
Using video games to study the acquisition and performance of psychomotor skills

Ozan Vardal

Doctor of Philosophy

University of York

Computer Science

May 2023

Abstract

Understanding how humans learn complex skills is a fundamental aim of cognitive science. Digital games offer promising opportunities to study cognitive factors associated with skill acquisition and performance, as they motivate longitudinal engagement and produce rich, multivariate data sets. By applying multivariate analysis techniques to data arising from gameplay, this thesis extended the literature on cognition as it pertains to psychomotor skill. We describe three studies that were conducted in this regard.

In the first study, we analyzed the relationship between the temporal distribution of play instances and performance in a commercial digital game (League of Legends). Using clustering techniques and big data, we demonstrated that players who cram gameplay into short time frames ultimately perform worse than those who space the same number of games over longer periods.

In the second study, we examined an experimental data set of participants who played Meta-T, a laboratory version of Tetris. Using Principal Components Analysis and regression techniques, we identified cognitive-behavioural markers of performance, such as action-latency and motor coordination. We also applied Hidden Markov models (HMM) to time series of these markers, showing that moment-to-moment dynamics in performance can be segmented into behavioural states related to latent psychological states.

In the third study, we investigated the neural correlates of behavioural states during performance. Using simultaneous MEG and behavioural recordings of participants playing Tetris, we segmented time series datasets of neural activity based on time stamps of behavioural epochs derived from HMMs. We compared behavioural epochs based on neural markers, showing that cognitive states derived from multivariate behavioural data correlate with neural activity in the alpha band power.

Taken together, this thesis advances our understanding of using digital game data to study cognition and learning. It demonstrates the feasibility of recording high-density neuroimaging data during complex behavioural tasks and obtaining reliable measures of internal neuronal states during complex behaviour.

Contents

| | |
|---|-----------|
| Abstract | 3 |
| List of Figures | 7 |
| List of Tables | 11 |
| Acknowledgements | 12 |
| Author's declaration | 14 |
| Introduction | 15 |
| 0.1 Psychological Studies of Expertise and Skill Acquisition | 16 |
| 0.1.1 The role of talent | 16 |
| 0.1.2 The role of training | 18 |
| 1 Literature Review | 23 |
| 1.1 Previous reviews | 23 |
| 1.2 Aims and rationale | 25 |
| 1.3 Search strategy | 26 |
| 1.4 Laboratory studies of skill acquisition | 27 |
| 1.4.1 Studies using Space Fortress | 27 |
| 1.4.2 Model-based studies of Space Fortress | 30 |
| 1.4.3 Studies using commercial games | 31 |
| 1.5 Observational studies of skill acquisition and expert performance . | 32 |
| 1.5.1 Longitudinal studies using telemetry data | 32 |
| 1.5.2 Cross-sectional studies using telemetry data | 33 |
| 1.6 Cross-sectional studies of expertise and performance | 35 |
| 1.6.1 Skill-based differences in cognition and behaviour | 35 |
| 1.6.2 Detailed behavioural analyses of individual players | 37 |
| 1.7 Future research | 38 |
| 2 Effects of Practice Distribution on Acquisition | 41 |
| 2.1 Introduction | 41 |
| 2.1.1 Effects of practice distribution on learning | 42 |
| 2.1.2 Distributed practice in digital games | 43 |
| 2.1.3 Contributions of studies using behavioural telemetry from action games | 46 |

