evaluation of a CogPonder model 48
3.4.1 Objectives and rationale 48
3.4.2 Dataset 49
3.4.3 Method 50
3.4.4 Model evaluation procedure 52
3.5 Results 54
3.6 Discussion 58
3.7 Implications 60
3.8 Limitations and future extensions 62

4 Training Cognition with Video Games **65**
Abstract 65
4.1 Introduction 66
4.2 Which video games improve cognition? 69

4.3 First and third person shooters (“action” video games) 74
4.4 Racing games 83
4.5 Real-time strategy games 84
4.6 Tetris 86
4.7 Casual mobile games 87
4.8 The neuroscience of video game play 89
4.8.1 Reward system 90
4.8.2 Spatial cognition and the hippocampal formation 92
4.8.3 Attentional networks and action video games 95
4.9 Concluding remarks 100
4.10 Future perspectives 102

5 Neural Correlates of Habitual Action Video Games Playing in Control-related Brain Networks 103

Abstract 103
5.1 Introduction 104
5.2 A graph-theoretic approach to cognitive control in cognitive neuroscience . . 106
5.2.1 The brain is intrinsically organized into networks. 106
5.2.2 The cognitive control brain networks 107
5.3 Measuring intrinsic networks can be studied during resting state. 110
5.4 Hypotheses 111
5.5 Data 113
5.6 Methods 114
5.6.1 Formal problem statement 114
5.6.2 Preprocessing 116
5.6.3 Data analysis pipeline 116
5.6.4 Evaluation of the classifier 119
5.7 Results 124

5.7.1 Participants can be accurately classified as AVGPs versus NVGP based on their resting state functional connectivities. 124
5.7.2 Resting-state functional connectivity differences between AVGPs and NVGPs are not circumscribed to a specialized brain network: they involve multiple networks and interplay between them. 125
5.7.3 Key results are robust to changes in the data analysis pipelines. . . . 128
5.8 Discussion 128
5.9 Limitations and future research 134
5.10 Conclusion 135
5.11 Supplementary Materials 136
5.11.1 Parcellations 136
5.11.2 Motion signals during resting state fMRI recording do not differentiate AVGPs from NVGPs 138
5.11.3 Classifying habitual AVGP using intrinsic functional connectivities depends on the parcellation technique as well as the connectivity metric 139
5.11.4 SHAP Analysis 141

General Discussion **146**
Defining cognitive control 147
Training and generalizing cognitive control 150
Future perspectives 151
Conclusion 153

References **155**

Appendices **186**

A A Formal Framework for Structured N-Back Stimuli Sequences **187**
Abstract 187
A.1 Introduction 188

A.1.1 Parameterizing the N-Back sequences . . . 189
A.1.2 Structured sequences . . . 190
A.2 Evaluating behavioral impacts of structural features . . . 193
A.2.1 Data . . . 193
A.2.2 Analysis . . . 193
A.3 Results . . . 194
A.4 Conclusion . . . 196

B Behaverse data model **198**
B.1 Introduction . . . 198
B.2 Challenges of behavioral data . . . 199
B.3 Data consistency levels . . . 202
B.4 Inconsistent data formats . . . 204
B.4.1 Unknown or inconsistent data level . . . 204
B.4.2 Inconsistent variable naming conventions . . . 204
B.4.3 Unknown values and units . . . 206
B.4.4 Conclusion . . . 206
B.5 Behavioral experiments require multiple types of data . . . 206
B.6 Behavioral, interaction data . . . 208
B.6.1 Source data, raw data and derived data . . . 210
B.6.2 Event data and trial data . . . 210
B.6.3 Key concepts for specifying trial data . . . 212
B.6.4 L1 data model . . . 217
B.7 Discussion . . . 221
B.8 Conclusion . . . 222

List of Figures

1.1 (Panel a) Introducing new tasks (task innovation) and constructs (concept innovation) is characterized by a burst followed by declining innovation. (Panel b) Task and Construct occurrences in publication abstracts are temporally decoupled. Time to operationalize constructs (blue) is the time between the first occurrence of a construct and the first co-occurrence of that construct with any tasks, while Time to conceptualize tasks (orange) is the time between the first occurrence of a task and the first co-occurrence of that task with any of the construct. (Panel c) The majority of the literature only used one task in their studies, showing a lack of multitask design of experiments. (Panel d) While the number of papers published each year increases exponentially, the number of tasks per study remains fairly constant across time. 25

1.2 Task-Construct hypergraph: representations of control-related constructs as hyperedges (vertical black lines) over a subset of tasks (nodes). Construct hypernomy is reflected as overlapping hyperedges (e.g., green regions), and task impurity as nodes scattered over multiple hyperedges (e.g., blue region). Distances between nodes are not meaningful. Nodes are reorganized for visual clarity and only a subset of the graph is displayed. 26

1.3 Associations between tasks and constructs minimally overlap across scientific disciplines. Rose plots show the relative association between constructs and tasks, with each color representing a different field. Lack of overlap between the "spikes" indicates disjoint operationalizations across fields. 29

1.4 Pairwise distances between the 25 most popular cognitive control tasks as measured by the symmetric Jensen-Shannon divergence of two multivariate normal distributions of their node attributes in the task-construct graph. Higher divergence indicates higher dissimilarity between corresponding scientific texts. Task-task distances may for example provide a data-driven proxy for predicting and explaining transfer effects in cognitive training research. 30

2.1 Overall architecture of CogEnv. CogEnv communicates with AndroidEnv via Protocol Buffer messages and manages access to the Behaverse events. O_t is the screenshot of the task at time t, O'_t is the extra observations extracted from the Behaverse events including information about the task and stimuli, r_t is the reward, and A_t is the agent's action. 35

2.2 Screenshot (O_t) of four Behaverse tasks. A) Digit Span (working memory), B) N-Back (working memory), C) Trail Making Test (cognitive flexibility), and D) Belval Matrices (matrix reasoning). See behaverse.org. 36

2.3 Hypothetical scenarios when comparing the performance of humans versus computational agents (see text). 38

3.1 The CogPonder framework. (A) An end-to-end model, termed "Operator", which on a given trial n takes an input X_n and outputs y_n. (B) CogPonder disconnects the Operator from its direct inputs and outputs and encapsulates the Operator inside a local virtual environment that is governed by the Controller (blue box). The Controller intercept both the inputs and outputs of the Operator, and determines what inputs are fed to the Operator and ultimately what output to be emitted on a given trial. Within a given trial n the Controller will repeatedly call the Operator, with each of these iterations being indexed by step s, until it decides to halt processing for trial n and to emit a response y_n. The halting is determined by a sample from a Bernoulli distribution parameterized by λ_s (decision diamond in the figure). 46
3.2 CogPonder learns to behave like humans. With increasing learning iteration (epochs) the loss decreases and asymptotes. This is true both when aligning CogPonder with the Stroop task (red curve) or with the N-back task (blue curve). Note that the two tasks were trained and tested separately. 55
3.3 CogPonder behavior is comparable to human behavior. CogPonder captures the overall pattern of average accuracy (left column of panels) and average response times (right column of panels) in both the Stroop task (upper row of panels) and in the N-back task (bottom row of panels) when grouping all types of trials ("All"). However, when separating trials by type ("congruent" and "incongruent" in the the Stroop task and "target" and "non-target" in the N-back task), some discrepancies are observed. Error bars show 95% confidence intervals. 56
3.4 CogPonder also mimics finer grained phenomena (e.g., response time distributions). 57

4.1 Intervention design to evaluate the causal impact of playing a specific type of video games on cognition (here termed experimental game). Participants are randomly assigned to play experimental video games or control video games. The training program typically requires at least 8 hours, and typically tens of hours of gameplay, distributed over weeks or months. Participants' cognitive skills are first evaluated on a battery of tests (pre-test) and tested again after completion of their training (post-test). If playing the experimental video games specifically improves the cognitive abilities assessed, then we expect the experimental group to improve more from pre- to post-test than the control group. 70

4.2 List of commercial video games used in cognitive training studies from Sala et al. (2018). This list contains a wide range of video game genres that have been used for training in the scientific literature (e.g., first person shooters, racing games, puzzle games, real-time strategy games, sports games) as well as non-video games (Space Fortress). Large differences in experiences between different game genres (a fast-paced multiplayer FPS is nothing like a slow paced, single player puzzle game) render the interpretation of any such results (positive, negative or null impact on cognition) quite difficult. This figure counts the number of publications cited in Sala et al. (2018) that used a particular video game (out of a total of 63 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games. 75

4.3 List of activities used as control treatment in video-game based training studies from Sala et al. (2018). Control treatments vary widely from playing video games to playing paper-and-pencil games; this makes it difficult to abstract the construct measured by such studies. This figure counts the number of publications cited in Sala et al. (2018) that used a particular video game or activity (out of a total of 63 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games. 76

4.4 List of action video games (all first person shooter games; FPS) used for cognitive training according to Bediou et al. (2018). Focusing on this specific video game genre substantially reduces the number of games titles but still represents a major portion of the scientific literature (contrast this with Figure 4.2). This figure counts the number of publications cited in Bediou et al. (2018) that used a particular video game (out of a total of 23 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games. 79

4.5 List of games used in the active control treatment when action video games were tested for cognitive training as tabulated by Bediou et al. (2018). This list includes only commercial video games (with the exception of the Sight Training program; contrast this with Figure 4.3). This figure counts the number of publications cited in Bediou et al. (2018) that used a particular video game (out of a total of 23 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games. . 80

5.1 Data analysis pipeline. All data were first preprocessed using a standard procedure (step 1). The same steps were applied irrespective of the AVGP/NVGP label of participants. This preprocessed data then served as input to the next steps which aimed to 2) train and 3) diagnose a AVGP versus NVGP classifier (see text for details). 117

5.2 AVGPs vs NVGPs classification accuracy as a function of parcellation and connectivity metric. The distribution of cross-validated out-of-sample prediction accuracies are displayed in orange for the actual data and in gray for a shuffled version of the data (to form an empirical null distribution; see text for details). Dots and diamonds represent the mean of the distribution; error bars represent the 95% confidence intervals. This figure shows that new participants can be accurately classified as AVGPs vs NVGPs based on their resting state functional brain connectivity with the best model reaching an accuracy of 72.6%. Classification accuracy varies however considerably with the specific parcellation and connectivity metric used. The black triangle on the X-axis shows the prediction accuracy using motion confounds; the observed accuracy (51%) was not significantly different from chance (see supplementary materials for details). 126

5.3 Permutation features importance of the top 6 AVGPs versus NVGPs classification models ordered by classification accuracy (see Figure 5.2). Each panel shows the 12 most important features (ordered by importance) for a given classifier, which is characterized by an atlas (i.e., Dosenbach2010 versus Gordon2014) and a connectivity metric (e.g., partial correlation, precision). Error bars represent 95% confidence intervals. 129

5.4 The effect of skipping skull stripping. It was necessary to skip the skull stripping step of the preprocessed T1w images of MRIQC because the scans were already defaced. The left panel in this figure shows a scan with skull stripping and the right panel, without skull stripping. As can be seen in this figure, by skipping skull stripping the recognition of the brain volumes became more accurate. 136

5.5 Dosenbach2010 networks. 137

5.6 Gordon2014 networks. 137

5.7 DiFuMo64 networks. 138

5.8 Bayesian model fitted to the choice of atlas (P), choice of connectivity metric (C), and prediction accuracy (y); See Formula Supp-1. We used full-rank coding of categorical variables (P and C), with C=`correlation` and P=`DiFuMo64` being the baseline references. 140

5.9 Comparing the choice of atlas and connectivity metric on classification performance. Error bars represent 2 standard deviations. We used full-rank coding of categorical variables with baseline reference being correlation for connectivity metrics (C=`correlation`) and DiFuMo for parcellation atlases (P=`DiFuMo64`). Intercept and baseline references are not shown. 142

5.10 Shap values for correct (green) and incorrect (red) classifications of participants as AVGPs or NVGPs. The plot reads from top to bottom, showing the impact of each connectivity to the model output (i.e., AVGP vs NVGP classification probabilities). Network features are ordered, from top to bottom, by their average importance (`mean(|SHAP|)`). 145

A.1 Classification performance for the base and extended models. AUC = Area Under the Curve 195

A.2 Relative importance of structural variables (V) on the prediction of participants' response accuracy. 196

B.1 From data collection to analysis. 1) Subjects interact with digital artifacts and produce data. 2) The resulting data ("source data") is typically stored in idiosyncratic formats, possibly determined by technical constraints of the digital artifacts. Furthermore, this "source data" may contain data that is not of direct relevance to researchers (e.g., technical information about the software) and important information may come from other sources (e.g., information about the study that is present only in the corresponding research paper). 3) It is typically necessary to extract the relevant data from the source data. Here we distinguish "event" data and "trial" data. Event data describes the behavioral data as a sequence of time stamped events, which have specific types (e.g., a mouse click) and data (e.g., the screen coordinates of the click). Trial data organizes those events following a task-pattern into a tabular form, where each row describes one trial. Further data files are necessary for example to describe the study. Note that it is typical for the data collection artifacts to already embed some data processing code and keep as source data only the "trial" data. 4) The most important type of behavioral data appears to be the event data from which different trial datasets may be extracted—this is in our opinion what should be viewed as the raw data and it will be valuable in the future to standardize behavioral event data and develop effective tools to deal with such data and extract trial-based data from them. 5) We define as Level 1 data, the data tables which are organized by trial. These are the tables we believe are most useful given current practices. In particular, we define the L1-Trial table, where each row contains complete and standardized information describing a particular trial (as is already currently the case, albeit inconsistently) and where the trial identifier is used as a primary key to additional, more detailed or specific tables (e.g., a table describing each of the mouse clicks that occurred during a trial). 6) The L1 data serves as the standardized input to data processing pipelines, which will derive additional tables (e.g., L2, L3), for example by transforming and summarizing data or aggregating across subjects 209

B.2 L1 Trial data. 1) In source data, relevant information may be scattered across multiple data files in a way that is not practical for subsequent processing. There are various design options to reorganize the source data into data structures that can be standardized and are easier to use. 2) One solution is to factor the data into many compact tables within a relational database system. While this solution has many technical advantages, it doesn't play well with current practices. 3) An alternative design solution—the one we chose for the current behaverse data model— defines a main "L1 Trial" table which is similar to what researchers already use today. However, in addition to providing the trial data, the L1 dataset contains additional, related tables (as in 2). Tables in L1 are related to each other by various primary keys, the most important one being the trial identifier within the Trial table. We believe that this solution is both of practical use for researchers and offers the possibility to augment the Trial table in a principled way to capture more of the richness of behavioral data than is typically the case. 220

List of Tables

1 TL;DR – Chapter 1 (CogText) 10

2 TL;DR – Chapter 2 (CogEnv) 11

3 TL;DR – Chapter 3 (CogPonder) 12

3 TL;DR – Chapter 3 (CogPonder) 13

4 TL;DR – Chapter 4 (Review) 14

4 TL;DR – Chapter 4 (Review) 15

5 TL;DR – Chapter 5 (ACNets) 16

5 TL;DR – Chapter 5 (ACNets) 17

4.1 Main 'classical' video game categories cited in the reviewed literature. These categories are based on the Video Game Questionnaire from the Bavelier lab. We provide in the supplemental materials the current version of the video game questionnaire and the selection criteria used in the Bavelier laboratory (version September 2019). The game categories it lists are motivated by research considerations and not by industry classifications. Yet, examples of games and our labels for game categories have evolved over the years in concert with the changing landscape of video games. 71

5.1 A Bayesian model comparison analysis shows that the choice of parcellation atlas affects classification accuracy most. In general, choosing Dosenbach2010 atlas and precision connectivity metric leads to the highest classification accuracy. Results from a "y ~ P * C" model (which reads "accuracy ~ atlas * metric") are shown in the table. Note that the table shows contrasts against the baseline reference of correlation connectivity metric and DiFuMo64 atlas. 141

A.1 List of structural variables (X) 191

A.2 List of constraints on structural variables and respective violation costs . . . 192

B.1 Data Consistency Levels. It is our understanding that current standards in behavioral sciences places us within levels 0 to 1. 203

General Introduction

It is said that humans are creatures of habit. But even habits are established and managed by a higher-order cognitive system—a human capacity expressed in innumerous situations that remains unmatched by any other species or artificial intelligence. My thesis aims to further our understanding of higher-order cognition. More specifically, I'm interested in our ability to be goal-driven and which enables us to produce complex, meaningful, context-dependent behavior, in uncertain environments, inhibit prepotent responses, monitor and manage the cross-talk between conflicting tasks.

The role this ability plays in daily life is evident, for instance, when making pizza! We first need to plan a sequence of tasks, from creating a shopping list, buying the ingredients, preheating the oven while proofing the dough, pausing the preparation of the toppings because the oven is beeping, and possibly multitasking to wash the dishes while cooking. Some skilled chefs can make great pizza on a stovetop burner rather than an oven, demonstrating that their ability to make pizza can generalize and transfer from one environment to another. Tasks like making pizza are complex because they require a variety of cognitive functions, including planning, multitasking, task switching, attention, flexibility, monitoring, handling feedback, practice, and generalization, to name just a few. Yet most people can routinely perform such complex tasks.

Goal-driven higher cognition is of utmost importance to humans as it determines many aspects of our lives (e.g., academic and professional success, social relationships, health). Unfortunately, we don't yet fully understand how this type of higher-order cognition works and how to improve it for the benefit of individuals and society. There are, however, many ideas, theories, and experimental work across multiple scientific fields that we can draw from.

Here, I will apply a multidisciplinary approach to clarify what this specific higher-order cognition is and how it operates in computational, quantitative terms. There are two primary motivations for me to focus on computational/quantitative accounts. First, they may provide principled ways towards understanding and developing interventions to improve humans' goal-directed cognition; a large body of work indicates this is indeed possible but we currently lack a clear theoretical framework to understand why and how those effects come about. Second, there have been important advances in artificial intelligence in recent years and these may benefit our understanding of human cognition. Conversely, the study of human cognition may, as it has several times in the past, lead to insights that benefit new developments in artificial intelligence.

The scientific concept that best characterizes what I referred to as "goal-driven higher-order cognition" is "cognitive control", as articulated in Badre (2020) and J. D. Cohen (2017). In this context, cognitive control is an umbrella term for a set of processes that generate and monitor plans and actions in pursuit of evolving goals, often in noisy environments. Other related terms used (sometimes interchangeably) in the scientific literature include for instance "executive functions", "attentional control", "executive control", and "self-regulation". For consistency and simplicity, I will only refer to "cognitive control" in the remainder of this thesis, acknowledging, as have many before, that this is a complex and to a large extent ill-defined concept.

Exploring cognitive control across cognitive sciences disciplines

This thesis is interdisciplinary and grounded in cognitive sciences. In particular, it applies principles and techniques from cognitive psychology, neuroscience, and artificial intelligence as these are key fields in which cognitive control related questions have been extensively investigated. This interdisciplinarity offers synergies that support the systematic study of cognitive control using modern tooling, and the development of artificial agents that may benefit from human-like control abilities by aligning to human cognitive functioning (Russell, 2020).

Cognitive Psychology and Neuroscience

In cognitive psychology, concepts that capture higher-order cognitive abilities such as cognitive control are difficult to define—and consequently also to quantify. This may in part be due to cognitive control being related to many other psychological constructs (see J. D. Cohen, 2017), and to its role in explaining task-dependent, contextual phenomena (Otto et al., 2013; Appendix A; Ralph, 2014). It may also be due to the more general limitation of psychological constructs being low-dimensional representations of distributed brain mechanisms (Jolly & Chang, 2019; Zink et al., 2021). Nevertheless, to understand cognitive control, psychologists have devised a variety of theoretical constructs and cognitive tasks (see Chapter 1 and Baggetta & Alexander, 2016) the relationships between which are not always very clear. This lack of a cohesive understanding calls for conceptual and empirical clarifications about what researchers mean by cognitive control and how to quantify it. Greater clarity and an integrated framework of cognitive control is required to advance the field.

In this regard, greater clarity may come from recent machine learning advances in natural language processing which have made it possible to analyze a large body of texts in order to identify and connect underlying ideas (Angelov, 2020; Beam et al., 2021; Dieng et al., 2020). Computational techniques such as ontologies and large language models can be leveraged to parse the ever-growing research on cognitive control in order to develop a cohesive framework that provides a holistic and pragmatic view of cognitive control that shows how cognitive control is conceptualized and operationalized in the scientific literature. This type of integrative work seems critical to make sense of currently disparate research that comprises many psychological constructs and computational models, several brain mechanisms, and multiple cognitive tasks.

An integrated and formal account of cognitive control would be invaluable for programs aiming to improve cognitive control abilities in humans. Given the role of cognitive control on daily functioning, long-term achievements, and psychological health (Diamond & Ling, 2019; Moffitt et al., 2011), for example, the possibility to improve cognitive control in a way that transfers to real life could have important implications across a wide range of use cases (e.g., rehabilitation, healthy aging, education, peak performance). The study of cognitive training and its consequences is also important from a theory perspective as interventional methods (as in cognitive training regimes) offer a means to causally test computational theories of cognitive control.

Despite the ubiquity of cognitive training studies (Bediou et al., 2018), we currently lack a satisfactory theory of how training on specific tasks generalizes to new ones (Moreau & Conway, 2014; Oei & Patterson, 2014b). It's not entirely clear which interventions impact the cognitive systems and how they do so—including what neural mechanisms in the brain enable cognitive control, how they are impacted by cognitive training, and how this impact causes the behavioral outcomes.

Currently, the main theories in this context revolve around one of two types of hypotheses. The first states that cognitive training interventions train multiple elementary cognitive processes and to the extent that new tasks rely on those same processes (or a subset of them), transfer effects will be observed on those new tasks (Oei & Patterson, 2014b). An alternative class of hypotheses state that cognitive training enhances domain-general abilities which are involved in virtually all cognitive tasks—among these domain-general abilities, cognitive, and attentional control are the most prominent (Anguera et al., 2013; Green & Bavelier, 2008). Which of these (if any) are true, remains an open question and part of the difficulty in making progress is the lack of theories that would allow predictions of how certain forms of training would or would not transfer to which other tasks.

The study of action video game training is of particular interest in cognitive control research. There is now a large body of research, including many training studies, that have established that playing specifically action video games causes improvements in performance across a broad range of cognitive tasks (Bediou et al., 2018)—some of which generalize to real-life abilities (Franceschini et al., 2012)—and there is also an increasing body of research investigating the neural mechanisms involved in video game play and their effects (see Chapter 4). These constitute a fertile ground to build cognitive control theories and bridging a gap between experimental psychology, cognitive neuroscience, and computational cognitive sciences. Brain function may for instance inspire new computational theories and behavioral experiments that involve cognitive control and generalization. In addition, action video games may offer cognitive neuroscientists a practical and safe means to causally study cognitive control and may also provide new cognitive control assessments tools that may be more effective and valid than traditional batteries of tasks. Finally, the idea that effective cognitive training requires specific complex tasks, such as action video games, and is mostly ineffective when using simple cognitive task (Owen et al., 2010) seems to imply that as a field we need to study cognition within those complex tasks rather than focusing solely on standard cognitive tests, like the Stroop task for example. This calls for a paradigm shift in studying cognitive control which may benefit from modern technological advances in artificial intelligence (Botvinick, 2022; Doebel, 2020; Perone et al., 2021; Zink et al., 2021).

To sum up, cognitive neuroscience and psychology face two main challenges: (a) gain greater clarity on the cognitive control constructs (what it is and how to measure it), and (b) understand what features of the cognitive system (i.e., the agent) and what features of the task (i.e., the environment) determine cognitive control, its functioning, and generalization in humans. Chapters 1, 2, and 3 aim to tackle these challenges.

Artificial intelligence

The field of artificial intelligence provides a unique perspective on human cognition. Recent advances in machine learning have dramatically changed our ability to build accurate and scalable models of human cognition that previously relied on minimal theoretical frameworks and limited data (Ho & Griffiths, 2022). That is, modern cognitive science requires not only understanding cognitive control from a neural and psychological basis (Lindsay, 2020) but also understanding the computational mechanisms and to build artificial agents that are aligned and comparable to human cognition (Botvinick, 2022).

Control in artificial intelligence

Since its conception, artificial intelligence researchers have sought to develop computational models that mimic human intelligence. Unsurprisingly then, cognitive control has been investigated in artificial intelligence early on (G. Miller et al., 1960).

What does cognitive control look like in AI? Ideas in AI related to cognitive control have taken many forms. In its most abstract conception, control has been associated with optimizing parameters of computational models to allow them to learn how to perform a task and achieve a specific goal (Bensoussan et al., 2020). This limited view of control can be nevertheless very powerful when it is implemented in advanced model architectures that allow for the emergence of complex behavior. Indeed, this approach has been very successful in designing generic artificial agents capable of performing many different, complex tasks (Reed et al., 2022; Yang et al., 2019).

There are, however, more elaborate views of cognitive control that have emerged over the past decade, inspired by research in computational cognitive science (Ho & Griffiths, 2022). One such view offers that humans may simultaneously entertain two internal systems when performing a task: a model-free system and a model-based system (Daw et al., 2011). In essence, the model-free system learns a policy (i.e., "how to act") that maps states (e.g., stimuli) to actions (i.e., "responses"). This system is fast but simple and task-specific and it may thus generate errors and limit generalization. The other system is model-based, meaning that in the process of learning a policy, the system exploits its understanding of how the world works (e.g., by incorporating beliefs about state-transition in the decision making process). This system is slower and more "effortful" but it may also be more flexible and lead to higher performance levels. What is interesting about this work is that it has been used to evaluate human behavior. The results of that work show that not only do humans rely on both systems

(Dolan & Dayan, 2013), but the extent to which they do so depends on how much resources they have (Otto et al., 2015). For example, by putting people in a stressful situation it can be observed that their reliance on the model-free system increases presumably because internal resources are deviated towards addressing the stressor (Otto et al., 2013).

Recent work shows that in addition to accounting for human phenomena, this idea of "two systems" may in fact be grounded in computational principles (Moskovitz et al., 2022). More specifically, this framework posits the existence of two systems where one of the systems aims to perform a task well, while the other system aims in addition to simplify itself (by minimizing its description length) an idea that resonates in psychology with the concepts like automation of behavior, habit formation and the reduction of effort with practice. A key motivation for a system to be implemented in this way is not only the long-term reduction of computational resources but also its ability to generalize to new tasks as simpler models will need to discard more minute elements that are specific to a task and may thus generalize more than the full model.

Other interesting ideas in this context includes what we call "recycling" (or the active attempt to match what was previously learned to a new situation rather than starting from scratch; Tomov et al., 2021) and "composition"—the idea that complex behavior may emerge from models that are composed of computationally specific building blocks (Yang et al., 2019). These are just a few of the many ideas that are relevant in this field and that offer new avenues for the study of cognitive control both in psychology and computer science.

The value and challenges of interdisciplinary research

It is clear from the literature reviewed above, that there is great scientific and practical value in aiming to bridge the gaps between psychological and computer sciences; computational models can inform psychological theories and vice versa.

It is important to note that both in psychology and in artificial intelligence, the concept of generalization is a major current scientific challenge. Humans are endowed with unique abilities to flexibly adapt their behavior and generalize what they've learned in one context to new, never-before seen situations (Tenenbaum et al., 2011). Playing action video games, for example, is thought to improve cognitive control abilities and generalize to a broad set of tasks, ranging from visual contrast perception (Chopin et al., 2019) to reading (Franceschini et al., 2017). The mechanisms underlying these human generalization abilities remain, however, largely unknown. Current artificial agents, on the other hand, have very limited generalization abilities despite their tremendous success in performing complex tasks well (Chollet, 2019). To be more specific, these models are able to generalize from a training dataset to unseen test datasets that follow the same distribution of data (e.g., a cat-dog classifier can classify new images of cats and dogs; i.e., these models are robust) but they cannot easily generalize to new tasks (e.g., a cat-dog classifier can't play chess; i.e., these models are not flexible). It appears then that there are great opportunities for psychology and artificial intelligence to join forces and develop new models of cognitive control that could help both better understand the human mind and develop the next generation of artificial agents.

A key step towards making this happen is to make it possible, and even easy, to compare human and artificial agents directly. There are many cases where this has been successfully done at the single task level (e.g., Daw et al., 2011; Otto et al., 2015, 2013). There is comparatively less work comparing human and artificial agents across multiple tasks (Mnih et al., 2015; Yang et al., 2019). Yet, as stated by Yang et al. (2019): "The brain has the ability to flexibly perform many tasks, but the underlying mechanism cannot be elucidated in traditional experimental and modeling studies designed for one task at a time." A virtual environment allowing human and artificial agents to perform the exact same battery of tasks would be highly valuable and support the integration of cognitive control theories across psychology and artificial intelligence. It may help ground cognition in computational terms (Mnih et al., 2015; e.g., which types of tasks can be performed by a given computational architecture and which cannot; Yang et al., 2019), provide new insights and concepts to both psychology and computer science (Christian & Griffiths, 2016; Laird et al., 2017; Stocco et al., 2021), offer benchmarks for human and artificial agents as well as their comparison (relative performance profiles;), lead to the development of new tasks (e.g., tasks that are diagnostic of types of artificial agents and that could be tested on humans), and perhaps new computational architectures that truely generalize (Chollet, 2019).

Current research

The main strategy in this thesis has been to establish a broader, interdisciplinary view of cognitive control that can be conceptually, computationally, and empirically studied and integrates work within and across scientific fields. In line with this strategy, the current work explores a diverse set of approaches that together aim to better delineate the fuzzy concept of cognitive control.

The thesis comprises five research articles. Each of these articles are summarized in the following information sheets and discussed as a whole in the general discussion. Together this work illustrates, I hope, the benefits of the synergy between experimental psychology, neuroscience, and artificial intelligence in the study of cognitive control and opens up interesting future research perspectives.

Information sheets

Table 1: TL;DR – Chapter 1 (CogText)

Title	**Linking Theories and Methods in Cognitive Sciences via Joint Embedding of the Scientific Literature: The Example of Cognitive Control**
Challenge	Gain clarity on what is meant by cognitive control in the scientific literature and how it can be measured empirically.
Context	Despite a large volume of publications, cognitive control remains a rather vague concept both theoretically and operationally (Baggetta & Alexander, 2016). Literature reviews by human domain experts have had limited success in bringing such clarity: they are not exhaustive, can't keep up with the rate of new publications, and may depict a biased, subjective perspective rather than an objective, quantitative view of the research field.
Why it matters	Greater clarity on cognitive control and its measurement are critical to advance the field and integrate currently disparate research branches.
Method	We conducted automated text analysis on a large corpus of scientific abstract (+500K) downloaded from PubMed. We used a state-of-the-art language model (GPT-3) to encode scientific texts and create a joint view of cognitive control related constructs and tasks. This method allows the grounding of theoretical constructs on cognitive tasks (in the sense that tasks are used to measure the constructs) as well as the grounding of tasks on cognitive constructs (in the sense that constructs are used to theorize behavior in tasks). It also offers a unique holistic view of cognitive control constructs and tasks within a single knowledge graph.
Results	The results confirm the complex nature of cognitive control, explain the difficulty of defining cognitive control and may lead to new theoretical and empirical insights. We conclude that cognitive control can't be assessed using a single task and should instead be measured using a battery of tasks (varying contexts and demands) or more complex tasks (e.g., video games). We also conclude that as a construct cognitive control may benefit from being decomposed into smaller, better defined constructs to make progress in the field.
Output	The article was accepted as a conference paper for the CogSci2022 conference, the preprint is published on ArXiv (Ansarinia et al., 2022) and will be submitted for publication soon. The dataset is available on huggingface.co/datasets/morteza/cogtext ,and the code is publicly available on github.com/morteza/CogText. The methods and implications are further described in Chapter 1.

Table 2: TL;DR – Chapter 2 (CogEnv)

Title	**CogEnv: A Virtual Environment for Contrasting Human and Artificial Agents across Cognitive Tests**
Challenge	Modeling the environment: develop a virtual environment that allows the direct comparison of human versus artificial agents and thus supports the integration of cognitive control theories across psychology and artificial intelligence.
Context	There have been important advances in artificial intelligence but those advances are not readily accessible to psychological scientists. Similarly, psychological scientists have developed tasks, concepts, and theories that might not be accessible or perceived as relevant by computer scientists. One impediment to a shared understanding is the lack of an interoperable environment in which both human and artificial agents can interact with the exact same tasks.
Why it matters	Being able to record and directly compare behavior from both human and artificial agents opens up many new possibilities. It may help ground cognition in computational terms (Mnih et al., 2015; e.g., which types of tasks can be performed by a given computational architecture and which can't; Yang et al., 2019), offer benchmarks for human and artificial agents as well as their comparison (relative performance profiles), lead to the development of new tasks (e.g., tasks that are diagnostic of types of artificial agents and that could be tested on humans), and new computational models. It also allows to train a given artificial agent on a battery of tasks and to study task correlation and transfer effects (i.e., training on one task leads to improved performance on other tasks depending on how "similar" the tasks are) that can be compared with and tested on human participants.
Method	We developed CogEnv, a virtual environment that lets us interface both human and artificial agents to perform the exact same computerized battery of cognitive tasks. A wide range of artificial agents can be tested with this battery, provided they follow a common protocol (i.e., use pixels/symbols as input, process reward signals, and emit action). The data collected from these agents is in the same shape and format as human data and can thus be processed using the exact same data analysis code that is typical in experimental psychology (thus facilitating the direct comparison of human and artificial agents). As a proof of concept, we successfully trained baseline RL agents to perform a battery of cognitive tasks for which we also collected human data.
Results	The overall framework is operational and appears very promising. A preliminary investigation illustrates the idea that the comparison of performance/error profiles of human versus baseline RL agents may reveal aspects of human cognitive control that are yet to be addressed by artificial agents.
Output	The article was accepted and published as a conference paper for the CCN2022 conference. The code is available at github.com/morteza/CogEnv. The method and implications of the proposed environment and expected performance profiles are further described in Chapter 2.

Table 3: TL;DR – Chapter 3 (CogPonder)

Title	**CogPonder: Towards a Computational Framework of General Cognitive Control**
Challenge	Modeling the agent: developing a shared account of response times for human and artificial agents using a new type of computational model that functionally decouples control from controlled processes.
Context	Computational models embody our theoretical understanding in an explicit and testable way. Current computational models of cognitive control are lacking in important ways. In psychology, cognitive control models tend to be designed for specific tasks (e.g., Stroop) which makes it hard to study cognitive control in general (e.g., across a battery of tasks, while playing video games or in real-life activities). Computer science, on the other hand, has recently been able to develop artificial agents that can perform complex tasks. However, computer scientists typically ignore resource limitations and how long it takes for an agent to make decisions and act (in some cases, the environment is "paused" for the agent computation to be completed). A defining (and measurable) property of human cognitive processing is that it takes time and that this amount of time varies depending on numerous factors in a meaningful way (De Boeck & Jeon, 2019; i.e., response time; see Ratcliff & Starns, 2013). The exertion of cognitive control impacts response times and this impact is a major source of information in psychological research (e.g., "task-switching costs"; Monsell, 2003). What is missing then is a new type of computational model of cognitive control that is flexible enough to be used in combination with any model (hence being able to address more complex tasks), which decouples control from operation in a way that might be theoretically meaningful and which offers computational scientists a means to add control mechanisms to their computational models.
Why it matters	The envisioned computational models would benefit psychology by offering a principled means to investigate cognitive control across a wide range of situations as well as the possibility to exploit innumerous complex models that have been developed in computer science. It would also benefit computer science by offering a principled and computationally practical (i.e., differentiable, modular) means to augment existing computational models with control abilities resulting in time varying responses. The comparison of response time profiles across human and artificial agents furthermore may offer insights benefitting both disciplines.

Table 3: TL;DR – Chapter 3 (CogPonder)

Method	We propose a general deep learning framework that functionally decouples control (generating varying response times) from the decision making processes (making choices). The framework involves a controller that acts as a wrapper around any computational models (that "perceive" the environment and generate "actions" on that environment) and controls when the model should stop its processing and output a choice (this is known as the halting problem). This model is inspired by the Test-Operate-Test-Exit (TOTE) architecture (G. Miller et al., 1960) that conceives control as a recurrent mechanism that ultimately halts a computational process once a specific condition has been met. We instantiated TOTE using PonderNet, a recent deep learning framework for adaptive computing. By controlling the halting, the fameworks allows to continuously control how much resources are dedicated to the decision making agent and jointly affects the choices (accuracy) and response speed of the system. We implemented CogPonder, a flexible, differentiable end-to-end deep learning model that can perform the same cognitive tests that are used in cognitive psychology to test humans. We then trained CogPonder to perform two cognitive control tasks (i.e., Stroop and N-back) while at the same time aligning it with human behavior. Next we compared the behavior of CogPonder (i.e., accuracy and response times distributions) with the behavior of humans.
Results	CogPonder can be trained to perform cognitive tests and generates behavior that is similar to human behavior across multiple experimental conditions. CogPonder therefore provides a means for further investigating both human cognition and the computational models. The proposed model is very flexible (i.e., CogPonder can wrap around any deep learning model so is unattached to specific model choices) and can be extended in many ways (e.g., using more advanced computational techniques to perform complex tasks). Most importantly, the proposed framework explicitly connects human behavior to artificial agents that produce human-like behaviors on a battery of cognitive control tasks. The framework thus provides interesting new insights and research opportunities for both psychological and computer science.
Output	The manuscript will be submitted for publication soon. The code is available at github.com/morteza/CogPonder. The method and results of the proposed computational model of response time are further described in Chapter 3.

Table 4: TL;DR – Chapter 4 (Review)

Title	**Training Cognition with Video Games**
Challenge	Clarifying the relationship between training cognitive control with action video games and its transfer effects by reviewing behavioral and brain evidence.
Context	Experience impacts brain functioning and structure and there is now considerable evidence that specific training regimes can improve cognitive control. In particular, playing action video games, as opposed to other kinds of games, has been shown to cause improvements across a broad range of cognitive abilities (Bediou et al., 2018). Although there is no satisfactory explanation of these effects yet, one prominent view states that video games improve cognitive/attentional control abilities and that this improvement in cognitive control explains the transfer effects (Green et al., 2012).
Why it matters	Training cognition in a way that transfers to real life has many practical implications (e.g., rehabilitation, healthy aging, education, peak performance). Understanding the underlying mechanisms would allow us to devise more effective interventions. The study of transfer effects is important because it offers a setting to test cognitive control theories in a non-trivial way. We currently have no satisfactory theory that could account for how training on one task would impact performance on a never seen before task. Understanding transfer requires developing computational models that can perform multiple tasks—this is a general goal that computational cognitive control models aim for. The study of training effects and their consequences is also important because they offer a means to causally test computational theories. Finally, the study of behavior during video game play poses interesting new questions to cognitive control scientists. Video games are complex interactive environments that engage cognitive systems in multiple, context dependent ways. Studying behavior during video game play may offer new insights on cognitive control that are relevant in the real world and that might not be apparent when using elementary cognitive tests.
Method	This chapter reviews the behavioral and neuroimaging literature on the cognitive consequences of playing various genres of video games.
Results	Our review highlights that different genres of video games have different effects on cognition. Action video games—as defined by first and third person shooter games—have been associated with greater cognitive enhancement, especially when it comes to cognitive control and top-down attention, than puzzle or life-simulation games. Playing action video games seems also to impact reward processing, spatial navigation, and reconfiguration of attentional control networks in the brain. Interpretations of the effects of playing action video games on behavior and the brain have been attributed to various psychological constructs, in particular attentional control, quick processing of sensory information, and rapid responses. These results suggest that cognitive training interventions need to be endowed with specific game mechanics for them to generate cognitive benefits, presumably by enhancing cognitive control abilities. We discuss what those game mechanics might be and call for a more systematic assessment of the relationship between video game mechanics and cognition. We also note that as video games become more and more advanced (i.e., mixing genres and game-play styles within the same video game), it will become increasingly difficult to study and understand their effects on cognition. This article lays a foundation for the study of cognitive and brain functioning using video games and illustrates the value of this approach to investigate general cognitive control.

Table 4: TL;DR – Chapter 4 (Review)

Output	The article has been published as a peer-reviewed book chapter (Cardoso-Leite et al., 2021). It is further provided in Chapter 4.

Table 5: TL;DR – Chapter 5 (ACNets)

Title	**Neural Correlates of Habitual Action Video Games Playing in Control-Related Brain Networks**
Challenge	Test the idea that action video game play affects neural functioning in ways that are compatible with cognitive control hypotheses according to which action video gaming improves cognitive control which in turn explains improved performance across a wide range of cognitive tests (i.e., transfer).
Context	On the one hand, research shows that playing action video games improves cognitive performance across a wide range of cognitive tasks, presumably by enhancing people's cognitive control abilities (Bediou et al., 2018). On the other, the cognitive neuroscience literature has highlighted integration of several functional brain networks as being important for cognitive control (Menon & D'Esposito, 2022). These two sets of theories have not yet been empirically confronted despite there being great value to do so. Indeed, there are competing hypotheses regarding the effects of action video gaming—some highlighting domain-general abilities (e.g., attention, cognitive control), others focusing on domain-specific ones (e.g., response speed). These alternative views make rather different predictions regarding changes in brain function (e.g., changes in specific functional networks vs changes in specific areas). Similarly, research on functional brain networks has highlighted numerous cognitive control networks. There are however some inconsistencies across such theories. Studying the impact of playing action video games provides a means to empirically test those theories and improve our understanding of how those networks work.
Why it matters	The study of the differences in functional brain networks between habitual action video game players and non-video game players can advance our understanding of both the mechanisms underlying the action video game training effects and the neural mechanisms supporting cognitive control in general. Confirming that action video game play affects cognitive control (via its functional neural underpinnings) has important implications for the study of cognitive training. It also has practical value as it would offer cognitive neuroscientists a new tool to causally study cognitive control. Finally, this type of work could lay a foundation towards bridging a gap between experimental psychology, cognitive neuroscience and computational cognitive sciences (brain function may for instance inspire new computational theories and behavioral experiments).

Table 5: TL;DR – Chapter 5 (ACNets)

Method	We curated a dataset collected by (Föcker et al., 2018). The dataset comprises resting-state fMRI data (7 minutes and 30 seconds, or 125 time points) and task-fMRI data from a total of 32 human subjects (16 habitual action video gamers and 16 non-gamers). The original study focused on task-fMRI; here we analyze the resting-state data. We developed a machine learning pipeline to investigate the differences between habitual action video gamers and non-video gamers in terms of their functional resting-state brain connectivities, focusing in particular on networks associated with cognitive control. We used a robust approach to preprocess, remove confounds, parcellate, aggregate networks, and extract resting-state functional connectivity measures from the BOLD signals. The whole pipeline was cross-validated, and several arbitrary choices in the preprocessing were considered as hyperparameters of the model (for example parcellation atlas and connectivity measure). We trained a classifier to discriminate unseen participants as action video gamers versus non-gamers based on their resting-state functional connectivities. We then investigated what features were responsible for the model prediction accuracy by applying a permutation feature importance test. Additionally, SHAP analyses were conducted to investigate the contribution of each feature to the output (not the accuracy) of the model.
Results	Our model is able to classify unseen participants as action video game players based only on their resting state functional connectivities with an accuracy of 72.6%. This high level of accuracy demonstrates the value of resting state functional data to study action video gaming. Interestingly, the performance of the classifier depended on the specifics of the method used (i.e., parcellation technique, type of connectivity metric), supporting the utility of the robust/exhaustive methodology employed in this study. Investigating why the classification was successful shows that there is in fact no specialized network that differs among the two groups of participants. Instead, it is the interplay between networks that matters most, and in particular the interplay between the cingulo-opercular and the sensorimotor networks and between the frontoparietal and the sensorimotor networks—a result that is robust to variations in parcellation and connectivity metric. These results do not support the view that individual networks are enhanced by action video game play and suggest instead a mechanism that involves a reconfiguration of a collection of networks. These results provide new insights and have clear implications for both theories of action video game training and for cognitive neuroscientific theories of cognitive control in the human brain.
Output	The article is being prepared for journal submission. The code is available on (github.com/morteza/ACNets)[https://github.com/morteza/ACNets]. The method and results are described in Chapter 5.

Chapter 1

Linking Theories and Methods in Cognitive Sciences via Joint Embedding of the Scientific Literature: The Example of Cognitive Control

Morteza Ansarinia, Paul Schrater, and Pedro Cardoso-Leite

Abstract

Traditionally, theory and practice of cognitive control are linked via literature reviews by human domain experts. This approach, however, is inadequate to track the ever-growing literature. It may also be biased, and yield redundancies and confusion.

Here we present an alternative approach. We performed automated text analyses on a large body of scientific texts to create a joint representation of tasks and constructs. More specifically, 385,705 scientific abstracts were first mapped into an embedding space using a transformers-based language model. Document embeddings were then used to identify a task-

construct graph embedding that grounds constructs *on* tasks and supports nuanced meaning of the constructs by taking advantage of constrained random walks in the graph. This joint task-construct graph embedding, can be queried to generate task batteries targeting specific constructs, may reveal knowledge gaps in the literature, and inspire new tasks and novel hypotheses.

1.1 Introduction

A key challenge in cognitive sciences, and in particular cognitive psychology and neuroscience, is to make sense of observable phenomena (i.e., behavior) in terms of theoretical constructs. Consider for instance cognitive control (CC)—a broad construct that comprises many components and engages multiple mechanisms which collectively aim to describe goal-directed behavior in a complex, uncertain world. CC is a major construct in cognitive sciences: In the year 2021 alone, PubMed indexed 974 papers with the term "cognitive control" in the title or abstract—an average of 3 papers per day. To understand CC, researchers have introduced a variety of theoretical constructs and conceived numerous cognitive tasks (see Baggetta & Alexander, 2016). However, the relationships between and within related constructs and tasks are not always clear. For example, because they are "measured" using the same set of tasks (e.g., Stroop, N-back, Digit Span, Stop-Signal, Task Switching), it seems reasonable to assume that cognitive control (Botvinick & Cohen, 2014), executive functions (Baggetta & Alexander, 2016), attentional control (Rey-Mermet et al., 2021), and self-regulation (Enkavi et al., 2019) are somewhat equivalent constructs; yet, they are not widely considered equal (Nigg, 2016).

Traditionally, the meaning and relationships between constructs and tasks are conceptualized in extensive literature reviews conducted by human experts. In this approach, researchers "manually" read, synthesize, and criticize the literature and write reviews or reports describing their understanding. Following such reviews, CC is viewed as interactions between generic core processes (e.g., inhibition, flexibility, working memory, and interference control in Diamond, 2013), interactive componential (Badre, 2011), tasks-specific processes driven by goals (Doebel, 2020; Logan, 2017), or optimal parameterization of naturalistic tasks (Botvinick & Cohen, 2014). This approach has been invaluable but it may also yield biased results (Beam et al., 2021; Brick et al., 2021) and seems inadequate to track the ever-growing literature and stay current. In this context, modern machine learning methods may provide useful and complementary insights.

When considering terms in the literature, there are two major impediments to creating consistent construct-task associations: *construct hypernomy* when conceptualizing CC and *task impurity* when operationalizing it. *Construct hypernomy* occurs when description of the same construct varies across different contexts due to the way it is assessed. It creates different meanings of the same concept. "Attentional Control", for example, likely means something different in Ahissar & Hochstein (1993) (as measured by low-level perceptual tasks) than it does in Burgoyne & Engle (2020) (as measured by complex cognitive tasks). *Task impurity*, on the other hand, refers to the idea that performance on a task loads onto multiple constructs (i.e., there is not a one-to-one mapping between constructs and tasks). Because of the impurity, no task taps into just one isolated construct. Performance in the Backward Digit Span, for instance, involves short-term memory, visual perception, sustained attention and working memory, to name just a few. The consequence is that constructs lack a consistent, groundable semantic content, corrupting interpretations of neural and cognitive research that depend on them.

Construct hypernomy and task impurity are quite common in CC research because complex concepts like cognitive control manifest themselves differently across different individuals and contexts (Burgoyne & Engle, 2020). For that, researchers often use multiple tasks in their studies and apply statistical methods such as latent factor analysis to discern underlying constructs. Nevertheless, the resulting latent models of CC are rarely agreed upon, as is the selection of tasks (Doebel, 2020; Enkavi et al., 2019; Nigg, 2016; see, for example, Rey-Mermet et al., 2021).

Ambiguous associations of constructs and tasks make it hard to interpret past results, hinder scientific progress and the development of effective interventions. With the advent of scalable machine learning, however, construct-task associations may be clarified. The goal of this paper is to approach the conceptual richness of a large body of scientific works and take advantage of recent context-aware language models in machine learning to clarify the association of CC tasks and constructs. More specifically, we collect and analyze scientific texts about CC tasks and constructs and encode text data into rich semantic embeddings using transfer learning. Transfer learning exploits the rich representations generated by natural language models trained to faithfully represent contextual meaning—unlike traditional bag-of-word or clustering techniques. Similarities between embedded representations are then used to build up a hypergraph (Battiston et al., 2021) that connects tasks and constructs.

First, we show that this hypergraph representation regrounds constructs on tasks and provides nuanced meaning of the constructs, ultimately demonstrating construct hypernomy. Second we show that pulling theoretical and experimental literature into overlapping components of a hypergraph may greatly benefit researchers: the joint task-construct embeddings can be queried to generate special-purpose task batteries, it may reveal knowledge gaps, inspire the design of new experiments and yield novel hypotheses regarding the structure and function of CC. This empirical and descriptive model of the literature, rather than expert-driven ones, may also be used in future applications to enhance knowledge searches (see Beam et al., 2021 for a comparison of a data-driven mapping of the literature and expert-driven knowledge frameworks like DSM for psychiatric illness and RDoC for for basic brain function).

1.2 Methods

Data. We created a lexicon of CC-related terms (172 terms, of which 72 were task names and 100 were construct names) based on the previously published work on cognitive control (Barch et al., 2009), Attentional Control (Bastian et al., 2020), Executive Functions (Baggetta

& Alexander, 2016; Diamond, 2013), and Self Regulation (Enkavi et al., 2019). Each term in the lexion was associated with a PubMed-specific search query by which papers with the term in their title or abstract were retrieved. This resulted in a dataset of loosely labeled documents, each labeled by one or more lexicon terms (n=522,972 hits, of which 385,705 were unique). For the purpose of the current analyses we only retained the title and abstract of the papers, along with the lexicon terms that were used to retrieve them. Having multiple labels per document was crucial to quantify the co-appearance of the terms in the literature. After the documents were collected we removed 14 terms from the lexicon because they yielded too few documents to support cross validation splits ($n < 5$).

Analysis. To understand the relationships among and between tasks and constructs, our goal is to build graphs that represent tasks and constructs as nodes and measure similarity/distance between them as edges. Graph G can be used to jointly infer embeddings of both construct and task nodes in a shared vector space, in that relative closeness of two nodes is estimated by the similarities of node attributes as well as the shared neighbors in the graph. Heterogeneous graph $G = (V_{tasks} \bigcup V_{constructs}, E)$ is defined by its two types of nodes, V_{tasks} and $V_{constructs}$, labeled by either a task or a construct term, while the weighted edges, E, represent the links between two or more nodes, reflecting similarity of the corresponding terms in the literature. Node attributes being relevant scientific texts, the existence and weight of a link between two nodes is predicted by the similarity of corresponding node attributes; the higher the similarity between node attributes, the higher the chance of the nodes being associated. The core problem becomes learning task and construct attribute embeddings that predict co-occurrence and semantic similarity measures. We used the following steps to create the graph G from the collected scientific texts.

The data collection resulted in a dataset of 385,705 unique, but loosely-labeled, abstract texts, all of which were then encoded into embeddings of 1024 dimensions using a pre-trained transformer language model (GPT-3 Ada for text similarity embedding; see Brown et al., 2020). The language model transformed raw texts into 1024-dimensional vectors, gpt3-embedding, representing semantic similarity between two or more pieces of text. Since keeping the original structure of the text was important for the model to understand the context, we did not preprocess the raw text. To convert text similarity into a shared topic representation (which improves relating task and construct text embeddings), we applied Top2Vec topic modeling (Angelov, 2020) to the gpt3-embedding which projected them into a space of 473 dimensions, i.e., topic-embeddings. Each column of the topic-embedding matrix represents a topic, and element ij shows the probability of assigning document i to the topic j. Realigning the gpt3-embedding into topic-embeddings improved the quality of the dataset for a number of reasons. First, it improves the quality of the labels in the dataset by discarding outlier documents. These are documents that belong to no topics of interest or are assigned to irrelevant topics (e.g., genetics)—after removing outlier documents, 293'014 unique documents remained for further analysis. Second, topic modeling allows one to extract a useful, interpretable representation of the documents, as each dimension of the topic-embedding shows the probability of assigning a document to a topic while being faithful to the contextual representation of the documents in the gpt3-embedding space. This generates a digraph between nodes representing lexicon terms and the topic-embedding vectors.

To convert this into a construct-task graph, we grouped lexical terms associated with construct and tasks to generate graph nodes. To compute topic-similarity between groups of lexical terms associated with each construct or task node, we fitted a multivariate normal distribution over the topic vectors of each node separately and then calculated the distance between all nodes as measured by the Jensen-Shannon divergence of those node-level distributions. This step added edges to the graph, G, with edges weighted by the inverse distance of nodes in the JS-divergence matrix.

To learn a representation of the graph that only preserves paths from tasks to constructs and vice versa, we then applied Metapath2Vec (1000 random walks of step size 100, accompanied by skip-gram Word2Vec embedding of size 128 and maximum window size of 5; as recommended in Ruch, 2020). The Metapath2Vec embedding encodes random walks of specific patterns in a heterogeneous graph, here patterns being alternating random walks between task and construct nodes.

Finally, by applying HDBSCAN soft clustering to the node attributes and thresholding the edges (discarding all the edges weighed within one standard deviation from median), we transform the graph G to a homogeneous hypergraph, i.e., nodes are now only of type task, while constructs are hyperedges that group a subset of tasks in overlapping clusters.

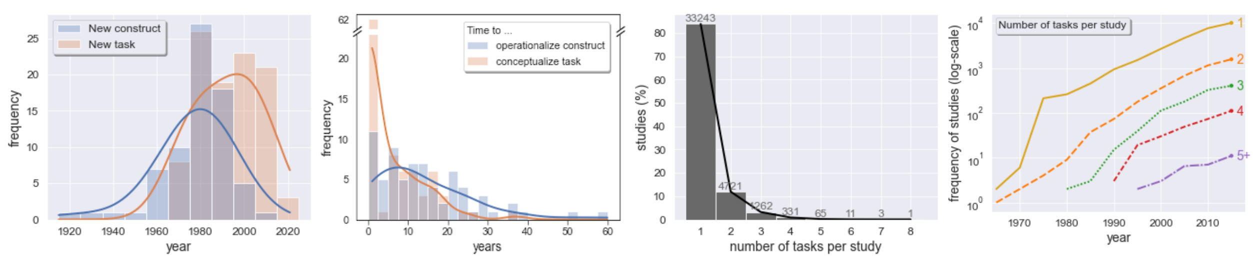
1.3 Results

We used a variety of data-driven approaches to collect and understand CC publications. Briefly, we (a) created an all-inclusive lexion of construct and task terms, (b) queried PubMed to collect relevant abstract texts, (c) vectorized all the raw texts using GPT-3 Sentence Similarity Embedding, an unsupervised pre-trained language model, (d) applied Top2Vec topic modeling technique to all the document embeddings together and identified dimensions of a useful latent space, i.e., topics. We then created a graphical representation of the lexicon terms, i.e., task-construct graph, and used them to predict the association between terms.

The richness of tasks and constructs in the literature. Although there are many task and construct terms, their relative frequencies differ widely. For example, “Stroop Task” is mentioned 8,003 times in the period 1973–2022 while “Delay Discounting Task” was only mentioned 466 times over the same period of time. The use of each term tends to increase over time. Interestingly the rate at which new constructs and tasks are introduced does not follow the same curve as the number of publications in the field; rather there seems to have been a peak of innovation for constructs around 1980 and for tasks around the year 2000 (panel a in Figure 1.1). Such patterns, visible in simple descriptive statistics (Figure 1.1), may provide interesting insights into understanding the maturity and vitality of a research field.

Regrounding constructs on tasks. It took on average 7 years for the constructs to be explicitly associated with a task (see panel b in Figure 1.1). The meaning of a theoretical construct may change across time and gain clarity and precision with new empirical measures and cognitive tasks being used by the research community to flesh out the construct. A core idea in this paper is that by evaluating how constructs are operationalized (i.e., linked to cognitive tasks) key insight can be gained about what a construct means. Grounding the definition of constructs on tasks provides a nuanced meaning of constructs that relies on observable measures. It also allows the computation of useful measures on constructs (e.g., specificity) and on between pairs of constructs (e.g., measures of redundancy, similarity, and distance). To investigate the relationships between cognitive constructs, we use hyperedges in the task-construct graph as a measure of similarity, indicating the extent to which a construct hyperedge can be reconstructed by neighboring tasks.

Construct hypernomy. The task-construct graph readily demonstrates construct hypernomy and task impurity in the CC literature. We first sought hypernomy as highly overlapping hyperedges of seemingly incompatible constructs, as well as a high degree of task nodes with neighboring constructs as a measure of the task impurity. Figure 1.2 illustrates overlapping hyperedges of the most popular constructs where hyperedges for cognitive control, Executive Control, Behavioral Control, Central Executive, and Attentional Control are overlapping and identical.



(a) (b) (c) (d)

Figure 1.1: (Panel a) Introducing new tasks (task innovation) and constructs (concept innovation) is characterized by a burst followed by declining innovation. (Panel b) Task and Construct occurrences in publication abstracts are temporally decoupled. Time to operationalize constructs (blue) is the time between the first occurrence of a construct and the first co-occurrence of that construct with any tasks, while Time to conceptualize tasks (orange) is the time between the first occurrence of a task and the first co-occurrence of that task with any of the construct. (Panel c) The majority of the literature only used one task in their studies, showing a lack of multitask design of experiments. (Panel d) While the number of papers published each year increases exponentially, the number of tasks per study remains fairly constant across time.

Task inconsistency across disciplines. A major source of hypernomy stems from descrip-

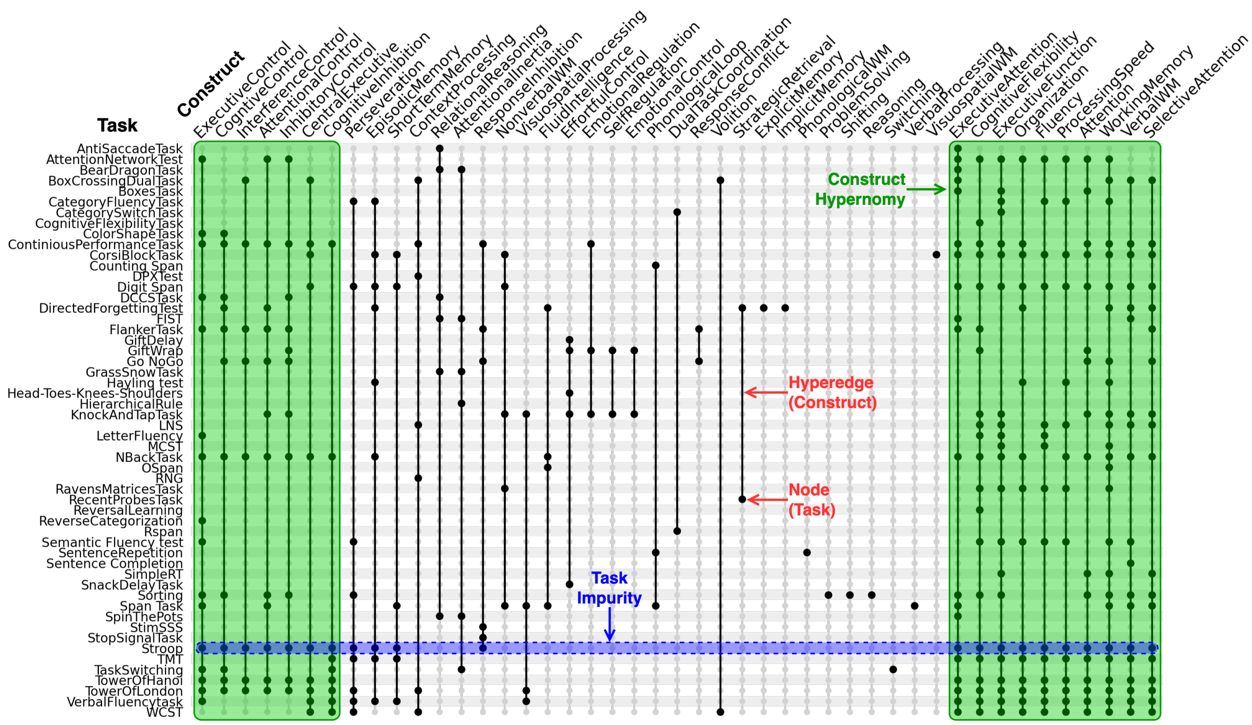


Figure 1.2: Task-Construct hypergraph: representations of control-related constructs as hyperedges (vertical black lines) over a subset of tasks (nodes). Construct hypernomy is reflected as overlapping hyperedges (e.g., green regions), and task impurity as nodes scattered over multiple hyperedges (e.g., blue region). Distances between nodes are not meaningful. Nodes are reorganized for visual clarity and only a subset of the graph is displayed.

tions and measurements of the constructs often being inconsistent across scientific communities. To test this idea, we sought to determine whether construct hyperedges, and their task associations, vary across four cognitive disciplines (psychology, neuroscience, cognitive science, and social science). Using the same method described in the analysis section we created four discipline-specific graph embeddings. The only difference was that publications were grouped by discipline, which was determined by searching for the terms "social", "psycho", "neur", or "cognit" in the journal titles. Constructs that have inconsistent task associations across the disciplines are hypernomic (Figure 1.3).

Refactoring tasks and constructs. Designing effective assessments of CC can be challenging for a number of reasons. Participants have limited time to spend on cognitive tasks. 1) If these tasks are poorly selected, performance on these tasks may not be very informative (e.g., measures are conceptually redundant); 2) If only one task is used, the inferential resolution of performance to construct is very limited. Thus in order to be able to make specific theoretical claims about CC it is necessary to use multiple, well-chosen tasks in experiments. This is currently not the case. As shown in Figure 1.1 (panel c), most research uses only one task. In fact, only 17 percent of publications used 2 or more tasks. The task-construct graph presented here may facilitate novel experimental designs of such multi-task, max-information experiments by providing a similarity-based space in which tasks can be identified, and grouped, by the overlapping subgraphs (i.e., constructs) that they belong to.

In the task-construct graph, two tasks are similar if they share identical neighbors, i.e., constructs. And tasks cover a set of constructs if their union set overlaps the corresponding hyperedges of the constructs. These principles equip researchers with sound and quantified methods to refactor tasks (e.g., discard redundant tasks, quantitatively measuring similarity of tasks via constructs, and performing set operations on a group of tasks). Such a refactored set of tasks controls the construct-redundancy of tasks and will shorten the time required to complete comprehensive assessments. It provides a method to design a task battery to effectively cover constructs (i.e., minimal redundancy while measuring different facets of the constructs).

Sparsity in the task space. There are numerous cognitive tasks in the literature; how these tasks relate to each other remains unclear. There are many cognitive control tasks that are rarely used (see Baggetta & Alexander, 2016), and even fewer used in combination with other tasks. Even when tasks were used together, their relationship might still be unclear. The question of how tasks relate to each other is key in the cognitive training domain where researchers aim to train cognitive abilities in general rather than performance on a specific task. In that context, a common point of disagreement is to predict and interpret transfer effects (i.e., how much training in task A improves performance in task B). A measure of distance between tasks based on their grounding on constructs may provide an objective foundation to understand these transfer effects—the task-construct graph embedding proposed here provides a means to compute such inter-task distances.

To quantify the distance between two cognitive tasks, we compute the Jensen-Shannon divergence between their node embeddings in the task-construct graph. Figure 1.4 shows, for example, that the Trail Making Task is relatively close to the Digit Span Task, suggesting its training effects transfer more easily to the Digit Span Task than to tasks such as the Discounting Task.

Distance between the task nodes can also allow us to identify gaps in the task space: gaps may be visible as disconnected graph components. Identifying such gaps may reveal opportunities to develop new useful tasks. Alternatively, there may only exist associations between groups of tasks and groups of constructs—i.e. the task-construct associations are not *atomic*. This reflects a lack of purity in the tasks or constructs or both that might be improved by refactoring constructs or decomposing tasks into components.

Querying the graph embedding for task batteries targeting specific cognitive constructs. Some studies use batteries of tasks that together address a research question and measure one or more constructs from several viewpoints. The process of building such task batteries can be facilitated by leveraging the task-construct graph embedding; one can query the graph for an array of tasks spanning a given set of constructs. The joint embedding translates queries into arithmetic operations in the embedding space (positive samples and negative samples), allowing for more explicit and visible decisions.

Query operations on the task-construct graph are made possible by using the underlying node embedding vectors extracted as a part of Metapath2Vec graph embedding. Queries include, for example, prioritizing tasks for a given construct, or a set of tasks for a set of constructs. To prioritize tasks for a construct, the task-construct graph looks for task nodes that are closest to the simple mean of the queried construct, e.g., in terms of sum of weighted node embeddings. And for a list of tasks for multiple constructs, find the minimum spanning tree that covers all the queried construct hyperedges. For example, if one queries `(Reward Processing + ReversalLearning - GoNoGo - SortingTask)`, one will get the recommendation to use the `BART`, `GiftDelay`, `BalanceBeam` (Baggetta & Alexander, 2016), and `StimSSS` (Enkavi et al., 2019) tasks, which are ordered by the cosine similarity between the mean vector of the query and the task vectors in the graph embedding model.

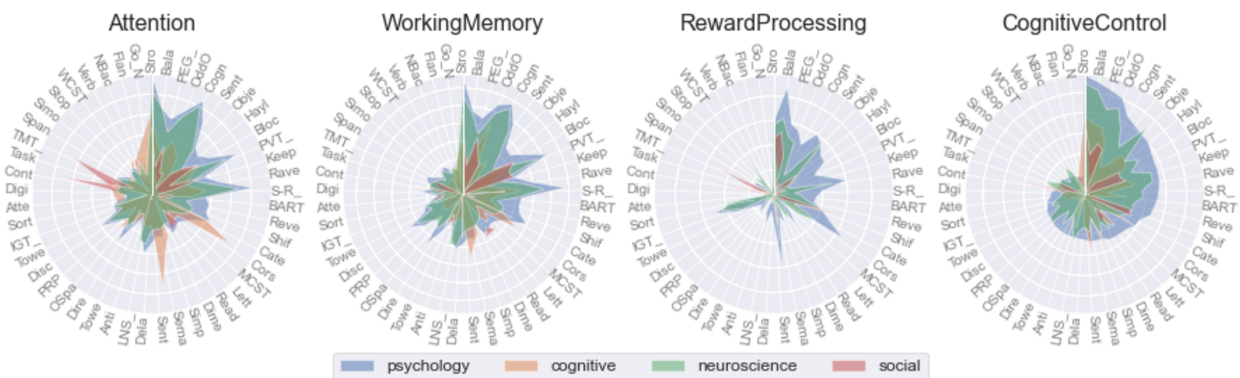


Figure 1.3: Associations between tasks and constructs minimally overlap across scientific disciplines. Rose plots show the relative association between constructs and tasks, with each color representing a different field. Lack of overlap between the “spikes” indicates disjoint operationalizations across fields.

1.4 Implications

Ambiguous meanings and relationships between cognitive tasks and constructs call for a more rigorous way to handle constructs—an obvious solution would be to adopt a more formal notation and refer to specific knowledge models (e.g., ontologies). The knowledge

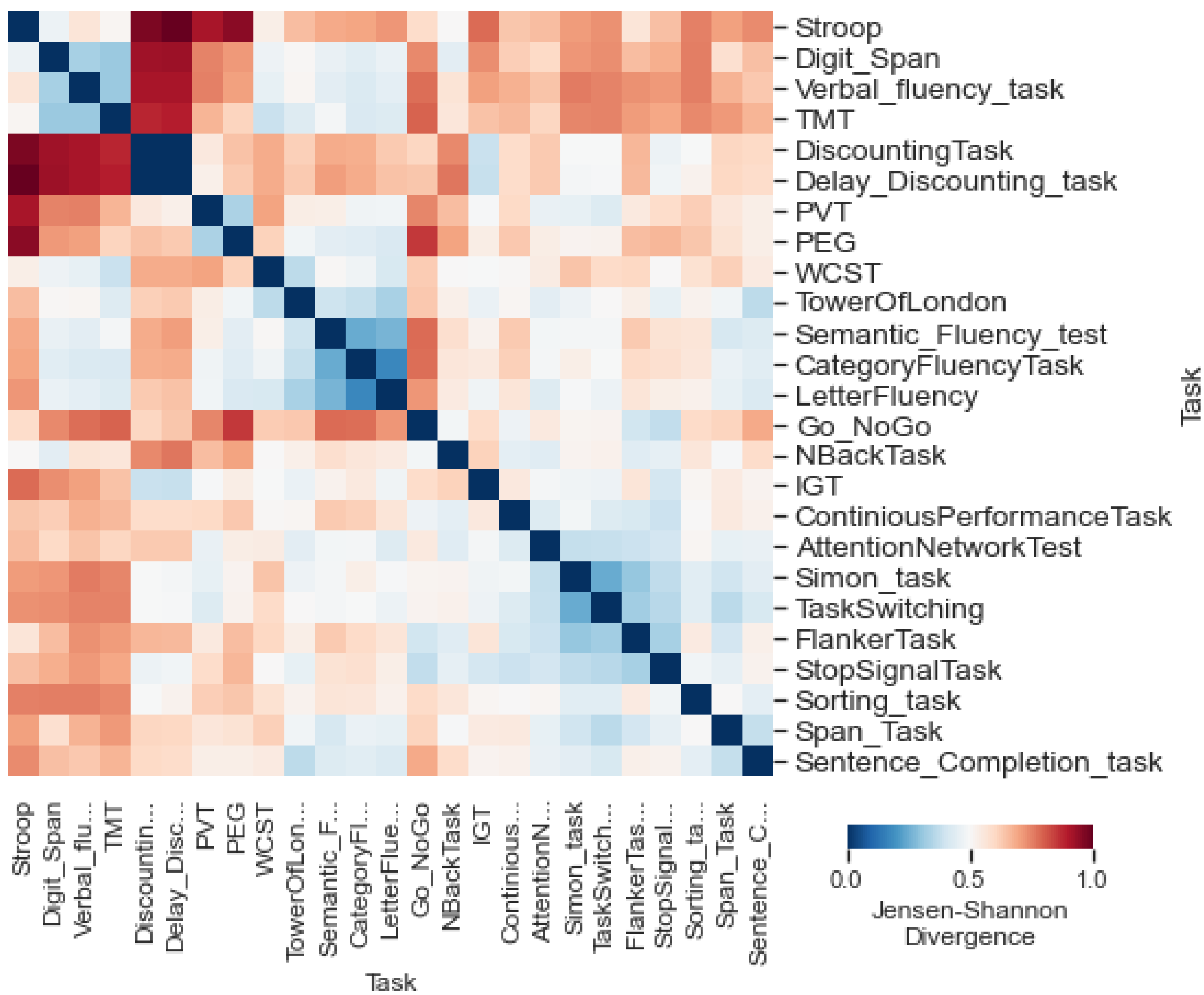


Figure 1.4: Pairwise distances between the 25 most popular cognitive control tasks as measured by the symmetric Jensen-Shannon divergence of two multivariate normal distributions of their node attributes in the task-construct graph. Higher divergence indicates higher dissimilarity between corresponding scientific texts. Task-task distances may for example provide a data-driven proxy for predicting and explaining transfer effects in cognitive training research.

model must be flexible enough to capture a wide range of association between constructs and tasks. The proposed task-construct graph embedding provides a useful representation of the cognitive control literature built upon topic embedding. In this representation, association of two entities, e.g., task-construct, relies on shared topics as well as the walks between them in a graph representation. By predicting links using topic embeddings of the nodes, we find most similar aspects of, for example, two constructs, a similarity that could be explainable in natural language.

A consistent, sound, and parsimonious framework of CC has been desired from the beginning. Yet, the growing number of publications and newly introduced constructs makes it impossible to integrate them into a bigger picture. While researchers may disagree on theoretical perspectives and thus on which terms to use, they generally might agree on the fact that if two constructs are "measured" by the same tasks, the constructs must be somewhat related. We proposed a joint embedding of constructs and tasks (based on scientific texts in a graph representation) to drive a more nuanced interpretation of the constructs by regrounding abstract constructs on the concrete set of observable tasks.

The proposed graph-based embedding enables explanatory reasoning driven by scientific texts. Unlike expert-driven models, the models reason regardless of the preferences in research; yet it is not clear whether other kinds of biases are addressed as the knowledge source and pre-trained language model are themselves produced by humans. By scaling up the knowledge model to a large body of available texts, the model is able to encapsulate even more aspects of cognitive control, and in general, multidisciplinary research.

Disagreements about the meaning of a construct are partly explained by differences in how we interpret responses to a particular task. By focusing on the co-occurrence of task and construct names in scientific texts, our approach implicitly makes strong assumptions about the relationship between abstract constructs and their imperfect but observable measures. The limitations of the present work can be partially addressed by expanding the hypergraph to include, for example, concepts such as brain mechanisms, research communities, and analysis techniques.

Explainable symbolic AI and machine learning have been long in debate to model knowledge. Regardless of the specific topic discussed here (i.e., cognitive control), the proposed model can be seen as an effort to connect symbolic modeling (as in ontologies) and machine learning (as in embeddings). Our method informs an ontology of scientific texts using context-aware embeddings that are extracted from a loosely-labeled body of scientific texts requiring minimal human input. It is an automated pipeline that only requires a lexicon, builds on large-scale language models and that can scale to millions of documents, making it a viable approach to meaningfully monitor the scientific literature continuously and extensively.

Chapter 2

CogEnv: A Virtual Environment for Contrasting Human and Artificial Agents across Cognitive Tests

Morteza Ansarinia, Brice Clocher, Aurélien Defossez, Emmanuel Schmück, and Pedro Cardoso-Leite

Abstract

Understanding human cognition involves developing computational models that mimic and possibly explain behavior; these are models that "act" like humans and produce similar outputs when facing the same inputs. To facilitate the development of such models and ultimately further our understanding of the human mind we created CogEnv—a reinforcement learning environment where artificial agents interact with and learn to perform cognitive tests and can then be directly compared to humans. By leveraging CogEnv, cognitive and AI scientists can join efforts to better understand human cognition: the relative performance profiles of human and artificial agents may provide new insights on the computational basis of human cognition and on what human abilities artificial agents may lack.

2.1 Introduction

Understanding the computations underlying human cognition is vital for scientific progress. Most efforts in cognitive sciences to understand how people perform cognitive tests focus on models that describe the data (e.g., factor analysis). There are only a few models describing the mechanisms underlying the performance of a task (i.e., models that "act" like humans and produce responses) and fewer still that can account for performance across many tasks.

One productive strategy has been to develop cognitive architectures (e.g., ACT-R, Anderson et al., 2004). Alternatively, recent developments in AI allow the application of flexible, generic architectures to solve a wide variety of problems. Their ability to do (or not do) so may reveal computational constraints underlying specific tasks (Yang et al., 2019). Reinforcement Learning (RL) seems particularly well suited to model performance in cognitive tests as they typically involve the presentation of a stream of stimuli and the execution of a discrete set of actions followed by a reward signal that may drive learning (Mnih et al., 2015).

Despite the relevance and potential of RL to model cognition, there is currently no easy way to train RL models on the same cognitive tests that are used to assess humans. Here we present CogEnv, a configurable multi-task environment for RL agents to emulate cognitive tests. Under the hood, CogEnv uses DeepMind's AndroidEnv (Toyama et al., 2021) to run the Behaverse cognitive assessment battery (see behaverse.org). Behaverse tasks are customizable at many levels, allowing the construction of a large number of randomized trials for training RL agents. In the following sections, we present the technical details of CogEnv and its ability to run RL agents on cognitive tests.

2.2 Technical specification

We simulate a real-time RL environment, where the environment, upon receiving an observation, invokes a callback method in the agent. We use AndroidEnv to run and manage the Behaverse cognitive assessment battery in a virtual Android device. A set of task-specific parsers then decodes screenshots, event streams, and system logs to extract numerical rewards and symbolic observations (see Figure 2.1). The reward and the observed state are then sent to the agent via the callback. CogEnv then waits for the agent to respond with an action, and issues a timeout if no response occurred within a duration specified by the cognitive test.

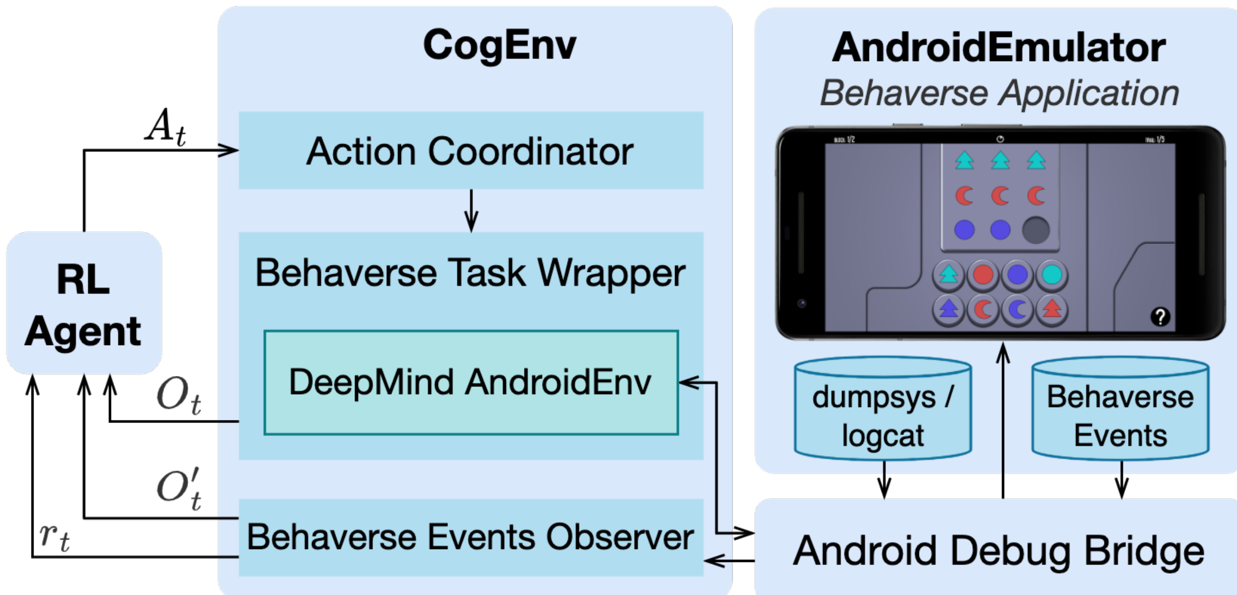


Figure 2.1: Overall architecture of CogEnv. CogEnv communicates with AndroidEnv via Protocol Buffer messages and manages access to the Behaverse events. O_t is the screenshot of the task at time t, O'_t is the extra observations extracted from the Behaverse events including information about the task and stimuli, r_t is the reward, and A_t is the agent's action.

2.2.1 Tasks

CogEnv currently runs four Behaverse tasks (see Figure 2.2) selected to cover main components of cognitive control (see Chapter 1). In the Belval Matrices test for example, agents are shown a matrix of symbols on a 3x3 grid, where one of the cells of the matrix has been removed, and they are tasked to identify the missing cell from a set of eight options (panel D of Figure 2.2). The Belval Matrices can randomly generate a large number of test items of varying difficulty and structure, which makes this test interesting for human and artificial learning studies.

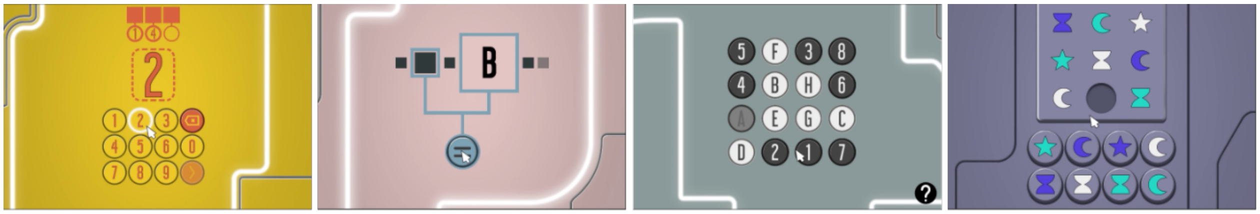


Figure 2.2: Screenshot (O_t) of four Behaverse tasks. A) Digit Span (working memory), B) N-Back (working memory), C) Trail Making Test (cognitive flexibility), and D) Belval Matrices (matrix reasoning). See behaverse.org.

2.2.2 Timing

CogEnv supports both step-lock (i.e., turn-based, where the environment pauses between two consecutive actions) and real-time mode (where the environment runs asynchronously from the agent). A real-time environment is necessary to study the timing of actions: in cognitive psychology, human behavior is typically evaluated in terms of both accuracy and speed.

2.2.3 Action space

Each test defines a discrete action space that is in fact bounded tap gestures on the buttons of the graphical interface. The Action Coordinator component (see Figure 2.1) automatically constructs a sequence of AndroidEnv gestures (TAP, TOUCH, and LIFT) that together perform the requested action as a set of movements in the emulated device.

2.2.4 Observation space

CogEnv asynchronously invokes and waits for the agent to act. The invocation is accompanied by a screenshot of the Behaverse screen, as well as the reward value and symbolic representations of the task state extracted from the logs and event streams.

2.3 Comparing humans and artificial agents

CogEnv allows us to compare human and artificial agents on the exact same cognitive tests, generating for both the same type of data that can be analyzed using a common data analysis pipeline. Figure 2.3 illustrates how such comparisons may yield new insights.

We collected data (accuracy and response time) from 200 human participants completing 20 items of the Belval Matrices (see Figure 2.3) and are currently training a selection of discrete control agents on the same test (i.e., DQN and R2D2 from the Acme Tensorflow library; see Hoffman et al., 2020; Toyama et al., 2021): agents are trained on 1000 randomly generated items and tested on a set of 20 unseen test items, the exact same used with human participants.

Contrasting human and artificial agents may yield one of the following main scenarii: **(A)** The artificial agent mimics the human performance profile well, suggesting it captures something fundamental about human cognition and that its study may help us better understand humans. **(B)** The artificial agent performs the task well but displays a different performance profile than humans. This could suggest that there are in fact several ways of solving the task and that the human performance profile has a characteristic computational signature. **(C)** The artificial agent performs like humans on some items but very differently on others. This may indicate that humans use a mixture of cognitive strategies or that the artificial agent needs to be augmented to perform human-like.

Whatever the case may be, it is clear that the comparison of human versus artificial agents, as well as the comparison among artificial agents provides a unique source of information that significantly augments our ability to make sense of human behavior in cognitive tests.

2.4 Conclusion

Cognitive tests play a central role in the study of human cognition. We introduced CogEnv, a framework that runs cognitive tests within a virtual environment that enables training and evaluating artificial agents in a way that is directly comparable to human studies. CogEnv also provides a way to study cognitive tests and how learning to perform well in one cognitive test might transfer to others.

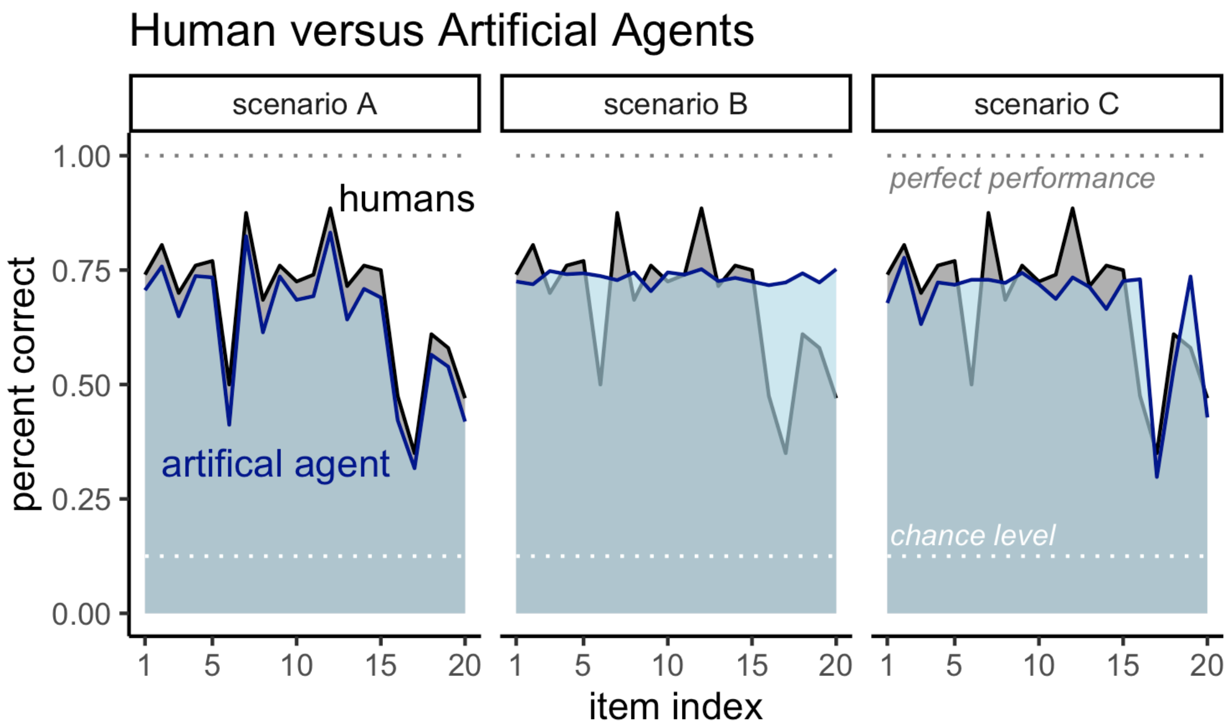


Figure 2.3: Hypothetical scenarios when comparing the performance of humans versus computational agents (see text).

Environments like CogEnv have proven quite useful in other fields, e.g., AnimalAI 3 for animal cognition (Crosby et al., 2020) and RecSym for recommendation systems (Ie et al., 2019). We believe that CogEnv can complement other approaches (e.g., cognitive architectures) and hope it will yield new insights on human cognition and help coordinate efforts across disciplines to better understand the computational foundations of cognitive performance.

Chapter 3

CogPonder: Towards a Computational Framework of General Cognitive Control

Abstract

Current computational models of cognitive control are lacking in important ways. In psychology, cognitive control models tend to be designed for specific tasks which makes it hard to study cognitive control in general (e.g., across a battery of tasks, playing video games, or in real-life activities). Computer science, on the other hand, has been able to develop artificial agents capable of performing complex tasks but typically ignores resource limitations and how long it takes for an agent to make decisions and act. Response time is of the essence in human cognition and varies meaningfully depending on numerous factors, including in particular cognitive control which supports adapting behavior to environmental constraints to achieve specific goals. Recent work further points to the fact that cognitive control models could equally greatly benefit the development of a next generation of intelligent agents in computer science. Here we propose CogPonder, a flexible, differentiable end-to-end general cognitive control framework that is inspired by the Test-Operate-Test-Exit (TOTE) architec-

ture (G. Miller et al., 1960) in psychology and by PonderNet (Banino et al., 2021) in computer science. CogPonder is a general deep learning framework that functionally decouples the act of control from the controlled decision making processes. The framework involves a controller that acts as a wrapper around any computational end-to-end model (that "perceive" the environment and generate "responses" on that environment) and controls when to stop processing and output a response (thus producing both a response and a response time). Here we implemented a simple instance of CogPonder and trained it to perform two classic cognitive control tasks (i.e., Stroop and N-back) while at the same time aligning its behavior to humans (i.e., similar responses and response times). The results show that across both tasks, CogPonder effectively learns from data to generate behavior that resembles the behavior of humans. This work thus demonstrates the value of this new computational framework of cognitive control and provides novel insights and research opportunities for both psychological and computer science.

3.1 Introduction

The scientific study of human cognition has largely focused on how long it takes people to perform tasks (e.g., press a key in response to a light, multiply two numbers or name the capital of Luxembourg) and on what factors impact those response latencies (e.g., intensity of the light, magnitude of the numbers, familiarity of the content). There is a long and rich history of research on response times and many computational models have been developed to account for response time phenomena (De Boeck & Jeon, 2019; Forstmann et al., 2016). Furthermore, the study of response times is particularly relevant because in contrast to other measures, such as percent correct or IQ, response times express a physical quantity in a ratio scale (Jensen, 2006) which allows the direct comparison of raw measurements.

An important class of response times models derives from the drift diffusion model (DDM; Ratcliff, 1978) which is specifically designed to model binary decision making. It considers both the response (what choice the person made) and the response time Ratcliff et al. (2016).

In this model, the stimulus triggers a stochastic ("noisy") signal which is accumulated until it eventually reaches an upper or lower threshold ("decision bounds")—the threshold that is reached determines the decision and the time when threshold is reached determines the response time. This type of model is appealing because it can account for a large range of behavioral data, has an intuitive computational interpretation (i.e., sequential probability ratio test) and seems to map well with neural decision-making signals (Forstmann et al., 2016; Gold & Shadlen, 2007). Furthermore, models like the DDM can be fit to behavioral data and the underlying model parameters provide useful and meaningful quantities that help better understand human cognition (e.g., the quality of the signal, people's biases for one option versus another). Indeed, with this model it becomes possible to make principled predictions about the effect of task parameters (e.g., instructions emphasizing speed versus accuracy) on behavior (e.g., decrease in both response times and accuracy) via their impact on model parameters (e.g., decrease of the decision bound parameter).

Of particular interest in this context are a family of tasks that relate to the psychological construct of cognitive control (Baggetta & Alexander, 2016). These tasks include for instance the Stroop task, Task-switching, the Go/No-go task, the Flanker tasks, and the N-back task, to name just a few. While cognitive control is a complex construct with a meaning that lacks consensus in the literature (see Chapter 1), one of its key properties is that it allows the cognitive system to regulate its processing to achieve particular outcomes (e.g., inhibit a prepotent response, maintain attentional focus), and this regulation of processes typically has a measurable impact on response times (i.e., control is effortful and takes time). Indeed, response times have long been the main variable of interest to cognitive control scientists, and computational models like the DDM have been used to capture these cognitive control effects on response times (Eisenberg et al., 2019; Pedersen et al., 2022; see e.g., Ratcliff et al., 2018).

Note that DDM is not a cognitive control model per se but rather a general two alternative decision making model. Adapting DDM to cognitive control settings would thus require additional machinery. Note also that there are computational models of cognitive control (e.g., Botvinick & Cohen, 2014) that could be coupled with DDM. However, these cognitive control models are typically custom-made for specific tasks, meaning that the model for the Stroop task cannot be readily transposed to the N-back task for example.

The models mentioned above constitute major achievements in psychology and they provide invaluable insights into the human mind. They are, however, imperfect. For instance, DDM-like models apply to a limited class of tasks. They are adequate for speeded two alternative choice tasks but not for multiple alternative choice tasks (Ratcliff et al., 2016) or tasks where the response is more complex than a choice (e.g., continuous tracking). Furthermore, these models do not in fact perform a task but instead generate data that looks like human data (i.e.,they are models of the data and not models of the cognitive processes). This is in contrast to "acting" models, like modern reinforcement learning models for instance, which may for instance receive the pixel values of images displayed on a computer screen as input and generate actions to play video games at human level performance (Mnih et al., 2015). Finally, models like the DDM are rather complex mathematical objects, without reliable closed-form solutions and are typically not differentiable. This makes it difficult to incorporate DDM in modern deep learning architectures that compute gradients to backpropagate errors and learn from data. These limitations are well-known and there are ongoing efforts to overcome them (e.g., Christie & Schrater, 2019; Rafiei & Rahnev, 2022).

In recent years there have been tremendous advances in machine learning, with computational agents learning to perform highly complex tasks better than humans (e.g., modern video games, Go, Stratego). These models are interesting because they are "acting" models and they are generic (i.e., the same model architecture can be used to learn to perform many different tasks). They are, however, also limited in important ways. First, these large models typically lack structure that would facilitate the interpretation of the underlying computations. This is in contrast to computational cognitive control models that employ an adequate *level of computational abstraction* but then lack the ability to perform complex tasks. Secondly, by and large, the machine learning community hasn't yet picked up on the concept of *cognitive control* and the idea that machine learning models could regulate themselves to adapt their computations to the level of complexity of the task to be performed or the amount of available resources (Moskovitz et al., 2022; Shenhav et al., 2017). A notable exception here is PonderNet (Banino et al., 2021) which we describe below. Finally, and related to the previous point, in contrast to researchers in psychology, researchers in machine learning have largely ignored response times, not only as a metric of interest (the time needed for a given, standard neural network to make a decision does not vary with the complexity of the task or the quality of the input; it depends only on the structure of the network), but also as a behavioral constraint for the artificial agent. There are many situations that require people to stop deliberating and commit to a decision. In RL models, it is common to place the agent in a sort of turn-based environment where its world stops, waiting for the agent to act (Ramstedt & Pal, 2019). Artificial agents that could control how long they deliberate would be able to adapt to changing environmental constraints.

To summarize, computational models in psychology and in computer science have different strengths and weaknesses. There could be great benefits for both fields to cross fertilize ideas and develop new types of computational control models. The work presented here is an attempt to move in that direction.

3.2 Desiderata for a general computational cognitive control framework

Our goal is to develop a computational framework for cognitive control models that would be valuable to both psychology and machine learning researchers and which combines the strength of their respective approaches. More specifically, we want our framework to have the following main features:

- **agency**: the model is able to perform the task at hand;
- **completeness**: the model accounts for both responses and response times;
- **versatility**: the same model can perform a wide range of tasks; this allows the study of performance across multiple tasks under a common computational framework;

- **modularity**: the model allows to augment any end-to-end computational model with cognitive control abilities; this allows both for great flexibility in model architectures and interpretability.
- **learnability**: the model is differentiable and can thus be integrated in state-of-the-art deep learning models and benefit from modern software (e.g., PyTorch, TensorFlow, or JAX) that use automatic differentiation for parameter optimization and GPUs for faster computing.
- **composition**: the model forms a building block of sorts and multiple such building blocks may be arranged in structures (e.g., sequence, hierarchy) to perform complex tasks; this allows for scalability while controlling complexity.

The inspiration for our model comes from two primary sources; PonderNet from machine learning and TOTE from psychology. The following is a description of both before we describe our framework named CogPonder.

3.2.1 PonderNet

PonderNet is a recently developed algorithm that adjusts the complexity of the computations executed by a neural network as a function of the complexity of the task and the input (Banino et al., 2021). With PonderNet, the same neural network uses fewer computational steps to perform simple tasks than complex ones. The rationale behind PonderNet is straightforward. In addition to learning to perform a specific task (using a reconstruction loss function), the network evaluates at each time step whether to stop or continue computation. This halting behavior is determined by learning a halting probability distribution that is constrained by a hyperparameter (a temporal regularization term encouraging fewer computation steps while exploring other possibilities). This approach is in stark contrast to traditional machine learning approaches where the complexity of the neural networks is determined by the size of the input, adjusted manually and set once and for all for a specific task.

PonderNet is interesting within the context of cognitive control. First, because PonderNet adjusts computational resources of a system based on the complexity of the task to be solved, it can be seen as a form of cognitive control. Second, by controlling the halting distribution, PonderNet highlights the conceptual importance of considering the time needed to perform a task (more exactly the number of computational steps). By doing so, PonderNet creates a bridge between the rich literature in experimental psychology grounded in the study of response times and the booming field of deep learning.

3.2.2 TOTE

PonderNet is reminiscent of the famous cognitive control model named TOTE (G. Miller et al., 1960), where TOTE stands for Test-Operate-Test-Exit. In TOTE, as in PonderNet, computations (or operations) unfold in cycles with tests evaluating on each cycle if a specific condition is met and consequently deciding whether to exit (halt) the process or trigger a new cycle of operations. As in PonderNet, the control mechanisms are functionally separated from the operators. Interestingly, the main motivation behind the TOTE model was to address complex human behavior. While this might be achieved with PonderNet by increasing the complexity of the underlying operator, in TOTE, the authors argue that complex behaviors could be modeled by organizing multiple TOTE units in sequences, hierarchies or other structures. Under this view, TOTE units are computational building blocks that can be assembled to generate complex behaviors. With the advent of modern computers and computational tools it is now possible to translate the ideas behind TOTE in computational models capable of performing complex tasks.

3.3 The CogPonder framework

The general idea behind the CogPonder framework is illustrated in Figure 3.1. The starting point for a CogPonder model instance is an end-to-end model, termed “Operator”, which on a given trial n takes an input X_n and outputs y_n (see panel “a” in Figure 2.1). This operator may for example be a deep neural network performing the Stroop task, in which

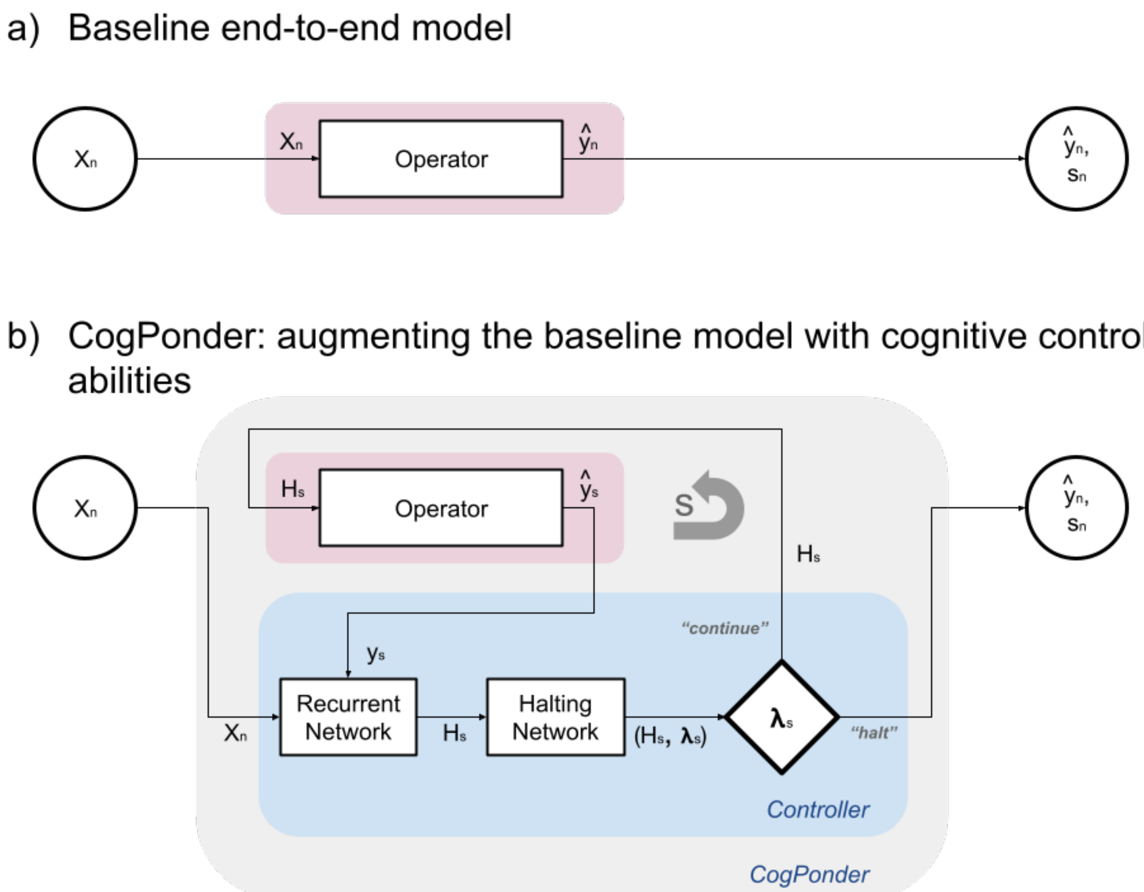


Figure 3.1: The CogPonder framework. (A) An end-to-end model, termed "Operator", which on a given trial n takes an input X_n and outputs y_n. (B) CogPonder disconnects the Operator from its direct inputs and outputs and encapsulates the Operator inside a local virtual environment that is governed by the Controller (blue box). The Controller intercept both the inputs and outputs of the Operator, and determines what inputs are fed to the Operator and ultimately what output to be emitted on a given trial. Within a given trial n the Controller will repeatedly call the Operator, with each of these iterations being indexed by step s, until it decides to halt processing for trial n and to emit a response y_n. The halting is determined by a sample from a Bernoulli distribution parameterized by λ_s (decision diamond in the figure).

case X might be a textual description of the stimulus or the pixel values of the screen and y might be a label of a color or a motor command to press a specific button.

The key idea behind CogPonder is to disconnect the Operator from its direct inputs and outputs and to encapsulate the Operator inside a local virtual environment that is governed by the Controller. The Controller intercept both the inputs and outputs of the operator, and determines what inputs are fed to the Operator and ultimately what output to be emitted on a given trial. There are many possible ways to implement the Controller, and different types of control the Controller could exert on the Operator. Here we consider the Operator as a blackbox (i.e., the Controller has no read or write access to the Operators internal parameters) and use a formulation that is very similar to PonderNet (possible extensions are discussed in "Limitations and possible future extensions"). More specifically, within a given trial n the Controller will repeatedly call the Operator, with each of these iterations being indexed by step s, until it decides to halt processing for trial n and to emit a response y_n. The number of iterations performed on trial n, s_n, reflects the response time for that trial.

Following PonderNet, the decision to "halt" or to "continue" iterating at step s is determined by a Bernoulli random variable Λ_s, with $\Lambda_s = 1$ meaning "halt" and $\Lambda_s = 0$ meaning "continue". The conditional probability of halting at step s, given that the process was not halted in the previous step is given by:

$$P(\Lambda_s = 1 | \Lambda_{s-1} = 0) = \lambda_s \qquad \forall\, 1 \leq s \leq S$$

where S is the maximum number of steps allowed before halting.

From this expression one can compute the unconditioned probability of halting at step s:

$$p_s = \lambda_s \prod_{j=1}^{s-1} (1 - \lambda_j)$$

Importantly, the value of λ_s is computed by the Controller at each timestep s, endowing it with the power to adjust the system's computational complexity and determining its response time distribution p_s. The other major quantity that the Controller needs to compute on each iteration is H_s, the input to the Operator (using X_n and H_{s-1}). Because λ_s is computed from H_s (see Figure 3.1), speed and accuracy are intrinsically coupled at the within-trial level. The Controller can be instantiated using neural networks and its parameters adjusted using standard methods and labeled data (see "Evaluation of a CogPonder model").

It is important to note that CogPonder is conceptually quite different from RTNet (Rafiei & Rahnev, 2022). In RTNet, the same input is passed multiple times through a neural network with each pass using slightly different weights (i.e., weights are not fixed but sampled from a distribution) and the output of each pass is accumulated in a special output layer until reaching a decision threshold (similar to DDM). In CogPonder a given model is wrapped by a controller, the model is iteratively fed different inputs (they are generated by the Controller) and the response time (number of computational steps) is determined by computational requirements rather than resulting from stochasticity that is injected in the system.

CogPonder is very similar but also different from PonderNet in the sense that CogPonder aims to align computational models with human behavior rather than adjusting computational resources of neural networks to the complexity of a particular task. CogPonder also aims to embrace the "building blocks" metaphor of TOTE and further our understanding of cognitive control (i.e., it aims to become a theoretical framework and not "only" a method).

3.4 Evaluation of a CogPonder model

3.4.1 Objectives and rationale

This work aims to be a proof of concept, demonstrating the value of CogPonder to both psychology and computer science research. The preliminary work presented below has two main objectives: demonstrate that the same CogPonder model instance can learn to perform two different cognitive control tasks from cognitive psychology; this is important because it shows tasks that have so far mostly been considered in isolation can now be investigated within a common computational framework. demonstrate that the behavior of a CogPonder model aligned to human behavior is able to capture important patterns in the human data; this is important because it shows that CogPonder might be useful to understand behavior and might also be used to run simulation ("what if") experiments.