| | | |
|----------|--|-----------|
| 2.1.4 | Aims of the present work | 47 |
| 2.2 | Materials and Methods | 48 |
| 2.2.1 | Task environment | 48 |
| 2.2.2 | Measures | 49 |
| 2.2.3 | Data and preprocessing | 50 |
| 2.3 | Results | 52 |
| 2.3.1 | Time Gap Clustering | 54 |
| 2.4 | Discussion | 60 |
| 2.4.1 | Optimising training for MOBAs | 63 |
| 2.4.2 | Conclusion | 64 |
| 3 | Tetris as an experimental platform for cognitive science | 65 |
| 3.1 | Introduction | 65 |
| 3.1.1 | What makes a good research game? | 66 |
| 3.1.2 | Previous games for research | 67 |
| 3.1.3 | Aims of the present work | 68 |
| 3.2 | Methods | 68 |
| 3.2.1 | Task environment | 69 |
| 3.2.2 | Measures | 70 |
| 3.3 | Results | 71 |
| 3.3.1 | Principal component analysis | 71 |
| 3.3.2 | Distinguishing top from bottom scorers | 75 |
| 3.3.3 | Relating skill components to overall score across games | 77 |
| 3.3.4 | Hidden Markov models of skill components | 78 |
| 3.4 | Discussion | 82 |
| 3.4.1 | Limitations and future work | 85 |
| 3.4.2 | Conclusion | 86 |
| 4 | Identifying latent perceptual states using behavioural and neural recordings during Tetris play | 87 |
| 4.1 | Introduction | 87 |
| 4.1.1 | Detecting latent states | 88 |
| 4.1.2 | Neural correlates of internal states | 90 |
| 4.1.3 | Aims of the present work | 91 |
| 4.2 | Methods | 91 |
| 4.2.1 | Participants | 91 |
| 4.2.2 | Behavioural data acquisition | 92 |
| 4.2.3 | Magnetoencephalography (MEG) | 92 |
| 4.2.4 | Structural magnetic resonance imaging | 94 |
| 4.2.5 | Analysis | 95 |
| 4.3 | Results | 98 |
| 4.3.1 | SVM decoding | 98 |
| 4.3.2 | Hidden Markov Model analysis | 99 |
| 4.4 | Discussion | 105 |
| 4.4.1 | Limitations | 108 |

Contents

| | | |
|----------|---|------------|
| 4.4.2 | Conclusion | 108 |
| 5 | Discussion | 109 |
| 5.1 | Chapter 2 | 110 |
| 5.2 | Chapter 3 | 112 |
| 5.3 | Chapter 4 | 114 |
| 5.4 | Conclusion | 115 |
| 6 | Appendix B | 117 |
| 6.1 | Table of variables used in Principal Component Analysis | 117 |
| | References | 121 |

List of Figures

| | | |
|-----|--|----|
| 1.1 | Screenshot of Space Fortress game (from Lee et al., 2015). | 28 |
| 2.1 | Trajectories of mean GPM (left panel), and KDA (right panel) of all players against match. Shaded regions indicate 95% confidence intervals. | 52 |
| 2.2 | Trajectories of mean GPM (left panel) and KDA (right panel) against match for players with different patterns of match spacing. Data in the figure are a subsample of players who initiate at a similar range of GPM and KDA (approximately surrounding the original sample median). Shaded regions indicate 95% confidence intervals. | 54 |
| 2.3 | Scatter plots of GPM and KDA against time range in days between first and 95th game respectively, with line of best fit. Axis plots show distributions of respective axis variables. | 54 |
| 2.4 | The left panel shows the two-dimensional projection of the observed 95 inter-match gaps in hours as generated by UMAP for the entire sample. The y and x axes represent the two dimensions individuated by UMAP. As opposed to Principal Component Analysis their associated values should be interpreted as coordinates on a plane rather than indicators of the magnitude of the two components. Each dot represents the history of inter-matches gaps for a single player while distance between dots indicates the degree of similarity between different patterns of spacing. The right panel shows the average evolution of inter-match gap in hours for the entire sample. The y axis indicates the time in hours elapsed since the previous match while the x axis indicates the order of the match. The solid line indicates the mean value while the shaded region shows the 95% confidence interval. The dotted red line separates the observation period (i.e., the first 95 matches) from the evaluation period (i.e., the last 5 matches). | 56 |

| | | |
|-----|---|----|
| 2.5 | The left panel shows the two-dimensional projection of the observed 95 inter-match gaps in hours as generated by UMAP for each spacing cluster across the entire sample. The y and x axes represent the two dimensions individuated by UMAP. As opposed to Principal Component Analysis their associated values should be interpreted as coordinates on a plane rather than indicators of the magnitude of the two components. Each dot represents the history of inter-match gaps in hours for a single player while distance between dots indicates the degree of similarity between different patterns of spacing. The right panel shows the average evolution of inter-match gap in hours for players in each spacing cluster. The y axis indicates the time in hours elapsed since the previous match while the x axis indicates the order of the match. The solid line indicates the mean value while the shaded region shows the 95% confidence interval. The dotted red line separates the observation period (i.e., the first 95 matches) from the evaluation period (i.e., the last 5 matches). | 58 |
| 2.6 | Trajectories of mean GPM (left panel) and KDA (Right panel) against match for players in our 4 autoencoder clusters. Data in the figure are a subsample of players who initiate at a similar range of GPM and KDA (approximately surrounding the original sample median). Shaded regions indicate 95% confidence intervals. | 59 |
| 3.1 | Scree plot of PCA. Points in the line show the variance explained by each successive principal component, where components are ordered from most to least variance explained. The vertical dashed line indicates the point of maximum curvature in the line, corresponding to our choice of number of components to be generated for our analysis. | 73 |
| 3.2 | Averaged component score trajectories across two exemplar games. The top and bottom panels show mean <i>disarray</i> , <i>well preparation</i> , <i>action efficiency</i> , and <i>decision-action latency</i> from the first and fifth game across the entire sample (including only those participants who played at least five games) respectively. Shaded regions show 95% confidence intervals of the mean. | 75 |
| 3.3 | Comparison of moment-to-moment performance between top and bottom scorers. Each panel depicts the mean trajectory of the respective performance component for bottom (red line) versus top scorers (grey line) for the first 50 tetromino drops from game 1. From top left going clockwise to the bottom left panel, the subplots show trajectories for mean <i>disarray</i> , <i>well preparation</i> , <i>decision-action latency</i> and <i>action inefficiency</i> respectively. Shaded regions depict 95% confidence intervals of the mean. | 76 |
| 3.4 | Scatter plots of OLS regressions of score on performance component. | 78 |