3.4.2 Dataset

Here we use a subset of the Self-Regulation Ontology dataset (publicly available and previously published in Eisenberg et al., 2019) which contains behavioral data from 521 of participants who completed computerized cognitive tests as well as questionnaires. In this study we consider only data from one human participant who completed two cognitive tasks: the Stroop test and the 2-back test. We chose these specific tasks because they have both been associated with the construct of cognitive control but are quite different in that they involve different types of stimuli (words versus letters), task instructions (name ink versus same/different), cognitive processes (involving the inhibition of a prepotent response versus updating memory) and responses options (2 versus 3 options).

In the Stroop task, participants were presented with a name of a color written in ink that was either congruent or incongruent with the word (e.g., the text "red" written in a blue color is incongruent, while the text "red" written in a red color is congruent) and they were instructed to report quickly and accurately the color of the ink (i.e., ignore the text) by pressing one of three keys (corresponding to the options red, green, blue). Each participant completed 24 practice trials and 96 test trials; here we consider only test trials.

In the N-back task, participants were presented with a stream of letters (e.g., "A", "X", "a") and they were instructed to report for each letter whether it was the same letter as the one presented N letters ago (irrespective of capitalization) by pressing one of two keys corresponding to "same letter" (i.e., target) and "different letter" (i.e., non-target). Each participant completed several versions of the N-back task; here we consider only the cases where N=2 (i.e., 2-back trials), which amounted to 342 trials.

In both tasks, we use the trial-level data for participants which includes a description of the stimulus (e.g., “A”), trial index, the participants response (e.g., the choice of the response option “red”) and time needed to make that response (i.e., response time, in milliseconds). For more details on the original datasets, see Eisenberg et al. (2019).

3.4.3 Method

Our goal is to train the same computational cognitive control model (i.e., “agent”) to perform both the Stroop and the 2-back tasks. In both cases, the model will receive as input a sequence of stimuli (i.e., color words or letters) and will generate a response to each stimulus (i.e., color words or same/different). Note that this is an “acting” model that is actually able to perform the task and not a “fitting” model that aims to fit patterns in the data. Note also that by responding to each stimulus, the data generated by the agent will have the same structure as the human data (i.e., trial-level data with a stimulus description, the choice made by the agent and the time it took the agent to make that decision).

In addition to training the agent to accurately perform the task, we want to *align* the agent with humans. By this we mean that we want to adjust the internal parameters of the computational model so that it will generate a behavior in response to stimuli that is similar to human behavior (e.g., similar response time distributions and accuracy levels).

This alignment is obtained by the following loss function, the value of which will be minimized during the training phase of the model (see “Model evaluation procedure”):

$$L_{\text{total}} = L_{\text{response}} + \beta L_{\text{time}} \tag{3.1}$$

This loss function comprises two terms which are weighted by the hyperparameter β. The first term aligns the agents choices with the choices made by human participants (“response reconstruction loss”):

$$L_{\text{response}} = \sum_{s=1}^{S} \mathcal{L}(\hat{y}_s, y) p_s \tag{3.2}$$

where $\mathcal{L}$ represents the cross entropy loss function.

The second term aligns the agent's response times (the distribution of the halting probability p_s) with the response times distribution of human participants d using KL divergence ("time regularization term"):

$$L_{\text{time}} = KL(p_s || d) \tag{3.3}$$

It is important to note that computers typically perform tasks much faster than humans do and that depending on the specific computer hardware (or software), the time needed to respond may vary considerably. This means that elapsed computation time is not the relevant variable to track and that we should instead track the number of computational steps (Cormen et al., 2022). In a given computational context (e.g., a particular task and performance constraint) this number may be stable despite the time needed to execute those steps varying significantly depending on the underlying hardware.

Equation 3.3 (L_{time}) requires computing the similarly (via KL divergence) between the distribution of halting times, which are expressed in number of steps, and participants response times distributions, which are expressed in milliseconds in our dataset. To compute this term it is necessary to either convert number of steps into milliseconds (e.g., using a hyperparameter that expresses the duration per step) or to convert the response times from milliseconds to number of steps (e.g., using a hyperparameter that expresses duration per step and dividing the response time by that duration). We used the second approach and manually determined an adequate value for the step duration hyperparameter (see "Model evaluation procedure").

3.4.4 Model evaluation procedure

3.4.4.1 Model architecture: CogPonder instantiation

Figure 3.1 (panel B) describes the general template for a CogPonder model. CogPonder is a framework that can be instantiated in many different ways. Here we chose a specific implementation to perform the Stroop and N-back tasks, noting nevertheless that other implementations are equally valid and that for other tasks more complex instantiations might be needed. Our goal is to demonstrate the value of the framework, not the value of this specific instantiation of the framework.

For the Operator in the model (see panel B in Figure 3.1) we used a simple neural network with one dense linear layer and ReLU activation. The Controller includes two separate networks: a recurrent network and a halting network. The recurrent network is a GRUCell that iteratively computes inputs to the Operator. At each iteration s it computes H_s and serves it as the input to the Operator. The halting network approximates the probability of halting at each time step ($lambda_s$). It is a fully connected linear layer with ReLU activation that receives as input H_s and determines the halting of the CogPonder model at a given time point s within a trial and the emission of the output for that trial.

Finally, the decision to halt or to continue processing is made at each processing step s within a given trial based on a biased coin flip (Bernoulli sample with probability of λ_s), which is emitted by the halting network (see panel B in Figure 3.1).

Note that the same model architecture was used to fit the Stroop and N-back tasks (separately) but there were slight differences between these two cases because the stimuli and responses are different in the two tasks. More specifically, a stimulus in the Stroop task is encoded using 2 inputs (color and word), while a stimulus in the N-back task requires 6 inputs (one-hot encoded letters). Similarly, in the Stroop task, the network needs to emit one of 3 choices while in the N-back only one of two choices. This being said, it is straightforward to extend these models so the exact same model architecture could apply to both cases.

3.4.4.2 Model training

Here we present preliminary work to align CogPonder to human data. CogPonder was fit to a single participant taken at random from the dataset and separately for the Stroop and the N-back tasks (i.e., different sets of parameters were adjusted for each task). Participants' data in each task represents a time series (i.e., trials are ordered and there is a dependency across trials). This data was split into 75% training set and 25% test set, corresponding to 72 train and 24 test trials in the Stroop task and 256 train and 86 test trials in the N-back task.

The training involved a maximum of 10000 epochs (i.e., loops over the dataset) which was stopped when no improvement was observed in minimizing total validation loss (early stopping with 0.01 patience on the validation L_{total}). We used stochastic gradient descent (Adam optimizer) to minimize L_{total} (see loss function in Equation 3.1). All model parameters within the Operator and Controller were adjusted simultaneously and using the same procedure with the exception of the step duration hyperparameter which for this preliminary analysis was set manually to 20ms. In total, 62 parameters were adjusted for the Stroop task and 239 parameters for the N-back task and it takes around 15 minutes to fit one participant on one task on an average laptop.

The evaluation of the model used the 25% of trials that were not used for training. Once the model parameters are set, the model can be used to generate behavior (i.e., responses and response times) in response to stimulus sequences. This artificial agent generated behavior can then be compared with human generated behavior using standard descriptive statistics such as average accuracy and average response time for example.

3.4.4.3 Step duration hyperparameter

As a first approximation we manually tested several values (10ms, 20ms, 50ms, 100ms) and selected the value of 20ms as this seemed to lead to the best alignment with human data and faster convergence of the model parameters. In a future iteration of this analysis, this hyperparameter will be estimated directly from the data using a dedicated validation set.

3.4.4.4 Non-decision time hyperparameter

In line with past computational models in psychology, we included in our model a non-decision time which reflects the sum of durations that affect the measured response time but are not related to the decision process per se (e.g., the time taken for light to be converted to action potentials in the retina). We assume that this non-decision time is approximately the fastest possible human response time for a given task. Thus, to remove this non-decision time from the recorded response times we subtracted the minimum response time from all data points, which resulted in response times being expressed in time steps ranging from 1 to $max(RT) - min(RT) + 1$. Compared to the raw response times, using these transformed response times resulted in faster convergence of the model parameters. In a future iteration of the analysis, non-decision time will be treated as a hyperparameter and estimated from the data.

3.5 Results

Our first goal is to determine if the same CogPonder model can learn to perform two different tasks using data from one human participant. Figure 3.2 shows the total loss (L_{total}, as defined in Equation 3.1) computed on the test data as a function of the number of epochs during the training phase. It is apparent from this figure that CogPonder does indeed learn in both tasks, with the loss reaching an asymptote after about 100 epochs (i.e., iteration through the training dataset). This figure also demonstrates that because of its design, CogPonder (like PonderNet) can take advantage of modern deep learning software to efficiently fit complex models.

Our second goal is to determine to what extent a CogPonder model *acts* like a human once it has been trained with human data. Because CogPonder is an acting model it generates trial-by-trial responses that have the same data shape and type as human responses. This

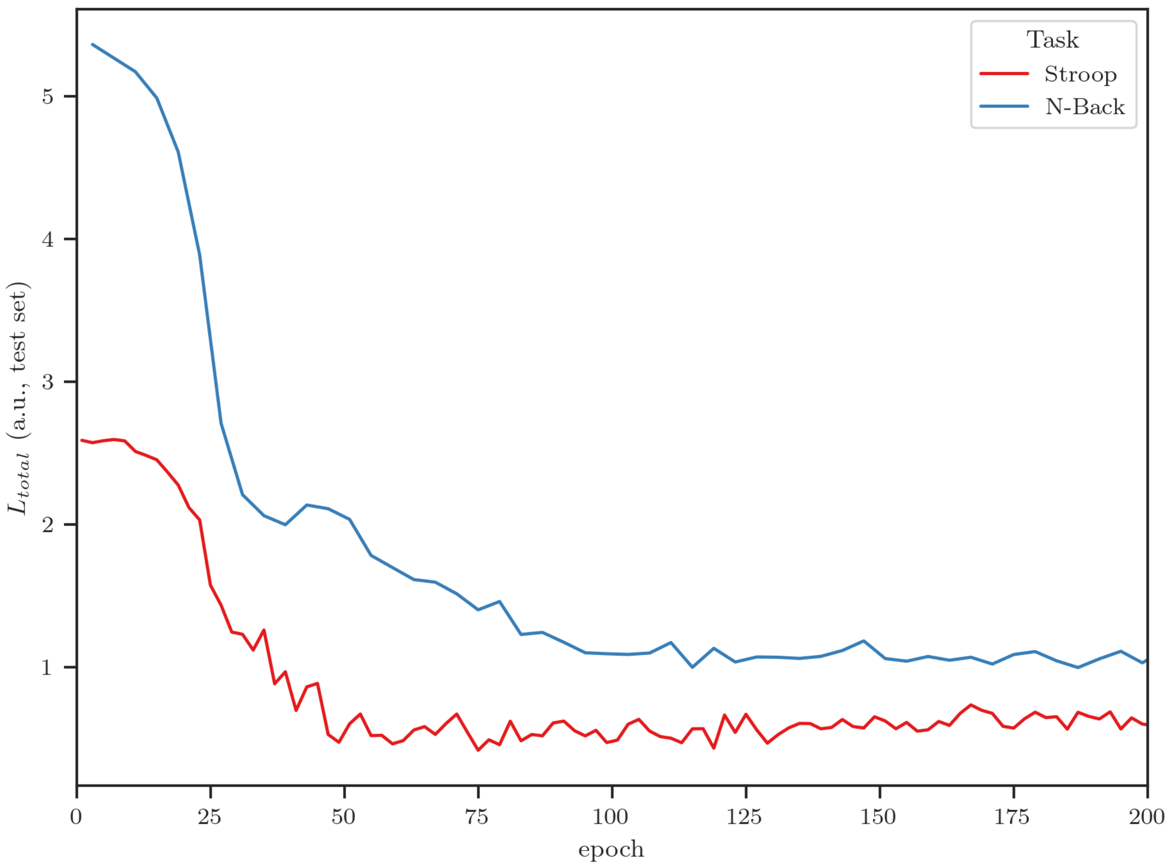


Figure 3.2: CogPonder learns to behave like humans. With increasing learning iteration (epochs) the loss decreases and asymptotes. This is true both when aligning CogPonder with the Stroop task (red curve) or with the N-back task (blue curve). Note that the two tasks were trained and tested separately.

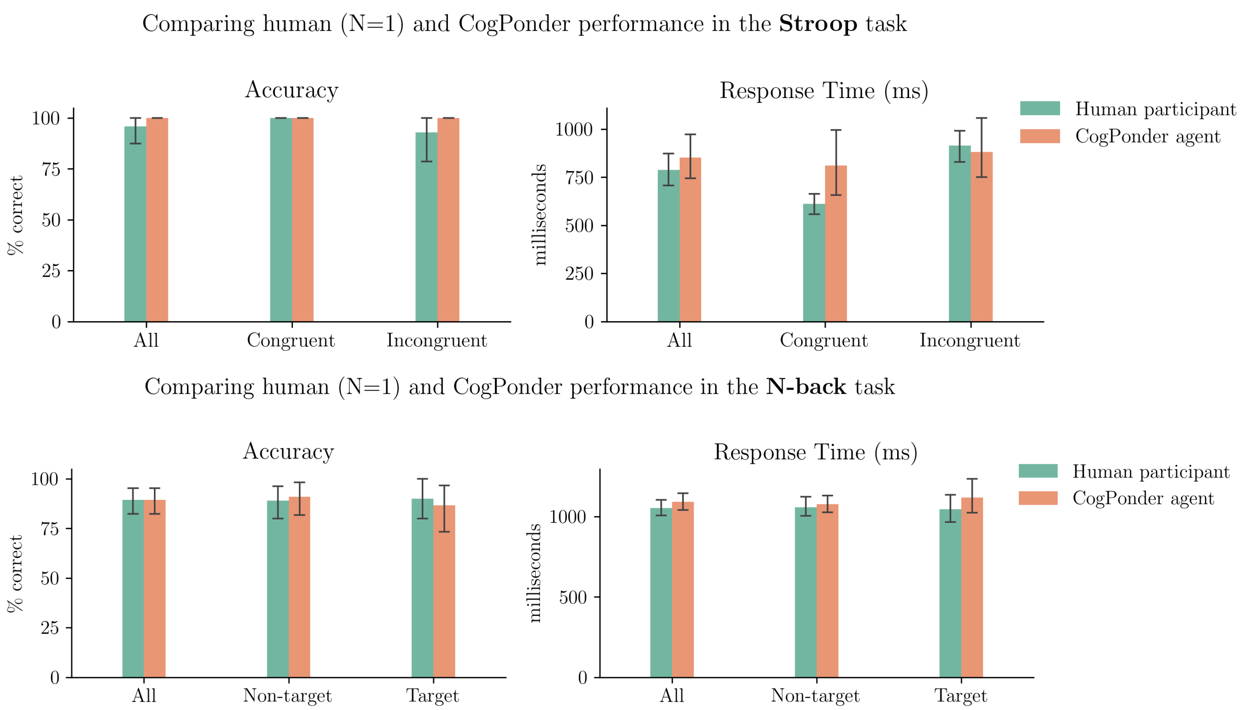
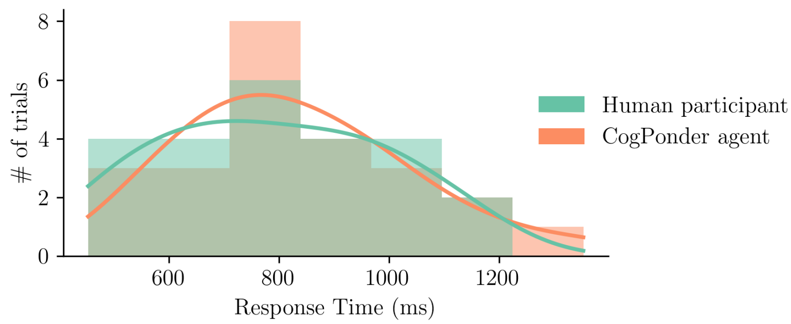


Figure 3.3: CogPonder behavior is comparable to human behavior. CogPonder captures the overall pattern of average accuracy (left column of panels) and average response times (right column of panels) in both the Stroop task (upper row of panels) and in the N-back task (bottom row of panels) when grouping all types of trials ("All"). However, when separating trials by type ("congruent" and "incongruent" in the the Stroop task and "target" and "non-target" in the N-back task), some discrepancies are observed. Error bars show 95% confidence intervals.

Comparing human (N=1) and CogPonder response time distributions in the **Stroop** task



Comparing human (N=1) and CogPonder response time distributions in the **N-back** task

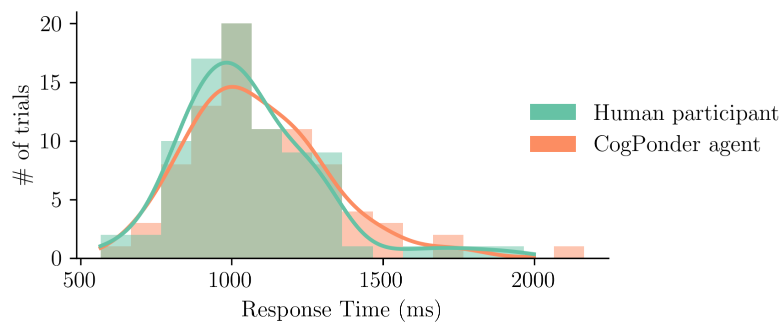


Figure 3.4: CogPonder also mimics finer grained phenomena (e.g., response time distributions).

allows to directly compare the behavior of CogPonder and human agents using the same descriptive statistics and data visualization code. As a first step, we compare the average accuracy and average response time of a human versus CogPonder agent in both the Stroop and N-back tasks (see Figure 3.3). It is apparent from Figure 3.3 that CogPonder is able to capture these broad patterns in the human data. In particular, for both the Stroop and N-back tasks, CogPonder produces a behavior with accuracy levels and response speeds that are in the same ballpark as human data when considering all types of trials ("All" label in the x-axis of Figure 3.3).

Next, we investigated to what extent CogPonder was able to reproduce finer grained human phenomena. To do so, we plotted average accuracy and average response time as a function of conditions (see Figure 3.3), as well as the distribution of response times for both the human and the CogPonder agent, separately for the Stroop and N-back tasks (see Figure 3.4). The "fits" are obviously not perfect. For example, while the human data shows a congruency effect in the Stroop task, whereby accuracy is lower and response times longer in incongruent trials than in congruent trials, no such effects are apparent in the CogPonder data. One should note however that the error bars are quite large and that it remains plausible that with a larger training dataset, CogPonder will be able to capture these Stroop effects. What is most encouraging is these results is the similarity between the response time distributions of the human participant and the CogPonder agent in both the Stroop and N-back tasks. Overall, these results suggest that CogPonder is able to mimic important markers of human behavior, which makes CogPonder a promising new approach to the study of human cognition.

3.6 Discussion

The present work is a first step towards developing CogPonder, a computational cognitive control framework that can be applied to a broad range of use-cases—including in particular batteries of cognitive tests. In this framework, cognitive control is envisioned as a model that wraps around any end-to-end operator model and controls both its inputs and outputs to achieve a desired performance profile. In this work we focused in particular on two classic experimental psychological tests (the Stroop and the N-back tests) and showed that a basic instance of CogPonder can be trained to align with human behavior and will then generate behavior that captures some key patterns in the human data, in particular average accuracy and response time as well as response time distributions. While these results are still preliminary and more work is needed to fully explore the capabilities of CogPonder, this work constitutes a proof of concepts and speaks for the value of the CogPonder framework.

CogPonder is unique in that it satisfies a number of important desiderata that are only partially satisfied by current models in cognitive sciences. First of all, CogPonder has *agency*—meaning it is an architecture that is able to perform tasks (e.g., make timed decisions when faced with particular stimuli). This is in contrast to models that focus on describing the structure of the data.

Second, CogPonder is *complete* in the sense that its behavior can have all the same dimensions as human behavior. This is in contrast to models that account only for the choices made by an agent but not their response times.

Third, and most importantly, CogPonder is *versatile* in the sense that it can in principle perform a wide range of tasks. In the present study, we focused only on two tasks but there is no reason this framework cannot account for a much broader range of tasks. This is in contrast to models that are tailored for individual tasks and limit our ability to use the model to understand cognitive control in general (i.e., across many tasks).

Fourth, the model is *modular* in the sense that different control models may be used to wrap any type of end-to-end model. This feature is important because it allows the development of models that are both flexible (i.e., can adapt to a large range of use cases), while at the same time offering interpretability (i.e., it's clear which effects can be attributed to the controller versus the operator). Fifth, CogPonder is *learnable* in the sense that the controller model is differentiable and can thus be incorporated into modern deep learning software that is highly effective to train large models on big datasets. This feature of CogPonder facilitates the use of CogPonder in practice, compared for example to models that require custom made code and fitting procedure. Finally, we believe, but haven't yet shown, that CogPonder allows for model *composition*. By this term we mean that CogPonder can be seen as a building block that models a local aspect of cognitive control and multiple CogPonder units may be chained or organized into hierarchical structures in order to achieve highly complex behavior while limiting the complexity of the overall computational model.

3.7 Implications

The present study shows that CogPonder can be applied to multiple tasks and is able to account for both responses and response times.

The implication for psychology is that CogPonder now offers new opportunities to study behavior and in particular cognitive control across a large range of tasks (e.g, beyond the Stroop test, beyond the two alternative choice family of cognitive tasks) using a common framework. This is important as it provides a common theoretical and computational ground to investigate human behavior. There are, in particular, two use-cases where we believe CogPonder will be particularly useful. The first use-case relates to simulations and the ability for CogPonder models to run "what if" experiments. More specifically, if we have computational models that can account for multiple cognitive tasks, one could use these cognitive models to develop new cognitive tasks that may be more diagnostic of certain model parameters or may help discriminate between competing computational models. The second use-case relates to cognitive training and transfer. There is currently a lack of quantitative theories that would allow one to predict how one person would perform a new task (given some historical data about that person), nor how exactly cognitive training would transfer to which other tasks and how much exactly performance should improve on those tasks. Multitask computational models of cognition are necessary to understand transfer and CogPonder is one way to develop such models.

Finally, it is also important to note that current models in computational psychology focus on modeling tasks that are relatively simple (e.g., the Stroop test) and are inadequate to model more complex human behavior (e.g., video game play). It is not obvious how models developed for the simpler tasks could be extended to grasp more of the complexity of human behavior. This is not the case for CogPonder. Because of its properties, it is rather conceptually straightforward to expand CogPonder to develop agents able to perform any task modern AI is able to solve. Thus an important achievement of CogPonder is its ability to break a "complexity of behavior ceiling" relative to existing approaches.

The present work also has numerous implications in computer science. As explained earlier, most current models in AI (i.e., deep learning, RL) have not yet caught up on the importance of response times and cognitive control as valuable modeling concepts. Currently, the focus in these fields is mostly on developing models that are able to perform difficult tasks with the highest possible level of accuracy, irrespective of the computational resources (for training and computation) and training data needed to achieve those accuracy levels. This strategy is clearly valuable and is quickly pushing the boundaries of AI. However, there is obviously the need to also develop computational models that can adjust their internal complexity to the complexity of a task to be solved (cf. PonderNet) and to the fluctuating demands of the environment. An artificial agent, acting and learning in the world, may not have the luxury of quasi infinite resources and unlimited time to act and may instead have to commit to quick, albeit less accurate decisions, the same way humans do. The CogPonder framework provides a principled way to extend modern end-to-end models developed with a focus on maximizing accuracy in a way that allows for graded, time-sensitive, and adaptive computation. Finally, AI aims to develop agents that are able to perform highly complex tasks (e.g., making pizza). A major challenge in this context is to control complexity so that models can be effectively trained using a reasonable amount of data. We believe that CogPonder, and in particular its potential for composition, may provide an interesting solution to this problem.

3.8 Limitations and future extensions

The current work is a proof of concept, and as such, it has obvious limitations that future work will address. First, there are improvements that can be made to implementation of the CogPonder model and its evaluation. For example, in the above work we set some hyperparameters manually instead of learning their values from data. Second, we trained the model on only one participant's data. In future iterations we will align the model to a larger set of participants and evaluate to what extent the CogPonder can capture inter-individual differences. Applying CogPonder to groups of participants may also require rethinking the CogPonder training procedure to allow for hierarchical as well as shared model parameters across participants. Third, we only tested two cognitive tasks, the Stroop and the N-back task, and only performed limited descriptive analysis to compare human and agent data on those tasks. In future work, we will more systematically explore CogPonder's ability to perform cognitive tests and develop finer grained analyses to assess its behavior. In particular, we aim to integrate CogPonder in the CogEnv virtual cognitive task environment (see Chapter 2) and develop automated data analysis pipelines that apply to both human and artificial data. Finally, although we showed that the same CogPonder model can be trained to perform different tasks, we have not yet investigated the relationships between those two trained model instance (e.g., are model parameters similar across the two tasks) nor have we trained a model to jointly perform both tasks (e.g., by including a task description as an input to the system). These steps seem crucial to assess the value of CogPonder as a theoretical model for cognitive control in psychology.

Although CogPonder is already a very flexible framework, there are several ways in which it could be further be extended, both inwards (i.e., changing the mechanics of CogPonder) and outwards (i.e., changing how CogPonder interfaces with other modules). In the current work, the Controller controls only one Operator. In more advanced versions, CogPonder could encapsulate and orchestrate multiple, perhaps competing Operators in parallel. Furthermore, in the current work, the Operator is conceived as a black box—a module that could be imported as is, without having to expose its internal workings and parameters. This is an interesting property from an engineering point of view as it clearly separates the development and testing of Operator models from the development and testing of Controller models. If, however, the Controller has reading and writing rights to the internals of the Operator, the Controller could be endowed with much greater control abilities (e.g., set or reset model weights, learn to continuously predict accuracy of the Operator based on the values of its internal parameters). Also, in the current work, the Controller focuses only on the current trial and on learning what inputs to provide to the Operator to achieve a desired outcome. But there are other roles that the Controller could play. For instance, the Controller could have a much more active role in the training of the Operator. This could be achieved for example by controlling the learning rate of the Operator but also by controlling what data to use for learning. For example, CogPonder could maintain an internal dataset—using historic ("episodic memory") or synthetic data (e.g., generated from a time-consuming process that the system aims to automate)—and train the Operator to perform well on that dataset. This type of mechanism would allow for offline ("replay") learning, and could be useful to achieve overall better performance with fewer new observations.

In addition to extensions that could be envisioned for the inner workings of CogPonder, there are also extensions in line with the "building-blocks" view of the TOTE model that might be worth investigating further. In the current implementation, CogPonder receives as input the stimulus description and outputs the response. It would make sense however to consider CogPonder as a piece of a larger system rather than the system as whole. Even in the case of simple response times, computational models in psychology have argued for the need to model not only the decision process but also other processes involved in the task (including for example, the transduction of photons to action potentials in the retina, the transmission of signals from the retina to the visual cortex, and the transmission of action potentials from the motor cortex to skeletal muscles)—these processes are typically lumped together and modeled as a non-decision process whose duration is added to the decision time to form the response time. In addition to providing more detailed accounts of simple tasks, composing

CogPonder networks into more complex neural architectures may provide a means to model planning and performance in complex sequential tasks. This is a key idea of TOTE: by organizing relatively simple TOTE building blocks into hierarchies and sequences it becomes possible to orchestrate and control complex sequences of behavior, such as preparing a pizza for example. CogPonder provides a principled way to build and train those building blocks; but much work is still needed to evaluate what exactly can be construed with them—we hope this preliminary work on CogPonder ignites interest in these exciting new lines of research, both in psychology and computer science.

Chapter 4

Training Cognition with Video Games

Pedro Cardoso-Leite, Morteza Ansarinia, Emmanuel Schmück, and Daphne Bavelier

Abstract

This chapter reviews the behavioral and neuroimaging scientific literature on the cognitive consequences of playing various genres of video games. The available research highlights that not all video games have similar cognitive impact; action video games as defined by first and third person shooter games have been associated with greater cognitive enhancement, especially when it comes to top-down attention, than puzzle or life-simulation games.

This state of affair suggests specific game mechanics need to be embodied in a video game for it to enhance cognition. These hypothesized game mechanics are reviewed; yet, we note that the advent of more complex, hybrid video games poses new research challenges and call for a more systematic assessment of how specific video game mechanics relate to cognitive enhancement.

4.1 Introduction

Across all ages, cognitive abilities play an important role in our quality of life and the type of life we lead. At young ages, executive functions, a cornerstone of cognitive abilities, are thought to determine educational achievement (Bull et al., 2008; e.g., Diamond, 2013; Geary et al., 2019; Goldin et al., 2014) and more generally to be "critical for success in school and life" (Diamond et al., 2007). Longitudinal studies in young children, for example report that cognitive abilities predict educational achievement attained two years later (Bull et al., 2008; Gathercole et al., 2004). Among executive functions, attentional control abilities have been of special interest as they mediate a various array of skills, from sustained attention in school to divided attention in team sports. In older adults, for example, attentional abilities correlate with driving accidents—the shrinkage in a persons' useful field of view, which is the spatial extent of their visual field to which they effectively pay attention, is strongly associated with a higher incidence of car accidents prior to the attentional test (Ball et al., 1993). The central role cognitive abilities play in our lives has led to many attempts to devise behavioral training programs to improve cognition, and in particular executive functions (Katz et al., 2018). While cognitive enhancement raises ethical concerns (similar to doping in sports), it also holds the promise for broad societal benefits (Bavelier & Green, 2019).

Numerous forms of cognitive training exist; yet, their efficiency and the underlying causal mechanisms remain controversial. This is the case, for example, of interventions attempting to improve fluid intelligence by training executive functions (Au et al., 2015; e.g., Jaeggi et al., 2008; Melby-Lervåg & Hulme, 2013). A key concern in the cognitive training literature is that training of specialized cognitive functions may lead to improvements in only those trained functions (i.e., "near transfer") and may not transfer to a broader range of tasks and situations (i.e., "far transfer"). While the necessary conditions for far transfer remain to be firmly established, variety in the training regime and the trained functions appear to be key factors (for an example in the domain of sports, see Güllich, 2018). An alternative perspective on the plasticity of cognitive abilities is to focus, not on targeted interventions designed by researchers, but rather to consider the impact of changes in our environment. The Flynn effect, or the rise of IQ scores through the 20th century, is one such example. With the advent of digital media, our lifestyle and cognitive activities—starting at the youngest ages—have radically changed over the past decades. For example, it has been argued that the excessive consumption of multiple media at the same time (e.g., texting while watching TV and browsing the web) may cause an attentional impairment in filtering out distraction (Ophir et al., 2009); although more recent data are less clear cut (Uncapher & Wagner, 2018; for reviews, see Wiradhany & Nieuwenstein, 2017). Whether those media-based environmental changes are for the better or for the worse remains highly debated (Bavelier et al., 2010; e.g., Ophir et al., 2009; Sparrow et al., 2011). Yet, investigating those effects holds the promise of bringing new insights into human brain plasticity and cognitive training.

Digital media occupy an increasingly large portion of our waking time. In the US, 8-12 year olds spend close to 6 hours on media each day (Rideout, 2016)—with similar trends being reported all over the world (e.g., Bodson, 2017; Waller et al., 2016). Digital media affect every aspect of our lives; these effects are complex and not fully understood yet (Bavelier et al., 2010). They depend not only on the specific medium being used but also how they are consumed and what content they deliver (e.g., Cardoso-Leite et al., 2016). Here we limit our scope by focusing on the effects of playing video games on cognition. This choice is motivated by three main points: (i) while the field of media and cognition is quite young, it is already clear that not all media use have the same impact on cognition implying that different media uses need to be considered separately (as stated earlier, media multitasking may be related to attentional deficits, while playing specific video games have instead be linked to attentional improvements); (ii) video games stand on their own by immersing players in extremely rich and complex experiences with high cognitive demands (a person watching television may spend hours without performing any significant action, while people playing video games may perform multiple meaningful decisions and actions per second); (iii) and finally, the literature concerning the impact of video game play on behavior, including cognition, is arguably one of the best documented today. We will focus only on the relationship between video game play and cognition and will not consider other aspects that might be equally important but are outside the scope of this work, such as the impact of violence, self-image, well-being, creativity, social functioning, or addiction (for reviews on such topics see, Gentile et al., 2017; Király et al., 2017; Stanhope et al., 2015).

Almost everyone plays video games now. Although the term video game raises the stereotypical image of the adolescent glued to his screen, there are now as many females, 50 or older, playing video games as there are boys under 18 playing video games. Interestingly, these two groups do not engage with the same genres of video games; older females mostly play puzzle or casual games, while boys play predominantly action-packed, role-playing games. This state of affairs highlights the need to pay close attention to video game genre or the type of experience different video games deliver. In 2015, both "tweens" (8-12 years old) and teens (13-18 year olds) in the US devoted on average about 1 hour and 20 minutes to playing video games each day; with boys playing substantially more than girls (Rideout, 2015). The relationship between video game play and cognition has been investigated in various large-scale correlation studies that collect data about children's gaming habits and various measures of interest (Adachi & Willoughby, 2013; Kovess-Masfety et al., 2016; Stanhope et al., 2015). One such study, conducted in Europe, reported that video game play was associated with enhanced intellectual, social and academic functioning (as rated by the child's teacher; Kovess-Masfety et al., 2016). Another associated gaming in 7-11 years old with faster response speeds, enhanced sustained attention and academic performance; but only for intermediate amounts of video game play per day (Pujol et al., 2016). A recent study on 3 to 7-year-old children furthermore documents that casual video game play at this young age may increase fluid intelligence (Fikkers et al., 2019). It thus appears that, at a macro-level, playing video games *in general* might have beneficial effects on cognition and educational achievement. However, in these studies, researchers typically don't evaluate the effects associated with *specific* genres of video games. Thus, the above macro-level effects actually represent an average over numerous micro-level effects induced for example by playing different genres of video games. Some of these micro-level effects may be negative and others positive.

The purpose of this chapter is to review the scientific evidence regarding the relationship between playing commercially available video games, as assessed behaviorally or through brain imaging and their potential impact on human cognitive abilities. While many unknowns remain on this topic, it appears clear today that among the many factors to consider is the specific genre of video game being played (Bediou et al., 2018; Powers et al., 2013; Powers & Brooks, 2014; e.g., Sala et al., 2018; Toril et al., 2014; P. Wang et al., 2016). Following this work, we review below our current understanding of the impact of video games first in general and then narrowing in on the specific game genre that appears most effective to improve cognition.

4.2 Which video games improve cognition?

Video games come in many different flavors; classifying video games in genres has proven elusive and there is no consensual taxonomy to date. Fifteen years ago when the research on the cognitive effects of video games gained significant traction, researchers seemed to commonly classify video games in a small set of video game genres (see Table 1). Since then, video games, video gamers and gaming has changed considerably and it seems that the video game classifications that have been used in this field are not adequate to characterize contemporary gaming (Dale et al., 2020; Dale & Green, 2017). This being said, because the current review focuses on past research that tended to use older games, and to keep in line with the cited literature, we will use the game genres as described in Table 1. Note that in this literature "action video games" has been used to refer to first and third person shooters, although some authors have made it more inclusive. In this review, "action video games" will strictly refer to first and third person shooters.

A wide range of commercial video games have been used in psychological research to evaluate their relationship to or impact on cognition (Bediou et al., 2018; Sala et al., 2018). Video game research has proceeded using a variety of study designs, including cross-sectional and intervention studies ("true experiments"). Among the latter, we find studies looking at short-lived effects on the scale of a few minutes and intervention studies looking at more long-lasting effects, from days to months or even years (see Figure 4.1) for the design of such intervention studies). True experiments are necessary to rule out the possibility that the observed group differences pre-date the video gaming activities, and thus assess the causal role of video game play.

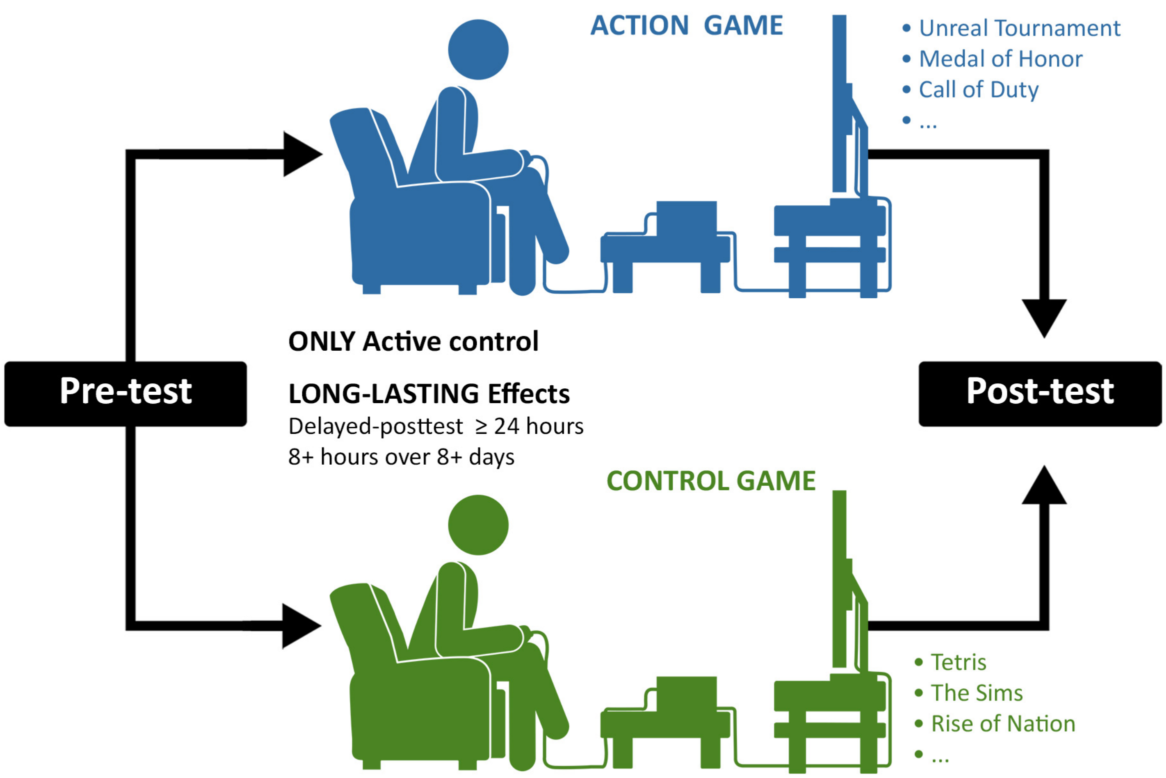


Figure 4.1: Intervention design to evaluate the causal impact of playing a specific type of video games on cognition (here termed experimental game). Participants are randomly assigned to play experimental video games or control video games. The training program typically requires at least 8 hours, and typically tens of hours of gameplay, distributed over weeks or months. Participants' cognitive skills are first evaluated on a battery of tests (pre-test) and tested again after completion of their training (post-test). If playing the experimental video games specifically improves the cognitive abilities assessed, then we expect the experimental group to improve more from pre- to post-test than the control group.

Table 4.1: Main 'classical' video game categories cited in the reviewed literature. These categories are based on the Video Game Questionnaire from the Bavelier lab. We provide in the supplemental materials the current version of the video game questionnaire and the selection criteria used in the Bavelier laboratory (version September 2019). The game categories it lists are motivated by research considerations and not by industry classifications. Yet, examples of games and our labels for game categories have evolved over the years in concert with the changing landscape of video games.

Category	Description	Examples
First & Third-person Shooters	game involving medium to long range weapon-based combat in first/third person perspective, against other players or AI characters.	Call of Duty, Overwatch, Unreal, Counterstrike
Real Time Strategy / Multiplayer Online Battle Arena	game in which the player manoeuver units to take control of the map and/or destroy enemy assets, usually in top-down view.	StarCraft, League of Legends, Age of Empire, Rise of Nations
Action Role Playing Game / Adventure Game	game involving varied action gameplay (e.g., shooting, close-combat, driving vehicles) in which the player controls a character that can be customized during the course of the game.	Uncharted, Mass Effect, Skyrim, Rise of The Tomb Raider
Sports or Driving Games	game that simulates real-life sports or driving a vehicle in the context of a competition.	Need for Speed, Mario Kart, NBA 2K12

Category	Description	Examples
Non-Action Turn-based Role Playing or Fantasy Games	game in which the player controls a character or party of characters that can be customized during the course of the game. Combat emphasizes decision making over rapid actions (i.e., turn-based or cooldown-based actions),.	World of Warcraft, Final Fantasy, Ultima, Pokemon
Turn-based Simulation, Strategy or Puzzle Games	turn based game centered around player decisions rather actions, involving strategic thinking and problem solving.	Solitaire, Bejeweled, Angry Birds, The Sims, Restaurant Empire, Rollercoaster Tycoon
Music Games	games centered around the interaction with a musical score, often involving rythm and memory.	Audiosurf, OSU!, Guitar Hero
Other	games that don't fit into any other category, or of unspecified type.	Cognitive training games, edutainment games, older 2D arcade games such as Pac-man or Zaxxon.

An important point to keep in mind in this literature is that all video games are not created equal as to their impact on cognition. Specific genres of video games have been shown to be effective in improving some aspects of cognition while others haven't. Studies that lump together all types of video game play are therefore at risk of blurring existing effects; for this reason, a number of studies adopt a more principled approach and focus on specific genres of video games.

A recent meta-analysis evaluated the impact of playing video games on cognition using a rather broad view of what counts as a video game (Sala et al., 2018). Figure 4.2 and Figure 4.3 use data from that meta-analysis and list the video games (or other activities) and their frequency of use in intervention studies aiming to enhance cognitive abilities. Figure 4.2 lists the games that were used for the experimental group, while Figure 4.3, lists games and activities that were used in active control groups. Several points are worth noting here. First, experimental and control activities vary widely. This variety makes it difficult to regroup these studies under one common research question as they each test different hypotheses. For example, when contrasting playing *Unreal Tournament* (FPS) vs. *Tetris* (Puzzle), one asks about the cognitive impact of action, first-person shooter games as compared to other games that also load highly on speed of processing and motor control; yet when contrasting playing *Tetris* (Puzzle) vs. *The Sims* (Life-Simulation Game), one rather asks about the possibility of training mental rotation by contrasting a game that requires such process and one that does not. Second, many of the activities listed are in fact not video games (e.g., paper-and-pencil games, watching videos). When contrasting, for example, playing a specific video game to playing paper-and-pencil games it is unclear if such studies evaluate the effectiveness of a specific video game, the impact of using a console, looking at a screen, or of digital media in general. Given the complexity of interpreting the outcome of grouping together and contrasting such a wide variety of activities, other meta-analyses investigating the impact of video game play on cognition have been more focused. The rationale here has been to group together video game genres that share features hypothesized to enhance cognition and to include only studies using other commercial video games as active control. Twenty years ago, researchers noticed by chance that study participants playing regularly first and third person shooters exhibited outstanding performance in attentional tasks (Bavelier & Green, 2016) and subsequently conducted an experimental study to test and verify the causal impact of playing those types of video games on attention (in contrast to a control group that played a different type of games; Green & Bavelier, 2003). These results led most of the field to focused on the impact of first and third person shooter games (e.g., *Unreal Tournament; Medal of Honor* (FPS)), also known as action video games, on cognition. Not surprisingly this is the most represented video game genre in the available literature. This is followed by racing games (e.g., *Mario Kart*, *Crazy Taxi*, *Need for Speed*) with rarer reports on real time strategy games (e.g., *StarCraft*, *Rise of Nations*, see Figure 4.2). While we will discuss below why these video game genres may be specifically well-tuned to change aspects of cognition, we now turn to the control games used in such studies. As illustrated by Figure 4.3, the video games most commonly used as controls are social simulation games such as *The Sims* (a life simulator game) and puzzle or visuo-motor coordination games (e.g., *Tetris*, *Ballance*, *Angry Birds*). This raises the possibility these genres have the least impact on cognition. Yet, it should be clear that different game genres might have different cognitive effects. Thus, depending on the study, the same game may be used for the experimental or for the control group. It appears from these figures that there is minimal overlap between the two lists (Figure 4.2) vs. Figure 4.3; see also Figure 4.4 and Figure 4.5)). A notable exception is *Tetris* which has been frequently used both as a control and as a cognitive training game, especially targeting mental rotation abilities. Below we review the literature for the main active game video game genres listed above.

4.3 First and third person shooters ("action" video games)

The game genre that has been most studied within the context of cognitive improvement is without a doubt First and Third Person Shooters ""Meta-analysis of Action Video Game Impact on Perceptual, Attentional, and Cognitive Skills"" (2018). This category of games has traditionally been called "Action Video Games" (AVG) in the field; however, the changing landscape of video games has made this nomenclature outdated and better classifications are

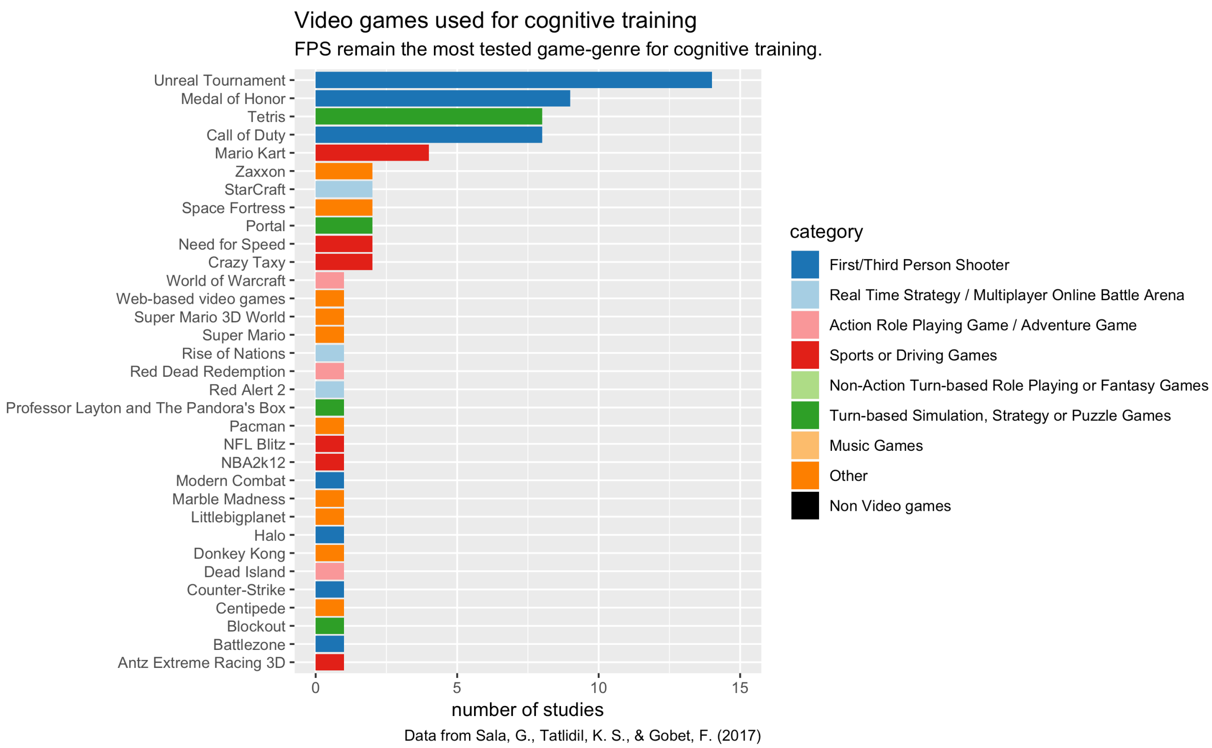


Figure 4.2: List of commercial video games used in cognitive training studies from Sala et al. (2018). This list contains a wide range of video game genres that have been used for training in the scientific literature (e.g., first person shooters, racing games, puzzle games, real-time strategy games, sports games) as well as non-video games (Space Fortress). Large differences in experiences between different game genres (a fast-paced multiplayer FPS is nothing like a slow paced, single player puzzle game) render the interpretation of any such results (positive, negative or null impact on cognition) quite difficult. This figure counts the number of publications cited in Sala et al. (2018) that used a particular video game (out of a total of 63 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games.

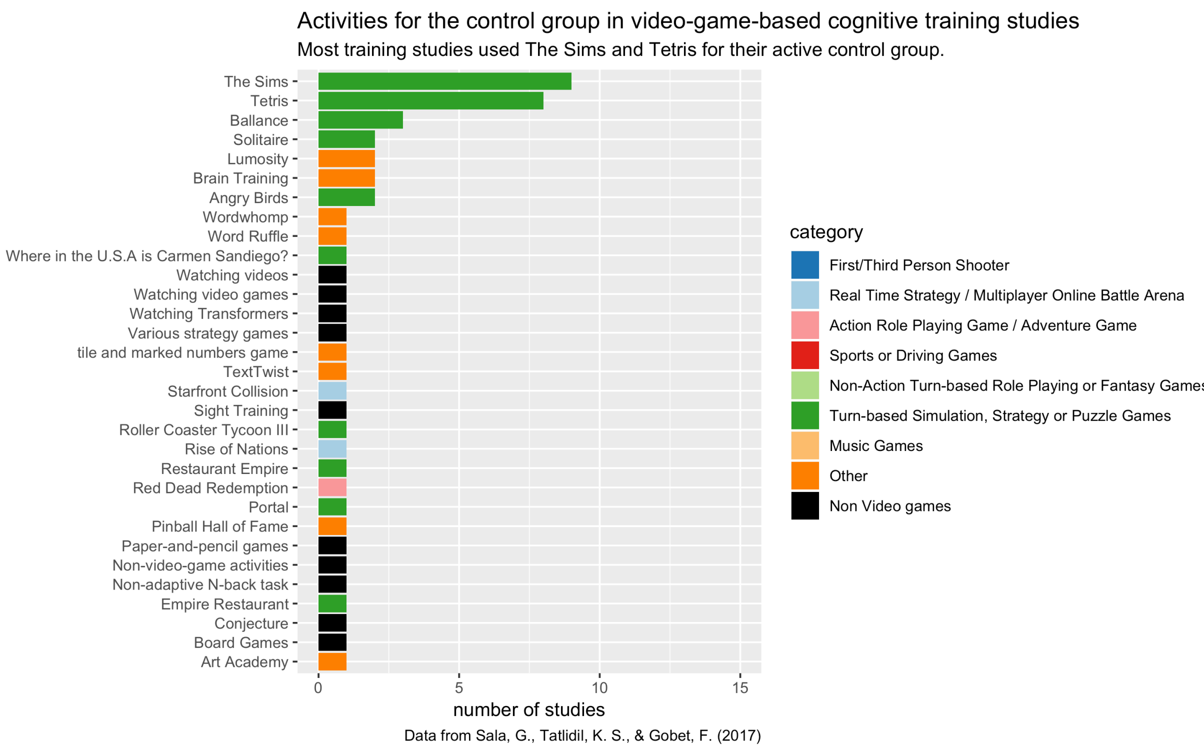


Figure 4.3: List of activities used as control treatment in video-game based training studies from Sala et al. (2018). Control treatments vary widely from playing video games to playing paper-and-pencil games; this makes it difficult to abstract the construct measured by such studies. This figure counts the number of publications cited in Sala et al. (2018) that used a particular video game or activity (out of a total of 63 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games.

needed (Dale et al., 2020; Dale & Shawn Green, 2017). First/Third Person Shooter games are (1) fast-paced, involving rapidly moving objects (e.g., projectiles) and transient events (e.g., explosion); they (2) require participants to distribute their attention to monitor events from central vision to the visual periphery; they (3) demand a high attentional focus by loading perceptual, cognitive and motor systems; and finally they contain (4) temporal and spatial uncertainty preventing full task automatization (Pedro Cardoso-Leite et al., 2020). Games in this category are typically violent and include titles like *Medal of Honor* and *Call of Duty*. It is critical to note that, contrary to what some have argued, action video games are not simply "any physically challenging video game in which reaction time plays a crucial role" (p1.; Karimpur & Hamburger, 2015). There are many games that require fast and accurate responding (e.g., fighting games, games like the *The World's Hardest Game*) that do not fulfill the criteria listed above.

Two types of studies investigated the relationship between action video gaming and cognition: correlation studies—where habitual first/third person shooter video game players (AVGP) are contrasted to individuals playing almost no video games at all (i.e., non video game players; NVGP)—and intervention studies—where individuals with only moderate video game play experience are asked to play either an action video game or a non-action video game for multiple hours distributed over weeks (see Figure 4.1). Correlational studies document significant differences between habitual AVGP and NVGP, leaving unclear the source of the difference. Intervention studies can clarify the causal role of video game play, as they evaluate whether game play changes performance between a baseline time before participants engage in the game play to a time after they have completed their game play training. Research on action video games has matured over the past 20 years and there is now a growing body of correlational and intervention studies—almost all of which however focus on healthy young adults. These intervention studies show for example that playing action video games rather than other forms of video games, causes improvements in visual perceptual abilities (Chopin et al., 2019), spatial cognition (Spence & Feng, 2010),some forms of memory (Pavan et al., 2019; Sungur & Boduroglu, 2012), and perhaps even academic topics such as reading

(Franceschini et al., 2015) or mathematical skills (Libertus et al., 2017). A recent meta-analysis has evaluated the impact of action video game on cognition subdividing outcome variables into one of 7 cognitive domains (Bediou et al., 2018): (1) perception, (2) top-down attention, (3) spatial cognition, (4) multitasking, (5) inhibition, (6) problem solving and (7) verbal cognition. Data from correlational studies show that habitual AVGP outperform NVGP in all of these domains with statistically significant effects for all but the less studied (6) problem solving category. Data from intervention studies show a similar trend, with AVG training causing numerically improved performance in all domains as compared to training with other commercial games. These effects are however smaller in size and less reliable than those observed in correlational studies, certainly calling for caution. Of all the domains studied, we note that top-down attention and spatial cognition seem most reliably improved by action video gaming interventions. The reduced effect sizes in intervention studies compared to correlational studies may be due to action video game players in the latter having substantially more gaming experience than the tens of hours typical of training studies. The reduced reliability on the other hand is due to both the effect sizes being smaller and to the reduced number of intervention studies per domain. As more studies are conducted, it will become clearer how much each specific domains may be positively impacted by playing action video games.

Most action video game studies focus on healthy young adults. A reason for this is that action video games are not adequate for children because of their violent content and they are not adequate for older adults because of their high difficulty level. While no experimental study would expose children to violent video games, some children do in fact play those age-inappropriate, violent games in their homes. In their meta-analysis Bediou et al. (2018) list three such cross-sectional studies focusing on the relationship between action video game and children's cognition. One such study tested typically-developing children and young adults, with ages ranging from 7 to 22 years, on three attentional tasks: the Useful Field of View (spatial attention), the Attentional Blink (temporal attention) and the Multiple Object Tracking task (sustained, dynamic attention; Dye et al., 2009). In addition, these

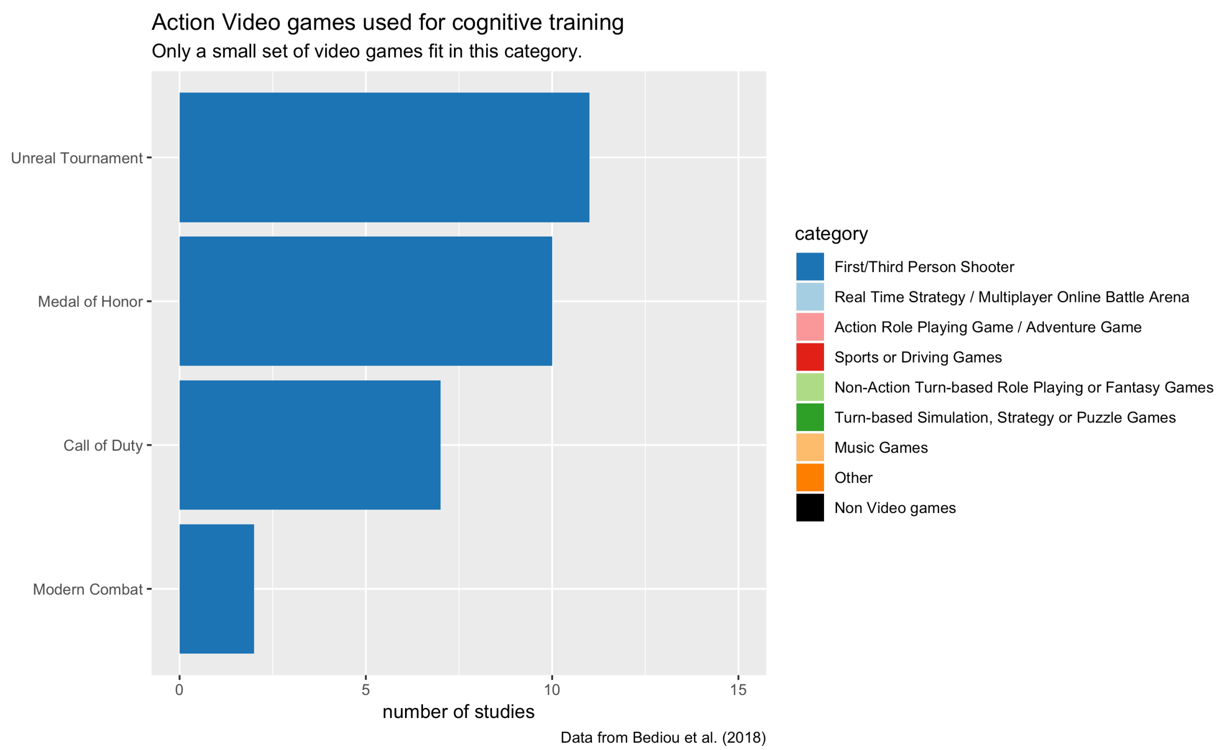


Figure 4.4: List of action video games (all first person shooter games; FPS) used for cognitive training according to Bediou et al. (2018). Focusing on this specific video game genre substantially reduces the number of games titles but still represents a major portion of the scientific literature (contrast this with Figure 4.2). This figure counts the number of publications cited in Bediou et al. (2018) that used a particular video game (out of a total of 23 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games.

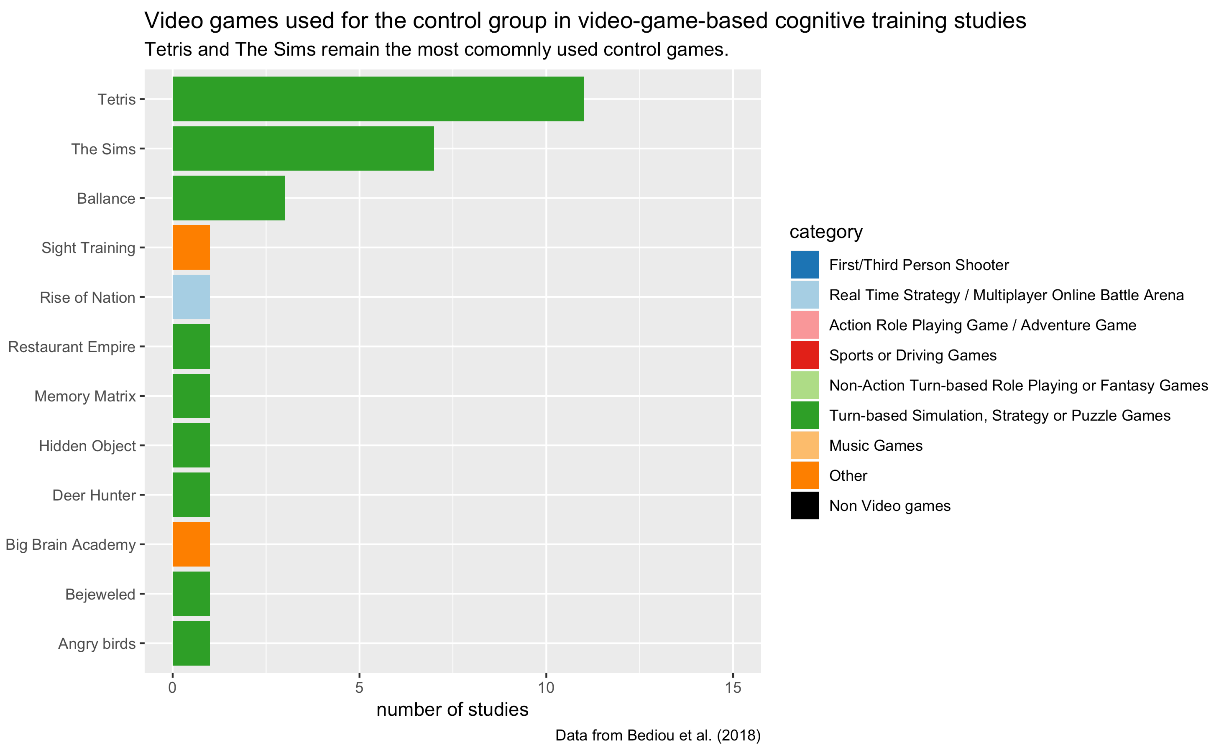


Figure 4.5: List of games used in the active control treatment when action video games were tested for cognitive training as tabulated by Bediou et al. (2018). This list includes only commercial video games (with the exception of the Sight Training program; contrast this with Figure 4.3). This figure counts the number of publications cited in Bediou et al. (2018) that used a particular video game (out of a total of 23 publications). Note that a publication could involve multiple experiments, each using potentially a different set of video games.

authors collected survey data about each participants' video gaming habits, allowing them to form two subgroups of participants: AVGP and NVGP. This type of data can be used to describe the time course of attentional development and evaluate how these time courses differ between AVGP and NVGP. The results show that AVGP presented a time course of attentional development that was accelerated compared to that of NVGP. The extent and onset of these group differences depended on the specific aspect of attention being considered. AVGP performed better than NVGP on the temporal attention task (i.e., attentional blink) starting at age 7-10, on the spatial attention (i.e., UFOV) at ages 11-13 and on the dynamic attention task (i.e., MOT) at ages 14-17. Such results confirm that various components of attention mature at different speeds and suggest they may be differentially affected by action video game play. Overall, the cross-sectional data collected on children present a pattern of results similar to what is observed in adults and indicate that action video games training may also be effective at younger ages.

To investigate the causal role of action video games on cognition in children, while avoiding exposing them to violent content, a few studies have selected commercial, age-appropriate mini-games that contain features similar to those attributed to action video games. Franceschini et al. (2013) have used this approach in 7-13 year old dyslexic Italian children to test the hypothesis that enhancing visual attention in Italian readers may in part alleviate reading difficulties. Children trained for 12 hours over two weeks either on action-like mini-games or control mini-games from *Rayman Raving Rabbids*. Note that *Rayman Raving Rabbids* comprises a large set of varied, small party games and thus does not technically fall in the first or third person shooter category. However, the authors rated each of the party-games from that collection as being action-video-game-like or not based on game features typically assigned to action video games. Mini games classified as action-video-game like were used for the experimental group while the mini games devoid of action mechanics were used for the control group. The training was distributed over about two weeks; those children assigned to play the action like mini-games displayed improvements in attention and in reading abilities, at least as measured by timed tasks of reading, as compared to a control group that played non-action-like games for the same amount of time (Franceschini et al., 2013). These first results were later confirmed in an English speaking sample of dyslexic children (Franceschini et al., 2017) and supported by a small sample, correlationational study on typically reading French adults (Antzaka et al., 2017). Yet, whether action or phonologically-based video games may help remediate dyslexia certainly remains controversial as other intervention studies have failed to find a positive impact on reading acquisition (Łuniewska et al., 2018). Moreover, a recent large sample correlational study that contrasted children who report playing video games to those that do not found a negative association between video game play and reading (Seok & DaCosta, 2019). The interpretation of this latter result remains difficult as it did not differentiate between game genres and had an overrepresentation of male children in the video game players group (indeed, if most video game players are boys, it's unclear if the effects relate to playing video games rather than other factors associated with boys being worse readers). Exploiting the proposal that action video game enhanced top-down attention, a recent study documents enhanced ability at performing optimal cue combination in 4-5 year old children after 7.5 hours of action-video-game-like mini-games (e.g., Fruit Ninja), as compared to playing control mini games (e.g., Puzzingo; training was distributed over 2 weeks; Nava et al., 2019). While the reviewed evidence points towards action video games having some efficacy in enhancing cognition, and especially attention in children, the empirical data is scarce and further studies are needed to confirm or infirm these results.

The use of action video gaming to train older adults' cognition is also quite rare (for a review, see Toril et al., 2014). One study had 65-91 year olds play either a first-person shooter (*Medal of Honor*), a puzzle game (*Tetris*) or an attention-training task (UFOV training) for six 90-minutes session or nothing (no-contact control group). Contrary to what was observed in younger adults, action video play did not improve attentional performance more than playing the puzzle game (Belchior et al., 2012). However, as pointed out by the authors, action video games might be too hard for older adults and training duration not long enough for them to learn how to play the game before the game could train their cognitive abilities. Indeed, players in the action video game group had to receive a step-by-step, PowerPoint-based explanation of the game by an experienced coach to make the difficulty level "manageable". Supporting this view, Boot et al. (2013) reported lower compliance in the action group than the other training groups in a sample of older adults. Because off-the-shelf action video games are designed to be challenging for adolescents and young adults already cognizant of the genre, they are likely too hard to be used with older adults (for a discussion, see section "Does Action Video Game Play Impact All Ages Equally?" in Bediou et al., 2018). Indeed, training with video games obey the same learning rules as training with any other forms of behavioral interventions (Stafford & Dewar, 2014). In particular, to be efficient, the training difficulty needs to be adapted to the learner's level, a concept pioneered as early as the 1900's by Vygotsky and his proposed "zone of proximal development". Thus, to train cognition in older adults it might be preferable to specifically design video games tailored for this population (Anguera et al., 2013).

4.4 Racing games

One of the most promising game genres for cognitive research are racing video games (Belchior et al., 2019; Cherney, 2008; L. Li et al., 2016; Wu & Spence, 2013). This is because they are typically less violent than first person shooter games; they are also easier to grasp by new gamers (Belchior et al., 2013) and easier to create for developers—which makes this genre ideal for cognitive training research (Anguera et al., 2013). Most importantly, this genre of video games can be easily adapted to capture the key mechanics of first or third person shooter games, and thus may offer similar cognitive benefits than first or third person shooter games do.

One study for example had young adults train for 10 hours on either an FPS (i.e., *Medal of Honor*), a racing game (*Need for Speed*) or a puzzle game (*Ballance*) and evaluated the impact of playing those games on visual search performance (Wu & Spence, 2013). Compared to training on the puzzle game, training on either the FPS or the racing game lead to improvements in divided attention and top-down attention control. Similarly, training on an

FPS (*Unreal Tournament 2004*) or on a racing game (i.e., *Mario Kart*) may both improve visuo-motor control; although the effects might not be strictly identical (L. Li et al., 2016).

Racing games have also been used to train older adults. One study had 65-86 year olds train for a total of 60 hours on either a racing game (i.e., *Crazy Taxi*) or a brain-training software (i.e., PositScience *InSight*) while others were part of a no-training control group (Belchior et al., 2019). The results suggest that both forms of training had modest transfer effects which for some were not present at post-test but only in the follow-up, 3 months later. Mental rotation, which was reported to improve with playing a racing game in younger adults (Cherney, 2008) does not seem to be affected in older adults.

While these studies suggest that using racing games might be viable pathway to cognitive enhancement, more data is needed to fully substantiate such a claim.

4.5 Real-time strategy games

A video game genre that has comparatively gained a lot of attention lately is real-time strategy video games. While older generations of strategy video games, not unlike chess, were mainly focused on strategic thinking and slow paced (i.e., "turn-based"), real-time strategy games include fast-paced action game mechanics. For example, in the real-time strategy game *StarCraft*, the player typically has control over multiple units in parallel, each of which requires frequent orders (e.g., move, attack, build) delivered through precise mouse clicks. Optimal play may require over 200 of such actions per minute (Lewis et al., 2011).

Using participants' self-reported video gaming habits data, Dale & Shawn Green (2017) formed four groups of participants and asked them to complete a large battery of cognitive tasks, including simple response time task, choice response time task, a go/no-go task, the Attentional Blink task, the Useful Field Of View, the Multiple Object Tracking and the Operation Span task. The four groups of participants (about n=14 per group) were habitual action video game players (AVGP), habitual real-time strategy players, people who rarely play video games (NVGP) and those who play more frequently but a wider range of game genres (i.e., "Tweeners"). Performance on the cognitive tasks differed between these groups. Overall, AVGP tended to perform best on these tasks and NVGP to perform worse with players in the real-time strategy and Tweeners groups performing somewhere in between these two groups. These cross-sectional results suggest that playing action video games but also real-time strategy games may improve performance on a variety of cognitive tasks.

To evaluate the causal effect of playing real-time strategy games on cognition, one study assigned 72 twenty year old (on average) women to play either one of two versions of *StarCraft* (a real-time strategy game) or *The Sims* (a slow pace life-simulator) for a total of 40 hours (completed on average in 43.7 days; Glass et al., 2013). The alternative versions of *StarCraft* differed in the amount of information players had to simultaneously keep track of and switch between. Before and after playing these video games, participants underwent a battery of cognitive tasks (including for example the Stroop task, Task Switching and the Operation Span task) selected to represent a latent construct of "cognitive flexibility". The results show that playing *StarCraft* improved cognitive flexibility more than playing *The Sims*. Additionally, the effects were strongest for the game version with higher load on cognitive flexibility.

Real-time strategy games have also been used for cognitive training in older adults (Basak et al., 2008). 70 year olds were randomly assigned to either play *Rise of Nations* (a slow paced real-time strategy game) for a total of 23.5 hours (distributed over 4 to 5 weeks) or to a no-training, no-contract control group (about 20 persons per group). Before and after the training (or non-training) all participants completed a battery of tasks covering what the authors call "executive control" (which included tasks like task-switching and the N-back) and "visuospatial skills" (e.g., mental rotation, attentional blink). The authors reported that playing *Rise of Nations* led to improved performance in the executive control but not in visuo-spatial skills (but see, Strenziok et al., 2014).

Studies investigating the association between real-time strategy game play and cognitive abilities in children are hard to find. One study had 3rd graders either play a fire-fighting real-time strategy game (*Fire Department 2: Fire Captain*) or read information about fire-fighting on a webpage for 40 minutes before taking a quiz about fire-fighting which included questions requiring to retrieve factual information, compare situations and solve problems (Chuang & Chen, 2007a, 2007b). Those who played the game performed better than the reading control group on fact retrieval and problem-solving items. However, it is rather unlikely that these effects are due to the game being a real-time strategy game (rather than say a puzzle game); instead it appears more plausible that learning about fire-fighting is more engaging and effective when that content is learned through active playing rather than by just reading.

4.6 Tetris

Tetris is arguably one of the most used video game in psychological research. It has been used to reduce cravings for food, drugs and other (Skorka-Brown et al., 2015), reduce intrusions of mental images related to traumatic events (Holmes et al., 2009) and to tone down the negative emotions associated with specific autobiographical memories (Engelhard et al., 2010). *Tetris* has also been used within the domain of cognitive training, sometimes as the experimental game and other times as the active control game (for a review, see Sala et al., 2018).

When used for cognitive training, *Tetris* is thought to train visuospatial cognition and more specifically mental rotation abilities as the game heavily relies on mental rotation. One study, for example, had 8-9 year old children either play *Tetris* (the experimental group) or *Where in the USA is Carmen Sandiego?* (a commercial game focusing on geography with minimal load on mental rotation; the active control group) for eleven 30-minutes sessions distributed over multiple days (De Lisi & Wolford, 2002). The results showed that playing *Tetris*, but not the control game, improved children's 2D mental rotation abilities as measured using a paper-and-pencil mental rotation test.

Studies on young adults, suggest that 6 hours of training on *Tetris* (as compared to a no-contact, no-training group) may improve performance on some visuospatial tasks (Okagaki

& Frensch, 1994; see also Boot et al., 2008; Terlecki et al., 2008). The effects however seem to be rather specific—training on a 2D *Tetris* version improved 2D mental rotation but not 3D mental rotation, while training on a 3D version of *Tetris* improved both (Moreau, 2013)—and several studies failed to observe improvements on 2D mental rotation after training on *Tetris* (Pilegard & Mayer, 2018; Sims & Mayer, 2002). *Tetris* has also been used for cognitive training in older adults, however not to train mental rotation but rather as a control-game. Yet, one study reported that in older adults playing *Tetris* may improve selective attention to the same extent as an action game or training on the attention task itself (Belchior et al., 2013), perhaps because for this age group, *Tetris* is already challenging and action video games are too difficult. The evidence supporting the usefulness of *Tetris* to improve cognition remains, therefore, mixed.

4.7 Casual mobile games

Casual mobile video game play is among the most common form of video gaming in the general population and it is increasingly popular among older adults (Chesham et al., 2017; Whitbourne et al., 2013). There have been several attempts to evaluate the impact of such video games on cognition; the results however are not always consistent. Note that we restrict here our review to commercial games and do not include the larger literature on computerized experimental psychology tasks, such as those developed by PositScience, Lumosity or tested by Owen et al. (2010).

One study for example (Oei & Patterson, 2013), had young adults train for a total of 20 hours over four weeks in various such games (*Hidden Expedition-Everest*, *Memory matrix 1.0*, *Bejewelled 2*, *Modern Combat: Sandstorm*, *The Sims 3*) and reported broad benefits (in various attentional and working memory tasks) only for the group playing the first-person shooter video game on mobile (i.e., *Modern Combat: Sandstorm*). Playing other, more casual video games did however improve performance on specific tasks (e.g., *Bejewelled 2* improving visual search) suggesting that casual video games might be used for targeted cognitive training interventions. However, using partly a different set of games and outcome measures, the same authors (Oei & Patterson, 2014a) reported no benefits of training for 20 hours on an FPS (*Modern Combat*), a real-time strategy (*Starfront Collision*) or a fast-paced arcade game (*Fruit Ninja*). Instead, they reported that slow-paced, physics game (*Cut the Rope*) lead to improvements in executive functions as indexed by performance in task-switching, flanker task and a go/no-go task. The authors provide various suggestions as to why their study failed to show improvements in the action-video-game trained group (e.g., differences in the experimental design). They also offer that the efficacy of the slow-paced physics game may be explained by that game involving cognitive processes that are important for executive functions (e.g., "strategizing, reframing and planning"). More research is needed to substantiate these claims.

A recurrent issue in this literature is to determine *a priori* and explain why training on a given game should improve performance on a given cognitive task. An interesting approach, grounded in Thorndike and Woodworth's principle of identical elements (Thorndike & Woodworth, 1901), consists in first evaluating the extent to which performance in various (casual) games correlate with performance on cognitive tasks, which are typically designed to isolate cognitive processes (Baniqued et al., 2013). Correlations between the two sets of measures may be caused by them involving the same set of underlying cognitive processes. Games that correlate with working memory and reasoning tasks may then be used to train those abilities. Using this approach, Baniqued et al. (2014) had participants play various categories of casual video games for 15 hours and measured their cognitive abilities across a large battery of tasks both before and after that training. The authors note that playing video games selected to tap into working memory and reasoning did not improve performance on working memory and reasoning tasks but instead improved performance on divided attention tasks (Baniqued et al., 2014). While this is undoubtedly an interesting and principled approach, more research is needed to solidify these results and gain insights into the differential effects of various game genres.

The literature reviewed above highlights the need to consider video game genres separately and argues for an empirical approach that contrasts specific commercial video game based on the mechanics it embodies rather than one that opposes any kind of video game to any kind of non-video game activity (Dale et al., 2020). While most evidence for the efficacy of video games for cognitive training currently rests on the use of action video games, future studies might reveal that other game genres are also (maybe differently) beneficial for cognition. Such studies may help to identify which game mechanics in video games are important to cause various cognitive improvements. An alternative, yet complementary route, consists in evaluating the neural processes involved in various forms of video game play as well as the consequences of video game play on the human brain. Below we review the literature on the neuroscience of video game play.

4.8 The neuroscience of video game play

Understanding what happens in the brain when people play video games, as well as the consequences that significant amounts of video game play has on brain structure and function may provide new insights to interpret the behavioral results described above. Playing video games has been associated with extensive neural alterations all over the brain, from sensorimotor regions to higher-order cortices such as prefrontal areas (Gong et al., 2019; Gong et al., 2015). For example, faster motor response times to visual stimuli in AVGP, compared to people who don't play video games, has been linked to increased white matter integrity in visual and motor pathways (Zhang et al., 2015), and AVGP in particular exhibited reduced brain activity during task preparation in the cuneus, middle occipital gyrus, and cerebellum which was interpreted to be indicative of increased neural efficiency (Gorbet & Sergio, 2018).

In the following sections, we briefly review the literature to highlight how video games affect brain organization, and how these functional and structural changes might in turn explain the reported behavioral consequences of playing video games. Yet, as discussed in the behavioral section above and exemplified in a recent review (Palaus et al., 2017), identifying the impact of video game play, as if it were a homogenous activity, on brain functions may be misguided.

Rather a more fruitful approach appears to focus on the information processing demands of the game play, and the exact processes engaged by the player. As a first step in that direction we consider below the impact of video game play on the brain systems linked to first reward system, and then spatial navigation before turning to the special case of action video games and the fronto-parietal networks of attention. Other brain systems (e.g., the motor system) may play important roles, but they will not be considered here.

4.8.1 Reward system

The brain's reward system is involved in learning and motivation. All successful video games tap into this system by using complex reward schedules to engage players for long play durations. Differences in the cognitive effects of training with various genres of video games might be related to differences in how these video game genres specifically activate the reward system. Although recent efforts attempt to characterize the specific cognitive effects of action video gaming involving the reward system (for a review, see Bavelier & Green, 2019), much remains to be uncovered as most research so far has focused on the relationship between video games and the reward system without differentiating what exact type of video game is being played. This being said, recent results show that the reward system may be a key player to consider when studying the effects of video games on the brain.

When contrasting playing a first person tank shooter game to watching a blank screen, Koepp et al. (1998) reported an increase in dopamine release in the ventral striatum (measured indirectly using Positron Emission Tomography) that correlated with the performance in the game (as measured by the highest game level reached by the participant) demonstrating that playing some video games can indeed causally affect the reward system.

Other studies investigated the potential long-term effects of video game play on brain function and structure. Kühn et al. (2011) observed that 14-year-old children who played frequently video games had a larger left striatum than same aged children who played infrequently, suggesting that prolonged video gaming may affect the structure of their reward system.

Furthermore, these changes in structure were accompanied by functional changes in that the frequent video players also displayed a larger BOLD activity than infrequent video players in response to losses during a gambling task. Similar studies conducted on adults provide somewhat different results. Kühn, Gleich, et al. (2014) observed that past video game experience correlated with gray matter volume in various brain areas (e.g., parahippocampal region) but not in the ventral striatum. These results may indicate that the effects of playing video games on the reward system may critically depend on the players age.

The evidence presented so far in this section is correlational implying that the observed brain differences may actually not be caused by video gaming but rather preexist and partially determine video gaming habits. There are however at least two studies that used an intervention design (contrasting video game training to a passive control group) in order to probe the direction of the causality effect (Kühn, Gleich, et al., 2014; Lorenz et al., 2015). Each of these studies had adults in the training group play a 3D platformer game (*Super Mario 64*) for 30 minutes per day over a period of two months and compared their changes in brain function and structure to those of a passive control group. Both studies reported that playing video games affected the size of various brain structures but did not, contrary to what was observed in the cross-sectional study on children, observe any structural changes in the striatum. The video game training did however affect the responsiveness of the ventral striatum to rewards. Lorenz et al. (2015) had their participants complete a task while under the fMRI scanner both before and after the video game training (for the intervention group) or before and after the waiting period (for the passive control group). The results show that for the participants in the control group the reward responsiveness in the ventral striatum decreased substantially from pre to post-test sessions while for the participants in the video gaming group this was not the case: participants trained on the 3D platformer video game exhibited similar activation levels in the ventral striatum in the pre and post-test session. The authors suggest these results may indicate a greater ability in the video game trained participants to maintain high levels of task motivation through the flexible control of the reward responsiveness of the striatum. They further hypothesize that this video-gaming induced effect on the reward system may be exploited for a broad range of uses cases.

Rewards schedules are a key component of all successful video games and it is still unclear how long term exposure to video games impacts the reward system. Current evidence supports the view that video games may alter the reward systems functioning as well as its structure (although, possibly only during childhood). While the results reported in this section may apply to all types of video games, the behavioral evidence clearly shows that it is necessary to distinguish various video game genres. The reward schedules implemented in different video game genres may have drastically different effects on the reward system, and through the reward system, on learning. There are ongoing efforts to clarify the possible mechanisms relating playing specifically action video games, the reward system and broad cognitive performance improvements (Howard-Jones & Jay, 2016; Miendlarzewska et al., 2016). More work is needed to formalize reward mechanisms in video games and assess the impact of different types of video games on the functional and structural properties of the human reward system.

4.8.2 Spatial cognition and the hippocampal formation

Video game play often requires discovering, and thus navigating, new worlds, be they landscapes, buildings or intergalactic spaces. Such video games are likely to engage the hippocampus whose role in memory and navigation is well established (Eichenbaum, 2017; for reviews see Lisman et al., 2017).

Frequent video gaming in adolescence and adulthood has been associated with volumetric changes of gray matter in the hippocampal region and its projections. Kühn, Gleich, et al. (2014) explored the correlation between gray matter volume and frequent gaming in adults, irrespective of the type of game being played. They measured gaming experience in a unit called *joystick years*, which reflects the lifetime amount of video game play, and evaluated to what extent joystick years was correlated with gray matter volume across all regions of the brain. Higher numbers of joystick years was associated with larger gray matter volume in both the occipital lobe and the hippocampal formation. Different gray matter volume in these two regions was proposed to reflect superior visuospatial expertise in video game players and to suggest that navigational exploration in early visual processing is affected by playing video games. Interestingly, recent findings also suggest a mediating role of the hippocampal formation during visual guidance (see Nau et al., 2018). Another correlational study reported a positive correlation between the amount of time spent on video games and gray matter volume in the hippocampus, in particular the entorhinal cortex that surrounds hippocampus (West et al., 2015). The navigation demands of most video games is in line with such changes in entorhinal cortex as this structure acts as a gateway to the hippocampus, and has been associated with spatial navigation, memory, and the perception of time (Bird & Burgess, 2008).

Changes in hippocampal volumes have been recently qualified as dependent on game genre and player strategies. Kühn, Gleich, et al. (2014) measured gray matter volume of the hippocampus and entorhinal cortex in relation to the lifetime amount of video game playing. Their results show that while playing puzzle and platformer games was associated with increased parahippocampal volume, playing action video games was associated with a decrease in parahippocampal volume (Kühn, Lorenz, et al., 2014). West et al. (2015) further qualified this effect as being related to particular cognitive strategies gamers might use for navigation, strategies that rely on different brain structures. One strategy that can be qualified as "spatial" involves constructing an internal cognitive map of the environment using landmarks and their relationships and then exploiting this map for navigation. The use of this strategy is thought to involve the hippocampus. An alternative, "non-spatial" strategy might instead rely on memorizing a fixed sequence of actions to be completed from a given location to reach a particular endpoint (e.g., when facing the entrance of the building, go left, then right, then left again). This second strategy therefore does not involve building internal representations but merely memorizing stimulus-response mappings. This non-spatial strategy is thought to involve the striatum. West et al. (2015) used a task where players navigated through a maze in the presence of landmarks that could be exploited to create an internal cognitive map. They then tested the same players on the same maze but removed the landmarks. Participants using a “spatial” strategy would be unable to use their internal maps in this situation as the landmarks were necessary to ground their cognitive map. Participants using a “non-spatial” strategy, on the other hand, would not be affected by this manipulation as they could still execute the memorized sequences of actions to reach the target. West et al. (2015) argue that the decrease in hippocampal volume observed in AVGP relative to NVGP may be accounted for AVGP relying more systematically on a non-spatial navigation strategy; in agreement with their hypothesis, AVGP performed better than NVGP when landmarks were removed, indicating that they exploited more systematically the non-spatial navigation strategy.

To further investigate the impact of spatial strategy during action video game play on hippocampal volume, West et al. (2018) conducted an intervention experiment comparing three groups of participants, one that was trained for 90 hours on action video games (e.g., *Call of Duty: Modern Warfare*), one that was trained for 90 hours on a 3D platformer video game (e.g., *Super Mario 64*) and a no contact group. Before and after the training entorhinal cortex, gray matter volume in the hippocampus was measured. Contrasting video game genres and play strategies shows that gray matter volume was reduced in the hippocampus after action video game training but only in participants using a non-spatial strategy. Yet, when a spatial strategy was used during training, action video game training resulted in increased hippocampal volume. Interestingly, among those trained on the 3D platformer, spatial learning was associated with increased gray matter volume in the hippocampus and non-spatial learning to increased gray matter volume in entorhinal cortex. The authors confirmed the impact these results in an additional training experiment which entailed training for 90 hours on action video games (e.g., *Call of Duty Modern: Warfare*). They note that it is only when the use of spatial strategy was encouraged during training that participants showed increased hippocampal formation volume. In conclusion, the neural impact of playing video games is mediated not only by the game genre but also by the very game play characteristics the player exhibit. This state of affair makes it clear that the impact of video game play on brain organization need to be qualified according to the processes the players engage while playing. As video games span widely different experiences, looking for the neural correlates of video game play in general is likely to remain an ill-posed research question. Finally, while the possibility to increase hippocampal volume through video games is promising to possibly address cognitive decline and in particular memory loss in aging, the directionality of the effects are yet not well understood. For example, reduction in gray matter volume was also observed after 5 days of intense mental calculation training (4 hours per day with two 10 minutes breaks), while at the same time performance being improved by the training (Takeuchi et al., 2011). Such results indicate that reductions in gray matter volume might not always be negative and/or reflect cognitive decline. Taking everything into account, genres and strategies affect how playing video games alters anatomical structures of the brain, calling for careful consideration of the way video games are designed, what content they present, and what strategies must be used to achieve the goals of the game.

4.8.3 Attentional networks and action video games

The strongest behavioral evidence regarding the impact of action video game training on cognition concerns increases in players' attentional resources over space, time, and objects as well as enhanced flexibility in the allocation of attention (Bavelier & Green, 2019). In this section we present functional and structural brain modifications that may underlie such attentional improvements.

Attentional functions are mediated by two main neural networks (Buschkuehl et al., 2012): a *ventral network of attention*, which encompasses the temporoparietal junction (TPJ) and ventral frontal cortex (VFC) and has been implicated in switching attention (as when redirecting attention towards a novel element in the environment); a *dorsal network of attention*, which consists of the dorso-lateral prefrontal cortex (DLPFC) and intra-parietal cortex and has been associated with strategic, goal-directed, top-down control over attention allocation. Coordination between the bottom-up and top-down networks has been associated with faster and more accurate responses to targets in a variety of cognitive tasks. Interventions targeting the dorso-lateral prefrontal cortex region, at least in children, enhances executive functions performance, including attentional control (Siniatchkin, 2017; e.g., J. Wang et al., 2018). Furthermore, these brain structures work in concert with the anterior cingulate cortex (ACC) which monitors and resolves conflicts, regulating in part the activity in the frontoparietal systems of attention (Petersen & Posner, 2012). Action video game play has been associated with more efficient neural activities in frontoparietal regions, and enhanced structural and functional connectivities in prefrontal networks, limbic system, as well as more posterior sensorimotor networks (Gong et al., 2017). This enhanced neural resource allocations in dorsal attentional network may contribute to the improved top-down attentional control and more efficient suppression of distracting information documented in AVGP (Bavelier et al., 2012). Attentional control can indeed optimize the selection of sensory information by two different mechanisms: by selecting more relevant signals, or by suppressing irrelevant signals and preventing noise to be transmitted to higher-order processes. Interestingly, AVGP not only benefit from enhanced attention to targets, they also show superior ability to suppress distractors (Bavelier et al., 2012). To track the fate of distractors during an attention-demanding visual task, several studies measured steady state visually evoked potentials (SSVEP), an imaging technique that uses periodic stimuli to frequency-tag neural responses in the visual cortex. Using this technique, both Mishra et al. (2011) and Krishnan et al. (2013) documented active suppression of distractors in AVGP, in line with enhanced selective attention. Since the SSVEP have the same frequency as the driving stimulus, it is possible to concurrently record responses to several stimuli if they are presented at different flickering rates. Mishra et al. (2011) measured SSVEP amplitudes, which are affected by selection and filtering processes in attention, in response to peripheral and foveal stimuli in a target detection task. While the SSVEP amplitude in response to attended targets was the same in AVGP and NVGP, SSVEP amplitude to distractors was decreased in AVGP relative to NVGP, suggesting enhanced filtering of irrelevant information. Similarly, Krishnan et al. (2013) compared SSVEP responses to targets and distractors in two groups of video game players, AVGP and role-playing video game players who served as their control group. Measuring signal-to-noise ratios of evoked potentials to both targets and distractors, Krishnan et al. (2013) showed that playing first person shooters could improve both the selection of targets and the suppression of distractors.

How bottom-up and top-down processes may change to both improve target selection and distractor suppression was assessed in an fMRI correlational study comparing AVGP relative to NVGP. Föcker et al. (2018) recorded fMRI scans while AVGP and NVGP participated in a cross-modal, endogenous Posner-cueing task. Young adults were presented with an auditory cue indicating the most likely location of a subsequent target on which participants were to perform a difficult, near-threshold visual discrimination task. This paradigm, closely modeled after Corbetta & Shulman (2002), allows one to separate neural responses to the auditory cues, which direct the attention allocation for the task to come, from the neural responses during the difficult visual task itself. The frontoparietal network, which is thought to mediate attention allocation, was more activated in NVGP than in AVGP when participants processed the cue and thus prepared for the task to come. This result may suggest that attention allocation is more efficient in AVGP than in NVGP. Interestingly, a small percentage of trials were in fact catch trials where only visual noise, but no visual target, was presented. In these catch trials, participants needed to withhold their response. AVGP outperformed NVGP on such trials exhibiting less false alarms. Moreover, only for AVGP did activation in the temporoparietal junction, middle frontal gyrus, and superior parietal cortex predicted their reduced false alarm rate, suggesting that these areas may operate and interact differently in AVGP compared to NVGP. Overall, these studies suggest that AVGP may benefit from better attentional control, or more flexibility in allocating attention, perhaps through a reconfiguration of the cross-talk between the main frontoparietal areas that mediate attention.

Whether these superior attentional skills result from alterations of processing in the goal-oriented, top-down attentional network, or rather from better filtering of irrelevant, potentially distracting information within early sensory cortices (or both) remains an open question. Neural markers of early attentional filtering were compared in EEG-based correlational studies contrasting AVGP and NVGP. Föcker et al. (2019), for example, tested if visual event related potentials (ERP) components differed between AVGP and NVGP in a high precision visual selective attention task. Faster response times and improved perceptual performance in AVGP was observed; yet, early markers of attentional selection such as the posterior N1 and the P1 were identical across groups. Differences between AVGP and NVGP were only observed in parietal generators such as the P2 and the anterior N1 components. As the P2 has been previously linked to task demands (Finnigan et al., 2011; Lefebvre et al., 2005), these results may indicate that AVGP are able to more effectively adapt attentional resources to varying tasks demands. A similar conclusion was reached by another intervention ERP study (Wu et al., 2012) that recruited 25 adults and recorded ERPs before and after 10 hours of video game training.

Participants with no video game experience in the previous 4 years, were randomly assigned to one of two training groups: the action group played *Medal of Honor: Pacific Assault* (FPS), whereas the control group played *Ballance*, a 3D puzzle game. Later, during the testing session, participants performed an attentional visual field task which assesses the ability to detect a target among distractors. As in Föcker et al. (2019), the two training groups exhibited comparable early sensory ERPs, in line with comparable comparable early attentional selection processes across training. Also, as in Föcker et al. (2019), the action trained group showed an increased P2 amplitude. Moreover, the amplitude of the P3 was also increased in the action trained group, possibly indicating enhanced attentional resources being allocated to the task (Kok, 2001). Overall, these results are in line with the proposal that the differences in attentional performance between AVGP and NVPG may reflect a functional reorganization of the goal-oriented, top-down, dorsal attentional network with distractor suppression being implemented at a central level, rather than through early perceptual filtering.

Furthermore, playing video games, irrespective of the specific game genre, seems to affect structural and functional properties of parts of the frontal cortex. A longitudinal training experiment study for example, evaluated the structural changes in the dorsolateral prefrontal cortex (DLPFC) resulting from two-month of training with *Super Mario 64*, a 3D platformer, non-action video game that requires navigational skills (Kühn, Gleich, et al., 2014). The results indicate that playing this video game induced structural changes by increasing the gray matter volume in the right DLPFC. Similarly, a correlational study reported that the self-reported weekly hours adolescents spent playing video games correlates positively with the thickness of their left DLPFC and left frontal eye fields (FEFs; Kühn & Gallinat, 2014)—cortical thickness is similar but not identical to gray matter volume (Winkler et al., 2010).

It has also been reported that relative to NVGP, AVGP have enhanced intra- and inter-network connectivities in the central executive network and salient network (Gong et al., 2016). These two networks are highlighted using fMRI measurements; the central executive network is associated with working memory, planning, and getting prepared to select an appropriate response to a stimulus, whereas the salient network with nodes in the subcortical reward system has been linked to salient stimuli detection as well as integrating emotional, sensory, and interoceptive signals (Menon, 2015). The central executive network typically contains the DLPFC and is engaged during attention-demanding tasks (Fox et al., 2006). Further analysis of large-scale networks with diffusion tensor imaging, which evaluates how strongly specific areas are connected, shows that those who spend more weekly hours playing action video games display an increased efficiency (as defined in graph theory) in local, global, and nodal levels of prefrontal, limbic, and sensorimotor networks (Gong et al., 2017). The local, global, and nodal efficiencies, respectively, reflect an increased fault tolerance across the network, improved information flow across the whole network, and the importance of a node, respectively. These neural regions are responsible for processing visual information, spatial orientation, motion perception, selective attention, and integrating multimodal stimuli. This finding supports the view that neural efficiency increases by mediating goal-oriented, top-down attentional processes as a consequence of automating visual sensorimotor tasks and delegating them to areas that handle low-level sensory processing.

While our understanding of the effects of playing video games on the human brain has improved considerably over the last decade, it remains nevertheless limited. Most of the literature reviewed is correlational in nature and based exclusively on adult participants. Studying young adults cross-sectionally is a cost-effective strategy to highlight candidate structures and generate and test hypotheses. Indeed, cross-sectional studies only involve a subject selection phase (using surveys for screening) and an assessment phase, while intervention studies require *in addition* multiple training sessions and a second assessment phase (to serve as a post-intervention test to be compared to the pre-intervention test). Intervention studies furthermore involve a high management cost to assure that participants don't drop out and complete the various steps of the study within the planned time frame. Cross-sectional studies are cost-effective to highlight interesting patterns; however, as for behavioral studies, this strategy needs to be complemented with intervention studies to establish causality and rule out the possibility that the neural differences between habitual action video gamers and non-gamers pre-dated the gaming experiences. Furthermore, the studies reviewed above were mainly conducted on young adults. However, the mechanisms involved may differ with age as the time course of brain plasticity is likely to differ across brain areas. It will thus be important to include pediatric samples in the future.

4.9 Concluding remarks

Research on the cognitive consequences of video game play has boomed over the past 15 years. As the range of video games tested widens, it becomes apparent that not all video games have the same cognitive impact. Rather, studies systematically contrasting specific game genres indicate that the content of the video game, the user interactions it requires, and attentional processes it engages are of paramount importance. This fact has two consequences. First, it makes little sense to ask about the cognitive impact of video game play; rather, it is important to recognize the variety of experiences video game play affords. Here we have reviewed game genres that have been used over the past 15 years using a game classification that might have been relevant for the covered research but is unlikely to upstand the drastic changes in game types, gamer profiles and gaming habits that have emerged since (Dale et al., 2020; Dale & Green, 2017); some initial work is being done to better characterize video gaming for cognitive research (Dale et al., 2020; Pedro Cardoso-Leite et al., 2020). Second, there is a need to build better theories on why playing certain video games but not others improve cognitive abilities; one route towards building such theories is to contrast commercial video games which differ by specific game mechanics or by specific content (Pedro Cardoso-Leite et al., 2020). Following this strategy, past research has focused mainly on contrasting "action video games" (i.e., mostly first and third person shooters) to other commercial video games (e.g., puzzle games). A recent meta-analysis supports a causal relationship between playing action video games and improvements in top-down attention and spatial cognition, with effects on other domains requiring further studies (Bediou et al., 2018). This is not to say, however, that this genre of video games is the only genre of interest for cognitive training. More recently, studies have investigated the effectiveness of racing games and real time strategy games, which may be suitable for a wider audience than action video games. While promising results have been reported, more research is needed to evaluate the efficacy of these alternative game genres and determine the mechanisms by which they may enhance cognition. The strategy of contrasting multiple game genres within the same study may be useful to both evaluate the relative efficacy of different game genres and to unveil the relevant game mechanics.

The study of how video games in general, and action video games in particular, engage and affect the brain has revealed network wide changes in reward, memory and attention brain circuits. This variety of effects suggests that the neural mechanisms responsible for the observed cognitive benefits is likely to go beyond the training of a few specific cognitive processes (Bavelier et al., 2012). Rather, aligned changes in memory, reward processing and mood, as well as attentional networks efficiency may result in faster processing speed, facilitating in turn a variety of cognitive processes. Future work is needed to unravel the link between the behavioral improvements noted after action video game play and their neural bases. Overall, while significant progress has been made over the past 15 years on our understanding of how to leverage video games for cognitive enhancement, there remain many unknowns in this young emerging field. First, the work so far makes it clear that different genres of video games have different effects on cognition, differences in game mechanics have been hypothesized but they remain to be fully tested to firmly document why playing action video games but not social simulation games may improve cognition, for example. Unpacking key game mechanics is central if we are to leverage lessons from action video game research to design therapeutic or educational video games. Second, although a theoretical framework around brain plasticity, attention and learning for the documented effects has been proposed, many of the mechanistic details remain to be worked out (Bavelier et al., 2012; Bavelier & Green, 2019). Third, our work has focused so far chiefly on cognition; understanding how to best induce plastic changes in other domains, such as emotion or social behavior, is equally important. Finally, most of the literature so far has focused on adults. As we now better understand the game mechanics that promote brain plasticity, the time has come to ask how to best use video games to foster children's development.

4.10 Future perspectives

Research over the past 15 years has focused mainly on establishing and validating the impact of action video games and probing the breadth of their impact on various cognitive constructs (e.g., top-down attention vs. bottom-up attention). Much remains to be done to catalogue and fully describe the impact of different video game genres on various aspects of behavior. Furthermore, our understanding of the taxonomy of video games needs to be improved so that we can move from vague high-level labels (e.g., "action video games") to objective, measurable indices (e.g. type of attention required; exact reward schedule implemented etc). In the future, we should be able to make quantitative predictions as to which video game to train on in order to enhance performance in one cognitive construct versus another. The challenges that lie ahead of us will require methodological and theoretical innovations as well as multi-lab and interdisciplinary team work.

Chapter 5

Neural Correlates of Habitual Action Video Games Playing in Control-related Brain Networks

Abstract

Playing action video games has been reported to lead to broad cognitive benefits, implying that this form of cognitive training may be exploited for positive societal impact. Although the underlying cognitive and neural mechanisms are not yet fully understood, current accounts revolve around the idea that playing action video games enhances cognitive control—a general ability modern cognitive neuroscience suggests is the result of the coordination of a multitude of brain networks that may be highlighted by recording functional brain connectivity of people at rest. In this study we use resting-state fMRI functional connectivities to train a machine learning model to classify people as habitual action video gamers or non-gamers and investigate which aspects of functional brain connectivity have the greatest effect on the prediction accuracy of the classification model. Our results show that this classification is indeed possible, with the best model reaching an accuracy level of 72.6%. This result is important for both theoretical and practical reasons, as it adds to a growing body of ev-

idence reporting long-term effects of action video gaming on the brain and demonstrates that resting-state imaging may be an effective research tool for studying cognitive training and transfer. Our results also show that what distinguishes action gamers from non-video game players most is not the activity in individual brain regions, nor the activity within individual specialized brain networks but rather the relationships between networks. This result is important in that it casts these cognitive training effects in the cognitive control framework in cognitive neuroscience, provides support to current theories of action video game training in psychology, and offers new insights into why action video game training generalizes to new cognitive tests. More specifically, our analyses highlight the importance of the interplay between cognitive control networks on the one hand (the fronto-parietal and cingulo-opercular networks) and the sensorimotor network on the other, suggesting that action video gaming may optimize cognitive control for the purpose of enhanced perception and rapid action. Overall, this work advances our understanding of the effects of action video gaming, of cognitive training and their transfer effects as well as the neural basis of cognitive control. We hope this work will contribute to the development of more effective cognitive training programs.

5.1 Introduction

Playing action video games has been shown to enhance a broad range of cognitive abilities—including the ability to switch between different tasks, filtering out irrelevant information, and focusing on important stimuli—while leaving other abilities unaffected (e.g., bottom-up attention) (Bediou et al., 2018). These results are important from a practical and theoretical point of view. Indeed, training cognition with action video games could be used for broad positive societal impact (see Chapter 4 for a review).

From a theoretical point of view, the mechanisms underlying the cognitive benefits of playing action video games are not yet fully understood. In psychology, the transfer effects of action video game play have been attributed to enhancements in task-specific processes (the "common demands hypothesis”; Oei & Patterson, 2014b), but also to domain-general abilities including reward processing (Nahum & Bavelier, 2020; West et al., 2015), cognitive control (Anguera et al., 2013; Benady-Chorney et al., 2020; R. West et al., 2020) and, most prominently, attentional control (Bavelier & Green, 2019; Föcker et al., 2019). To simplify, we will use “cognitive control” as an umbrella term to encompass related concepts (e.g., “executive control”, “attentional control”, “cognitive flexibility”) and conceptualize it broadly as “the coordination of mental processes and action in accordance with current goals and future plans” (Menon & D’Esposito, 2022). We purposefully ignore certain nuances and state that a main family of hypotheses pinpoint changes in cognitive control as the key consequence of action video game play that causes transfer effects to a broad range of cognitive tasks.

In cognitive neuroscience, playing video games has been associated with numerous changes in brain structure—e.g., increased gray matter in the caudate nucleus and decreased gray matter in the hippocampus (West et al., 2018) and brain function—the specifics of these changes however depend on the type of game being played and how it is played (for a review see Chapter 4). One study for example, used functional resonance imaging (fMRI) to record the brain activity of participants while they performed an attention demanding visual detection task in the presence of distractors. When contrasting habitual action video game players (AVGPs) with people who don’t play video games (i.e., non-video game players; NVGPs), it was clear that the frontoparietal brain network, a key neural actor in attention control, was *less* activated by increased attentional demands in AVGPs than in NVGPs (Bavelier et al., 2012). This type of result has been interpreted as implying increased top-down attentional control abilities in AVGPs compared to NVGPs: because the attentional system is *more* effective in AVGPs, their BOLD response increases *less* with increasing attentional demands (Bavelier & Green, 2019; Green & Bavelier, 2012).

The empirical evidence, both in experimental psychology and cognitive neuroscience is rich and the theoretical accounts too complex to be accurately depicted here. It is however fair to say that the main hypotheses regarding the transfer effects of action video games involve domain-general cognitive abilities (i.e., cognitive control) which are assumed to be subserved by networks of brain areas (e.g., the frontoparietal attentional network) rather than by a single brain area (e.g., the left prefrontal cortex). It appears then that a brain-wide systems approach would be invaluable to the study of action video game training and their transfer effects. There have been recently important advances in applying graph-theoretical tools to cognitive neuroscience that are now providing new insights about brain function in general and cognitive control in particular (Menon & D'Esposito, 2022; Zink et al., 2021). By applying these new approaches to the study of action video gaming we hope to tell apart competing hypotheses and better understand the underlying mechanisms as well as human cognitive control systems in general.

5.2 A graph-theoretic approach to cognitive control in cognitive neuroscience

5.2.1 The brain is intrinsically organized into networks.

It has become increasingly clear in cognitive neuroscience that the traditional, modular approach (where cognitive function X is performed by brain area Y) is limited (R. Poldrack, 2006); and that instead we need to reason in terms of known systems and networks that interact with each other to generate intelligent behavior (Hutzler, 2014). This is particularly true in the case of cognitive control, where the scientific evidence was unable to pinpoint a single cognitive control area and instead highlighted multiple control networks (Menon & D'Esposito, 2022; Zink et al., 2021).

For example, a large body of work recording fMRI while humans perform a variety of visuospatial attentional tasks has highlighted two attentional systems: a dorsal frontoparietal system involved in top-down attentional control (e.g., maintaining attentional focus on a stimulus) and a more ventral system responsible for bottom-up attention (e.g., detecting a danger) (Corbetta & Shulman, 2002). These two systems are also known as the dorsal (DAN) and ventral (VAN) attentional networks respectively. It is important to note that although these two networks are specialized and functionally separate, their coordination is required for adaptive behavior and thus the two systems must interact. More specifically, in this particular model, the ventral system is thought to act as a circuit breaker, interrupting activity in the top-down system when an important signal calls for immediate attention.

In recent years, many computational approaches have been developed to directly model brain activity as a timeseries of interacting brain networks (as opposed to previous work inferring networks from snapshots of average co-activation patterns) and to adopt a more systematic study of the relationships between brain networks and cognition across many tasks. Using such graph-theoretic approaches on multi-task fMRI datasets (Cole et al., 2013), on resting-state datasets (Dosenbach et al., 2008, 2010) or both (Dadi et al., 2020), researchers have identified several brain networks as playing key roles in cognitive control (see below). It is important to note that these networks do not represent the ground truth yet; there are inconsistencies across methods, some subjectivity in the choice of hyperparameters and limitations in the current computational approaches (e.g., a given brain area can be assigned to only one network by most standard methods). As our methods and datasets will improve, so will the validity and accuracy of the highlighted functional networks.

5.2.2 The cognitive control brain networks

Multiple brain networks, relevant to the current study, have been identified in the literature and are presented below. These networks are part of a parcellation atlas which assigns brain voxels to a brain region, and brain regions to networks. Alternative methods led to alternative parcellations, meaning that a given brain region may be assigned to different networks depending on the parcellation or even not be assigned at all, and some networks exist only in some parcellations but not others.

5.2.2.1 The Dosenbach2010 atlas

In a cross-task analysis of 10 cognitive tasks, Dosenbach et al. (2010) identified 160 regions over the whole brain that were consistently active during cognitive control tasks (also see Dosenbach et al., 2007). Those regions served as seeds to extract a graph from the resting-state fMRI. Edges of the graph were weighted by the correlation between respective resting-state time-series and then thresholded to identify six networks, to which they assign specific roles based on their involvement in cognitive tasks. Once this atlas is applied, activities in 160 seeds are mapped to one of those six networks, which we describe next. The fronto-parietal network (FPN) includes regions in the dorsolateral prefrontal cortex, inferior parietal lobe, dFC, ventral anterior prefrontal cortex, and IPS (for more details see, supplementary material). FPN is thought to be involved in the rapid adjustments to real-time changes in tasks demands. The cingulo-opercular network (CON) includes regions in the anterior prefrontal cortex, ventral prefrontal cortex, basal ganglia, anterior insula, adjoining fronto-insular cortex, thalamus, precuneus, superior temporal, temporoparietal junction, and dorsal anterior cingulate cortex. CON is thought to be involved in maintaining attention and stable task sets. The sensorimotor network (SMN) includes regions in precentral gyrus and mid insular, supplementary motor area (SMA), preSMA, superior parietal. SMN is involved in integration of sensory information and motor movements. The occipital network includes regions in primary (V1) and secondary visual cortices (V2). Occipital network is involved in visual processing, The cerebellum network includes regions in lateral, medial, and inferior cerebellum. Cerebellum is thought to be indirectly related to task performance and may be involved in generating error codes (Fiez, 1996). The default mode network (DMN) includes ventromedial prefrontal cortex, ventrolateral prefrontal cortex, inferior temporal, post cingulate gyrus, and angular gyrus. DMN is activated in the absence of attentional demands. It may not directly be involved in cognitive control, but may influence cognitive functions indirectly (Anticevic et al., 2012; Brandman et al., 2020; Greicius & Menon, 2004).

5.2.2.2 The Gordon 2014 atlas

The Gordon2014 atlas is a surface-based parcellation that was derived from boundary maps of BOLD activations in two resting-state fMRI datasets. This atlas identifies 13 cortical networks: The cingulo-opercular network (cf. CON in Dosenbach2010), The fronto-parietal network (cf. FPN in Dosenbach2010), DorsalAtt, (aka DAN); centered on the intraparietal cortex and superior frontal cortex, is involved in top-down goal-directed selection of stimuli and responses. Regions of the dorsal network show sustained activation when subjects are cued to attend to a feature of stimulus (attention set). VentralAtt, (aka VAN); centered on the temporoparietal cortex and inferior frontal cortex, is specialized for the detection of behaviorally relevant stimuli, particularly when they are salient or unexpected. The default mode network (cf. DMN in Dosenbach2010), The cingulo-parietal network (aka CPN) includes regions in anterior cingulate cortex, ventral and dorsal parts of the precuneus, inferior temporal cortex, and lateral parietal cortex, and superior frontal cortex. This network has been often observed when the subjects do not perform any task [i.e., resting; toro2008]. The sensorimotor network of the hand (SMNhand), The sensorimotor networks of the mouth (SMNmouth), The salience network (SN) includes a set of regions with hubs in dorsal anterior cingulate and ventral anterior insular cortices. It receives inputs from limbic and sensory regions and is often attributed to monitoring and dynamic switching. The auditory network includes regions in superior temporal gyrus, and is thought to process auditory information. The visual network is located in the occipital lobe, and is thought to process sensory inputs originating from the eyes. The retrosplenial temporal network (aka RTSC) is located immediately behind the corpus callosum. The function of this region isn't fully understood yet. It is thought to be involved in coordinating perceptual and memory functions because of its proximity to visual and hippocampal areas. Unassigned set of regions. The regions that were not assigned to any networks were not excluded from further analysis, but rather labeled as "unassigned".

This atlas is particularly important in the context of studying the effects of action video gaming because it comprises the two attentional networks that are often cited in this context

(Corbetta et al., 2008; Föcker et al., 2018); namely the dorsal and the ventral attention networks (DAN and VAN). For a list of coordinates of regions and corresponding networks see supplementary materials.

5.2.2.3 The DiFuMo atlas

In addition to the two parcellation atlases listed above, we included in this study a more recent data-driven atlas, called DiFuMo, which has been developed on a large structural and functional dataset rather than prior research on cognitive control (Dadi et al., 2020). The reasons to include DiFuMo is that DiFuMo may be less biased by theoretical considerations and may highlight networks that are more stable because they are grounded on a larger dataset.

DiFuMo differs slightly from Dosenbach2010 and Gordon2014 atlases as it is a probabilistic functional parcellation that is extracted from thousands of task-fMRI and rs-fMRI scans, with different versions of DiFuMo, identifying varying numbers of regions (i.e. 64, 128, 256, 512, or 1024 regions). Hence, voxels across the whole brain are mapped to either 64, 128, 258, 512, or 1024 regions. We used the mapping for 64 regions, each of which was mapped to seventeen networks proposed by Yeo et al. (2011). For each region, DiFuMo provides an anatomical name, MNI152 coordinates, the mapping of regions to networks defined in Yeo et al. (2011), and the ratios of white matter, gray matter, and CSF. We mapped voxels to regions and then applied the mappings to map regions to networks. Coordinates of the DiFuMo regions and their corresponding assignment to brain networks is provided in the supplementary materials.

5.3 Measuring intrinsic networks can be studied during resting state.

While task fMRI is frequently used to identify brain activities that are attributed to cognitive functions, spontaneous brain activities during rest (intrinsic networks) show substantial overlap with task-driven networks, both in their spatial organization and functional roles (Kraus et al., 2021; Varoquaux, 2020)—provided resting state brain activity is recorded for long enough (Birn et al., 2013). If action video gaming impacts brain function, this impact should be manifest not only during the performance of cognitive tasks, but also during rest (A. L. Cohen et al., 2008; Kraus et al., 2021). Moreover, domain general processes like cognitive control and attention, which are thought to be altered by action video game play, are processes that are common to many tasks and therefore one would expect that long-term coactivation of their corresponding brain networks during gaming to alter functional resting connectivity (R. A. Poldrack et al., 2015).

The similarity between task-induced and intrinsic networks, makes resting-state recordings an invaluable tool to understand long-term effects of action video gaming on cognitive control networks. First, resting-state data may offer an effective way to measure individual differences in executive functions (Reineberg et al., 2015), cognitive control performance (FPN, Salience Network, CON, and DMN; see Menon & D'Esposito, 2022 for a review), attention (VAN and DAN; see Corbetta & Shulman, 2002 for a review), and numerous other behavioral dimensions (Seguin et al., 2020). This could for example be useful to rapidly evaluate the efficiency of new cognitive training programs and evaluate to what extent they will transfer to new tasks. A second reason resting-state data is an interesting method in this context relates to the controversy around expectation effects (action gamers performing better because they believe they should perform better) rather than genuine cognitive improvements being responsible for some of the observed performance differences between AVGPs and NVGPs (Parong et al., 2022; Tiraboschi et al., 2019). Resting-state data might provide a means to assess such differences, untainted by prior task experience or expectation effects.

5.4 Hypotheses

The graph theoretic approach to cognitive control that we just presented allows us to cast cognitive theories in more explicit terms. According to the common demands theory one might expect to see changes only at the level of specific, specialized brain regions, but no changes at a systems level and possibly no changes that would not be visible in resting-state functional connectivity data. Alternatively, there is a class of theories predicting changes beyond the isolated brain region. Some researchers might for example expect to see changes specifically in the top-down attentional control system (DAN) but not for example in the bottom-up attentional system (VAN). This type of result would be in line with the notion that a domain-general subsystem (e.g., top-down attention) is enhanced by action video game play. Finally, some researchers may expect the effects of action video games to go beyond individual networks and affect cognitive control more broadly. This hypothesis would translate into changes in inter-network connectivity differences between AVGPs and NVGPs. A main goal of the present study is to test these three families of hypotheses (which are not mutually exclusive). Discriminating between these macro-hypotheses will not only help us understand the effects of action video games but also the breath of generalization effects as the broader the effect on the brain networks, the broader one would expect those changes to manifest as improved behavioral performance across a wider range of cognitive tasks.

In addition to these macro-hypotheses, numerous more detailed predictions can be made. Among the six networks of the **Dosenbach2010** atlas, we specifically expect FPN, CON, and SMN to be diagnostic of AVGP, as these networks have been frequently highlighted in that literature. For instance, AVGPs have been reported to both be able to focus their attention better than NVGPs and to be less disrupted by distractors, while at the same time being more capable to switch between tasks (Bediou et al., 2018). This phenomenology suggests more effective CON (for sustained performance) and FPN (for flexibility) networks. In addition, AVGPs have also been shown to outperform NVGPs on sensorimotor tasks (Gozli et al., 2014). This increased behavioral performance may be linked to superior cognitive control abilities but could also result from changes in the SMN network itself. Changes in other networks of the Dosenbach2010 seem less likely (e.g., DMN, Cerebellum). It appears then that these three networks, FPN, CON and SMN, as well the relationships between them, may best characterize the functional connectivity differences between AVGPs and NVGPs.

Among the 13 networks of the **Gordon2014** atlas, we expect AVGPs and NVGPs to differ mostly on the dorsal attentional network (DAN) and the frontoparietal networks (FPN). We expect no differences between AVGPs and NVGPs with respect to the remaining networks. In addition to these network-specific effects, one can make predictions about differences in inter-network relationships between AVGPs and NVGPs. Indeed, there is growing evidence that FPN and CON become more integrated with increased task demands and that their integration correlates with task performance (J. R. Cohen et al., 2014), even at the trial-by-trial level (Shine et al., 2016; Shine & Poldrack, 2018). This being said, how exactly cognitive control is achieved within a neural network perspective is not yet fully understood (Menon & D'Esposito, 2022; Zink et al., 2021) and the results of this study may perhaps contribute to that understanding.

5.5 Data

For the purpose of this study, we used an unpublished resting-state fMRI dataset that was collected in a previous study (Föcker et al., 2018). The dataset included a total of 32 subjects (16 AVGPs and 16 NVGPs) who participated in a resting-state fMRI session after completing several cognitive tasks in the scanner. The aim of the original study was to investigate attentional control in action video gamers. In that study, researchers excluded from their analyses 1 NVGP for being a music expert, and 2 AVGPs for being high media multitaskers (see Föcker et al., 2018 for details). In this study, we decided to exclude none of the participants and to use the entire cohort of 32 subjects.

The fMRI data were acquired using a Siemens TrioTim 3T scanner with an eight-channel head coil, 4mm isotropic resolution, 125 time points, TE/TR = 30/3000 ms, flip angle = 90°. Anatomical T1w images were defaced prior to the preprocessing to ensure participants' privacy. Overall, the resting-state dataset included a time series of 7 minutes and 30 seconds per subject.

All the participants were volunteers and gave informed consent. In accordance with the

Declaration of Helsinki, the Research Subject Review Board of the University of Rochester approved the study.

A noteworthy point about the design of the study is that participants attended the resting-state fMRI scanning session after completing a task-fMRI session in which an auditory Posner-cueing task was used (see Föcker et al., 2018). It is therefore possible that this task may have somewhat contaminated the subsequent resting-state functional connectivities (Hasson et al., 2009; Lor et al., 2022; Tailby et al., 2015). In our particular case, the auditory Posner-cueing paradigm was designed to engage perceptual and attentional processes, both of which are thought to differ between AVGP and NVGP (Föcker et al., 2018). Hence, observing AVGPs versus NVGPs differences in resting-state activities involving the auditory cortex may either reflect differences in intrinsic brain function and/or differences in task-related brain activation patterns that persist after completion of the task. It is therefore important to be cautious when interpreting the present results and to replicate this study using additional datasets.

5.6 Methods

5.6.1 Formal problem statement

The goals of this study are (a) to evaluate whether intrinsic brain functioning (as assessed using resting-state fMRI data) differs between habitual action video game players and non-video gamers and (b) whether the observed differences (if there are any) are compatible with current theories of action video game training effects.

We trained a computational model to classify people as habitual action video gamers (AVGP) or non-action video gamers (NVGP) using their resting-state functional connectivity data. We expect the ability of the model to correctly classify unseen participants as AVGP vs NVGP to exceed the chance level. If this is indeed the case, we will further investigate the fitted model to understand the causes of its performance (e.g., by identifying the most diagnostic resting-state functional connectivities in the model). Our hypothesis is that both inter- and intra-network connectivities contribute to classification performance.

The classification problem we want to tackle can be formulated as follows:

$$X \in \mathbb{R}^{|\text{subjects}| \times |\text{networks}| \times |\text{timepoints}|}$$

$$y \in \{\text{AVGP}, \text{NVGP}\}$$

$$\hat{y} = f(X, \theta)$$

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} |y - \hat{y}|$$

Where X is the resting-state functional connectivity matrix of the networks (see "Network Aggregation" section below for details), y is the true label of the subject (either AVGP or NVGP), f is a classification model that receives as input X and outputs $\hat{y}$—a prediction of y (label). The classification model has parameters θ, which are learned from data while minimizing $y - \hat{y}$. These model parameters include the choice for a particular parcellation atlas and connectivity metric as well as model weights.

Given this setting, the hypotheses of this study are (H1) resting-state connectivity differences allows the robust classification of AVGP vs NVGP, and (H2) difference between AVGP and NVGP involve both specialized networks (i.e., within network connectivity) and the cross-talk between brain networks (i.e., between-networks connectivity). If we consider the connectivity pattern as a graph with brain networks as its nodes, and connectivity between networks as its edges, then the two hypothesis can be formally expressed as follows:

$$\text{(H1)} \quad \hat{\theta}_{\text{nodes}} \cup \hat{\theta}_{\text{edges}} \in \text{Control Networks}$$

$$\text{(H2)} \quad |\hat{\theta}_{\text{nodes}}| < |\hat{\theta}_{\text{edges}}|$$

5.6.2 Preprocessing

Considering that even minor changes to the preprocessing steps can affect the result of the analysis (Lindquist et al., 2019), we used a reproducible pipeline for the entire preprocessing stage. Specifically, we opted for MRIQC (v21.0.0rc2; Esteban et al., 2017) for data quality checks and fMRIPrep (v20.2 LTS; Esteban et al., 2019) for preprocessing, without making any modifications to the default parameters. The only exception was that we skipped the skull stripping because the scans were already defaced for privacy reasons (see Figure 5.4).

For each participant, the preprocessing pipeline resulted in 125 images of size 646464 isomorphic 4mm voxels in the MNI152NLin2009cAsym common space (Ciric et al., 2021). The preprocessing pipeline extracted an additional set of motion-based artifacts which was further removed from the signals by applying confound regression during the parcellation step (described below). Note that the extracted motion signals did not differ between AVGPs and NVGPs. Indeed, the performance of a AVGP vs NVGP classifier using those motion signals did not exceed chance level (chance level=50%, mean validation accuracy=51%, SD=18%, 100-repeated 4-fold cross-validated; see supplementary materials).

All the additional preprocessing decisions were made automatically based on the "simple" denoising strategy in the Nilearn package (v0.9, Abraham et al., 2014) which recommends high pass filtering at 0.1 Hz, 6 degree head motion correction, basic CSF component removal, demeaning, no global signal removal, no scrubbing, no compcor correction, and no ICA-AROMA (Abraham et al., 2014; see Fox et al., 2005; Team, 2022 for details). We also examined whether the removed confounds, motion as well as other signals, differed between AVGP and NVGP groups. We observed no significant difference between AVGP and NVGP with respect to the removed confounds (see supplementary materials).

5.6.3 Data analysis pipeline

The complete data analysis pipeline is illustrated in Figure 5.1. All data were first preprocessed using a standard procedure (step 1 in Figure 5.1, see "Preprocessing" for details).

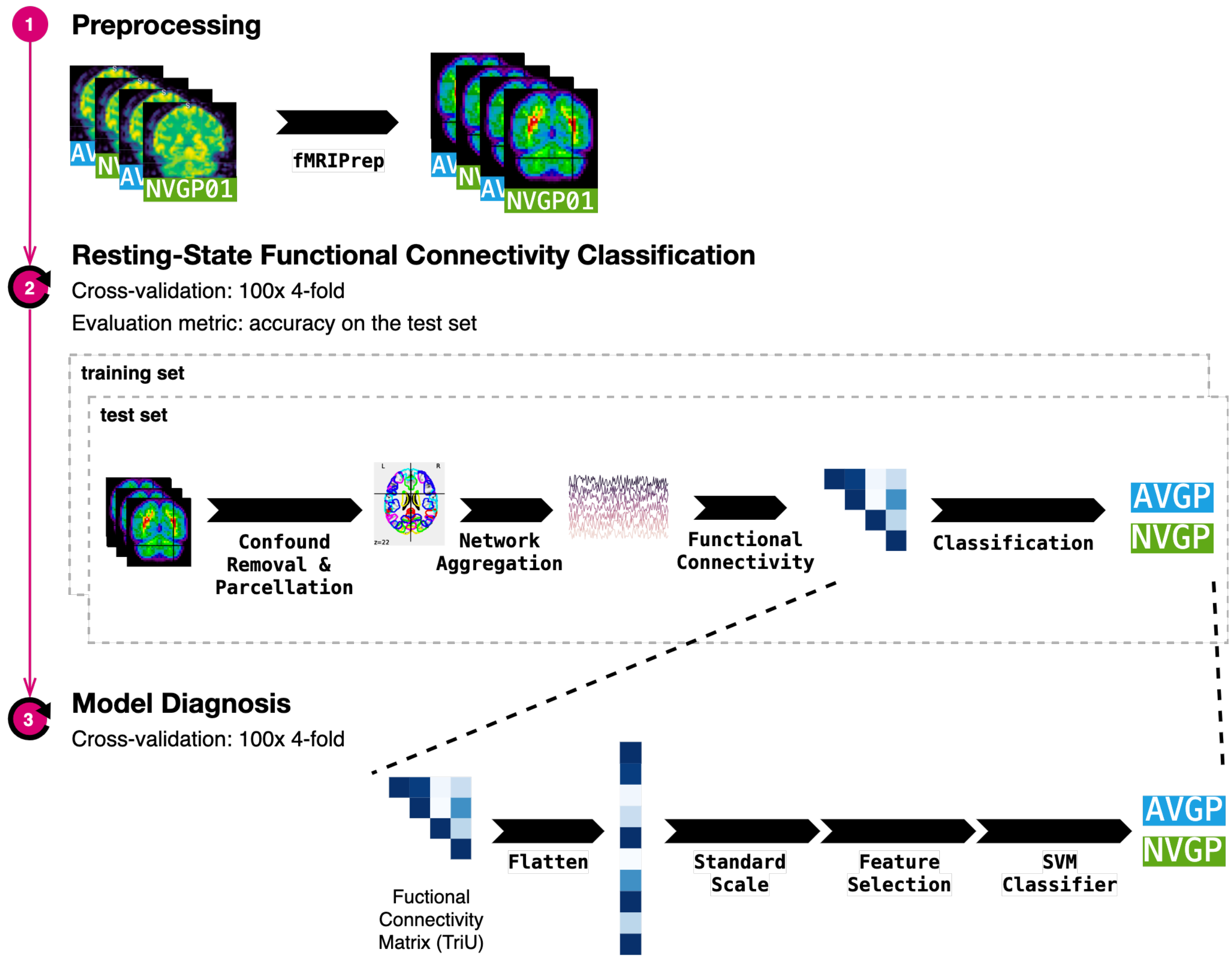


Figure 5.1: Data analysis pipeline. All data were first preprocessed using a standard procedure (step 1). The same steps were applied irrespective of the AVGP/NVGP label of participants. This preprocessed data then served as input to the next steps which aimed to 2) train and 3) diagnose a AVGP versus NVGP classifier (see text for details).

The same steps were applied irrespective of the AVGP/NVGP label of participants. This preprocessed data then served as input to the next step which aimed to train a AVGP versus NVGP classifier (step 2 in Figure 5.1).

To train our model to classify participants as AVGP versus NVGP, we first split the data into a training set and a test set (by randomly assigning participants to either subset). Next we performed a sequence of operations on the training set, which include confound removal, parcellation (i.e., mapping time-series of voxels to time-series of regions according to a given parcellation atlas), network aggregation (i.e., mapping time-series of regions to time-series of networks as defined in the atlas), connectivity extraction (i.e., calculate connectivity metrics for the network time-series) and ultimately the classification model see 5.1. For the classification model we used a support vector machine (SVM with linear kernel and L1 regularization) as this type of model is often used as a first baseline. Following best practices in machine learning (R. A. Poldrack et al., 2020) we computed the accuracy of the classification on the test dataset (i.e., on data from participants that were not used to train the model). This is to ensure that the model will generalize to other participants and is not overfitting the training data. Finally, the whole procedure was repeated 100 times to ensure the metrics were representative of the data and not of a specific random split of the data.

The next step of the data analysis pipeline (step 3 in Figure 5.1) takes as input the fitted model and aims to diagnose what features of the input data are responsible for the observed classification accuracy. More specifically, we used permutation importance to assess the contribution of functional connectivity features on the models' prediction accuracy. In this procedure, the importance of a given feature is quantified by how much the prediction accuracy of a model decreases as a result of randomly shuffling the values of that feature. In addition to permutation importance, we also applied SHAP analyses—a more recent machine learning technique used to interpret fitted models. While permutation importance focuses on the models accuracy, SHAP focuses on what features are responsible for the models output (i.e., classifying a person as an AVGP regardless of whether that person is or is not an AVGP). The results of the SHAP analyses are presented in the supplementary materials.

These were the broad data analysis steps involved in this study. Below we present further details about each step.

5.6.4 Evaluation of the classifier

The cross-validated pipeline was trained on 75% of the data (24 subjects) and evaluated on the remaining (8 subjects). The training/testing step was repeated 100 times on randomized splits of the data (hence 100-repeated 4-fold stratified and shuffled cross validation). As a result of this repeated cross-validation, the prediction performance of the model was measured by the distribution of 100 accuracies on the test sets.

The cross-validated steps included parcellation (three candidates), factoring voxels to networks (see below), calculating functional connectivity metrics (five candidates), flattening the upper triangular connectivity matrix, normalization, model-based feature selection (selecting half of the features based on linear L1-regularized SVM coefficients), and a classifier (linear L1-regularized SVM).

For each cross-validation split, a new model was created, separately trained on the training set, before recording its prediction accuracy on the test set. To optimize hyper-parameters of the pipeline, we used grid search tuning on the training set (75% of the entire dataset or 24 subjects) with 5-fold cross validation. The hyper-parameters included whether to standardize features or not, and the SVM regularization parameter, all of which were evaluated by the classification accuracy on the validation folds. Test splits were not used to tune or train the model.

5.6.4.1 Parcellation

Grouping data from voxels into meaningful brain regions allows both to reduce the complexity and noise in the data but also to inject semantics in the data (i.e., brain regions and networks are more meaningful than isolated voxels; Varoquaux & Craddock, 2013). Because there is no consensus yet on which parcellation atlas is the best (Salehi et al., 2019), we opt for using three different parcellation atlases as the parameter of the classification model: 1) Dosenbach2010, 2) Gordon2014, 3) DiFuMu64 (see the supplementary material for a list of parcellation parameters).

To create a reduced and more meaningful spatial representation of brain function we aggregated voxels into regions according to the selected parcellation atlases (Dosenbach2010, Gordon2014, and DiFuMo64; see "Introduction" for more details and motivations on selecting these atlases). This step first produced region-wise time-series (step 1) and then network-wise time-series (step 2).

We first used the maximum likelihood method to estimate time-series of the defined regions in the atlases (i.e., parcels or spatial maps) from a set of preprocessed voxel-wise time-series.

$$\text{(Step 1)} \quad \hat{U}_r = \underset{U_r}{\operatorname{argmin}} \| Y - U_r V_{p \to r} \|$$

$$Y \in \mathbb{R}^{t \times p}, U_r \in \mathbb{R}^{t \times r}, V_{p \to r} \in \mathbb{R}^{p \times k}$$

$$\mathbf{t} \text{ time points}, \mathbf{p} \text{ voxels}, \mathbf{r} \text{ regions}, \mathbf{n} \text{ networks}$$

$\hat{U}_r$ here represents the maximum-likelihood estimate of the region-wise time-series, Y is the observed voxel-wise preprocessed time-series, U_r is the tested region time-series, and $V_{p \to r}$ is the mapping of each voxel to regions from the atlas. The atlases and data instances were both resampled to a 2mm resolution. We used Nilearn (v0.9) to mask the brain and resample images. Ultimately, this step yielded parcel-wise subject-level time-series for the regions in each atlas.

We then aggregated regions into networks in order to obtain a representation that is semantically relevant, and produced network-wise time-series. The reason to aggregate regions into networks was twofold. First, regions may become active during several cognitive functions which makes it challenging to attribute regions to specific cognitive functions (R. Poldrack, 2006). Second, one region may belong to multiple networks, so they may become active in different contexts and processes. By assigning semantics to the networks (rather than regions), the model would be simpler (yet less comprehensive), which makes it possible to interpret the results in terms of general cognitive functions that are commonly related to cognitive control (e.g., attention, inhibition, multitasking, or working memory to name a few) rather than sparse activation in regions (Dadi et al., 2019; Varoquaux & Craddock, 2013). Smaller number of features is also important for computational and statistical traceability of the model (e.g., 7 networks instead 135 networks in Dosenbach2010 atlas) . For instance, empirical benchmarks show that the baseline classification algorithm that we use (binary SVM) works best when there are fewer features (A. Li, 2022).

In order to estimate the network time-series, we applied the same maximum likelihood methods as the one used to aggregate voxel-wise time-series into region-wise time-series.

$$\text{(Step 2)} \quad \hat{U}_n = \underset{U_n}{\operatorname{argmin}} \|\hat{U}_r - U_n V_{r\to n}\|$$

$$U_n \in \mathbb{R}^{t\times n};\ V_{r\to n} \in \mathbb{R}^{n\times r}$$

$$\mathbf{n} \text{ networks}$$

$\hat{U}_n$ represents time-series for each networks of a given atlas, $\hat{U}_r$ is the estimated region-wise time-series extracted in the previous step 1, U_n is the candidate network-wise time-series, and $V_{r\to n}$ is the mapping of the regions to networks as defined by the parcellation atlas. For every network in the atlas, this step resulted in one time-series.

Aggregating voxels into networks allows to compute a functional connectivity matrix that shows relationships between networks rather than between regions or voxels. The diagonal values of the network functional connectivity matrix would further represent within-network activities.

5.6.4.2 Functional connectivity metrics

Given the network-level time-series, we calculated functional connectivity matrices that measure the relationship between networks. We computed five alternative resting-state functional connectivities metrics: covariance, Pearson's correlation, partial correlation, tangent projection of covariance, and precision (sparse inverse covariance). For n networks, the connectivity matrix would contain n^2 values. As the connectivity matrices were symmetric, we flattened the upper triangular part of the matrix (including diagonal values) and used the resulting vector as the input to the classification task.

5.6.4.3 AVGP vs NVGP classifier

As the final step of the pipeline, we fitted a binary classifier that receives participants' vectorized functional connectivity matrices and predicts their label (AVGP or NVGP). Choices of parcellation atlases and connectivity metrics were then contrasted in terms of prediction accuracy on the out-of-sample test set.

More specifically, we trained an L1-regularized linear SVM classifier after standardization (removing the mean and scaling) and model-based feature selection, for which hyper-parameters were optimized based on the training set (see Figure 5.1 and cross-validation section for details). We trained the model on 75% of the data, and validated it on the remaining 25% (8 subjects). The classification was independently trained 100 times and in each iteration the prediction accuracy of the model was evaluated on the test set. This resulted in 100 numerical values that represent the goodness-of-fit for a given set of parameters and hyperparameters (i.e., atlas name and connectivity metric).

5.6.4.4 Model diagnostics

Feature ranking is a common first step when aiming to explain machine learning models. To measure and rank the contribution of each resting-state functional connectivity to the classification performance, we used cross-validated permutation importance. Permutation importance is a model-agnostic technique where the importance of a feature is measured by the change in the accuracy when the feature is shuffled (Molnar, 2022). However, permutation importance is more appropriate for datasets with uncorrelated features—this is not the case here since spatial dependence between adjacent and overlapping brain regions might result in multicollinearity between network connectivities. To partially address this limitation, we used repeated cross-validated permutation importance techniques to not only extract the feature importance but to infer confidence intervals for the measured importance. We repeated the permutation procedure 100 times, yielding 100 measurements for each train/test split. This procedure was repeated 1000 times with 4-fold cross validation to compute confidence intervals on feature importance.

Permutation importance measures the impact of individual features on the performance of the model; it may still suffer from interaction between features (McGovern et al., 2019). This limitation is mainly addressed in techniques such as multi-pass permutation importance where the correlation between features is broken by keeping previously assessed features permuted while assessing the new features. This provides an improved interpretation of model performance, yet for models that produce suboptimal predictions, interpreting the output of the model rather than its performance may provide a deeper understanding of how individual features and their interaction contribute to a prediction. Therefore, we also performed an additional feature importance analysis using SHAP values (SHapley Additive exPlanations). While permutation importance methods focus on the impact of features on a model's performance, SHAP values focus on understanding what features are responsible for the output of the model, irrespective of whether the prediction is correct or not. Additionally, when using SHAP, the correlation between features is broken by considering the effects of all the other features and interactions between features. As we only applied SHAP analysis to one specific model (i.e., the model with highest prediction accuracy), the results of the SHAP analysis are presented separately in the supplementary materials. We expected to see similar ranking of features in both the permutation importance test and the SHAP analysis.

Finally, we anticipated that the prediction accuracy of the classification model would be affected by particular combinations of parcellation atlases and connectivity metrics. There are indeed different lines of evidence and reasoning that led to the development of those parcellations and metrics and these may be more or less relevant for the purpose of AVGP vs NVGP classification. The Dosenbach2010 atlas, for instance, results from an attempt to identify networks that enable cognitive control—this atlas may therefore be more relevant in our analysis than atlas that were developed for other purposes. To assess which parcellation, connectivity metric and their combination were most effective in terms of classification accuracy, we used Bayesian model comparison. The details of this analysis are provided in the supplementary materials.

5.7 Results

5.7.1 Participants can be accurately classified as AVGPs versus NVGP based on their resting state functional connectivities.

We trained machine learning models to classify unseen participants as either AVGPs or NVGPs (see Methods). The best predictive model classified participants with a 72.6% accuracy (95% CI [69.9, 75.4]), which is substantially above the 50% chance level (i.e., train/test splits were stratified and half of the participants in the sample were action video gamers). These results are robust and cannot be attributed to chance or overfitting. Indeed, the model performance was validated on unseen participants and it was unable to make accurate predictions on random data. More specifically, when randomly shuffling participants group membership within a bootstrapped permutation test, the model yielded an average classification performance of 50% (95% CI [47, 53])—considering this distribution of bootstrapped classification accuracies as an empirical null distribution, the probability of observing a classification accuracy of 72.6% is only p=0.015. These results are important because they clearly show that action video gamers and non-gamers have different functional brain connectivity patterns during rest.

As explained earlier, the specific data analysis results of fMRI data may vary considerably depending on details of the data analysis pipeline. To ensure that our results are robust, we systematically evaluated multiple parcellations and connectivity metrics. The model with the highest classification accuracy used the Dosenbach2010 parcellation atlas and the partial correlation connectivity metric (see Figure 5.2).

This type of analysis begs several additional questions. A first question asks to what extent particular choices of parcellation or connectivity metrics impact the model's classification accuracy (e.g., are some atlases more effective than others?). Figure 5.2 displays the classification accuracy for each combination of parcellation and connectivity metric used in this study. It appears from this figure that both of these choices do indeed have a major influence on the prediction accuracy, with parcellation playing a major role (i.e., overall, the Dosenbach2010 atlas yields higher accuracy levels than DiFuMo64) and connectivity metric a somewhat lesser role (i.e., partial correlations are more effective than simple correlations). These effects were quantified and confirmed using Bayesian model comparisons (for details, see supplementary materials).

A second question we want to address is to what extent the interpretation of the results depends on specific methodological choices. That is, beyond their impact on classification accuracy, do specific data analysis choices affect the conclusions about which aspects of brain function differ among AVGPs and NVGPs. This question will be addressed in the next section.

5.7.2 Resting-state functional connectivity differences between AVGPs and NVGPs are not circumscribed to a specialized brain network: they involve multiple networks and interplay between them.

The previous results show that resting-state fMRI data can be used to accurately classify participants as AVGPs vs NVGPs. Now we want to investigate which aspects of the resting state data are responsible for that prediction accuracy. For example, functional brain

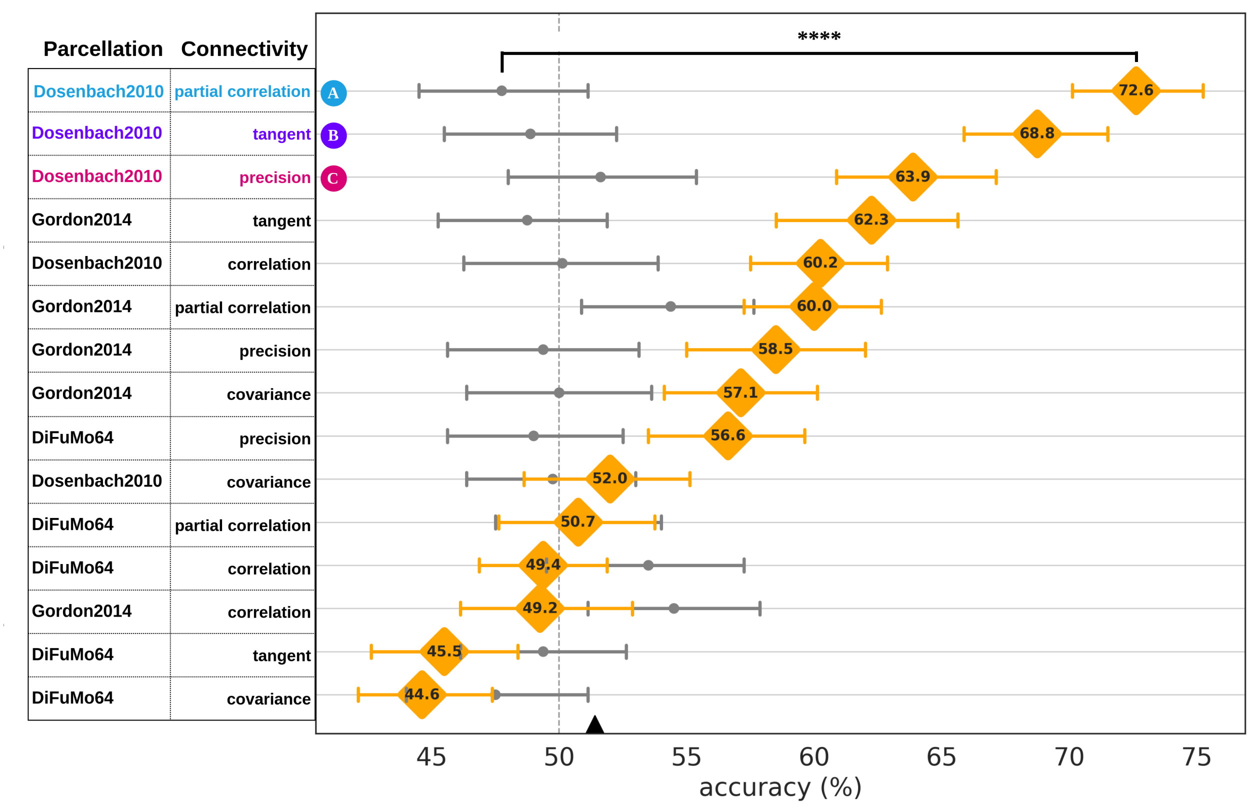


Figure 5.2: AVGPs vs NVGPs classification accuracy as a function of parcellation and connectivity metric. The distribution of cross-validated out-of-sample prediction accuracies are displayed in orange for the actual data and in gray for a shuffled version of the data (to form an empirical null distribution; see text for details). Dots and diamonds represent the mean of the distribution; error bars represent the 95% confidence intervals. This figure shows that new participants can be accurately classified as AVGPs vs NVGPs based on their resting state functional brain connectivity with the best model reaching an accuracy of 72.6%. Classification accuracy varies however considerably with the specific parcellation and connectivity metric used. The black triangle on the X-axis shows the prediction accuracy using motion confounds; the observed accuracy (51%) was not significantly different from chance (see supplementary materials for details).

networks have been identified as being responsible for attentional control (e.g., Corbetta et al., 2008). If habitual action video gaming alters a specific network one would expect that network to be an important feature in a classification model. Habitual action video gaming could however have broader effects on brain function and alter multiple networks or even the relationships between those networks.

To determine how each network and connectivity between networks contributes to the model's classification accuracy, we performed permutation importance analysis on the 6 top performing classifiers—those that perform better than chance level. The permutation feature importance method assigns an importance score to each input feature by evaluating how much randomly shuffling the values of that features would decrease the model's classification accuracy (for details, see section "Model diagnostics").

The permutation importance results are displayed in Figure 5.3. When focusing on the best model (in terms of classification accuracy)—that is the model that uses the Dosenbach2010 parcellation and the partial correlation metric—it is clear that the connectivity between the cingulo-opercular network and the sensorimotor network (CON-SMN) is the most important feature. The second most important feature is the connectivity between the fronto-parietal network and the sensorimotor network (FPN-SMN).

It is interesting, and perhaps surprising even, that the best performing model is one where the connectivity within individual brain networks that have previously been associated with cognitive control, in particular FPN and CON, is discarded (i.e., the within-network connectivity is quantified only when using the tangent or precision connectivity metric). Connectivity within networks, more specifically within CON, is only ranked third in the third best performing model (i.e., when using the tangent as the connectivity metric on the Dosenbach2010 atlas); in all other cases, the influence of individual networks seems negligible.

Overall, it appears that the relationships between networks play a much bigger role in discriminating AVGPs from NVGPs than the networks themselves (e.g., the importance of FPN is negligible). In particular, the present analysis suggests that habitual action gaming may affect how cognitive control networks (FPN and CON) interface with the sensorimotor network (SMN).

5.7.3 Key results are robust to changes in the data analysis pipelines.

Are these results robust to changes in parcellation and connectivity metric? Answering this question is somewhat challenging because different atlases identify different networks with different semantic interpretations which thus leads us to somehow compare apples to oranges. This being said, when considering the cases using the Dosenbach2010 atlas, it appears that the results are very reliable (see Figure 5.3). Indeed the top two features—which involve inter-network connectivities—are the same across variations in connectivity metric. When considering the cases using the Gordon2014 parcellation, the results highlight again the importance of relationships between networks. However, the specific networks are somewhat different. In particular, in these cases we observe that the connectivity between the Auditory network and the FPN network has the highest impact on classification accuracy (note that Dosenbach2010 does not include an Auditory network). The consistency of the results across variation of connectivity metric is however greatly reduced when using Gordon2014 parcellation rather than Dosenbach2010. One of the factors that determines this consistency is the model's prediction accuracy (i.e., models closer to chance performance will yield less consistent feature importance ranks) and thus our interpretation of the results should weight feature importance ranks by the models' classification accuracy.

5.8 Discussion

In this study we have shown that using resting-state functional brain connectivity it is possible to reliably classify new participants (i.e., participants whose data were not used to train the classifier) as a habitual action video gamer player (AVGPs) or a non-video game player (NVGPs). This result is important for several reasons. First, these differences in resting-

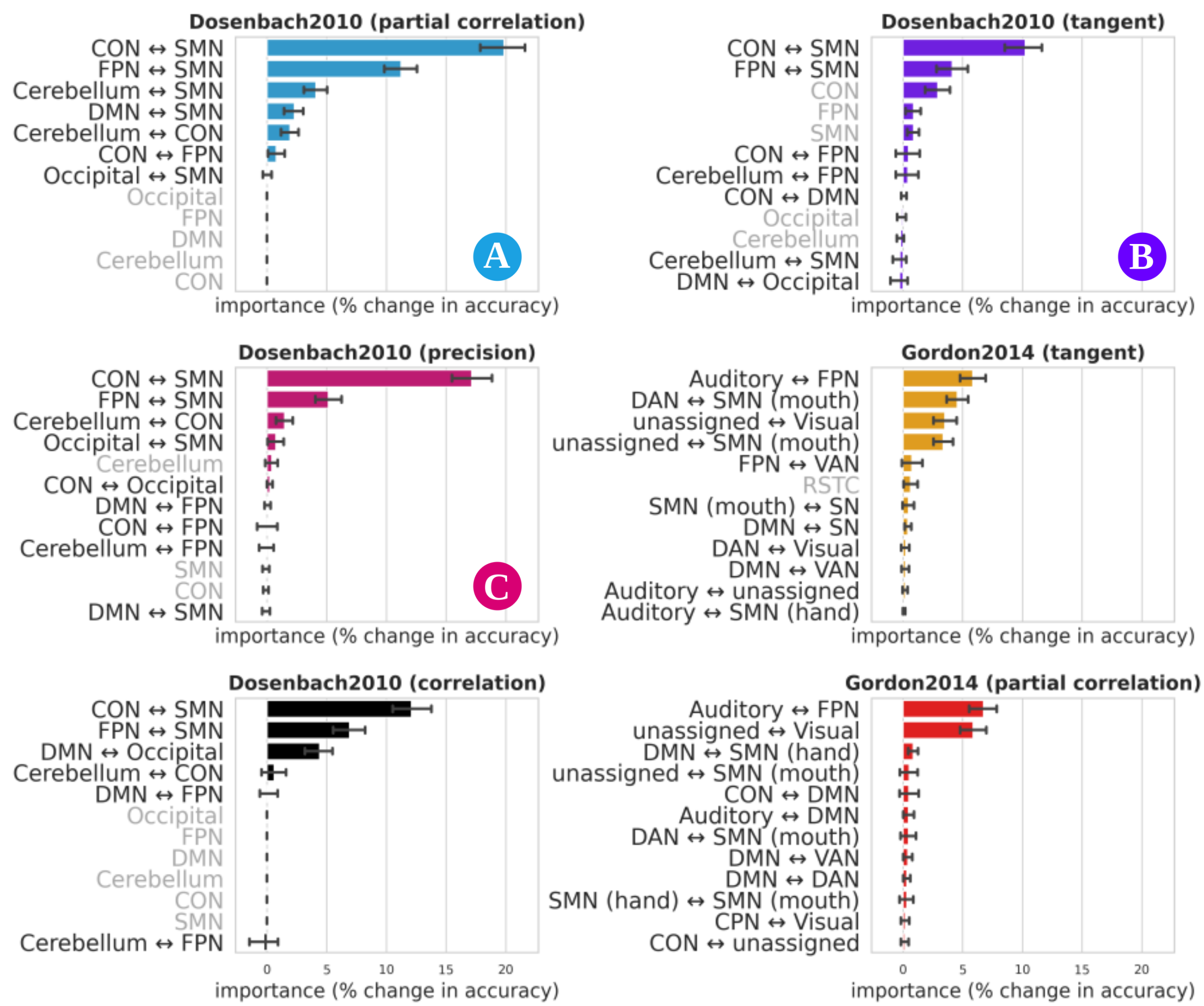


Figure 5.3: Permutation features importance of the top 6 AVGPs versus NVGPs classification models ordered by classification accuracy (see Figure 5.2). Each panel shows the 12 most important features (ordered by importance) for a given classifier, which is characterized by an atlas (i.e., Dosenbach2010 versus Gordon2014) and a connectivity metric (e.g., partial correlation, precision). Error bars represent 95% confidence intervals.

state data provide additional support to the growing literature documenting the correlates and consequences of action video game play, and offer new insights regarding the underlying neural mechanisms. Second, this result supports the notion that resting-state data may be used to study the correlates and consequences of action video game play (and possibly other forms of media consumption) on brain function in a way that is both time-effective and less contaminated by potential expectation and placebo effects (Boot et al., 2011). Finally, this result suggests that resting-state brain connectivity data may be an invaluable tool in the quest to develop effective cognitive training programs. The rapid measurement of changes in brain connectivity may be able to detect subtle training-induced effects (with the specific pattern of brain changes being likely related to the breath of transfer). In addition, resting-state connectivity can easily be measured repeatedly (for example to assess dose-response curves; Chopin et al., 2019). This is in stark contrast to traditional behavioral measures where participants may get better at a cognitive test each time they are exposed to that same test, confounding the benefits of the training program with the learning effects on a specific cognitive test (Green et al., 2019, 2014).

The second main result of this study concerns the overall patterns of brain connectivity that are important in classifying new participants as AVGPs versus NVGPs and how these patterns relate to current theories of cognitive training and transfer using action video games. We group current theories into three main families. The first family assigns action video gaming effects to improvements in specific brain areas and predicts no AVGP vs NVGP differences in resting-state connectivity. The second family of hypotheses, states that action video gaming is associated with improvements in specific functional networks (for example, a more effective dorsal fronto-parietal network supporting top-down visuo-spatial attentional control). Finally, the third family of hypotheses states that action video games affects cognitive control more broadly, which manifests in changes in the relationships between functionally specialized brain networks (i.e., a reconfiguration of brain networks, a more efficient coordination of multiple networks). Our results show very clearly that the main differences in brain connectivity between AVGPs and NVGPs are at this higher-level, inter-network connectivity level. This result is incompatible with views that attribute action video game effects exclusively to specific cognitive processes, or to specific domain-general cognitive functions and also provides some insights about why playing action video games may yield broad transfer effects.

The third key result of this study concerns methodology. Previous work has shown that the results of brain imaging analysis can vary substantially depending on details of those analyses (Botvinik-Nezer et al., 2020). To yield more robust conclusions, we adopted a data analysis strategy that involved testing many combinations of parameters and choices and evaluating the impact of those combinations on the end results (Dadi et al., 2019). In line with past work, we observe indeed that some results are highly dependent on specific methodological choices while others are more robust. More specifically, we tested three parcellation atlases and five connectivity metrics. Our results show that the choice of parcellation atlas has a major impact on a machine learning model's ability to accurately classify participants as AVGPs versus NVGPs: Dosenbach2010 parcellation atlas yielded overall better classification performance than either Gordon2014 and DiFuMo parcellation atlases. This result may seem surprising because DiFuMo is grounded in a much larger data collection than Dosenbach2010. We speculate that the Dosenbach2010 performs best in this context because it is grounded in a more careful selection of tasks. Alternatively, DiFuMo may perform worse because by aggregating data from multiple contexts without formally accounting for context (e.g., within a hierarchical model), DiFuMo may wash out some important distinctions. Regarding connectivity metric, their impact on classification accuracy is also clear, although perhaps less dramatic. For example, quantifying relationships between brain regions or networks led to higher classification accuracy when using partial correlation rather than simple correlations. This result may indicate that although the correlation between two nodes may be high due to external factors (all nodes are co-activated), what seems to matter most is the specific association between nodes that cannot be accounted for by other nodes. More work is needed to understand why some metrics perform better than others. This is not a trivial question and it implies that before a satisfactory response is found, future research should adopt a robust methodology and test multiple connectivity metrics rather than arbitrarily picking a specific one. This being said, our results show rather consistently that the best atlas for our purposes is Dosenbach2010, and that features that are highlighted as important among the best performing models are consistent across variations of connectivity metric. This consistency across parameter variations increases the confidence in the results we report in the next section.

The fourth and final set of results of this study concerns the specific networks and inter-network relationships highlighted by our analyses. Within our set of models, those using the Dosenbach2010 parcellation performed best and some of those using Gordon2014 performed above chance level. When using Dosenbach2010, the most important features in the data to accurately classify participants as AVGPs vs NVGPs were the relationships between the cingulo-opercular network (CON) and the sensori-motor network (SMN) on the one hand, and the relationship between the fronto-parietal network (FPN) and the SMN on the other hand. The FPN and CON networks are hypothesized to work in tandem to provide both the stability and the flexibility required for adaptive cognitive control. More specifically, CON is associated with task-set maintenance that promotes long-term stable control while FPN has been associated with moment-to-moment control that is demanded for flexible, stimulus-driven control. Interestingly, in our results, the direct relationship between these two networks is not discriminative of AVPGs vs NVPGs; what is discriminative, however, is the relationships between these two networks and the sensorimotor network. That is, the predictive performance of this classifier relied mostly on the interplay between control networks and lower perceptual networks rather than activities within a specific brain network. One potential explanation for the observed interplay may lie in the computational mechanisms involved in the connectivities between control networks and lower perceptual networks. Previous research has suggested that the integration of information from multiple brain networks is crucial to successfully exert cognitive control over behavior, as it allows for the flexible use of various sources of information to make predictions (Jiang et al., 2018). For example, the control networks may help prioritize certain sources of information and guide the allocation of computational resources, while the lower sensorimotor networks may provide detailed sensory input and fast motor response. This dynamic interplay between control and perceptual brain networks may be a key factor in the ability of AVGPs to achieve high levels of performance.

The Gordon2014 atlas yielded overall lower classification accuracies and a reduced consistency in the feature importance ranks across connectivity metrics. Yet, this atlas is particularly interesting in the present context because it comprises two networks that are often cited in the context of action video gaming: the dorsal attentional system (DAN) that is responsible for top-down attention and the ventral attentional system (VAN) that is responsible for bottom-up attention. On their own, neither of these two networks seem important to classify participants as AVGPs vs NVGPs during rest. This result seems at odds with other results using task-related fMRI and may (e.g., Bavelier et al., 2012). There are however several potential explanations for this pattern of results. Perhaps there are in fact differences, but they are just less important. Perhaps a better parcellation would yield stronger effects and perhaps there are network relationships that are apparent only during task performance and not during rest. More work is needed to tell these apart.

The most important feature when using the Gordon2014 atlas was the relationship between the fronto-parietal network and the auditory network. To the best of our knowledge, this relationship was not to be expected. We believe that it does not reflect a stable difference between AVGPs and NVGPs but rather is a temporary consequence of the specific task participants completed just prior to the resting-state recording (an attention demanding, auditory Posner-cueing task). This is a very interesting result per se as it suggests that cognitive tasks have a different short-term impact on resting-state connectivity depending on participants' gaming status. It also makes the point that post-task resting state connectivity reconfiguration effects may be an interesting new type of measurement to consider for the study of cognitive control, cognitive training and transfer effects.

5.9 Limitations and future research

The generalizability of our current conclusions are limited by the dataset we have used. Indeed, in this study we used only a single dataset, which included a limited number of participants and a relatively short resting-state recording period. The methods developed in this study can however easily accommodate additional datasets and we leave it for future work to replicate and extend the present results.

In addition, in the current dataset, the resting-state data was recorded after participants completed a cognitive task. It is possible that performing that task tainted the resting-state brain activity (Lor et al., 2022). More specifically, participants completed a demanding attention task that required paying attention to auditory cues—the highlighted CON-SEN connectivity when using the Dosenbach2010 atlas and the Auditory-FPN connectivity when using the Gordon2014 atlas may therefore not be intrinsic to participants resting state brain activity but instead reflect AVGPs versus NVGPs neural differences during task performance. To clarify this point, the current analysis must be replicated on a separate dataset where no cognitive task is completed prior to recording resting-state fMRI.

It seems plausible to us that the differences we report here on functional brain connectivity among AVGPs and NVGPs is caused by playing action video games. At this stage, such a statement is however speculative. It will be necessary to run an actual training study to establish a causal relationship between playing action video games, increased inter-network connectivity and behavioral transfer effects. It is also possible that long-term effects of playing action video games which may be observed when studying habitual gamers (like in this study) are rather different from the short term effects that one might observe in cognitive training studies. This calls for caution when interpreting results and for studies combining multiple methods and types of participants.

In this study we established that brain connectivity differed between AVGPs and NVGPs; we did not establish however that these same connectivity indicators simultaneously account for changes in behavioral performance. It could be that brain metrics that are useful to discriminate AVGPs from NVGPs are different from the brain metrics that explain high versus low behavioral performance. Furthermore, while past research has demonstrated a strong overlap between task-induced and resting-state brain connectivity, the possibility remains of there being important differences. Some differences in brain function between AVGPs and NVGPs may only emerge during task performance while some differences observed during resting-state may vanish when people engage in a specific task. Again, we leave these important questions for future research.

Our results are in line with a growing body of work in highlighting the value of using graph-theoretic approaches to study brain function and its relationships with cognition (Zink et al., 2021). While there has been tremendous progress in this approach over the past decade, more work is still needed. Of particular value is recent theoretical work aiming to explain cognitive control from a network perspective (Menon & D'Esposito, 2022). This type of work is important not only to the study of the effects of action video gaming, but more generally to our understanding of how the human brain enables intelligent behavior.

5.10 Conclusion

By unveiling the mechanisms underlying the effects of playing action video games on brain function we can further our understanding of transfer of cognitive training and devise more effective training programs for positive societal impact. The results of this study show that new participants can be accurately classified as habitual action video game players or non-video game players based on their resting-state functional brain connectivity. What distinguishes the brain connectivity most between these two groups of people are not changes in isolated brain regions or even functional networks but rather the cross-talk between multiple networks, in particular between cognitive control networks on the one hand and a sensorimotor network on the other. These results are important because they suggest that the broad cognitive transfer effects observed after training with action video games may result from a reconfiguration of cognitive control networks.

5.11 Supplementary Materials

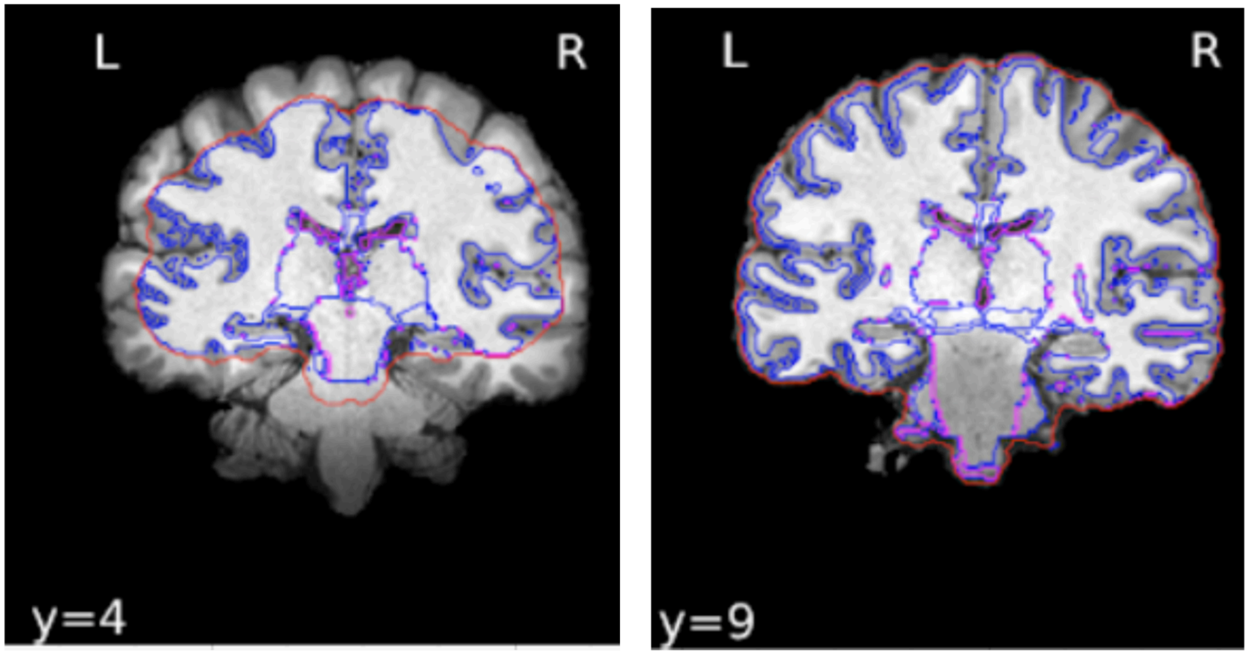


Figure 5.4: The effect of skipping skull stripping. It was necessary to skip the skull stripping step of the preprocessed T1w images of MRIQC because the scans were already defaced. The left panel in this figure shows a scan with skull stripping and the right panel, without skull stripping. As can be seen in this figure, by skipping skull stripping the recognition of the brain volumes became more accurate.

5.11.1 Parcellations

5.11.1.1 Dosenbach2010 parcellation atlas

Figure 5.5 shows the networks as defined in Dosenbach2010 parcellation atlas. For a full list of regions, their MNI coordinates, and corresponding networks, see (Dosenbach et al., 2010; Nilearn Team, 2022b).

5.11.1.2 Gordon2014 parcellation atlas

Figure 5.6 shows the networks as defined in Gordon2014 parcellation atlas. For a full list of regions, their MNI coordinates, and corresponding networks, see (Gordon et al., 2016).

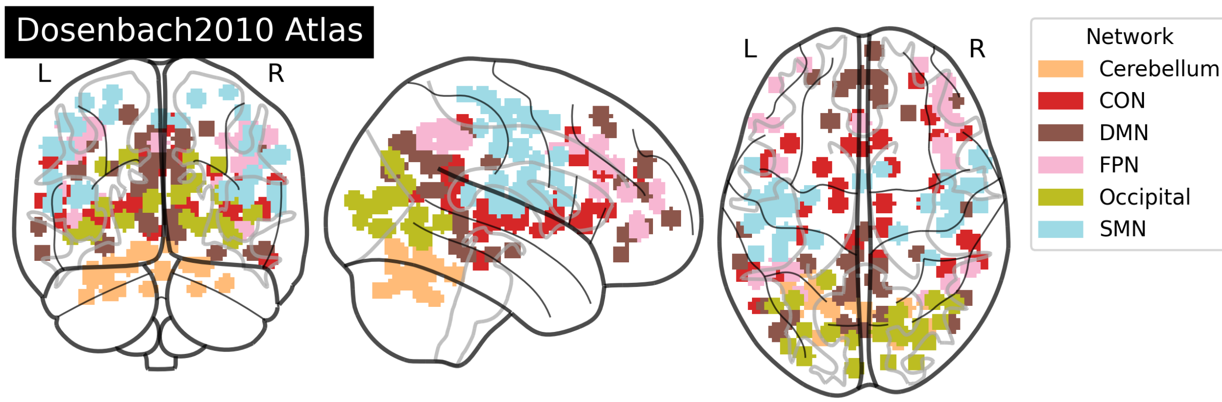


Figure 5.5: Dosenbach2010 networks.

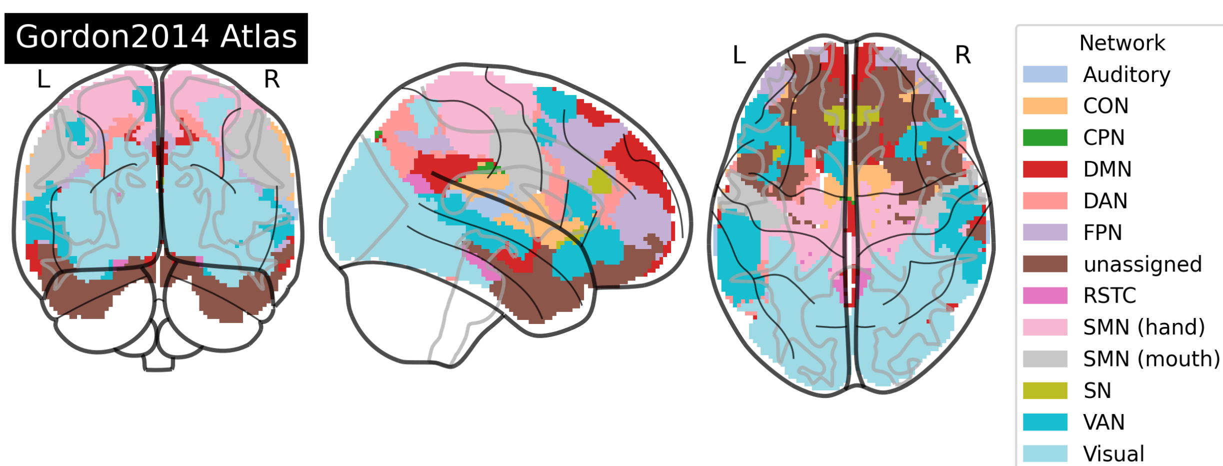


Figure 5.6: Gordon2014 networks.

5.11.1.3 DiFuMo64 parcellation atlas

Figure 5.7 shows the networks as defined in DiFuMo64 parcellation atlas. For a full list of regions, their MNI coordinates, and corresponding *Yeo2011-17* networks, see (Dadi et al., 2020; Nilearn Team, 2022a; Yeo et al., 2011).

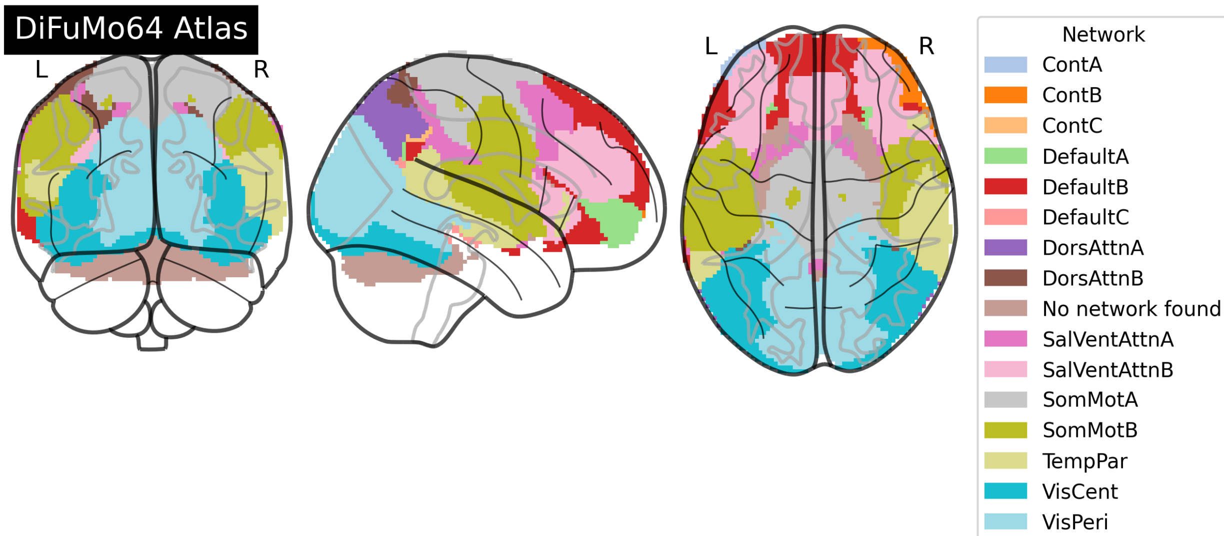


Figure 5.7: DiFuMo64 networks.

5.11.2 Motion signals during resting state fMRI recording do not differentiate AVGPs from NVGPs

Participants' motion is a major confound in the analysis of resting-state functional connectivity. It can create spurious functional connectivity particularly when there are systematic differences between groups of participants (Powers & Brooks, 2014). Previous research has highlighted sensorimotor differences between AVGPs and NVGPs (Gozli et al., 2014); these differences may be masked or confounded with behavior induced brain activations during resting-state. Hence, before interpreting group differences in functional connectivity it is important to assess participants' movement behavior so that functional connectivity group differences can be accurately interpreted as genuine differences in brain function rather than as movement-induced artifacts.

To ensure that group differences in functional connectivity can be attributed to the cognitive functions rather than motion, we extracted motion-related data (6 variables; see Fox et al. (2005) for more details) and used that data to train a support vector machine (SVM) to classify people as AVGPs vs NVGPs. The rationale of this analysis is that if the motion data differs among these two groups of participants, it should be possible to classify participants as AVGPs vs NVGPs based on their motion patterns. If instead, there are no differences in motion behavior between these two groups of participants, the classifier should perform at chance level.

We trained a binary support vector machine (linear L1-regularized SVM) to classify participants as AVGP or NVGP based on their motion confounds. The accuracy of the classifier was evaluated on out-of-sample test data (100-repeated 4-fold cross validation). The results show that the performance of the classifier is not significantly different from chance (accuracy = 51%; see Figure 5.2). This suggests that motion confounds in habitual action video gamers and non-gamers are equivalent and that group differences in functional brain connectivity are unlikely related to group differences in motion behavior. Following standard practice, we removed the motion confounds from the resting-state signals (see the "Preprocessing" section).

5.11.3 Classifying habitual AVGP using intrinsic functional connectivities depends on the parcellation technique as well as the connectivity metric

In this study we used a robust methodology, testing multiple parcellations and connectivity metrics. Here we want to quantify how these different choices impact the results (i.e., the accuracy of the AVGPs vs. NVGPs classifier). To do so, we used a Bayesian model that estimated the effect of parcellation choice and connectivity metric choice on classification accuracy. The Bayesian model aimed to fit the data using the following formula:

$$y \sim P + C + P : C$$

where y represents prediction accuracy (in percent), P is categorical variable representing the choice of parcellation atlas (three levels including Dosenbach2010, Gordon2014, and DiFuMo), C is a categorical variable representing the choice of connectivity metric (five levels including Pearson's correlation, partial correlation, tangent, covariance, and precision), and $P : C$ is the interaction between parcellation and connectivity metric.

We used the evaluation scores from the cross-validated classification pipeline described in Methods section (100-repeated 4-fold cross-validation), which resulted in 100 measurements per each combination of P and C (in total, 1500 data points for y). We then used the Bambi package (v0.9.2; Capretto et al. (2022)) to fit the Bayesian model depicted in Figure 5.8. As shown in the graph, we contrasted all choices for P against *DiFuMo* as the baseline reference, and choices for C against *correlation* as the baseline reference. To estimate the posterior distributions, we used NUTS ("no U-turn sampler") with 4 chains, 500 tuning samples (discarded before sampling from posteriors), and 2000 samples drawn from the posterior.

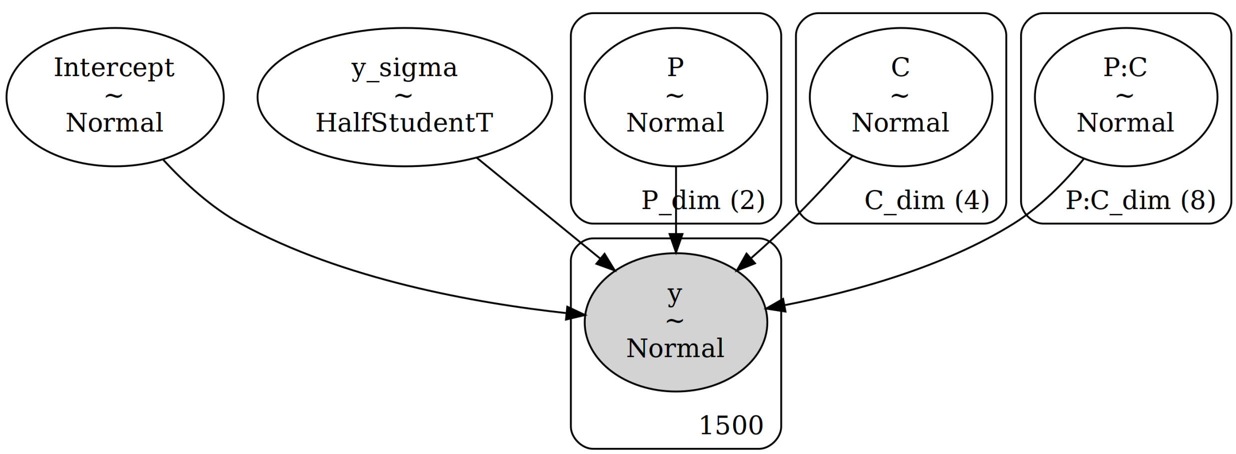


Figure 5.8: Bayesian model fitted to the choice of atlas (P), choice of connectivity metric (C), and prediction accuracy (y); See Formula Supp-1. We used full-rank coding of categorical variables (P and C), with C=`correlation` and P=`DiFuMo64` being the baseline references.

The results are shown in the Table 5.1 and Figure 5.9 below. Overall, they show that choices of parcellation atlas and connectivity metric have a big impact on the results. For our purposes, the best parcellation atlas is Dosenbach2010 and the best connectivity metric is the partial correlation.

Table 5.1: A Bayesian model comparison analysis shows that the choice of parcellation atlas affects classification accuracy most. In general, choosing Dosenbach2010 atlas and precision connectivity metric leads to the highest classification accuracy. Results from a "y ~ P * C" model (which reads "accuracy ~ atlas * metric") are shown in the table. Note that the table shows contrasts against the baseline reference of correlation connectivity metric and DiFuMo64 atlas.

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
Intercept	2.948	0.125	2.72	3.19	0.002	0.002	2610	4279	1
P:C[Gordon2014, tangent]	1.346	0.25	0.858	1.785	0.004	0.003	3358	5024	1
y_sigma	1.235	0.023	1.193	1.279	0	0	12350	5695	1
P:C[Gordon2014, covariance]	1.006	0.247	0.557	1.48	0.004	0.003	3576	5031	1
P:C[Dosenbach2010, tangent]	0.987	0.249	0.504	1.45	0.004	0.003	3715	4887	1
P:C[Dosenbach2010, partial_correlation]	0.874	0.247	0.406	1.338	0.004	0.003	3532	5360	1
P[Dosenbach2010]	0.873	0.175	0.531	1.194	0.003	0.002	2931	4186	1
P:C[Gordon2014, partial_correlation]	0.746	0.249	0.318	1.256	0.004	0.003	3408	5267	1
C[precision]	0.583	0.176	0.248	0.903	0.003	0.002	3343	4276	1
P:C[Gordon2014, precision]	0.157	0.248	-0.287	0.651	0.004	0.003	3516	4153	1
C[partial_correlation]	0.113	0.175	-0.215	0.436	0.003	0.002	3114	5076	1
P[Gordon2014]	-0.008	0.178	-0.34	0.333	0.003	0.002	2791	4114	1
P:C[Dosenbach2010, covariance]	-0.283	0.246	-0.729	0.199	0.004	0.003	3611	5076	1
P:C[Dosenbach2010, precision]	-0.295	0.248	-0.764	0.165	0.004	0.003	3510	4848	1
C[tangent]	-0.307	0.174	-0.652	-0	0.003	0.002	3289	5000	1
C[covariance]	-0.376	0.175	-0.725	-0.061	0.003	0.002	3361	4674	1

5.11.4 SHAP Analysis

The permutation feature importance method presented in the main text identifies the importance of individual features in the machine learning model that predicted AVGPs vs NVGPs. Alternatively, SHAP (SHapley Additive exPlanations) values are a method for explaining the *output* of a machine learning model. They provide a breakdown of the contribution of each feature to the model's output, taking into account the interactions between features. The main difference between permutation importance and SHAP values is that permutation importance only considers the effect of a single feature on model performance, while SHAP values consider the effects of all features and their interactions. Additionally, permutation importance is a measure of feature importance, while SHAP values are a method for explaining model predictions.

Thus, here we ask a somewhat complementary question to the feature importance analysis:

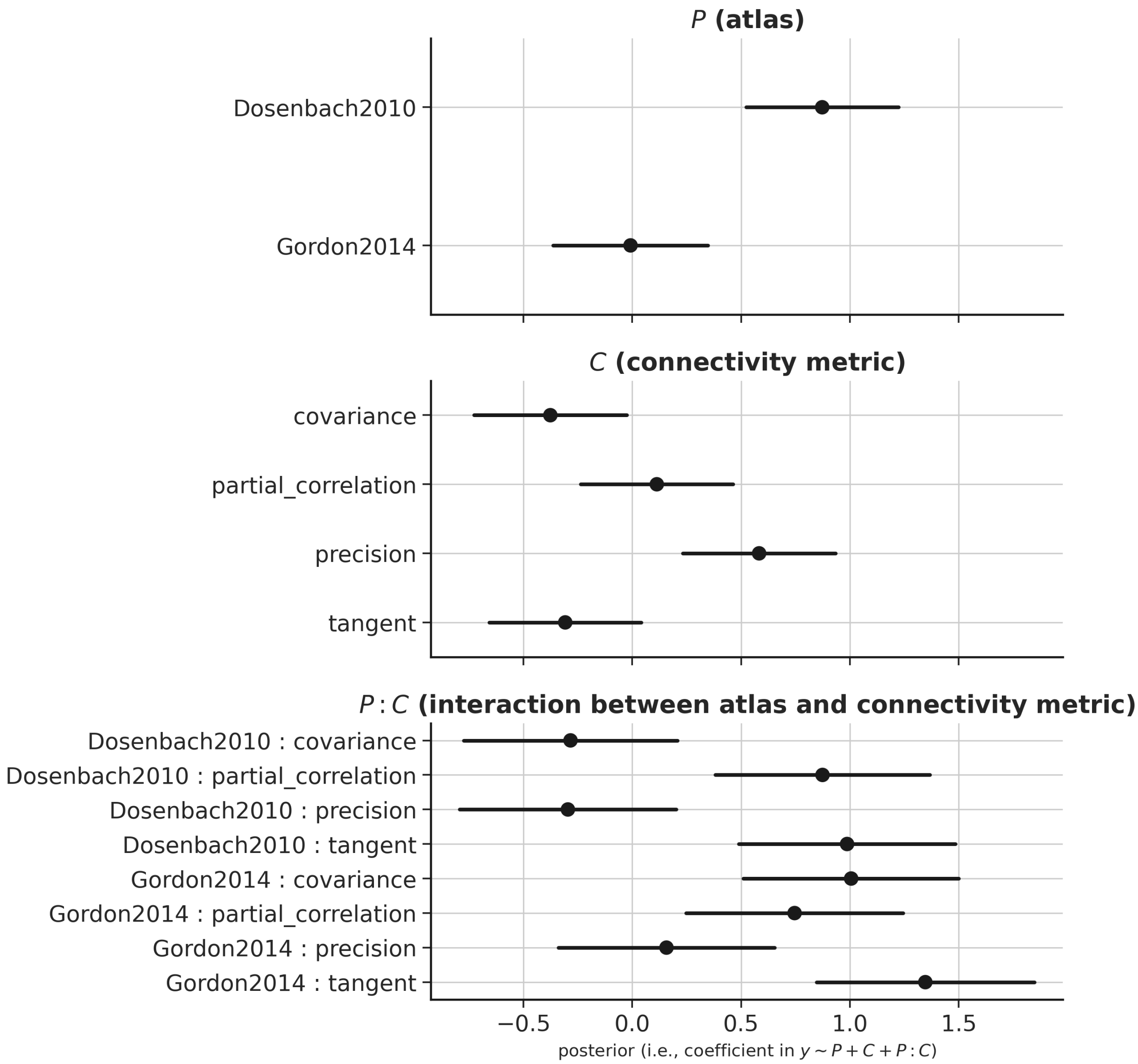


Figure 5.9: Comparing the choice of atlas and connectivity metric on classification performance. Error bars represent 2 standard deviations. We used full-rank coding of categorical variables with baseline reference being correlation for connectivity metrics (C=`correlation`) and DiFuMo for parcellation atlases (P=`DiFuMo64`). Intercept and baseline references are not shown.

what role do features play in the choices made by the classifier? For example, which features determine most misclassifications?

We applied SHAP analysis to assess the importance of individual features on classification output (e.g., in binary classification, probabilities of assigning a given observation to two possible outcomes) while considering the effects of other features and their interactions (Lundberg & Lee, 2017). Note that we only report here the results of the SHAP analysis on the best performing classification model (i.e., Dosenbach2010 model with partial correlation connectivity metric; see main text for details). The results of this analysis are illustrated in Figure 5.10. As in the permutation importance analysis, the three most important features in SHAP are the CON-SMN, FPN-SMN, and CON-FPN connectivities.

Next, we ask which features contribute most to misclassifying participants. To address this, we used SHAP values to investigate all the predictions regardless of their correctness and differentiate "important" features from "misleading" ones. This is enabled by calculating the contribution of features to misclassified predictions (misses) and comparing the ranking of features against the ranking in correctly classified predictions (hits). In our case, SHAP values for misclassified predictions can identify the connectivities that may be responsible for misclassifying non-video game players (NVGPs) as video game players (AVGPs) – potentially through superior cognitive abilities that result in NVGP connectivity patterns being more similar to those of AVGPs, possibly through compensating cognitive abilities such as expertise in music, sports, or other types of video games (see Föcker et al., 2018 for details on subjects' expertise).

As shown in the Figure 5.10, misclassified outputs also relied on a similar set of features as correct classification. But the ranking of features based on their importances is slightly different between correct and incorrect predictions. For the correct predictions, the order of importance as measured by absolute mean SHAP values matches the ranking of features produced by permutation feature importance (see Figure 5.2 in the main text); yet for incorrect predictions, the order is not the same, suggesting some other network connectivities (rather than CON-SMN and FPN-SMN) may interfere. One important disparity between the correct classifications and incorrect ones is the connectivity between FPN-CON, which shows a stronger contribution to the prediction output of the misclassified subjects (it is ranked 6 in correct predictions but ranked 3 in incorrect ones). This compensatory role of the connectivity between two control networks (frontoparietal and cingulo-opercular networks) may imply improved cognitive control in some non-video game players or, conversely, could imply automatization (hence reduced connectivity) between CON and FPN in some non-video game players. However, more research is needed to fully understand the role of CON-FPN in habitual action video game players and cognitive control.

In brief, this specific analysis showed that CON-SMN, FPN-SMN, and FPN-CON connectivities contribute the most to the prediction, regardless of its correctness. This result provides additional support for the previously presented results in the main text that habitual action video gaming may impact cognitive functioning by influencing the cross-talk between control and sensorimotor networks rather than activities within individual networks. This implies that attentional and cognitive control, if in fact targeted by playing action video games, relies on a distributed set of large-scale brain networks, each with distinct cognitive functions.

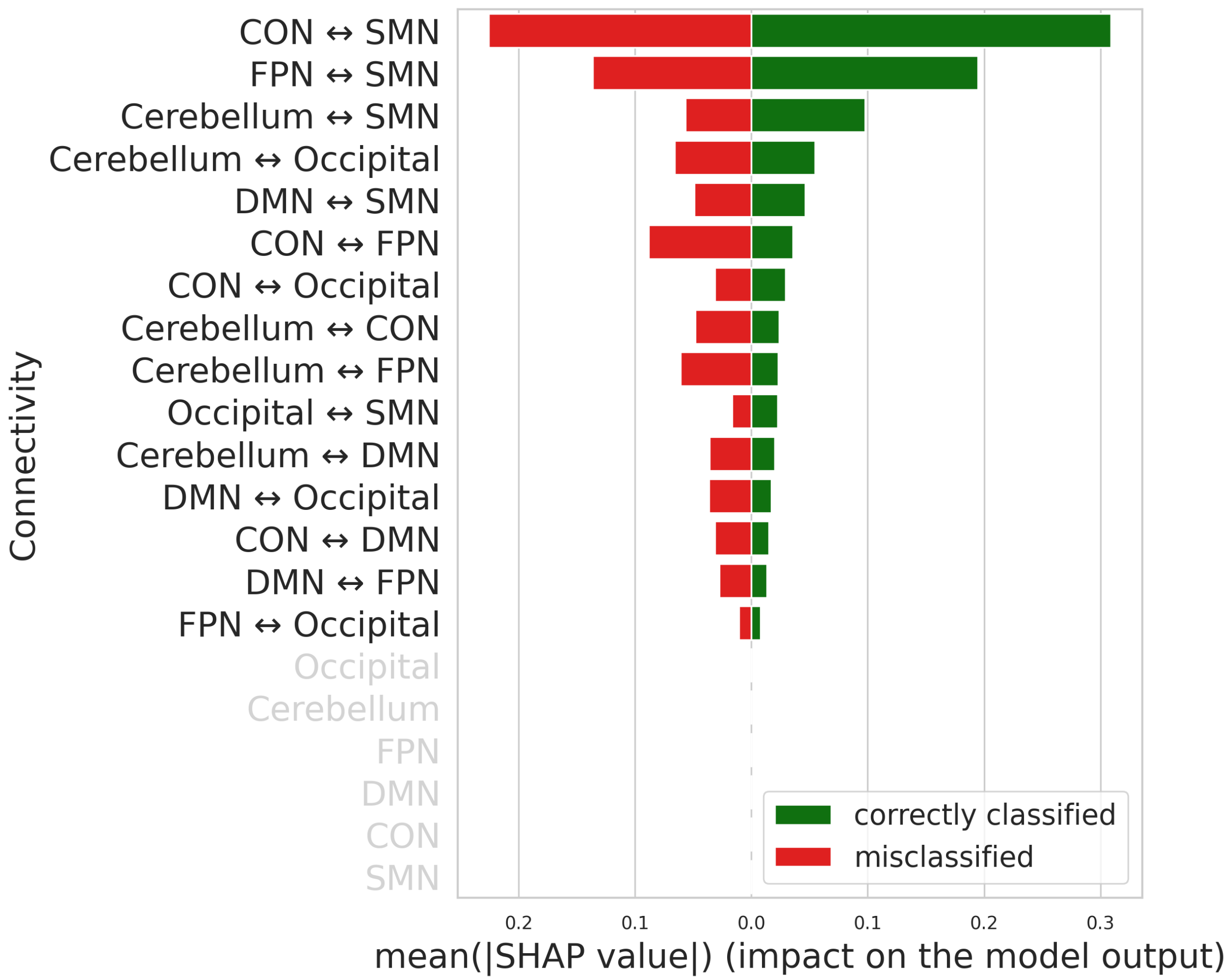


Figure 5.10: Shap values for correct (green) and incorrect (red) classifications of participants as AVGPs or NVGPs. The plot reads from top to bottom, showing the impact of each connectivity to the model output (i.e., AVGP vs NVGP classification probabilities). Network features are ordered, from top to bottom, by their average importance (`mean(|SHAP|)`).

General Discussion

On the importance of cognitive control research

Psychology is tasked to make sense of what humans do, and what humans do depends on what happens in their immediate environment (G. Miller et al., 1960). One ability that is of utmost importance to human functioning is to exercise cognitive control which enables pursuing goals in a changing world, avoiding prepotent responses, and effectively generalizing prior experiences to new situations. Due to its ubiquitous presence in everything we do, cognitive control plays a crucial role in our daily lives, long-term achievements, and health. Accordingly, the possibility to enhance cognitive control in a way that transfers to real life situations could have important implications.

Progress towards developing effective cognitive control training programs is however limited by the lack of a formal, quantitative definition of cognitive control. The main challenges that this thesis aims to address are (a) to gain greater clarity on the cognitive control constructs (what it is and how to measure it), and (b) to understand what features of the cognitive system (i.e., the agent) and what features of the task (i.e., the environment) determine cognitive control, its functioning, and generalization.

On the importance of a multidisciplinary view of cognitive control

To address these challenges, this thesis relies on the multidisciplinary synergy within cognitive sciences, primarily between artificial intelligence, psychology, and cognitive neuroscience.

This synergy is apparent at several levels. First, we apply artificial intelligence techniques as mere tools in our toolbox to interpret human data. In this sense, modern machine learning models provide new insights on human cognition as they are applied to behavioral data, scientific documents, and neuroimaging data. Second, a richer form of interdisciplinary synergy allows us to build bridges between disciplines, to develop new computational models that instantiate cognitive control and generalize across tasks, furthering our understanding of cognitive control in humans.

Defining cognitive control

On the importance of defining and quantifiying cognitive control

Concepts that capture higher-order cognitive abilities such as cognitive control are difficult to define—and consequently to quantify. To understand those cognitive abilities, previous research has devised a variety of theoretical constructs and cognitive tasks, the relationships between which are not always clear. Chapter 1 (CogText) is an attempt to quantitatively assess this lack of a cohesive understanding by using recent advances in artificial intelligence. More specifically, we performed a large-scale text analysis to create a knowledge graph that relates theoretical constructs and empirical tasks about cognitive control. The rationale of this analysis is that constructs are related to each other to the extent they are assessed using a similar set of cognitive tasks and, conversely, cognitive tasks are similar to the extent they are thought to involve similar cognitive constructs. As expected, the knowledge graph confirms the complex nature of cognitive control and illustrates two specific phenomena that may explain the difficulty of defining cognitive control: task impurity (tasks measuring multiple constructs) and construct hypernomy (multiple ways of defining and measuring constructs). These results have several implications for the study of cognitive control. First, greater theoretical clarity is needed on cognitive control—this may be achieved by adopting a more formal approach grounded in computational modeling. Second, there is currently no single task capable of assessing cognitive control on its own, indicating a need for better assessment environments. This could entail, for instance, assessing cognitive skills using a battery of tasks (varying contexts and demands) or to develop better, perhaps more complex tasks (e.g., video games). Finally, because cognitive control is not associated with a single cognitive function but rather involves interactions with many cognitive functions in multiple tasks, it is likely that cognitive control is associated with a range of large-scale brain networks as opposed to a single brain area or network.

On the importance of an interoperable battery of tasks for humans and artificial agents (CogEnv)

There have been significant advancements in both artificial intelligence and psychology, but they have not yet been fully integrated. This may be due to a lack of appreciation for their relevance, or the lack of tools to directly compare the behavior of humans and artificial agents. While there are many examples in the scientific literature of human behavior being compared to specific artificial agents, this type of comparison is typically done at the level of a single task, using a limited set of computational agents. Furthermore, these comparisons are not developed in a way that allows for reuse or extension.

To study cognitive control and other cognitive processes, it is necessary to be able to systematically compare the behavior of humans and artificial agents across multiple tasks. A tool that allows humans and artificial agents to perform the same set of tasks and directly compare their behavior would greatly benefit our understanding of cognitive control in both psychology and computer science.

To facilitate testing and integration of multidisciplinary theories in an interoperable environment, Chapter 2 provides a virtual environment called CogEnv. CogEnv lets both humans and artificial agents perform the same battery of cognitive tasks, providing data that can be directly compared in typical psychological experiments. As a proof of concept, we trained baseline RL agents to perform a battery of cognitive control tasks, and also collected human data for comparison. The overall framework is operational and appears promising. A preliminary investigation suggests that comparing the performance and error profiles of human versus baseline RL agents may reveal aspects of human cognitive control that have not yet been addressed by artificial agents.

On the importance of artificial models that act and functionally decouple control from the controlled act (CogPonder)

The goal of CogEnv is to allow for the direct comparison of computational agents and humans performing exactly the same cognitive tasks and thus to promote more systematic progress in our understanding of cognitive control. Yet, to make progress it is necessary to develop artificial agents that are capable of performing multiple tasks and which provide insights on cognitive control. To this end, in Chapter 3, we developed CogPonder, a computational framework for general cognitive control. . It is a flexible, differentiable end-to-end deep learning model that decouples the act of control from the controlled act and that can learn to perform the same cognitive tests that are used in cognitive psychology to test humans.

The goal of CogEnv is to allow for the direct comparison of computational agents and humans performing the same cognitive tasks, thus promoting more systematic progress in our understanding of cognitive control. To achieve this goal, it is necessary to develop artificial agents that are capable of performing a battery of cognitive tasks while being constrained by computational requirements of cognitive control.

To this end, Chapter 3 presents CogPonder, a computational framework for general cognitive control. CogPonder is a flexible, differentiable end-to-end deep learning model that separates the act of control from the controlled act, and can be trained to perform the same cognitive tests used in cognitive psychology to test humans. We implemented an instance of CogPonder and trained it to perform two cognitive tasks, aligning its behavior with that of humans collected in a previous study. The results show that after training, CogPonder behaves similarly to humans across both tasks in terms of accuracy and response time distributions. These results demonstrate the potential of the CogPonder framework to provide interesting new insights and research opportunities for both psychology and computer science.

Training and generalizing cognitive control

On the importance of cognitive training to study and test cognitive control

Research on the effects of complex tasks, such as video games, on cognitive training may benefit from and contribute to the proposed broad view of cognitive control. Chapter 4 reviews the literature on the effects of different genres of video games on cognition. Action video games, such as first- and third-person shooter games, are particularly interesting because they have been specifically associated with greater cognitive enhancement compared to other types of video games, such as puzzle or life-simulation games. The transfer effects of action video game playing to a range of cognitive tasks have been linked to improvements in reward processing, spatial navigation, and most notably for the context for this thesis, top-down attention and cognitive control.

This review highlights that cognitive training interventions using video games need to be endowed with specific game mechanics to generate cognitive benefits, potentially by enhancing cognitive control abilities. We discuss the potential game mechanics that could be used and call for more systematic research on the relationship between video game mechanics and cognition. We also note that as video games become more advanced and mix different genres and gameplay styles, it will become increasingly difficult to study and understand their effects on cognition. This article lays the foundation for the study of cognitive and brain functioning using video games and illustrates the value of this approach for investigating general cognitive control.

On the importance of studying brain function to understand cognitive control (ACNets)

The study of differences in functional brain networks between action video game players and non-video game players can advance our understanding of the mechanisms underlying the training effects and the neural mechanisms supporting cognitive control in general. In Chapter 5, we show that it is possible to reliably classify new participants as habitual action video game players or non-video game players based on their resting-state functional connectivity. Furthermore, an analysis of the features that are most important for this classification accuracy reveals that what differentiates habitual video game players from non-video game players is not the connectivity within specialized functional brain networks, but rather the relationships between networks, supporting current theories of action video game training that attribute their benefits to domain-general abilities. The results also show that the most important inter-network relationships in this context involve control-related and sensorimotor networks, specifically, the relationships between the cingulo-opercular and the sensorimotor networks, and between the fronto-parietal and the sensorimotor networks.

Because these results suggest that action video game play affects cognitive control, they have important implications for the study of cognitive training. Furthermore, by demonstrating that resting-state data contains information related to habitual action video gaming, these results suggest that resting-state data could be a valuable tool for studying cognitive training effects and their potential for transfer, potentially leading to the development of more effective cognitive training programs. Additionally, these results have practical value for cognitive scientists studying cognitive control, as they imply that action video game training may be a new tool for causally studying cognitive control.

Future perspectives

There are a number of limitations to the work presented in this thesis. One major limitation of CogEnv is its scope. For example, in its current form, CogEnv ignores the real-time nature of most cognitive environments by providing a turn-based mechanism that suspends the environment until the agent finishes its computations and generates an action—a situation such as this is unlikely to occur in real life. Possible directions for addressing this limitation include extending the capabilities of CogEnv by establishing a larger library of computational models that can interface with the environment perhaps in real-time, such as CogPonder (or more broadly real-time reinforcement learning Ramstedt & Pal, 2019), and creating a more comprehensive set of cognitive tests (e.g., tests that are explored in Chapter 1, Enkavi et al., 2019, and the Behaverse cognitive assessment battery; see behaverse.org). Additionally, there is potential for further refinement of the data processing and analysis pipelines, and the adoption of more standardized data formats that support multiple tasks and experimental designs (see for example Behaverse data model in Appendix B).

Another area for future research could be the development of more realistic and ecologically valid cognitive tasks and experiments (cf. Discussion in Chapter 3 and Chapter 5), such as the use of video games or other rich and challenging environments, as well as the use of more complex and dynamic scenarios. By testing cognitive models in more realistic environments, we can better understand the limitations and capabilities of these models, and improve their performance.

This may also require a more general computational account of cognitive control. By developing a computational model of cognitive control that is applicable to a wide range of tasks, we may be able to better understand and improve cognitive control in both humans and artificial agents. One particularly interesting approach involves the integration of artificial intelligence techniques into cognitive modeling. This could be facilitated by applying scalable machine learning (e.g., deep learning) to complex cognitive models, in order to create more accurate and comprehensive simulations of the human brain and mind. By testing sophisticated cognitive models at scale, we can better understand the limitations and capabilities of these models, and improve their ability to explain human phenomena. Another useful approach consists of developing better taxonomies, concepts, and tasks—ideally via collaborative efforts of researchers in neuroscience, experimental psychology, and artificial intelligence—which may lead to more comprehensive and consistent models of cognitive functions. For example, psychologists can provide detailed ontologies of cognitive processes, while neuroscientists provide insights into the underlying brain mechanisms that support those processes, and computer scientists develop new algorithms and technologies for modeling and testing those processes.

A multidisciplinary synergy may also be achieved through direct comparisons of human and artificial agents. By using advanced artificial intelligence techniques, it is possible to create artificial agents that mimic human cognitive processes. By comparing the performance of these agents to human subjects on a variety of cognitive environments, we can better understand the similarities and differences between human and artificial cognition, and develop more accurate and comprehensive models of the human mind. The use of previously unexplored experimental techniques is another important direction for future research in cognitive control. In particular, the combination of functional magnetic resonance imaging (fMRI) and resting-state fMRI, along with the use of multitask batteries and interventional experimental designs, can provide valuable insights into the mechanisms of cognitive control. Integrating task-driven and resting-state fMRI data has the potential to inform us about the neural basis of cognitive control, and this information can be used to develop scientific theories of cognitive control and identify potential neural markers of cognitive control abilities. By including multitask batteries to assess transfer effects, it is possible to determine how the brain enables the generalization of prior performance on one task to another. This type of insight may contribute to unveiling the mechanisms underlying cognitive control, and to develop theories about how cognitive control abilities are acquired, how and when they generalize, and why some interventions are successful and others are not.

Conclusion

Taken together, the current work explores approaches from a variety of cognitive science disciplines that aim to better understand the concept of cognitive control. I presented cases in which neuroscience, experimental psychology, and artificial intelligence can collaborate to advance our understanding of cognitive control and the challenge of generalizing this capacity to new contexts (i.e., transfer effect). In the age of ubiquitous computing and large datasets, bridging the gap between behavior, brain, and computation has the potential to fundamentally transform our understanding of the human mind and inspire the development of truly intelligent artificial agents.

References

Abelson, R. P. (1995). *Statistics as principled argument.* Lawrence Erlbaum Associates, Inc.

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Muller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine Learning for Neuroimaging with Scikit-Learn. *arXiv*.

Adachi, P. J. C., & Willoughby, T. (2013). More Than Just Fun and Games: The Longitudinal Relationships Between Strategic Video Games, Self-Reported Problem Solving Skills, and Academic Grades. *Journal of Youth and Adolescence*, *42*(7), 1041–1052. https://doi.org/10.1007/s10964-013-9913-9

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences*, *90*(12).

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An Integrated Theory of the Mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Angelov, D. (2020). *Top2Vec: Distributed Representations of Topics*. https://doi.org/10.48550/ARXIV.2008.09470

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., & Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, *501*(7465), 97–101. https://doi.org/10.1038/nature12486

Ansarinia, M., Schrater, P., & Cardoso-Leite, P. (2022). Linking theories and methods in cognitive sciences via joint embedding of the scientific literature: The example of cognitive control. *arXiv Preprint arXiv:2203.11016*. https://arxiv.org/abs/2203.11016

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., & Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends in Cognitive Sciences*, *16*(12), 584–592. https://doi.org/10.1016/j.tics.2012.10.008

Antzaka, A., Lallier, M., Meyer, S., Diard, J., Carreiras, M., & Valdois, S. (2017). Enhancing reading performance through action video games: The role of visual attention span. *Scientific Reports*, *7*(1), 14563. https://doi.org/10.1038/s41598-017-15119-9

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*(2), 366–377.

Badre, D. (2011). Defining an ontology of cognitive control requires attention to component interactions. *Topics in Cognitive Science*, *3*(2).

Badre, D. (2020). *On Task: How Our Brain Gets Things Done*.

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and Operationalization of Executive Function: Executive Function. *Mind, Brain, and Education*, *10*(1), 10–33. https://doi.org/10.1111/mbe.12100

Ball, K., Owsley, C., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1993). Visual attention problems as a predictor of vehicle crashes in older drivers. *Investigative Ophthalmology & Visual Science*, *34*(11), 3110–3123.

Banino, A., Balaguer, J., & Blundell, C. (2021). *PonderNet: Learning to Ponder*. https://doi.org/10.48550/ARXIV.2107.05407

Baniqued, P. L., Kranz, M. B., Voss, M. W., Lee, H., Cosman, J. D., Severson, J., & Kramer, A. F. (2014). Cognitive training with casual video games: Points to consider. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.01010

Baniqued, P. L., Lee, H., Voss, M. W., Basak, C., Cosman, J. D., DeSouza, S., Severson, J., Salthouse, T. A., & Kramer, A. F. (2013). Selling points: What cognitive abilities are tapped by casual video games? *Acta Psychologica*, *142*(1), 74–86. https://doi.org/10.1016/j.actpsy.2012.11.009

Barch, D. M., Berman, M. G., Engle, R., Jones, J. H., Jonides, J., MacDonald, A., Nee, D. E., Redick, T. S., & Sponheim, S. R. (2009). CNTRICS final task selection: Working memory. *Schizophrenia Bulletin*, *35*(1), 136–152. https://doi.org/10.1093/schbul/sbn153

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging*, *23*(4), 765–777. https://doi.org/10.1037/a0013494

Bastian, C. C. von, Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kałamała, P., Meier, M. E., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M., Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T., Gustavson, D. E., Koch, I., … Wiemers, E. A. (2020). *Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations* (No. 10.31234/osf.io/x3b9k). PsyArXiv. https://doi.org/10.31234/osf.io/x3b9k

Batou, A., & Soize, C. (2013). Calculation of Lagrange Multipliers in the Construction of Maximum Entropy Distributions in High Stochastic Dimension. *SIAM/ASA Journal on Uncertainty Quantification*, *1*(1), 431–451. https://doi.org/10.1137/120901386

Battiston, F., Amico, E., Barrat, A., Bianconi, G., Ferraz de Arruda, G., Franceschiello, B., Iacopini, I., Kéfi, S., Latora, V., Moreno, Y., Murray, M. M., Peixoto, T. P., Vaccarino, F., & Petri, G. (2021). The physics of higher-order interactions in complex systems. *Nature Physics*, *17*(10), 1093–1098. https://doi.org/10.1038/s41567-021-01371-4

Bavelier, D., Achtman, R. L., Mani, M., & Föcker, J. (2012). Neural bases of selective attention in action video game players. *Vision Research*, *61*(C), 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., & Dye, M. W. G. (2010). Children, Wired: For Better and for Worse. *Neuron*, *67*(5), 692–701. https://doi.org/10.1016/j.neuron.2010.08.035

Bavelier, D., & Green, S. (2016). Brain Tune-Up from Action Video Game Play. In *Scientific*

American. https://www.scientificamerican.com/article/brain-tune-up-from-action-video-game-play/. https://doi.org/10.1038/scientificamerican0716-26

Beam, E., Potts, C., Poldrack, R. A., & Etkin, A. (2021). A data-driven framework for mapping domains of human neurobiology. *Nature Neuroscience*, *24*(12), 1733–1744. https://doi.org/10.1038/s41593-021-00948-9

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). 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

Belchior, P., Marsiske, M., Sisco, S. M., Yam, A., Bavelier, D., Ball, K., & Mann, W. C. (2013). Video game training to improve selective visual attention in older adults. *Computers in Human Behavior*, *29*(4), 1318–1324.

Belchior, P., Marsiske, M., Sisco, S., Yam, A., & Mann, W. (2012). Older adults' engagement with a video game training program. *Activities, Adaptation & Aging*, *36*(4), 269–279. https://doi.org/10.1080/01924788.2012.702307

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

Benady-Chorney, J., Aumont, É., Yau, Y., Zeighami, Y., Bohbot, V. D., & West, G. L. (2020). Action video game experience is associated with increased resting state functional connectivity in the caudate nucleus and decreased functional connectivity in the hippocampus. *Computers in Human Behavior*, *106*, 106200. https://doi.org/10.1016/j.chb.2019.106200

Bensoussan, A., Li, Y., Nguyen, D. P. C., Tran, M.-B., Yam, S. C. P., & Zhou, X. (2020). *Machine Learning and Control Theory*. https://doi.org/10.48550/ARXIV.2006.05604

Bird, C. M., & Burgess, N. (2008). The hippocampus and memory: Insights from spatial processing. *Nature Reviews Neuroscience*, *9*(3), 182–194. https://doi.org/10.1038/nrn2335

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., & Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*, *83*, 550–558. https://doi.org/10.1016/j.neuroimage.2013.05.099

Bodson, L. (2017). *Regards sur les activités quotidiennes des jeunes résidents (Regards du STATEC Nr. 15).* Institut national de la statistique et des études économiques (statec).

Boot, W. R., Blakely, D. P., & Simons, D. J. (2011). Do Action Video Games Improve Perception and Cognition? *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00226

Boot, W. R., Champion, M., Blakely, D. P., Wright, T., Souders, D. J., & Charness, N. (2013). Video Games as a Means to Reduce Age-Related Cognitive Decline: Attitudes, Compliance, and Effectiveness. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00031

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

Botvinick, M. M. (2022). Realizing the promise of AI: A new calling for cognitive science. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.08.004

Botvinick, M. M., & Cohen, J. D. (2014). The Computational and Neural Basis of Cognitive Control: Charted Territory and New Frontiers. *Cognitive Science*, *38*(6), 1249–1285. https://doi.org/10.1111/cogs.12126

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., … Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, *582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Brandman, T., Malach, R., & Simony., E. (2020). *The Surprising Role of the Default Mode Network* [Preprint]. Neuroscience. https://doi.org/10.1101/2020.05.18.101758

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., Janowsky, J. S., Taylor, S. F., Yesavage, J. A., Mumenthaler, M. S., Jagust, W. J., & Reed, B. R. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, *130*(4), 746–763. https://doi.org/10.1037/0096-3445.130.4.746

Brick, C., Hood, B., Ekroll, V., & de-Wit, L. (2021). Illusory essences: A bias holding back theorizing in psychological science. *Perspectives on Psychological Science*.

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). Language models are few-shot learners. *arXiv*.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*(3), 205–228. https://doi.org/10.1080/87565640801982312

Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, *29*(6).

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience*, *2 Suppl 1*, S167–79. https://doi.org/10.1016/j.dcn.2011.10.001

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2022). Bambi: A simple interface for fitting bayesian linear models in python. *Journal of Statistical Software*, *103*(15), 1–29. https://doi.org/10.18637/jss.v103.i15

Cardoso-Leite, P., Ansarinia, M., Schmück, E., & Bavelier, D. (2021). Training cognition with video games. In *The Oxford Handbook of Developmental Cognitive Neuroscience*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198827474.013.38

Cardoso-Leite, P., Kludt, R., Vignola, G., Ma, W. J., Green, C. S., & Bavelier, D. (2016). Technology consumption and cognitive control: Contrasting action video game experience with media multitasking. *Attention, Perception, & Psychophysics*, *78*(1), 218–241.

Cherney, I. D. (2008). Mom, Let Me Play More Computer Games: They Improve My Mental Rotation Skills. *Sex Roles*, *59*(11-12), 776–786. https://doi.org/10.1007/s11199-008-9498-z

Chesham, A., Wyss, P., Müri, R. M., Mosimann, U. P., & Nef, T. (2017). What Older People Like to Play: Genre Preferences and Acceptance of Casual Games. *JMIR Serious Games*, *5*(2), e8. https://doi.org/10.2196/games.7025

Chollet, F. (2019). *On the Measure of Intelligence* (No. arXiv:1911.01547). arXiv. https://arxiv.org/abs/1911.01547

Chopin, A., Bediou, B., & Bavelier, D. (2019). Altering perception: The case of action video gaming. *Current Opinion in Psychology*, *29*, 168–173. https://doi.org/10.1016/j.copsyc.2019.03.004

Christian, B., & Griffiths, T. (2016). *Algorithms to live by: The computer science of human decisions* (First international edition). Henry Holt and Company.

Christie, S. T., & Schrater, P. (2019). Understanding the timing of cognitive processes with a variable rate neural code. *2019 Conference on Cognitive Computational Neuroscience*. https://doi.org/10.32470/CCN.2019.1397-0

Chuang, T.-Y., & Chen, W.-F. (2007a). Effect of Computer-Based Video Games on Children: An Experimental Study. *2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07)*, 114–118. https://doi.org/10.1109/DIGITEL.2007.24

Chuang, T.-Y., & Chen, W.-F. (2007b). Effect of Digital Games on Children's Cognitive Achievement. *Journal of Multimedia*, *2*(5).

Ciric, R., Thompson, W. H., Lorenz, R., Goncalves, M., MacNicol, E., Markiewicz, C. J., Halchenko, Y. O., Ghosh, S. S., Gorgolewski, K. J., Poldrack, R. A., & Esteban, O. (2021). *TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.02.10.430678

Cohen, A. L., Fair, D. A., Dosenbach, N. U. F., Miezin, F. M., Dierker, D., Van Essen, D.

C., Schlaggar, B. L., & Petersen, S. E. (2008). Defining functional areas in individual human brains using resting functional connectivity MRI. *NeuroImage*, *41*(1), 45–57. https://doi.org/10.1016/j.neuroimage.2008.01.066

Cohen, J. D. (2017). Cognitive Control. In *The Wiley Handbook of Cognitive Control* (pp. 1–28). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118920497.ch1

Cohen, J. R., Gallen, C. L., Jacobs, E. G., Lee, T. G., & D'Esposito, M. (2014). Quantifying the Reconfiguration of Intrinsic Networks during Working Memory. *PLOS ONE*, *9*(9), e106636. https://doi.org/10.1371/journal.pone.0106636

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*(9), 1348–1355. https://doi.org/10.1038/nn.3470

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2022). *Introduction to algorithms* (Fourth edition). The MIT Press.

Crosby, M., Beyret, B., Shanahan, M., Hernández-Orallo, J., Cheke, L., & Halina, M. (2020). The animal-AI testbed and competition. In H. J. Escalante & R. Hadsell (Eds.), *Proceedings of the NeurIPS 2019 competition and demonstration track* (Vol. 123, pp. 164–176). PMLR.

Dadi, K., Rahim, M., Abraham, A., Chyzhyk, D., Milham, M., Thirion, B., & Varoquaux, G. (2019). Benchmarking functional connectome-based predictive models for resting-state fMRI. *NeuroImage*, *192*, 115–134. https://doi.org/10.1016/j.neuroimage.2019.02.062

Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K. J., Wassermann, D., Thirion, B., & Mensch, A. (2020). Fine-grain atlases of functional modes for fMRI analysis. *NeuroImage*, *221*, 117126. https://doi.org/10.1016/j.neuroimage.2020.117126

Dale, G., & Green, C. S. (2017). Associations Between Avid Action and Real-Time Strategy Game Play and Cognitive Performance: A Pilot Study. *Journal of Cognitive Enhancement*, *1*(3), 295–317. https://doi.org/10.1007/s41465-017-0021-8

Dale, G., Joessel, A., Bavelier, D., & Green, C. S. (2020). A new look at the cognitive neuroscience of video game play. *Annals of the New York Academy of Sciences*, *1464*(1), 192–203. https://doi.org/10.1111/nyas.14295

Dale, G., & Shawn Green, C. (2017). The Changing Face of Video Games and Video Gamers: Future Directions in the Scientific Study of Video Game Play and Cognitive Performance. *Journal of Cognitive Enhancement*, *1*(3), 280–294. https://doi.org/10.1007/s41465-017-0015-6

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

De Boeck, P., & Jeon, M. (2019). An Overview of Models for Response Times and Processes in Cognitive Tests. *Frontiers in Psychology*, *10*.

De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *The Journal of Genetic Psychology*, *163*(3), 272–282. https://doi.org/10.1080/00221320209598683

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). THE EARLY YEARS: Preschool Program Improves Cognitive Control. *Science*, *318*(5855), 1387–1388. https://doi.org/10.1126/science.1151148

Diamond, A., & Ling, D. S. (2019). Review of the Evidence on, and Fundamental Questions About, Efforts to Improve Executive Functions, Including Working Memory. In *Cognitive and Working Memory Training* (pp. 143–431). Oxford University Press. https://doi.org/10.1093/oso/9780199974467.003.0008

Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439–453. https://doi.org/10.1162/tacl_a_00325

Doebel, S. (2020). Rethinking Executive Function and Its Development. *Perspectives on Psychological Science*, *15*(4), 942–956. https://doi.org/10.1177/1745691620904771

Dolan, R. J., & Dayan, P. (2013). Goals and Habits in the Brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. https://doi.org/10.1016/j.tics.2008.01.001

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., Fox, M. D., Snyder, A. Z., Vincent, J. L., Raichle, M. E., Schlaggar, B. L., & Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078. https://doi.org/10.1073/pnas.0704320104

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., Nelson, S. M., Wig, G. S., Vogel, A. C., Lessov-Schlaggar, C. N., Barnes, K. A., Dubis, J. W., Feczko, E., Coalson, R. S., Pruett, J. R., Barch, D. M., Petersen, S. E., & Schlaggar, B. L. (2010). Prediction of Individual Brain Maturity Using fMRI. *Science*, *329*(5997), 1358–1361. https://doi.org/10.1126/science.1194144

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing Speed of Processing With Action Video Games. *Current Directions in Psychological Science*, *18*(6), 321–326. https://doi.org/10.1111/j.1467-8721.2009.01660.x

Eichenbaum, H. (2017). The role of the hippocampus in navigation is memory. *Journal of Neurophysiology*, *117*(4), 1785–1796. https://doi.org/10.1152/jn.00005.2017

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), 2319. https://doi.org/

10.1038/s41467-019-10301-1

Engelhard, I. M., van Uijen, S. L., & van den Hout, M. A. (2010). The impact of taxing working memory on negative and positive memories. *European Journal of Psychotraumatology*, *1*. https://doi.org/10.3402/ejpt.v1i0.5623

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12).

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, *12*(9), e0184661. https://doi.org/10.1371/journal.pone.0184661

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Fiez, J. A. (1996). Cerebellar Contributions to Cognition. *Neuron*, *16*(1), 13–15. https://doi.org/10.1016/S0896-6273(00)80018-5

Fikkers, K. M., Piotrowski, J. T., & Valkenburg, P. M. (2019). Child's Play? Assessing the Bidirectional Longitudinal Relationship between Gaming and Intelligence in Early Childhood. *Journal of Communication*, *69*(2), 124–143. https://doi.org/10.1093/joc/jqz003

Finnigan, S., O'Connell, R. G., Cummins, T. D. R., Broughton, M., & Robertson, I. H. (2011). ERP measures indicate both attention and working memory encoding decrements in aging: Age effects on attention and memory encoding ERPs. *Psychophysiology*, *48*(5), 601–611. https://doi.org/10.1111/j.1469-8986.2010.01128.x

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

Föcker, J., Mortazavi, M., Khoe, W., Hillyard, S. A., & Bavelier, D. (2019). Neural Correlates of Enhanced Visual Attentional Control in Action Video Game Players: An Event-Related Potential Study. *Journal of Cognitive Neuroscience*, *31*(3), 377–389. https://doi.org/10.1162/jocn_a_01230

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*(1), 641–666. https://doi.org/10.1146/annurev-psych-122414-033645

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051. https://doi.org/10.1073/pnas.0604187103

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Franceschini, S., Bertoni, S., Ronconi, L., Molteni, M., Gori, S., & Facoetti, A. (2015). “Shall We Play a Game?”: Improving Reading Through Action Video Games in Developmental Dyslexia. *Current Developmental Disorders Reports*, *2*(4), 318–329. https://doi.org/10.1007/s40474-015-0064-4

Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012). A Causal Link between Visual Spatial Attention and Reading Acquisition. *Current Biology*, *22*(9), 814–819. https://doi.org/10.1016/j.cub.2012.03.013

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

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16. https://doi.org/10.1002/acp.934

Geary, D. C., Berch, D. B., & Mann Koepke, K. (2019). Introduction: Cognitive Foundations for Improving Mathematical Learning. In *Cognitive Foundations for Improving Mathematical Learning* (pp. 1–36). Elsevier. https://doi.org/10.1016/B978-0-12-815952-1.00001-3

Gentile, D. A., Bailey, K., Bavelier, D., Brockmyer, J. F., Cash, H., Coyne, S. M., Doan, A., Grant, D. S., Green, C. S., Griffiths, M., Markle, T., Petry, N. M., Prot, S., Rae, C. D., Rehbein, F., Rich, M., Sullivan, D., Woolley, E., & Young, K. (2017). Internet Gaming Disorder in Children and Adolescents. *Pediatrics*, *140*(Supplement 2), S81–S85. https://doi.org/10.1542/peds.2016-1758H

Glass, B. D., Maddox, W. T., & Love, B. C. (2013). Real-Time Strategy Game Training: Emergence of a Cognitive Flexibility Trait. *PLoS ONE*, *8*(8), e70350. https://doi.org/10.1371/journal.pone.0070350

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Goldin, A. P., Hermida, M. J., Shalom, D. E., Elias Costa, M., Lopez-Rosenfeld, M., Segretin, M. S., Fernandez-Slezak, D., Lipina, S. J., & Sigman, M. (2014). Far transfer to language and math of a short software-based gaming intervention. *Proceedings of the National Academy of Sciences*, *111*(17), 6443–6448. https://doi.org/10.1073/pnas.1320217111

Gong, D., He, H., Liu, D., Ma, W., Dong, L., Luo, C., & Yao, D. (2015). Enhanced functional connectivity and increased gray matter volume of insula related to action video game playing. *Scientific Reports*, *5*(1), 9763. https://doi.org/10.1038/srep09763

Gong, D., He, H., Ma, W., Liu, D., Huang, M., Dong, L., Gong, J., Li, J., Luo, C., & Yao, D. (2016). Functional Integration between Salience and Central Executive Networks: A Role for Action Video Game Experience. *Neural Plasticity*, *2016*(4), 1–9. https://doi.org/10.1155/2016/9803165

Gong, D., Ma, W., Gong, J., He, H., Dong, L., Zhang, D., Li, J., Luo, C., & Yao, D. (2017). Action Video Game Experience Related to Altered Large-Scale White Matter Networks. *Neural Plasticity*, *2017*, 1–7. https://doi.org/10.1155/2017/7543686

Gong, D., Yao, Y., Gan, X., Peng, Y., Ma, W., & Yao, D. (2019). A Reduction in Video Gaming Time Produced a Decrease in Brain Activity. *Frontiers in Human Neuroscience*, *13*, 134. https://doi.org/10.3389/fnhum.2019.00134

Gorbet, D. J., & Sergio, L. E. (2018). Move faster, think later: Women who play action video games have quicker visually- guided responses with later onset visuomotor-related brain activity. *PLoS ONE*, *13*(1), e0189110. https://doi.org/10.1371/journal.pone.0189110

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, *26*(1), 288–303. https://doi.org/10.1093/cercor/bhu239

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

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

Green, C. S., & Bavelier, D. (2012). Learning, Attentional Control, and Action Video Games. *Current Biology*, *22*(6), R197–R206. https://doi.org/10.1016/j.cub.2012.02.012

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. https://doi.org/10.1038/nature01647

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., Kramer, A. F., Vinogradov, S., Ansorge, U., Ball, K. K., Bingel, U., Chein, J. M., Colzato, L. S., Edwards, J. D., Facoetti, A., Gazzaley, A., Gathercole, S. E., Ghisletta, P., Gori, S., Granic, I., Hillman, C. H., Hommel, B., Jaeggi, S. M., … Witt, C. M. (2019). Improving Methodological Standards in Behavioral Interventions for Cognitive Enhancement. *Journal of Cognitive Enhancement*, *3*(1), 2–29. https://doi.org/10.1007/s41465-018-0115-y

Green, C. S., Strobach, T., & Schubert, T. (2014). On methodological standards in training and transfer experiments. *Psychological Research*, *78*(6), 756–772. https://doi.org/10.1007/s00426-013-0535-3

Green, C. S., Sugarman, M. A., Medford, K., Klobusicky, E., & Daphne Bavelier, null. (2012). The effect of action video game experience on task-switching. *Computers in Human Behavior*, *28*(3), 984–994. https://doi.org/10.1016/j.chb.2011.12.020

Greicius, M. D., & Menon, V. (2004). Default-Mode Activity during a Passive Sensory Task: Uncoupled from Deactivation but Impacting Activation. *Journal of Cognitive Neuroscience*, *16*, 1484–1492. https://doi.org/10.1162/0898929042568532

Güllich, A. (2018). Sport-specific and non-specific practice of strong and weak responders in junior and senior elite athletics – A matched-pairs analysis. *Journal of Sports Sciences*, *36*(19), 2256–2264. https://doi.org/10.1080/02640414.2018.1449089

Hasson, U., Nusbaum, H. C., & Small, S. L. (2009). Task-dependent organization of brain regions active during rest. *Proceedings of the National Academy of Sciences*, *106*(26), 10841–10846. https://doi.org/10.1073/pnas.0903253106

Ho, M. K., & Griffiths, T. L. (2022). Cognitive Science as a Source of Forward and Inverse Models of Human Decisions for Robotics and Control. *Annual Review of Control, Robotics, and Autonomous Systems*, *5*(1), 33–53. https://doi.org/10.1146/annurev-control-042920-015547

Hoffman, M., Shahriari, B., Aslanides, J., Barth-Maron, G., Behbahani, F., Norman, T., Abdolmaleki, A., Cassirer, A., Yang, F., Baumli, K., et al. (2020). Acme: A research framework for distributed reinforcement learning. *arXiv Preprint arXiv:2006.00979*. https:

//arxiv.org/abs/2006.00979

Holmes, E. A., James, E. L., Coode-Bate, T., & Deeprose, C. (2009). Can Playing the Computer Game "Tetris" Reduce the Build-Up of Flashbacks for Trauma? A Proposal from Cognitive Science. *PLOS ONE*, *4*(1), e4153. https://doi.org/10.1371/journal.pone.0004153

Howard-Jones, P. A., & Jay, T. (2016). Reward, learning and games. *Current Opinion in Behavioral Sciences*, *10*, 65–72. https://doi.org/10.1016/j.cobeha.2016.04.015

Hutzler, F. (2014). Reverse inference is not a fallacy per se: Cognitive processes can be inferred from functional imaging data. *NeuroImage*, *84*, 1061–1069. https://doi.org/10.1016/j.neuroimage.2012.12.075

Ie, E., Hsu, C., Mladenov, M., Jain, V., Narvekar, S., Wang, J., Wu, R., & Boutilier, C. (2019). Recsim: A configurable simulation platform for recommender systems. *arXiv Preprint arXiv:1909.04847*. https://arxiv.org/abs/1909.04847

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.

Jaeggi, S. M., Buschkuehl, 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

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences* (1st ed). Elsevier.

Jiang, J., Wagner, A. D., & Egner, T. (2018). Integrated externally and internally generated task predictions jointly guide cognitive control in prefrontal cortex. *eLife*, *7*, e39497. https://doi.org/10.7554/eLife.39497

Jolly, E., & Chang, L. J. (2019). The flatland fallacy: Moving beyond Low–Dimensional thinking. *Topics in Cognitive Science*, *11*(2), 433–454. https://doi.org/10.1111/tops.12404

Juvina, I., & Taatgen, N. A. (2007). Modeling control strategies in the N-Back task. *Pro-*

ceedings of the Eight International Conference on Cognitive Modeling, 73–78.

Karimpur, H., & Hamburger, K. (2015). The Future of Action Video Games in Psychological Research and Application. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01747

Katz, B., Shah, P., & Meyer, D. E. (2018). How to play 20 questions with nature and lose: Reflections on 100 years of brain-training research. *Proceedings of the National Academy of Sciences*, *115*(40), 9897–9904. https://doi.org/10.1073/pnas.1617102114

Király, O., Tóth, D., Urbán, R., Demetrovics, Z., & Maraz, A. (2017). Intense video gaming is not essentially problematic. *Psychology of Addictive Behaviors*, *31*(7), 807–817. https://doi.org/10.1037/adb0000316

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., Brooks, D. J., Bench, C. J., & Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, *393*(6682), 266–268. https://doi.org/10.1038/30498

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557–577.

Kovess-Masfety, V., Keyes, K., Hamilton, A., Hanson, G., Bitfoi, A., Golitz, D., Koç, C., Kuijpers, R., Lesinskiene, S., Mihova, Z., Otten, R., Fermanian, C., & Pez, O. (2016). Is time spent playing video games associated with mental health, cognitive and social skills in young children? *Social Psychiatry and Psychiatric Epidemiology*, *51*(3), 349–357. https://doi.org/10.1007/s00127-016-1179-6

Kraus, B. T., Perez, D., Ladwig, Z., Seitzman, B. A., Dworetsky, A., Petersen, S. E., & Gratton, C. (2021). Network variants are similar between task and rest states. *NeuroImage*, *229*, 117743. https://doi.org/10.1016/j.neuroimage.2021.117743

Krishnan, L., Kang, A., Sperling, G., & Srinivasan, R. (2013). Neural Strategies for Selective Attention Distinguish Fast-Action Video Game Players. *Brain Topography*, *26*(1), 83–97. https://doi.org/10.1007/s10548-012-0232-3

Kühn, S., & Gallinat, J. (2014). Amount of lifetime video gaming is positively associated with entorhinal, hippocampal and occipital volume. *Molecular Psychiatry*, *19*(7), 842–847.

https://doi.org/10.1038/mp.2013.100

Kühn, S., Gleich, T., Lorenz, R. C., Lindenberger, U., & Gallinat, J. (2014). Playing Super Mario induces structural brain plasticity: Gray matter changes resulting from training with a commercial video game. *Molecular Psychiatry*, *19*(2), 265–271. https://doi.org/10.1038/mp.2013.120

Kühn, S., Lorenz, R., Banaschewski, T., Barker, G. J., Büchel, C., Conrod, P. J., Flor, H., Garavan, H., Ittermann, B., Loth, E., Mann, K., Nees, F., Artiges, E., Paus, T., Rietschel, M., Smolka, M. N., Ströhle, A., Walaszek, B., Schumann, G., … The IMAGEN Consortium. (2014). Positive Association of Video Game Playing with Left Frontal Cortical Thickness in Adolescents. *PLoS ONE*, *9*(3), e91506. https://doi.org/10.1371/journal.pone.0091506

Kühn, S., Romanowski, A., Schilling, C., Lorenz, R., Mörsen, C., Seiferth, N., Banaschewski, T., Barbot, A., Barker, G. J., Büchel, C., Conrod, P. J., Dalley, J. W., Flor, H., Garavan, H., Ittermann, B., Mann, K., Martinot, J.-L., Paus, T., Rietschel, M., … Gallinat, J. (2011). The neural basis of video gaming. *Translational Psychiatry*, *1*(11), e53–e53. https://doi.org/10.1038/tp.2011.53

Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, *36*(11). https://doi.org/10.18637/jss.v036.i11

Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*, *38*(4), 13–26. https://doi.org/10.1609/aimag.v38i4.2744

Lefebvre, C. D., Marchand, Y., Eskes, G. A., & Connolly, J. F. (2005). Assessment of working memory abilities using an event-related brain potential (ERP)-compatible digit span backward task. *Clinical Neurophysiology*, *116*(7), 1665–1680. https://doi.org/10.1016/j.clinph.2005.03.015

Lewis, J. M., Trinh, P., & Kirsh, D. (2011). A corpus analysis of strategy video game play in starcraft: Brood war. *Cognitive Science*, *33*.

Li, A. (2022). Scikit-Learn-Extra Documentation—Comparison of EigenPro and SVC on Digit Classification. In *Comparison of EigenPro and SVC on Digit Classification*. https://scikit-learn-extra.readthedocs.io/en/stable/auto_examples/eigenpro/plot_eigenpro_synthe

Li, L., Chen, R., & Chen, J. (2016). Playing Action Video Games Improves Visuomotor Control. *Psychological Science*, *27*(8), 1092–1108. https://doi.org/10.1177/0956797616650300

Libertus, M. E., Liu, A., Pikul, O., Jacques, T., Cardoso-Leite, P., Halberda, J., & Bavelier, D. (2017). The Impact of Action Video Game Training on Mathematical Abilities in Adults. *AERA Open*, *3*(4), 233285841774085. https://doi.org/10.1177/2332858417740857

Lindquist, M. A., Geuter, S., Wager, T. D., & Caffo, B. S. (2019). Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human Brain Mapping*, *40*(8), 2358–2376. https://doi.org/10.1002/hbm.24528

Lindsay, G. W. (2020). Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*, *14*, 29. https://doi.org/10.3389/fncom.2020.00029

Lisman, J., Buzsáki, G., Eichenbaum, H., Nadel, L., Ranganath, C., & Redish, A. D. (2017). Viewpoints: How the hippocampus contributes to memory, navigation and cognition. *Nature Neuroscience*, *20*(11), 1434–1447. https://doi.org/10.1038/nn.4661

Logan, G. D. (2017). Taking control of cognition: An instance perspective on acts of control. *American Psychologist*, *72*(9), 875–884. https://doi.org/10.1037/amp0000226

Lor, C. S., Zhang, M., Karner, A., Steyrl, D., Sladky, R., Scharnowski, F., & Haugg, A. (2022). *Pre- and post-task resting-state differs in clinical populations* [Preprint]. Neuroscience. https://doi.org/10.1101/2022.09.20.508750

Lorenz, R. C., Gleich, T., Gallinat, J., & Kühn, S. (2015). Video game training and the reward system. *Frontiers in Human Neuroscience*, *9*, 40. https://doi.org/10.3389/fnhum.2015.00040

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774).

Curran Associates, Inc.

Łuniewska, M., Chyl, K., Dębska, A., Kacprzak, A., Plewko, J., Szczerbiński, M., Szewczyk, J., Grabowska, A., & Jednoróg, K. (2018). Neither action nor phonological video games make dyslexic children read better. *Scientific Reports*, *8*(1), 549. https://doi.org/10.1038/s41598-017-18878-7

Martin, R. C. (Ed.). (2009). *Clean code: A handbook of agile software craftsmanship*. Prentice Hall.

McGovern, A., Lagerquist, R., John Gagne, D., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. *Bulletin of the American Meteorological Society*, *100*(11), 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Menon, V. (2015). Salience Network. In *Brain Mapping* (pp. 597–611). Elsevier. https://doi.org/10.1016/B978-0-12-397025-1.00052-X

Menon, V., & D'Esposito, M. (2022). The role of PFC networks in cognitive control and executive function. *Neuropsychopharmacology*, *47*(1), 90–103. https://doi.org/10.1038/s41386-021-01152-w

"Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills": Correction to Bediou et al. (2018). (2018). *Psychological Bulletin*, *144*(9), 978–979. https://doi.org/10.1037/bul0000168

Miendlarzewska, E. A., Bavelier, D., & Schwartz, S. (2016). Influence of reward motivation on human declarative memory. *Neuroscience & Biobehavioral Reviews*, *61*, 156–176. https://doi.org/10.1016/j.neubiorev.2015.11.015

Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior.*

Miller, K. M., Price, C. C., Okun, M. S., Montijo, H., & Bowers, D. (2009). Is the N-Back Task a Valid Neuropsychological Measure for Assessing Working Memory? *Archives of*

Clinical Neuropsychology, *24*(7), 711–717. https://doi.org/10.1093/arclin/acp063

Mishra, J., Zinni, M., Bavelier, D., & Hillyard, S. A. (2011). Neural Basis of Superior Performance of Action Videogame Players in an Attention-Demanding Task. *Journal of Neuroscience*, *31*(3), 992–998. https://doi.org/10.1523/JNEUROSCI.4834-10.2011

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533. https://doi.org/10.1038/nature14236

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (Second edition). Christoph Molnar.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Moreau, D. (2013). Differentiating two- from three-dimensional mental rotation training effects. *Quarterly Journal of Experimental Psychology (2006)*, *66*(7), 1399–1413. https://doi.org/10.1080/17470218.2012.744761

Moreau, D., & Conway, A. R. A. (2014). The case for an ecological approach to cognitive training. *Trends in Cognitive Sciences*, *18*(7), 334–336. https://doi.org/10.1016/j.tics.2014.03.009

Moskovitz, T., Miller, K., Sahani, M., & Botvinick, M. M. (2022). *A Unified Theory of Dual-Process Control*. https://doi.org/10.48550/ARXIV.2211.07036

Nahum, M., & Bavelier, D. (2020). Video games as rich environments to foster brain plasticity. In *Handbook of Clinical Neurology* (Vol. 168, pp. 117–136). Elsevier. https://doi.org/10.

1016/B978-0-444-63934-9.00010-X

Nau, M., Julian, J. B., & Doeller, C. F. (2018). How the Brain's Navigation System Shapes Our Visual Experience. *Trends in Cognitive Sciences*, *22*(9), 810–825. https://doi.org/10.1016/j.tics.2018.06.008

Nava, E., Föcker, J., & Gori, M. (2019). Children can optimally integrate multisensory information after a short action-like mini game training. *Developmental Science*, e12840. https://doi.org/10.1111/desc.12840

Nigg, J. T. (2016). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

Nilearn Team. (2022a). Nilearn documentation (v0.9) - fetch DiFuMo brain atlas. In *nilearn.datasets.fetch_atlas_difumo*. https://nilearn.github.io/stable/modules/generated/nilearn.da

Nilearn Team. (2022b). Nilearn documentation (v0.9) - Load the Dosenbach et al. ROIs. In *nilearn.datasets.fetch_coords_dosenbach_2010*. https://nilearn.github.io/stable/modules/generated

Oei, A. C., & Patterson, M. D. (2013). Enhancing Cognition with Video Games: A Multiple Game Training Study. *PLoS ONE*, *8*(3), e58546. https://doi.org/10.1371/journal.pone.0058546

Oei, A. C., & Patterson, M. D. (2014a). Playing a puzzle video game with changing requirements improves executive functions. *Computers in Human Behavior*, *37*, 216–228. https://doi.org/10.1016/j.chb.2014.04.046

Oei, A. C., & Patterson, M. D. (2014b). Are videogame training gains specific or general? *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00054

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology*, *15*(1), 33–58. https://doi.org/10.1016/0193-3973(94)90005-1

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, *106*(37), 15583–15587. https://doi.org/

10.1073/pnas.0903620106

Otto, A. R., Raio, C. M., Chiang, A., Phelps, E. A., & Daw, N. D. (2013). Working-memory capacity protects model-based learning from stress. *Proceedings of the National Academy of Sciences*, *110*(52), 20941–20946. https://doi.org/10.1073/pnas.1312011110

Otto, A. R., Skatova, A., Madlon-Kay, S., & Daw, N. D. (2015). Cognitive Control Predicts Use of Model-based Reinforcement Learning. *Journal of Cognitive Neuroscience*, *27*(2), 319–333. https://doi.org/10.1162/jocn_a_00709

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., Howard, R. J., & Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778. https://doi.org/10.1038/nature09042

Palaus, M., Marron, E. M., Viejo-Sobera, R., & Redolar-Ripoll, D. (2017). Neural Basis of Video Gaming: A Systematic Review. *Frontiers in Human Neuroscience*, *11*, 248. https://doi.org/10.3389/fnhum.2017.00248

Parong, J., Seitz, A. R., Jaeggi, S. M., & Green, C. S. (2022). Expectation effects in working memory training. *Proceedings of the National Academy of Sciences*, *119*(37), e2209308119. https://doi.org/10.1073/pnas.2209308119

Pavan, A., Hobaek, M., Blurton, S. P., Contillo, A., Ghin, F., & Greenlee, M. W. (2019). Visual short-term memory for coherent motion in video game players: Evidence from a memory-masking paradigm. *Scientific Reports*, *9*(1), 6027. https://doi.org/10.1038/s41598-019-42593-0

Pedersen, M. L., Alnæs, D., van der Meer, D., Fernandez-Cabello, S., Berthet, P., Dahl, A., Kjelkenes, R., Schwarz, E., Thompson, W. K., Barch, D. M., Andreassen, O. A., & Westlye, L. T. (2022). Computational Modeling of the n-Back Task in the ABCD Study: Associations of Drift Diffusion Model Parameters to Polygenic Scores of Mental Disorders and Cardiometabolic Diseases. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2022.03.012

Pedro Cardoso-Leite, Augustin Joessel, & Daphne Bavelier. (2020). Games for enhancing cognitive abilities. In Jan L. Plass, Richard E. Mayer, & Bruce D. Homer (Eds.), *Hand-*

book of game-based learning. The MIT Press.

Perone, S., Simmering, V. R., & Buss, A. T. (2021). A Dynamical Reconceptualization of Executive-Function Development. *Perspectives on Psychological Science*, *16*(6), 1198–1208. https://doi.org/10.1177/1745691620966792

Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*(1), 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Pilegard, C., & Mayer, R. E. (2018). Game over for Tetris as a platform for cognitive skill training. *Contemporary Educational Psychology*, *54*, 29–41. https://doi.org/10.1016/j.cedpsych.2018.04.003

Poldrack, R. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2020). Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*, *77*(5), 534. https://doi.org/10.1001/jamapsychiatry.2019.3671

Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., & Bilder, R. M. (2011). The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00017

Poldrack, R. A., Laumann, T. O., Koyejo, O., Gregory, B., Hover, A., Chen, M.-Y., Gorgolewski, K. J., Luci, J., Joo, S. J., Boyd, R. L., Hunicke-Smith, S., Simpson, Z. B., Caven, T., Sochat, V., Shine, J. M., Gordon, E., Snyder, A. Z., Adeyemo, B., Petersen, S. E., … Mumford, J. A. (2015). Long-term neural and physiological phenotyping of a single human. *Nature Communications*, *6*(1), 8885. https://doi.org/10.1038/ncomms9885

Powers, K. L., & Brooks, P. J. (2014). Evaluating the Specificity of Effects of Video Game Training. In F. C. Blumberg (Ed.), *Learning by Playing* (pp. 302–330). Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199896646.003.0021

Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review*, *20*(6), 1055–1079. https://doi.org/10.3758/s13423-013-0418-z

Pujol, J., Fenoll, R., Forns, J., Harrison, B. J., Martínez-Vilavella, G., Macià, D., Alvarez-Pedrerol, M., Blanco-Hinojo, L., González-Ortiz, S., Deus, J., & Sunyer, J. (2016). Video gaming in school children: How much is enough?: Video Gaming. *Annals of Neurology*, *80*(3), 424–433. https://doi.org/10.1002/ana.24745

Rafiei, F., & Rahnev, D. (2022). *RTNet: A neural network that exhibits the signatures of human perceptual decision making* (p. 2022.08.23.505015). bioRxiv. https://doi.org/10.1101/2022.08.23.505015

Ralph, J. (2014). *Statistical manipulation and control strategies of the n-back task* [Thesis]. Rensselaer Polytechnic Institute.

Ramstedt, S., & Pal, C. (2019). *Real-Time Reinforcement Learning* (No. arXiv:1911.04448). arXiv. https://arxiv.org/abs/1911.04448

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59.

Ratcliff, R., Huang-Pollock, C., & McKoon, G. (2018). Modeling Individual Differences in the Go/No-go Task with a Diffusion Model. *Decision (Washington, D.C.)*, *5*(1), 42–62. https://doi.org/10.1037/dec0000065

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., & Starns, J. J. (2013). Modeling Confidence Judgments, Response Times, and Multiple Choices in Decision Making: Recognition Memory and Motion Discrimination. *Psychological Review*, *120*(3), 697. https://doi.org/10.1037/a0033152

Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-Maron, G., Gimenez, M., Sulsky, Y., Kay, J., Springenberg, J. T., Eccles, T., Bruce, J., Razavi, A., Edwards, A., Heess, N., Chen, Y., Hadsell, R., Vinyals, O., Bordbar, M., & de Freitas, N. (2022). *A Generalist Agent* (No. arXiv:2205.06175). arXiv. https://arxiv.org/abs/2205.06175

Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., & Banich, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage*, *104*, 69–78. https://doi.org/10.1016/j.neuroimage.2014.09.045

Rey-Mermet, A., Singmann, H., & Oberauer, K. (2021). Neither measurement error nor speed-accuracy trade-offs explain the difficulty of establishing attentional control as a psychometric construct: Evidence from a latent-variable analysis using diffusion modeling. *PsyArXiv*.

Rideout, V. (2016). Measuring time spent with media: The Common Sense census of media use by US 8- to 18-year-olds. *Journal of Children and Media*, *10*(1), 138–144. https://doi.org/10.1080/17482798.2016.1129808

Rideout, V. (2015). *The Common Sense Census: Media Use by Tweens and Teens* (p. 104). Common Sense Media.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Ruch, A. (2020). Can X2vec save lives? Integrating graph and language embeddings for automatic mental health classification. *Journal of Physics: Complexity*, *1*(3).

Russell, S. J. (2020). *Human compatible: Artificial intelligence and the problem of control*. Penguin Books.

Sala, G., Tatlidil, K. S., & Gobet, F. (2018). Video game training does not enhance cognitive ability: A comprehensive meta-analytic investigation. *Psychological Bulletin*, *144*(2), 111–139. https://doi.org/10.1037/bul0000139

Salehi, M., Greene, A. S., Karbasi, A., Shen, X., Scheinost, D., & Constable, R. T. (2019).

There is no single functional atlas even for a single individual: Functional parcel definitions change with task. *NeuroImage*, 116366. https://doi.org/10.1016/j.neuroimage.2019.116366

Seguin, C., Tian, Y., & Zalesky, A. (2020). Network communication models improve the behavioral and functional predictive utility of the human structural connectome. *Network Neuroscience*, *4*(4), 980–1006. https://doi.org/10.1162/netn_a_00161

Seok, S., & DaCosta, B. (2019). Video Games as a Literacy Tool: A Comparison of Players' and Nonplayers' Grades, Reading Test Scores, and Self-Perceived Digital Reading Ability. *Society for Information Technology & Teacher Education International Conference*, 777–781.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a Rational and Mechanistic Account of Mental Effort. *Annual Review of Neuroscience*, *40*, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., Moodie, C. A., & Poldrack, R. A. (2016). The Dynamics of Functional Brain Networks: Integrated Network States during Cognitive Task Performance. *Neuron*, *92*(2), 544–554. https://doi.org/10.1016/j.neuron.2016.09.018

Shine, J. M., & Poldrack, R. A. (2018). Principles of dynamic network reconfiguration across diverse brain states. *NeuroImage*, *180*, 396–405. https://doi.org/10.1016/j.neuroimage.2017.08.010

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology*, *16*(1), 97–115. https://doi.org/10.1002/acp.759

Siniatchkin, M. (2017). Anodal tDCS over the left DLPFC improved working memory and reduces symptoms in children with ADHD. *Brain Stimulation*, *10*(2), 517. https://doi.org/10.1016/j.brs.2017.01.509

Skorka-Brown, J., Andrade, J., Whalley, B., & May, J. (2015). Playing Tetris decreases drug and other cravings in real world settings. *Addictive Behaviors*, *51*, 165–170. https:

//doi.org/10.1016/j.addbeh.2015.07.020

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science (New York, N.Y.)*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Spence, I., & Feng, J. (2010). Video Games and Spatial Cognition. *Review of General Psychology*, *14*(2), 92–104. https://doi.org/10.1037/a0019491

Stafford, T., & Dewar, M. (2014). Tracing the Trajectory of Skill Learning With a Very Large Sample of Online Game Players. *Psychological Science*, *25*(2), 511–518. https://doi.org/10.1177/0956797613511466

Stanhope, J. L., Owens, C., & Elliott, L. J. (2015). Stress Reduction: Casual Gaming versus Guided Relaxation. *Human Factors and Applied Psychology Student Conference HFAP Conference*.

Stocco, A., Sibert, C., Steine-Hanson, Z., Koh, N., Laird, J. E., Lebiere, C. J., & Rosenbloom, P. (2021). Analysis of the human connectome data supports the notion of a "Common Model of Cognition" for human and human-like intelligence across domains. *NeuroImage*, *235*, 118035. https://doi.org/10.1016/j.neuroimage.2021.118035

Strenziok, M., Parasuraman, R., Clarke, E., Cisler, D. S., Thompson, J. C., & Greenwood, P. M. (2014). Neurocognitive enhancement in older adults: Comparison of three cognitive training tasks to test a hypothesis of training transfer in brain connectivity. *NeuroImage*, *85 Pt 3*, 1027–1039. https://doi.org/10.1016/j.neuroimage.2013.07.069

Sungur, H., & Boduroglu, A. (2012). Action video game players form more detailed representation of objects. *Acta Psychologica*, *139*(2), 327–334. https://doi.org/10.1016/j.actpsy.2011.12.002

Tailby, C., Masterton, R. A. J., Huang, J. Y., Jackson, G. D., & Abbott, D. F. (2015). Resting state functional connectivity changes induced by prior brain state are not network specific. *NeuroImage*, *106*, 428–440. https://doi.org/10.1016/j.neuroimage.2014.11.037

Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Fukushima, A., & Kawashima, R. (2011). Working Memory Training Using Mental Calculation Impacts

Regional Gray Matter of the Frontal and Parietal Regions. *PLoS ONE*, *6*(8), e23175. https://doi.org/10.1371/journal.pone.0023175

Team, N. (2022). *Confound Strategies (Nilearn v9.2)*. https://nilearn.github.io/stable/modules/generate

Teeters, J. L., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., Denisov, G., Siegle, J. H., Olsen, S. R., Martin, C., Chun, M., Tripathy, S., Blanche, T. J., Harris, K., Buzsáki, G., … Sommer, F. T. (2015). Neurodata Without Borders: Creating a Common Data Format for Neurophysiology. *Neuron*, *88*(4), 629–634. https://doi.org/10.1016/j.neuron.2015.10.025

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*, *22*(7), 996–1013. https://doi.org/10.1002/acp.1420

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. II. The estimation of magnitudes. *Psychological Review*, *8*(4), 384–395. https://doi.org/10.1037/h0071280

Tiraboschi, G. A., Fukusima, S. S., & West, G. L. (2019). An Expectancy Effect Causes Improved Visual Attention Performance After Video Game Playing. *Journal of Cognitive Enhancement*, *3*(4), 436–444. https://doi.org/10.1007/s41465-019-00130-x

Tomov, M. S., Schulz, E., & Gershman, S. J. (2021). Multi-task reinforcement learning in humans. *Nature Human Behaviour*, *5*(6), 764–773. https://doi.org/10.1038/s41562-020-01035-y

Toril, P., Reales, J. M., & Ballesteros, S. (2014). Video game training enhances cognition of older adults: A meta-analytic study. *Psychology and Aging*, *29*(3), 706–716. https://doi.org/10.1037/a0037507

Toyama, D., Hamel, P., Gergely, A., Comanici, G., Glaese, A., Ahmed, Z., Jackson, T., Mourad, S., & Precup, D. (2021). AndroidEnv: A reinforcement learning platform for android. *arXiv Preprint arXiv:2105.13231*. https://arxiv.org/abs/2105.13231

Uncapher, M. R., & Wagner, A. D. (2018). Minds and brains of media multitaskers: Current findings and future directions. *Proceedings of the National Academy of Sciences*, *115*(40), 9889–9896. https://doi.org/10.1073/pnas.1611612115

Varoquaux, G. (2020). Estimating brain functional connectivity and its variations from fMRI [Pdf]. In *Estimating brain functional connectivity and its variations from fMRI*. https://www.normalesup.org/~varoquau/HDR_Gael_Varoquaux.pdf.

Varoquaux, G., & Craddock, R. C. (2013). Learning and comparing functional connectomes across subjects. *NeuroImage*, *80*, 405–415. https://doi.org/10.1016/j.neuroimage.2013.04.007

Waller, G., Willemse, I., Genner, S., Suter, L., & Süss, D. (2016). *JAMES - Jeunes, activités, médias – enquête Suisse*. Haute école des sciences appliquées de Zurich.

Wang, J., Tian, J., Hao, R., Tian, L., & Liu, Q. (2018). Transcranial direct current stimulation over the right DLPFC selectively modulates subprocesses in working memory. *PeerJ*, *6*, e4906. https://doi.org/10.7717/peerj.4906

Wang, P., Liu, H.-H., Zhu, X.-T., Meng, T., Li, H.-J., & Zuo, X.-N. (2016). Action Video Game Training for Healthy Adults: A Meta-Analytic Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00907

West, G. L., Drisdelle, B. L., Konishi, K., Jackson, J., Jolicoeur, P., & Bohbot, V. D. (2015). Habitual action video game playing is associated with caudate nucleus-dependent navigational strategies. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1808), 20142952. https://doi.org/10.1098/rspb.2014.2952

West, G. L., Konishi, K., Diarra, M., Benady-Chorney, J., Drisdelle, B. L., Dahmani, L., Sodums, D. J., Lepore, F., Jolicoeur, P., & Bohbot, V. D. (2018). Impact of video games on plasticity of the hippocampus. *Molecular Psychiatry*, *23*(7), 1566–1574. https://doi.org/10.1038/mp.2017.155

West, R., Swing, E. L., Anderson, C. A., & Prot, S. (2020). The Contrasting Effects of an Action Video Game on Visuo-Spatial Processing and Proactive Cognitive Control.

International Journal of Environmental Research and Public Health, *17*(14), 5160. https://doi.org/10.3390/ijerph17145160

Whitbourne, S. K., Ellenberg, S., & Akimoto, K. (2013). Reasons for Playing Casual Video Games and Perceived Benefits Among Adults 18 to 80 Years Old. *Cyberpsychology, Behavior, and Social Networking*, *16*(12), 892–897. https://doi.org/10.1089/cyber.2012.0705

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10). https://doi.org/10.18637/jss.v059.i10

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., & Glahn, D. C. (2010). Cortical Thickness or Grey Matter Volume? The Importance of Selecting the Phenotype for Imaging Genetics Studies. *NeuroImage*, *53*(3), 1135–1146. https://doi.org/10.1016/j.neuroimage.2009.12.028

Wiradhany, W., & Nieuwenstein, M. R. (2017). Cognitive control in media multitaskers: Two replication studies and a meta-Analysis. *Attention, Perception, & Psychophysics*, *79*(8), 2620–2641. https://doi.org/10.3758/s13414-017-1408-4

Wu, S., Cheng, C. K., Feng, J., D'Angelo, L., Alain, C., & Spence, I. (2012). Playing a First-person Shooter Video Game Induces Neuroplastic Change. *Journal of Cognitive Neuroscience*, *24*(6), 1286–1293. https://doi.org/10.1162/jocn_a_00192

Wu, S., & Spence, I. (2013). Playing shooter and driving videogames improves top-down guidance in visual search. *Attention, Perception, & Psychophysics*, *75*(4), 673–686. https://doi.org/10.3758/s13414-013-0440-2

Yang, G. R., Joglekar, M. R., Song, H. F., Newsome, W. T., & Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, *22*(2), 297–306. https://doi.org/10.1038/s41593-018-0310-2

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Zhang, Y., Du, G., Yang, Y., Qin, W., Li, X., & Zhang, Q. (2015). Higher integrity of the motor and visual pathways in long-term video game players. *Frontiers in Human Neuroscience*, *9*, 695. https://doi.org/10.3389/fnhum.2015.00098

Zink, N., Lenartowicz, A., & Markett, S. (2021). A new era for executive function research: On the transition from centralized to distributed executive functioning. *Neuroscience & Biobehavioral Reviews*, *124*, 235–244. https://doi.org/10.1016/j.neubiorev.2021.02.011

Appendix A

A Formal Framework for Structured N-Back Stimuli Sequences

Morteza Ansarinia, Dominic Mussack, Paul Schrater, and Pedro Cardoso-Leite

Abstract

Numerous cognitive tasks, like the n-back, employ sequences of stimuli to target particular cognitive functions. These sequences are generated to satisfy specific criteria but the generation process typically induces unintentional statistical structure in the sequences which may not only affect performance but also alter the strategies participants use to complete the task.

Here we propose that the generation of stimulus sequences can be conceptualized as a soft constraint satisfaction problem and offer experimental evidence demonstrating the impact of local sequence features on human behavior. Our approach to sequence generation provides a means to better control and assess sequence structures, which in turn could help clarify the cognitive and neural processes involved in cognitive tasks.

A.1 Introduction

With more than 1600 hits on PubMed, the n-back task is one of the most popular tasks in cognitive psychology today. It is widely used not only to evaluate working memory capacity but also as a training protocol to improve working memory and possibly fluid intelligence Jaeggi et al. (2008). In the n-back task, participants are presented a sequence of stimuli and have to determine for each stimulus if it matches or not the stimulus presented n-steps ago. Stimuli that match are called "targets", those that don't match are called "distractors", with close misses (i.e., distractors that would be targets under a slightly different N) are called "lures". While the task is widely considered a working memory task, it does not correlate well with other "gold-standard" working memory tasks, such as the complex span task (Jaeggi et al., 2010; K. M. Miller et al., 2009).

Previous studies have raised concerns that the n-back task may be solved using multiple strategies, not all of which rely purely on working memory processes (Ralph, 2014). There are numerous variants of the n-back task, but even within a variant participants could use various strategies. One source of variation in the n-back task that is potentially biasing participants' strategies are the statistical properties of the sequences of stimuli used for the n-back task, which are typically uncontrolled for and differ across studies (Braver, 2012). For example, (Ralph, 2014) showed that various statistical properties of n-back sequences may favor a reactive cognitive control strategy whereby people's performance relies on detecting stimulus familiarity rather than on active information updating in working memory. Because statistical properties of stimulus sequences seem to bias cognitive control strategies and hence cause heterogeneous behavioral and neurophysiological outcomes it is necessary to characterize those statistical properties and develop methods to generate adequate sequences.

Here, we propose an approach that allows researchers to parameterize interesting features of the n-back sequences which may affect behavior. We then evaluate the predictive effect of such uncontrolled parameters on behavioral outcomes. Results from this research may have implications on the way the n-back task is put into practice to study working memory or improve cognitive skills. While our focus here is on the n-back, the principles presented below apply to a broader range of cognitive paradigms.

A.1.1 Parameterizing the N-Back sequences

While n-back sequences are usually thought of as an ordered set of i.i.d. generated and sequentially independent stimuli, in practice the sequences of stimuli are neither objectively nor subjectively independent. Objective local structure is introduced by design constraints like a fixed number of target or stimulus set size, while subjectively people are highly sensitive to local sample structure in sequences. For example, unconstrained sampling from a uniform distribution to generate sequences may lead to frequent local repetitions of stimuli (i.e., "lumpiness,"; Abelson, 1995). In the n-back task, such local patterns could encourage people to identify targets solely based on stimulus familiarity rather than to use their working memory, as this strategy may in this case lead to high performance at low cognitive cost. Here we define a few basic measures known to be important for the perception of local structure in sequences, and show how to use these measures to parameterize families of sequences.

N-Back sequences are typically generated by randomly sampling M stimuli from a vocabulary set V (e.g., a set of 8 letters) with the constraint of having a specific number of targets (T) in the sequence, given the fixed value of N for the intended n-back version. Researchers typically manipulate N and T to study behavioral and neural correlates of working memory; other parameters are treated as nuisance variables.

A common procedure to generate n-back sequences involves two steps: first a sequence of stimulus-role placeholders (e.g., D=distractor, T=target) is generated; then particular stimuli are sampled from the vocabulary to fulfill those roles. For example, the first step might generate the sequence `DDDTDTDT` while the second step would instantiate particular stimuli (e.g.,`ABCEDEA`). Generating n-back sequences using this procedure is problematic however because the resulting sequences are typically highly skewed with some stimuli being presented much more frequently than others and frequently presented stimuli having a higher probability of being targets (Ralph, 2014). Moreover, lures are more likely to trigger false alarm responses and to require proactive control processes.

The lack of control for parameters such as lures and lumpiness may compromise results interpretations and generate scientific confusion because such parameter may affect cognitive strategies and consequently increase behavioral and neurophysiological data heterogeneity (Juvina & Taatgen, 2007). Ralph (2014) urged researchers to carefully control frequency distribution of stimuli, stimulus repetition, the fraction of targets and the fraction of lures, and the number of different stimuli in the vocabulary set in order to have a better handle on cognitive strategies. However, generating sequences that fulfill multiple criteria may not always be possible or practical using standard, brute-force approaches; there might for instance be cases where no such sequence exists. Furthermore, future research may require the addition or removal of criteria and such changes would typically require rewriting new sequence generators.

In the following section we conceptualize the generation of structured sequences for the n-back as a constraint satisfaction problem. This approach has several key advantages: a) it provides an implementation blueprint that accommodates a wide range of use cases b) it supports the softening of constraints to ensure approximate solutions can be found within a practical timespan; c) it supports compositional control of constraints that is well suited for hypothesis testing and d) by taking advantage of the Maximum Entropy optimization framework and Conditional Random Fields model, it is possible to move from an intuitive definition of constraints to the space of probability distributions that are invaluable for modeling and data analysis (Batou & Soize, 2013).

A.1.2 Structured sequences

A sequence is an ordered set of M stimuli sampled from a vocabulary of V stimuli that satisfies specific criteria. A sequence of stimuli that (approximately) satisfies a set of specific constraints on parameters or features is a qualified sequence.

The problem of generating a qualified sequence can be reduced to a soft constraint satisfaction problem, P:

$$P = \langle X, D, C, W \rangle$$

where X is a set of structural variables to be controlled (see Table **??**), D is the set of distributions over the variables, C is the set of constraints expressed as expected values for X (see Table **??**), and W is a cost function that uses the constraints to map a sampled sequence to a real value (Table **??**); it represents the degree to which a particular sequence violates the constraints in C. Generating a qualified sequence for the n-back task can be formulated as minimizing the aggregated cost of violating the constraints. Note that some constraints in the n-back task cannot be relaxed; for example, constraints which include the expected value of the N, must be fully satisfied for the sequences to be valid.

Table A.1: List of structural variables (X)

Variable	Description
x_N	N, number of trials to look back for a target.
x_t	Targets ratio describes the number of target trials in a sequence regardless of the stimulus.
x_s	Skewness is maximum deviation of stimuli frequency from uniform distribution.
x_l	Lures ratio represents the number of distractors which would be targets for $N-1$ or $N+1$.
x_v	Vocabulary size is the number of all unique stimuli to be presented.
x_{tl}	Recent targets ratio represents the number of targets in recent trials.
x_{ll}	Local lures ratio describes the number of lures in recent trials.
x_{vl}	Local vocabulary size is the number of unique stimuli presented in recent trials.
x_{ul}	Lumpiness is the maximum number of repetitions in a sequence.
x_{sl}	Local skewness is the number of unique items shown in recent trials.

Variable	Description
x_g	Gap is the number of trials since the last time the same stimulus appeared.

Table A.2: List of constraints on structural variables and respective violation costs

Constraints (C)	Violation Cost (W)
$E[x_n] = N$	$W_n \sim \begin{cases} 0 & x_n = N \\ \infty & x_n \neq N \end{cases}$
$E[x_t] = T \times trials$	$W_t \sim 1 - \mathcal{N}(T \times trials, 1)$
$E[x_{tl}] = \frac{T \times w}{trials}$	$W_{tl} \sim 1 - \mathcal{N}(\frac{T \times w}{trials}, 1)$
$E[x_l] = L \times trials$	$W_l \sim 1 - \mathcal{N}(L \times trials, 1)$
$E[x_{ll}] = \frac{L \times w}{trials}$	$W_{ll} \sim 1 - \mathcal{N}(\frac{L \times w}{trials}, 1)$
$E[x_v] = \lvert V \rvert$	$W_v \sim 1 - \mathcal{N}(\lvert V \rvert, 1)$
$E[x_{vl}] = min(\lvert V \rvert, w)$	$W_{vl} \sim 1 - \mathcal{N}(min(\lvert V \rvert, w), 1)$
$E[x_{ul}] = w$	$W_{ul} \sim 1 - \mathcal{N}(w, 1)$
$E[x_s] = \frac{trials}{\lvert V \rvert}$	$W_s \sim 1 - \mathcal{N}(\frac{trials}{\lvert V \rvert}, 1)$
$E[x_{sl}] = max(1, \frac{w}{\lvert V \rvert})$	$W_{sl} \sim 1 - \mathcal{N}(max(1, \frac{w}{\lvert V \rvert}), 1)$
$E[x_g] = \frac{trials}{w}$	$W_g \sim 1 - \mathcal{N}(\frac{trials}{w}, 1)$

We have argued that sequence structure may affect cognitive performance and that consequently such features need to be controlled. We argued for the use of the constraint satisfaction framework as a principled approach to evaluate and generate qualified sequences. This approach operates on structural variables which may or may not affect human behavior and thus may or may not require stringent control.

To evaluate the relevance of the structural variables highlighted above for the n-back task we will analyze an existing dataset which did not explicitly manipulate or control for these structural variables. If these structural variables are informative about participants' n-back performance it follows that they are scientifically relevant and should be explicitly listed and constrained for both sequence generation and performance evaluation.

A.2 Evaluating behavioral impacts of structural features

A.2.1 Data

We used a previously published n-back dataset from (Cardoso-Leite et al., 2016). This dataset contains n-back data from 60 healthy adults (M=20.68, SEM=0.42) completing both the 2-back and 3-back versions of the n-back paradigm. For each version participants completed 3 sequences of 30 trials each which resulted in a grand total of 360 n-back sequences and 10'800 trials. On each trial, stimulus identity, reaction time and accuracy were recorded. For more details about this dataset, see (Cardoso-Leite et al., 2016).

A.2.2 Analysis

To evaluate the need to control for structural variables we fit and contrast two nested models that predict participants accuracy on a trial-by-trial basis, using a different set of predictor variables.

The *base model* uses the common approach of relating performance to descriptors of the sequence as a whole (i.e., x_n, x_v, and x_t) as well as the current stimulus (i.e., target or distractor) to predict the accuracy of the response to the current stimulus.

The *extended model* includes in addition all the structural variables listed in Table **??** (e.g., x_l, x_u, x_s). These structural variables are computed not on the sequence as a whole but rather on the recent stimulus history (8 previous stimuli, excluding the current stimulus). This approach exploits local variation along the dimensions of the structural variables to evaluate the impact of those variables on accuracy.

The data was subdivided into a training (80%) and a test set (20%). Both models were fit to the same training set using the imbalanced Partial Least Squares (PLS) method; this method was chosen because most responses were correct (92%) and the predictor variables are not mutually independent. Both models were then evaluated by their ability to account for test data using the area under the curve (AUC) as the model performance metric. The reliability of the AUC was further characterized using bootstrapping (1000 repetitions).

Two main conclusions can be drawn if the extended model outperforms the base model: a) structural variables affect behavior and hence need to be controlled by the sequence generator, b) even when they are controlled at the level of a sequence as a whole, local variations in structural variables may already be enough to affect behavior and it might be necessary to use trial-by-trial estimates of local properties to analyze human behavior and brain activity.

A.3 Results

Figure A.1 shows the ROC curves for the two fitted models. The base model predicts response accuracy above chance level (AUC=59.51; $CI_{95\%}$ = [54.81, 64.21]). The addition of structural variables as predictors in the extended model improves model performance substantially (AUC=68.56; $CI_{95\%}$ = [65.76, 71.36]).

To determine which variables drive the performance accuracy of the extended model, we ran a model-based variable importance analysis using the *Boruta* package in R (Kursa & Rudnicki, 2010). These importance scores were calculated using random forest method alongside shadow features, which are copies of original features but with randomly replaced values; this serves to remove the importance of a feature while nevertheless maintaining their distribution of values unchanged.

This analysis shows that the structural features computed on the recent history contributes most to the predictability of participants' accuracy. Figure A.2 shows the relative importance of the predictor variables used by the extended model.

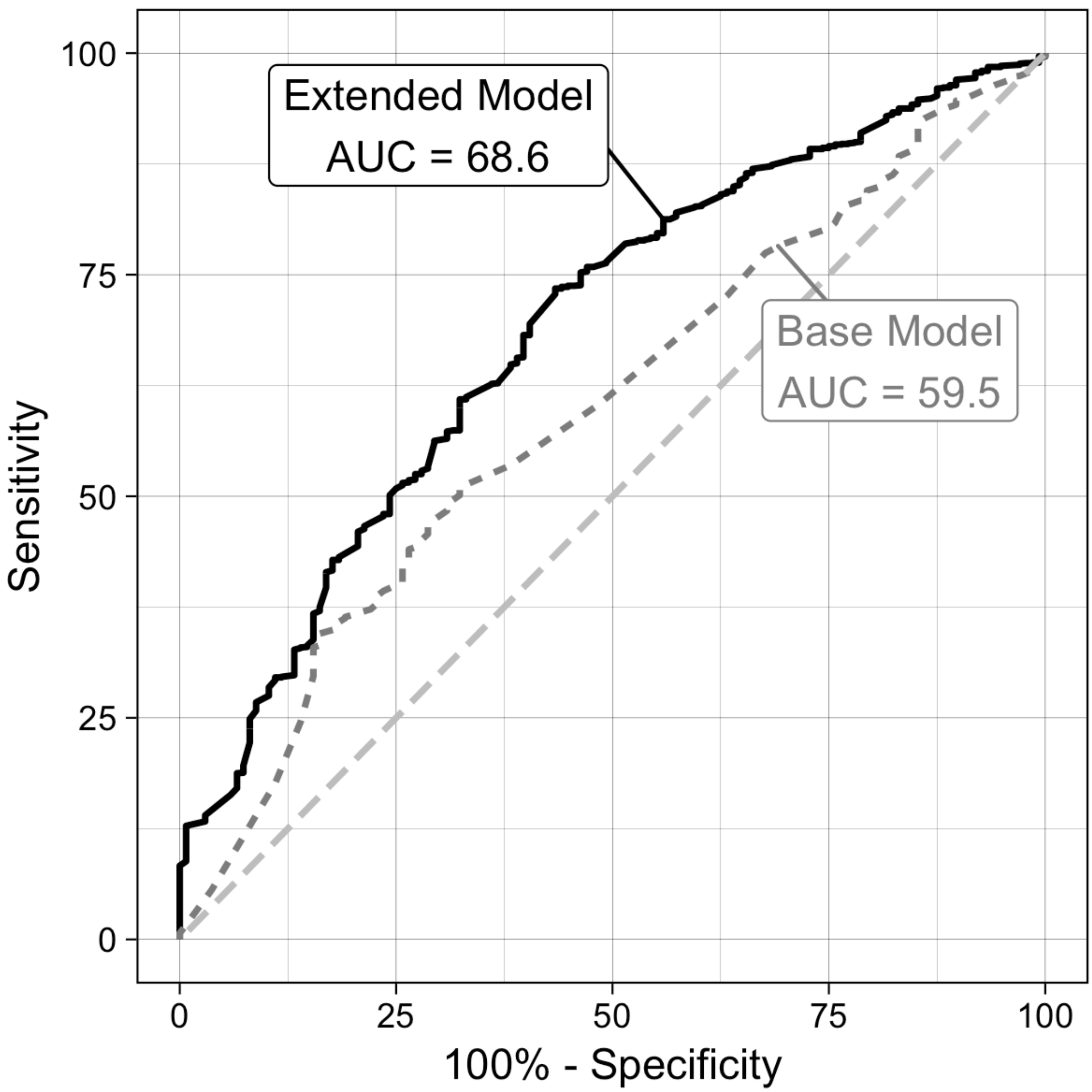


Figure A.1: Classification performance for the base and extended models. AUC = Area Under the Curve

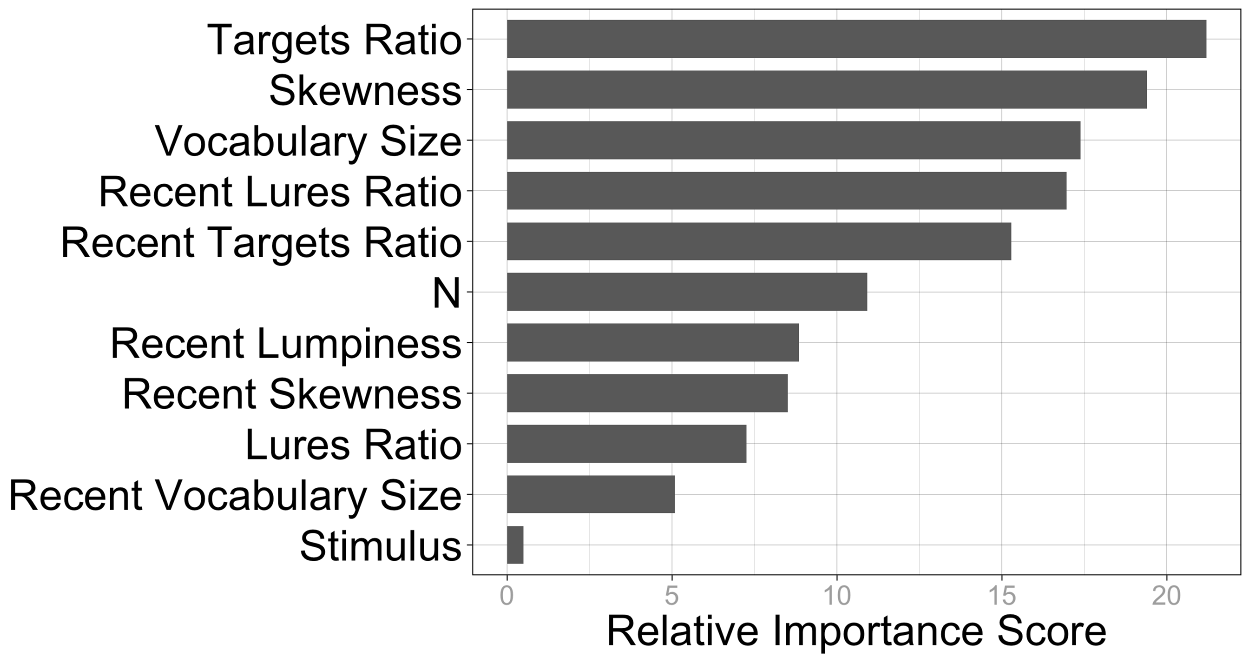


Figure A.2: Relative importance of structural variables (V) on the prediction of participants' response accuracy.

Although a direct causal relationship cannot be inferred from the results, higher contribution of recent trials in the extended model (i.e., higher relative importance of x_{vl}, x_{tl}, x_{ll}, and x_{sl} than their global counterparts, x_v, x_t, x_l, and x_s) suggests that behavioral responses are partially guided by a more fine-grained set of structural features.

A.4 Conclusion

In sum, we propose a compositional framework to parameterize and exploit interesting features of the n-back sequences and evaluate behavioral effects of the features of random sequences. We developed two predictive models to compare the importance of these structural features.

Methods that are commonly used to generate n-back sequences use independent random sampling for each trial and cannot control all the influential features. Instead of an independent random sampling process, we proposed a framework to reformulate generating the n-back sequences as a soft constraint satisfaction problem. This approach can be used to formalize the effect of structural patterns in other cognitive tasks that present random sequences of stimuli.

Appendix B

Behaverse data model

Aurélien Defossez, Morteza Ansarinia, Brice Clocher, Emmanuel Schmück, Paul Schrater, and Pedro Cardoso-Leite

B.1 Introduction

Experimental psychologists have been collecting behavioral data for over a century now. As psychological sciences and related fields are maturing, it has become increasingly clear that the field needs to establish and converge on standards and standard operating procedures.

Data is essential to science. The recent rise of the open science movement and the increased propensity to share and reuse data, as well as the need to integrate results across multiple studies (e.g., within meta-analyses) has revealed many shortcomings in the way we currently process our datasets and has motivated several initiatives aiming to make these datasets easier to find and use. Prominent examples include BIDS (Brain Imaging Data Structure, bids.neuroimmaging.io; see Gorgolewski et al., 2016), which focuses on brain imaging data and NeuroData without border (Teeters et al., 2015) which tackles neurophysiological data.

Behavioral data, however, has received comparatively less attention, perhaps because at glance sight it appears simpler than those large imaging datasets. We argue that behavioral data is in fact more complex than meets the eye and that defining clear standards for behavioral data may benefit all fields that rely on such data.

Standardizing how we define, name, format, organize, describe and store behavioral data can provide multiple benefits, including:

- efficiency (e.g., less work, reuse of code, automated software);
- robustness (e.g., less errors because of ambiguous idiosyncrasies);
- transparency (e.g., fewer hidden choices in the code and data);
- quality (e.g., via automated checks of data quality, consistency and completeness);
- usability (e.g., via clear documentation, ready-to-use data).

Note also that non-standardized data formats call for non-standardized data analyses which may obfuscate results at a time where more papers are published than anyone can read. By contributing and using data standards, we may accelerate scientific progress in psychological sciences, as seems to have been the case in other fields (for examples, see Teeters et al., 2015).

Here we present key ideas, concepts and principles that guided us in creating the Behaverse data model (v2020.12.1); the more detailed, somewhat opinionated and continuously updated specification of this data model is accessible at behaverse.github.io/data-model. While there have been significant efforts to make behavioral data easier to share and find, our focus here is on structuring behavioral datasets to both reveal the essential structure common to behavioral data and make them easier to (re)use.

B.2 Challenges of behavioral data

There are key challenges to systematizing behavioral data.

First, behavioral data is highly diverse, as it includes body movement, gaze, key presses, mouse clicks, written output and speech to name just a few. We currently have no clear standards for each of these measurement types, no standards that would be consistent across measurement types and no standards on how to relate multiple measurement types (both conceptually and practically). Hence, while we are technically able to record rich, multivariate behavioral datasets, we lack the conceptual and software tools to effectively exploit that richness.

Second, to interpret behavioral data it is necessary not only to characterize the behavior itself but also the context in which that behavior occurred. Taking as an example the most basic of cognitive tests, a particular key press is interpreted as being a response to a particular stimulus within a particular task that evaluates to "correct" or "incorrect"—the key press on its own, however, is not very informative. Note that this is not necessarily the case for other types of measurements (e.g., functional connectivity between two brain areas). Hence, the accurate description and effective processing of behavioral data requires rich annotations of the task and its underlying theoretical constructs, the stimulus and the person's state. Major efforts have been made in this direction (e.g., R. A. Poldrack et al., 2011); however, current solutions haven't yet matured enough to be an integral and standard part of the behavioral data analysis process.

Third, and related to the previous point, the way we describe behavioral data is limited by our understanding of what a task is. Indeed, although "tasks" or "tests" are the cornerstones of experimental psychology and related fields, we do not have a theory of tasks (which could for instance characterize the structural relationships between any two tasks) or even a clear framework on how to name or think about fundamental concepts like "instructions", "feedback" or "trial", let alone how to convert them into usable data structures—this applies not only to concepts in psychology but also more general concepts like "raw data". This lack of clarity on concepts that are pervasive in behavioral data have led to the discarding of what seems to us to be critical information (e.g., task instructions not being recorded anywhere) and is at least partially responsible for the large inconsistencies one may find today across publicly shared datasets (e.g., names, meanings and units of measurement). Hence, there is a clear need to better conceptualize tasks, clarify concepts and converge on standards.

Finally, the current practices and software tools used today for behavioral data analyses seem inadequate to handle the rich and complex data structures that seem necessary to accurately describe behavior. Without a clear understanding of those data structures we can't create effective tools that exploit that richness; but without effective tools there is no incentive for researchers to invest effort in structuring their data accordingly. Hence, until we have clear standards, well-structured rich datasets, effective data analysis software and a demonstration of added value, most researchers will understandably continue to work the way they've done in the past. Hence, while we should aim for better standards and tools, we still need to take into account current practices and tools and offer solutions that can be useful today.

The challenges we just described are considerable and overcoming them will require sustained efforts over many years. Our goal here is to contribute to overcoming these challenges and improve the way we describe and organize behavioral data. The solutions we propose here focus on three dimensions:

- **clarity.** Below we describe various ways in which current datasets are inconsistent. We then present and define several key concepts for behavioral data, the most important of which being perhaps the notion of a "trial" which we define as an instance of a "task-pattern". Rows in a "trial table" are then formed by extracting data from event data according to a task-pattern (using a query-like process) and each row in the "trial table" needs to contain all the information that is necessary to evaluate that trial (i.e., determine whether the response was correct or not). We also define different types of data tables (e.g., "L1" data) as well as canonical data tables (see behaverse.github.io/data-model).
- **consistency.** There are many choices to make when structuring data. These include, for instance, which naming conventions to adopt (e.g., "RT" versus "response_time"), which specific names to use for a particular concept (e.g., "subjects" versus "participants") and in what units to express certain variables (e.g., "seconds" versus "milliseconds"). While many of these choices may be arbitrary, it is vital for achieving the overarching goal of consistency to actually make these choices and document them in

a clear way (Martin, 2009)—we have started this process and documented our choices publicly (see behaverse.github.io/data-model).

- **usability.** Our particular choices for structuring behavioral data is motivated by the desire to make this data model useful and compatible with the tools and processes most researchers already use today. More specifically, we focus on tabular data (rather than more complex data structures) and aim for a good balance between human readability and computer/data efficiency. As we describe below, behavioral data involves many different types of data which could be compactly stored in a wide range of related tables. Such tables would however be much harder to process for humans as the information about a particular trial would now be distributed over multiple tables. Instead, we define, a primary "trial table" that contains all of the high level information about a trial (in line with current practices), and whose primary key serves to connect additional, possibly subtrial data (e.g., the timestamp of each of the images presented during that trial). To keep this paper short, we focus here only on what we believe to be central ideas; more content and specifics are available in the accompanying website (behaverse.github.io/data-model).

B.3 Data consistency levels

In this section we describe how typical behavioral data currently available in public repositories look like and detail various issues that make it hard to reuse them. Behavioral data from experiments in psychology or related fields are currently scattered across multiple locations, including researchers' personal webpages or various public repositories (e.g., osf.io)—which over the past decade have made it much easier to find relevant datasets. Exploring these datasets quickly reveals large differences in how behavioral datasets are formatted, named, organized, described and shared—sometimes even within the same lab. Unfortunately, finding a behavioral dataset today is no guarantee that it will be usable at all and it seems that in most cases substantial work would be necessary to understand and use them.

Table B.1: Data Consistency Levels. It is our understanding that current standards in behavioral sciences places us within levels 0 to 1.

Level	Description
0	The dataset is incomplete; critical information is missing (e.g., description of what the variables mean).
1	All datasets are formatted in a unique way and can't be joined without reformatting.
2	Datasets can be joined when they originate from the same task ”variant” (e.g., a 2-back task using digits)
	but not from distinct variants (e.g., a 2-back versus a 3-back task).
3	Datasets can be joined across all variants of a a task (e.g., all N-back tasks).
4	Datasets can be joined within a family of tasks (e.g., all CPT-like tasks).
5	Datasets can be joined across several task families.
6	All datasets can be joined.

To qualify the current state and future progress in behavioral data standardization we devised a *data consistency* scale which describes 7 levels of consistency, defined by the type of table joints—or merging of different data tables—that a data model supports (see Table A.1). Next, to get a rough sense of the data consistency level in cognitive psychology, we selected three popular cognitive tests—the digit-span task, the N-back task and the AX-CPT task. We then searched, downloaded and reviewed recent datasets from osf.io. Our goal here is not to make claims about the quality of the specific data samples we chose or of the research conducted using that data (hence, we keep them anonymous). Our goal is also not to be exhaustive and have a definite characterization of the current state of affairs. Instead, we want to point out the diversity and inconsistencies that currently exist in such datasets and describe the various issues that one encounters right after discovering what seems to be a relevant dataset. Below we describe these issues in the order one would encounter them.

B.4 Inconsistent data formats

Most data sets seem to be in csv format. However, we also found several Excel files and proprietary formatted data which could not be read at all. Oftentimes, data is shared as a single data file (containing the data for all participants) or in multiple files that all have the same structure (e.g., one file per participant). These datasets rarely provide a codebook to explain the meaning and possible values in their datasets and it would therefore be necessary to manually go over other available materials (e.g., the corresponding research paper) to attempt to uncover that information.

B.4.1 Unknown or inconsistent data level

Behavioral data come in various levels of granularity. Some data sets might contain each response given by every participant while others may only include aggregated data for each person (e.g., one row per participant versus one row per trial). It is typically impossible to know which level of data granularity the shared data offers before actually opening and inspecting the data files.

It is also very common that data tables mix data that are from different sources or levels of granularity. For example, a data table might include trial-level data for each participant (i.e., a row for each response the participant gave) but at the same time have a column that indicates the age and gender of the participants (e.g., the values "21" and "female" repeated across all rows within a given participant) or even summary statistics (e.g., d'prime), whereby it can sometimes be ambiguous as to whether those summary statistics were computed on the trial-level and then joined to the trial-level data or whether they were computed using other data.

B.4.2 Inconsistent variable naming conventions

Naming variables is notoriously hard and unsurprisingly, there are numerous inconsistencies in variable names (Martin, 2009). We found inconsistencies in naming conventions across but also within datasets. Some data sets use lower-case "snake_case" (e.g., "n_correct") others use upper-case snake-case (e.g., "N_Level"). Some use CamelCase (e.g., "TrialList") or a mixture between CamelCase and snake_case (e.g., "V_FalseAlarm") or still something else (e.g., "TrialList.Sample"). Some variables may be in all uppercase (e.g., "CUE_ACC") or include information about the coding scheme (e.g., a column named "FEMALE=1"). While one may argue that such conventions are more or less arbitrary, it stands to reason that a given convention should be used consistently across a given dataset. This is not the case in the random sample of studies we've reviewed as within the same table we could find for example "Span_amount", "CorrectAnswer" and "TrialList.Sample".

We also note the variability with which the same construct is named and coded. For example, most if not all datasets have a variable to refer to individual participants in a study. Common variable names to refer to participants are "id", "Subject" and "SubjectID". The use of "id" may however be ambiguous (id could perhaps refer to trial index). Sometimes the values that this variable takes is an integer (e.g., 15), sometimes it's a concatenation of something that seems to be a study or condition name and an integer (e.g., "A_15"). Coding schemes for the subject variable may be somewhat arbitrary but there might be an issue when there are multiple datasets. For example are "A_15" and "B_15" different people or are they the same person (participant 15) that completed two different tasks ("A" and "B")?

Another variable that is common in behavioral data sets refers to individual trials within an experiment. Again we observed quite some variability. While it is common to use the name "trial" or "id", we also found datasets where the trial index variable was missing and seemed thus to be implicit in the order of the rows of the table and other cases where the "trial" variable was not used to refer to the index of the trials but rather to describe a type of trial (e.g., "start", "nontarget", "v_target").

B.4.3 Unknown values and units

Another common issue, which might be resolved by the use of codebooks, is the absence of information about the possible values a variable can take and what units a variable is expressed in. For example, it is very common for data sets in experimental psychology to include response time data. It is typically not possible to determine if they are expressed in milliseconds, seconds or minutes before inspecting the data and using domain knowledge to infer the units.

B.4.4 Conclusion

A quick review of publicly available datasets reveals substantial inconsistencies in the way individual researchers/research groups (including ourselves) structure their data. Such inconsistencies are inconsequential for researchers working on their own data but limit the reuse of data by other researchers and the aggregation across data sets, even for datasets collected using very similar tasks.

In what follows we first describe some key properties of behavioral data before introducing the behaverse data model we currently use.

B.5 Behavioral experiments require multiple types of data

Data from cognitive psychology experiments are often shared in the form of a single table where each row refers to an individual trial completed by a person. While it is convenient to only have one file for data-analysis, this "simplicity" is in fact illusory and valuable data is currently hidden within the associated paper, computer code (or still other documents), if not missing altogether.

Typical behavioral data collection scenarios involve collecting data that are semantically distinct but intrinsically linked by virtue of the data collection situation. Consider for instance a typical cognitive psychology experiment. A research group invites participants to their lab to complete a computerized version of the “digit-span” test twice. What type of information could one expect this study to collect? Below is a non-exhaustive list of the kinds of data that are or should be recorded:

1. Information about the study (e.g., who conducted the study, when and where; what was the intentions; is the study approved by an ethics committee; what was the funding source); this information is typically idiosyncratically present in manuscripts but should be structured in a standard way, for example, in a “Study” table.
2. Information about the participants. This can include variables like birth date, gender, or nationality. Part of this information may be in the manuscript (e.g., “we recruited participants from city X”) and part of it may be in the trial data (e.g., the “age” and “gender” variables that are in the trial-level data). It is important to note that some information about participants is fixed (e.g., birth date) while other information may be context dependent and linked to the actual moment of data collection (e.g., age). Static information about the participant should be stored in a “Subject” table, while dynamically changing information (e.g., age) might be stored in a “Session” table.
3. Information about the activity participants engaged with. In cognitive tests, this would include for instance the name of the task, task parameters, the instructions given to participants. This information is typically buried in a research paper and often incomplete (e.g., the actual task instructions, although essential, are rarely listed in full). More and more often, the actual code that was used to run the activity is made available as well—but it may require significant work to uncover task parameters from code. Information about the task or activity should be organized in an “Activity” table.
4. Information about the hardware being used and of participants’ physical environments. For example, this could indicate particular brands and models of tablets or computers, versions of OS and software.
5. Information related to the interactions between the participant and the computer/environment, in particular information about what stimulus was shown, when

and where and what inputs participants made.

6. Information about events that occurred while participants were engaged in the activity. For example, this could include information about the quality of the data collection process (e.g., average frame rate) or observations made during the experiment (e.g., experimenter notes that a participant seems to be falling asleep); this type of information might be stored in a lab or personal notebook.
7. Information about participants progress through the study (e.g., list of participants having completed one test but not the other, data and time of completion of tasks, order of task completion).

The list above is not exhaustive but includes the main types of data that could in principle be collected in all behavioral experiments. The point we want to make here is that a data collection campaign comprises in fact multiple data tables and each data table has its own type (i.e., specific requirements, formats).

Our goal in this document is not to go over each of these data types and review existing solutions (although such an enterprise would certainly be useful). Our primary focus in this document is on the data type (5) which we'll refer to as the actual **behavioral** data. In our opinion, this is the data type that has received the least attention and presents the largest inconsistencies across studies. It is also the type of data that is most relevant for behavioral data analysis and which would most benefit from standardization.

B.6 Behavioral, interaction data

There is a lack of clarity on the meaning of terms that are commonly used in behavioral data (e.g., what constitutes "raw data"? what is a "trial"? what is a "task"). In behaverse.org/data_model we define several of those terms and other conventions we use in the behaverse data model. In what follows, we attempt to present the big picture view of behavioral data and clarify essential terms.

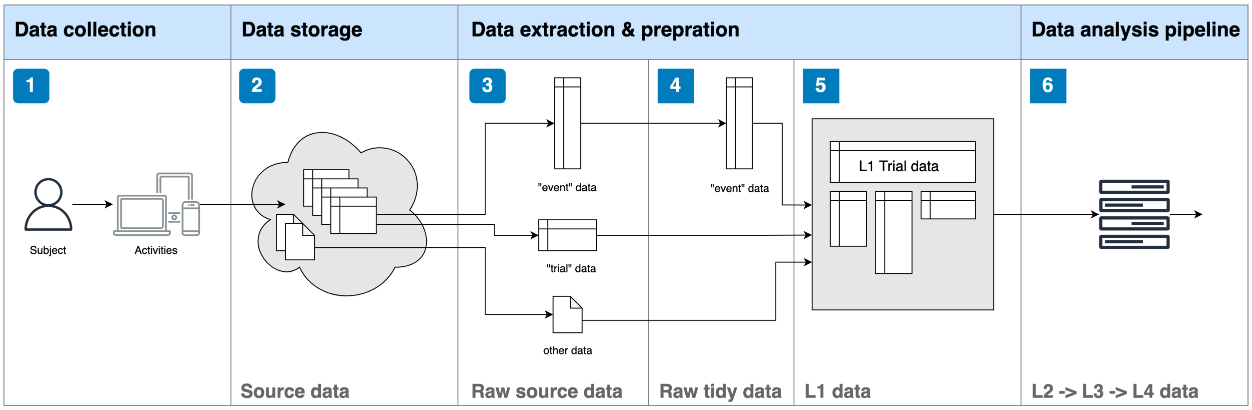


Figure B.1: From data collection to analysis. 1) Subjects interact with digital artifacts and produce data. 2) The resulting data ("source data") is typically stored in idiosyncratic formats, possibly determined by technical constraints of the digital artifacts. Furthermore, this "source data" may contain data that is not of direct relevance to researchers (e.g., technical information about the software) and important information may come from other sources (e.g., information about the study that is present only in the corresponding research paper). 3) It is typically necessary to extract the relevant data from the source data. Here we distinguish "event" data and "trial" data. Event data describes the behavioral data as a sequence of time stamped events, which have specific types (e.g., a mouse click) and data (e.g., the screen coordinates of the click). Trial data organizes those events following a task-pattern into a tabular form, where each row describes one trial. Further data files are necessary for example to describe the study. Note that it is typical for the data collection artifacts to already embed some data processing code and keep as source data only the "trial" data. 4) The most important type of behavioral data appears to be the event data from which different trial datasets may be extracted—this is in our opinion what should be viewed as the raw data and it will be valuable in the future to standardize behavioral event data and develop effective tools to deal with such data and extract trial-based data from them. 5) We define as Level 1 data, the data tables which are organized by trial. These are the tables we believe are most useful given current practices. In particular, we define the L1-Trial table, where each row contains complete and standardized information describing a particular trial (as is already currently the case, albeit inconsistently) and where the trial identifier is used as a primary key to additional, more detailed or specific tables (e.g., a table describing each of the mouse clicks that occurred during a trial). 6) The L1 data serves as the standardized input to data processing pipelines, which will derive additional tables (e.g., L2, L3), for example by transforming and summarizing data or aggregating across subjects

B.6.1 Source data, raw data and derived data

We consider as source data, all the data that is saved by the data collection artifact (e.g., computerized cognitive test) in its original structure and format (e.g., a single data file in a proprietary data format; multiple json files). Source data can contain all sorts of data. It includes the raw data but may also include metadata (e.g., information about the artifact itself) as well as derived data (e.g., a performance score computed from the raw data). Source data is typically in idiosyncratic formats and not usable as is.

Not all source data is raw data; and raw data needs not be source data. There are certain operations that can be performed on the raw source data to extract and constitute a dataset that is more usable without that dataset losing the "raw data" status. For example, if a source file is saved as a csv (comma separated values) file, converting that csv file into a tsv (tab separated values) file, is a trivial operation that has no consequences on the outcome of the study. On the other hand, filtering out some data based on performance or rounding numeric values are operations that may impact the outcome of subsequent analyses; hence the data that results from applying those operations can no longer be considered "raw".

Operations we consider to preserve "rawness" are selection by type (not by value), removal of duplicates, renaming of variable names for clarification, change of units, reordering of rows and columns and referencing/indexing (e.g., numbering rows of a certain type) and reversible file format conversion (e.g, csv to tsv). In short, as long as the information in the data is equivalent to the information in the raw source data, in our opinion, that data can be said to be raw.

B.6.2 Event data and trial data

Two common ways to structure behavioral data are by event or by trial (source data may contain either event data or trial data or both). Event data lists particular events that occurred during a study (e.g., a person pressed a key, a stimulus was displayed on the screen) with a timestamp (i.e., when did that event occur) and information describing the event (e.g., where on the screen did the click occur, how long did it last). The event data format is common in cases where behavior is related to other, time varying measures (e.g., in fMRI or EEG studies); it is much less common in behavioral sciences where information about when particular events occured is often discarded. In those fields, it is much more common to structure the behavioral data by trial, meaning, as a table where each row corresponds to a “trial” and each column to a variable describing what happened during that trial (e.g., for trial_index = 3, correct = TRUE).

It is important to note that beyond the shape factor, trial data and event data are quite different. Event data may describe events as they occurred and are thus more objective (e.g. a click occurred at timestamp 6.824). Trial data, on the other hand, are fundamentally tainted by the experimenter who needs to define (typically implicitly) a “task-pattern” which defines which events to select from the flow of events that occurred during the study and how to aggregate and/or transform them in order to constitute a row in the Trial table.

Let’s take an example to make this point clearer. In a N-back task, participants are shown letters, one at a time, and asked to report whether the letter that is currently displayed is the same as the letter shown N steps earlier. Let’s further compare a 2-back and a 3-back test that use the exact same sequence of letters. The event data from these two tasks may look virtually identical (they have events describing the occurence of letters and key presses). The trial data, on the other hand should look differently because for the 2-back test we use a different “task-pattern” than in the 3-back test. For example, in the first case we might describe the stimulus of the first two trials as “3-1-3” and “1-3-4”, while the same sequence of events in the 3-back task only forms one trial whose stimulus could be described as “3-1-3-4”.

Figure B.1 shows various steps in the lifetime of a dataset, ranging from its collection to the aggregation of summary statistics across participants. The format and structure of the source data is subject to various engineering constraints and specific to particular data collection software systems; it is therefore unlikely that we’ll converge on standards for source data that would apply to all use-cases any time soon. However, we could aim to define standards for raw event and trial data which could be readily used as input for data analyses pipelines and shared on public data repositories.

Here we focus on describing the L1 data, leaving for later standardization efforts of event data. This choice is motivated by our belief that standardizing trial data will be of most practical value to the research community.

B.6.3 Key concepts for specifying trial data

The data format that seems most useful and characterizes many shared behavioral datasets displays one row per "trial"—we call this the "Trial table". For example if an experiment tested 50 participants and each participant completed 200 trials, the Trial data table would contain 10'000 rows in total (assuming all the data was in a single table).

It is important to note at this stage that the term "trial" is not used in a consistent manner in the literature and the corresponding data files. The following section aims to highlight and clarify this issue.

B.6.3.1 The meaning of "trial"

Different meanings are associated with "trial". Firstly, "trial" may be used to refer to iterations of a chunk of code that is executed repeatedly (or equivalently a sequence of stimulation and input recording events). For example, a trial may consist of the presentation of an image on the screen and the recording of a keypress made by the user after the appearance of that visual stimulus. Secondly, "trial" may be used as an index to refer to individual rows in a data table. For example, each time the user presses a key we add a line to a data table that indicates which stimulus was shown and which button the user pressed. Thirdly, "trial" may refer to an instance or sample of a specific experiment in the statistical sense. For example, we want to determine if a particular coin is biased and repeatedly throw that coin and record the outcome; each throw represents a trial of that particular experiment. Finally, "trial" may be used to refer to a period of time or "episode" during the experiment (e.g., "the participant blinked during the second trial", "there was a 5 minutes break between trials 50 and 51"). In the most basic cognitive tests, all three meanings are congruent and thus interchangeable. But as experimental designs increase in complexity, even slightly, those notions are no longer equivalent and it becomes necessary to use more precise terminology.

Let's take a simple example to illustrate this point. Imagine a task where a letter is shown for 1 second and participants have to press one of two keys in response to that letter during the subsequent second—this code loop then repeats 100 times. In condition-1, participants are asked to press the right key each time they see the letter X and to press the left key otherwise (a "Sustained Attention to Response Task" like test Robertson et al. (1997)). In condition-2, users are asked to press the right key each time they see the letter X but only if it was preceded by the letter A and to press the left key otherwise (the AX-CPT task; Braver et al. (2001)). Finally, in condition-3, both tasks are to be completed at the same time: a single letter is successively shown on the screen, but there are now two sets of buttons, one per task.

While the same code can be used to run these three conditions, from the perspectives of the participant and researcher, they are different in important ways. In condition-1, we would expect the stimulus description to refer to a unique letter, while in condition-2, a stimulus would refer to pairs of letters (this information is necessary to determine in each case whether participants' responses were correct or not). Furthermore, if condition-1 and condition-2 use the same sequence of letters, the resulting number of trials will be different across the two conditions. Consequently, in this example, a "trial" in the code-loop sense no longer maps directly to a "trial" in the table index sense as information from two different code-loop trials is now contained in a single table-index trial. Next, if we consider the second experimental condition, one might assume that an experimenter will be interested only in those instances where a letter X was shown and it was preceded by another letter. If those instances define "trials" in the statistical sense, then trials should count only these specific instances. For example, if we assume that there were 100 code-loop trials (i.e., presentations of letters) but only 5 of those presented the letter X then there could at most be 5 trials (in the statistical sense) in that experiment, and thus only 5 rows in the corresponding data table. Finally, if we focus on condition-3, we see that for a given letter, there are two "trials" (one per task) occurring at the same time. Trial in this (and other cases) can therefore no longer be used to refer to a time period—to refer to particular, temporally distinct and non-overlapping time periods in an experiment we recommend to use "episode" instead. In condition-3, we could then have the same episode index correspond both to the 5th trial of the first task and the first trial of the second task.

The example above illustrates that "trial" can be used in inconsistent ways and that it is necessary to clarify its meaning. Within the behaverse data model we use the statistical definition of trial and define a trial with a corresponding task-pattern (see below). For indexing rows in a table we use a more generic "id" variable and for indexing particular time periods in a study we use "episode".

B.6.3.2 The task-pattern

Consider again the example experiment presented earlier where under two different conditions, letters were presented successively and participants were required to press one of two keys in response to those letters. The event data from both of these conditions could virtually be identical, with the same type of events being recorded each time a stimulus is shown or key is pressed. However, the corresponding trial data would look rather differently across both sets of conditions.

One can think of the trial data as something that is "created" from the event data (+ some other stuff). Indeed, one could write "extraction" code that would parse the event data looking for specific sequences of event types, extract the data corresponding to those event types and process and shape them into a row of the trial table—we call this code the "extractor" and save its parameters together with its trial data.

The specific sequence of event types, used by the extractor to query the event data, is what we call the task-pattern (in analogy to pattern in regular expressions). A task-pattern is typically of the form {stimulus-set; action-set}. In condition-1 of our example task, the stimulus-set might be all letters, while in condition-3 it might be all pairs of successively presented letters or all pairs of letters where the second letter is the letter "X" (depending on the experimenter's intention). In both cases, the action-set is any of the two possible button clicks that occur within 1 second after the stimulus. Task-patterns can of course be more complex; the key idea here is that the definition of a trial of a particular type is determined by a task-pattern. In the behaverse data model, when we index a trial, we index trials for a given task-pattern.

There are two points we want to emphasize here. Firstly, while the event data can be seen as an objective description of what actually happened during a study (e.g., the letter "A" shown on the screen center at 10:42:01"631"; the left arrow key was pressed at 10:42 02'246"), the trial data necessarily reflects the experimenters view of what that data means (e.g., the key press is a response to the letter, the response time is computed as the difference of times stamps and equals 0.615 seconds, and the response is correct given the current task rule). In fact, a different trial dataset could be generated from the same event dataset. The take-home message then, is that a) we need to store the event data as this data is privileged and more objective/raw than the trial data, and b) for a given trial dataset we need to maintain information about its provenance (e.g., the name of the task-pattern or extractor-code used to go from event data to trial data). Secondly, we believe that the concept of task-pattern is important beyond the context of data extraction and might be useful to characterize tasks for computational modeling or to implement artificial agents capable of performing tasks.

B.6.3.3 Evaluation

The task-pattern defines what constitutes a valid trial within a given experiment; it defines a subset of all possible stimulus and input sequences. Each element in this set of valid trials is mapped to a value. For example, it is very common in cognitive psychology for the response on a given trial to evaluate to "correct" or "incorrect". The value function or "evaluation" can be seen as a set of rules which are typically (implicitly) described in the task instructions (e.g., [to be correct:] "if you see the letter X press this key, otherwise press that key"); the value function may also be defined relative to an idealized policy—the particular way the experimenter believes participants should map stimuli (sequences) to action (sequences) within the context of the study.

B.6.3.4 Runtime extraction and evaluation

It is important to note that the software we use to present stimuli to participants and record their actions typically encodes information that reveals our intentions and may in fact distort the data. For instance, some researchers might not record event data and instead create the trial data directly as events unfold in time—their code instantiates an "extractor". This will typically discard data (e.g., when did a trial start) which makes it impossible to later reconstruct the time course of events as they occurred. Furthermore, that same code also typically includes evaluation code, as this might be necessary within the experiment itself, for example to display participants a correct/incorrect feedback signal for a given response.

It can be convenient and sometimes necessary to have these data processing functions embedded in the data collection code and operate during runtime on the events as they occur. However, one should also be wary of the fact that this code may contain errors. If we record only the output of those processes, i.e., runtime generated trial data but no event data, it might be impossible to detect and ultimately correct those errors.

B.6.3.5 Trial data versus L1-data

When describing the data that is extracted from the event data we used both the terms L1-data and Trial data in the sections above. These two terms, however, are not synonymous. Rather, L1-data refers to the state of the data (typically multiple tables) within a stage of the data analysis pipeline (see Figure B.1). Trial-data, on the other hand refers to a specific type of data table where each row contains data from a single trial as defined above. In the next section we'll review the structure of the L1-data, and discuss what other tables besides the Trial table may exist within L1.

B.6.4 L1 data model

Behavioral data (e.g., from computerized cognitive tests) are typically shared in a tabular format (e.g., one csv file per task), where rows typically correspond to individual “trials” and columns refer to different types of variables that describe that trial (e.g., response time). This, however, is insufficient. Firstly, it is already the case that the single-table trial-data does not include all necessary information. For example, it is typically necessary to read the paper about that data to learn about task parameters that did not vary across trials (e.g., the duration of stimulus presentations). Extracting that data and putting them in a consistent format would facilitate subsequent data usage. Secondly, behavioral data contains information that can be grouped into different semantic categories. These subcategories may have nested structures which do not play well with a simple single-table format but may instead be properly organized into multiple sets of tidy tables. More specifically, we define the following semantic data categories for the L1 data:

1. **Context:** provides context information for a particular trial, such as, identifiers for a study, a session, a participant and task.
2. **Task Information:** describes the tasks participants were exposed to (e.g., instructions, task parameters).
3. **Extraction Information:** describes how event data was converted into trials.
4. **Stimulus Information:** describes what stimuli were presented to participants.
5. **Options Information:** describes the different options participants had for responding on a given trial.
6. **Input information:** describes the actions participants made (e.g., a button click).
7. **Response Information:** describes the meaning of participants inputs within the context of the task (e.g., option “match”).
8. **Evaluation:** describes the value associated with participants’ responses (e.g., this response was correct); this value is not necessarily communicated back to the participants.
9. **Feedback Information:** describes if and how participants received explicit informa-

tion about their response or performance (e.g., green check after a correct response); this data describes physical events shown to the participants. Note that one may have the case where a "green check" feedback is shown to participants after an incorrect response (i.e., evaluation and feedback are distinct constructs).

10. **Outcome Information:** describes the consequences of the participants' action in the test. For example, in a serial ordered search task, participants are asked to open boxes to search for a token. Opening a box has the outcome of revealing its content and changing the state of the world (e.g., it reveals an empty box). While an outcome may implicitly contain feedback information, it is not necessarily the case. On the other hand feedback is solely meant to convey participants information about their performance. Outcome and feedback and evaluation are distinct constructs. In our box opening example, a participant may correctly click on an empty box (evaluation), see a green check (feedback), and see that the box is in fact empty (outcome).
11. **Reward Information:** participants sometimes get a reward in tests; this could for example take the form of points, money or even food.
12. **Experimental Design Information:** provides additional, optional data or features that the experimenter believes will be useful to interpret participant's responses (e.g., tagging certain trials in the N-back task as being "pre-lure" or "post-lure" with the intention to contrast performance on these two types of trials).
13. **Hardware information:** provides information about the hardware that was used to collect the data (e.g., this keypress was collected from keyboard #2).
14. **Technical Runtime information:** provides information about how well the trial was executed from a technical point of view (e.g., were there unexpected lags?).
15. **Information about additional data:** provides information about additional measures that might have been collected during the study (e.g., brain imaging data).

Each of these categories could have its own table with additional tables associated to them because there are typically different subtypes of data for each of these (for example, there are different kinds of possible stimuli and each kind of possible stimulus could have its own table).

There are two points we want to make here. First, behavioral data, as we hope to have demonstrated, is more complex than typically assumed; it involves a myriad of interconnected data tables. Second, current practices and data analysis tools do not address this complexity and instead focus on an easier to handle subset of the data (i.e., only the data that is strictly necessary for a particular analysis).

In order to get a more comprehensive and consistent handle on all of the behavioral data while at the same time remaining compatible with current practices and tools we opted for a particular set of design principles to organize the multiple L1 tables (see Figure B.2).

The first principle is to keep a trial table which is similar to what is already customary in the field. Each row in this table describes one trial and columns may contain summary information about particular aspects of that trial. For example, in a digit-span task where the stimulus is a sequence of digits presented at a certain rate one may summarise the stimulus for a given trial as "3;4;5;1". We define standards and conventions for that trial table to achieve consistency across datasets (see behaverse.github.io/data-model).

The second principle is to separate information depending on whether or not it is common or specific (e.g., to a task) and whether it describes the trial as a whole or particular events that occurred during the trial. Taking again the example of the digit-span test, "3;4;5;1" describes the stimulus at the trial level and is thus present in the trial table. The timestamp of the digit 5 during that trial is specific to an event and is thus present in the stimulus table which describes all the stimuli that occurred within each trial.

The third principle is that the trial table serves as the master table with the id of each row in that table serving as the key to link all the tables within L1. For example, knowing from the Trial table that "3;4;5;1" was presented on trial_id 2378, one can find within the Stimulus table the list of stimuli shown during that trial together with the properties of those stimuli (e.g., timestamp, location, duration).

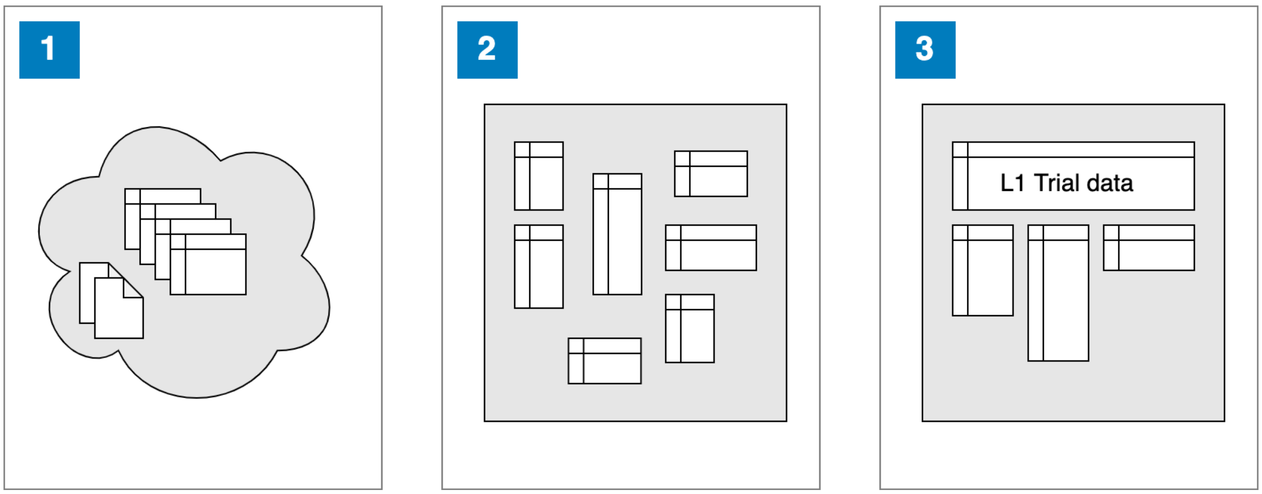


Figure B.2: L1 Trial data. 1) In source data, relevant information may be scattered across multiple data files in a way that is not practical for subsequent processing. There are various design options to reorganize the source data into data structures that can be standardized and are easier to use. 2) One solution is to factor the data into many compact tables within a relational database system. While this solution has many technical advantages, it doesn't play well with current practices. 3) An alternative design solution—the one we chose for the current behaverse data model— defines a main "L1 Trial" table which is similar to what researchers already use today. However, in addition to providing the trial data, the L1 dataset contains additional, related tables (as in 2). Tables in L1 are related to each other by various primary keys, the most important one being the trial identifier within the Trial table. We believe that this solution is both of practical use for researchers and offers the possibility to augment the Trial table in a principled way to capture more of the richness of behavioral data than is typically the case.

We believe that this design strikes a good balance between the somewhat contradictory requirements (e.g., the efficiency of a fully relational database versus human readability and ease of use); it is compatible with the way researchers are already structuring their trial data and offers a principled way to organize related data that is currently ignored but shouldn't.

B.7 Discussion

The standardization of behavioral data structures may not be the most exciting endeavour for a researcher—after all, great scientific advances were made without such standards, researchers can analyse data without following standards and it may seem to many that time spent on such mundane issues is time diverted from doing actual research. While there certainly is some truth to those statements, we believe that developing good standards for structuring behavioral data holds the promise for significantly improving the quantity and quality of behavioral research and may lead to novel insights.

As have argued many before us (e.g., Gorgolewski et al., 2016), standardizing data structures may increase research quality by clarifying concepts that are understood or used differently by different people. When those standards are public, they contribute to make science more open, transparent and reproducible. Finally, the use of standards can guide the development of various software tools that are specifically designed to take advantage of those standards.

There are a few examples that demonstrate how sometimes even simple data organization principles can lead to the development of an elegant and efficient software ecosystem that greatly facilitates the analysis of data. In the R community, for example, the notion of "tidy" data (e.g., "tidy data"; Wickham, 2014) has led and contributed to the development of the suite of tools known as the "tidyverse" (Wickham et al., 2019) which has had a massive impact on data science. Similarly, in the neuroimaging community, the BIDS' way of organizing imaging data has had profound positive effects for the field as whole, facilitating the sharing and reuse of imaging data but also leading to the development of software tools to check for example the integrity of data but also efficient and standardized data analysis pipelines (e.g., fmriprep.org; Esteban et al., 2019). What these examples show is that the development of standards for structuring data can lead to the development of tools and data analysis standards that greatly benefit the field. It is our hope that by contributing to standardizing behavioral data, equally impressive progress can be achieved in behavioral sciences.

In this document, we focused only on a few key concepts; other ideas are presented in greater detail in the projects' website (behaverse.github.io/data-model) which holds an updated version of the behaverse data model. Many questions remain unanswered, various aspects of behavioral data to be explored and numerous decisions to be taken. Ultimately, the value of this or any other data model will require demonstrating that it can indeed represent rich behavioral data across a variety of settings in a consistent way and that it offers concrete benefits to the researchers using those standards.

B.8 Conclusion

Behavioral data is fundamental in cognitive sciences and there is clearly a need for standards to organize such data so it can be efficiently analyzed, shared and reused. Here we emphasized several key issues and presented constructs we believe are essential for structuring behavioral data and which currently seem to be used inconsistently.

Much remains to be discussed. To keep this document short and decrease the likelihood of its content becoming obsolete as our standards evolve, we decided to focus here only on key points and refer the reader to the online documentation of the behaverse data model (see behaverse.github.io/data-model).