| | | |
|-----|---|----|
| 3.5 | HMM state transition matrix. The figure shows the inferred transition matrix of the model, describing the probability of participants switching from one state (y axis) to the next (x axis). . . | 81 |
| 3.6 | Bar chart of state fractional occupancies. Each bar shows the fractional occupancy of the respective state, defined as the fraction of total time on task spent in that state across the entire data set. | 82 |
| 3.7 | Violin plot of principal component distributions across states. Each "violin" depicts the distribution of the corresponding performance component across all tetromino drops within occurrences of a given state. The black bar in the center of the violin is a box plot, with the center showing the median observed value in the distribution. The coloured portion of the violin shows the kernel density estimate of the distribution. The tails of the violin extending beyond the inner box plot show the range of extreme outliers. | 83 |
| 4.1 | Depiction of Meta-T user interface. The left side of the screen shows the current board, including the tetromino that is currently being controlled above it. The top right side of the screen shows the <i>next</i> tetromino that will be played after the current one is dropped. The player is also presented with the current game number, their current score, number of cleared lines, and the current level. Image taken from Lindstedt and Gray (2019). | 93 |
| 4.2 | Partial depiction of MEG-MRI co-registration process. The left panel shows the 3D rendering of the participant's cortex following import of their preprocessed MRI data into the Brainstorm analysis environment. The right panel shows the alignment of the participant's digitized head surface with their 3D head model. | 94 |
| 4.3 | Distributions of component score at each tetromino drop for each component. From top left going clockwise to bottom left, the plots show histograms of z-scored well preparation, action inefficiency, decision-action latency, and change in disarray respectively. | 97 |
| 4.4 | SVM decoding of left versus right button inputs for participant R3154. The top left two panels show the averaged 244 channel MEG time series from -400ms to 400ms for left and right translation inputs respectively. The bottom left two panels show corresponding heat maps of cortical source estimates in picoAmpere at 0.085 seconds after the input was registered by Meta-T. The right most panel shows the averaged percentage accuracy of the classifier at each millisecond of the MEG time series. The red vertical dotted line on each line plot shows the time point corresponding to the presented cortical activity. | 99 |

List of Figures

- 4.5 **Overview of three state HMM temporal dynamics.** The **top left panel** shows the inferred transition matrix of the model, describing the probability of switching between each pair of states. **The top right panel** shows the fractional occupancy for each state, defined as the fraction of total time spent in that state. **The bottom panel** shows the distribution of maximum fractional occupancy across acquisitions in the sample. That is, for each data acquisition in the sample, the maximum fractional occupancy represents the fraction of total time spent in the state that the participant occupied for the most amount of time for that acquisition. 100
- 4.6 **Time series of posterior probabilities and observed performance in an example game.** The **top panel** shows the time series of state posterior probabilities, describing the inferred probability of each state being active given the data, across all tetromino drops. **The bottom panel** shows the time series of all four performance components, displayed as z-scores, across all tetromino drops. 102
- 4.7 **Violin plot of performance component distributions across states.** Each "violin" depicts the distribution of the corresponding performance component across all tetromino drops within occurrences of a given state. The black bar in the center of the violin is a box plot, with the center showing the median observed value in the distribution. The coloured portion of the violin shows the kernel density estimate of the distribution. The tails of the violin extending beyond the inner box plot show the range of extreme outliers. 103
- 4.8 **Distributions of RMS cortical amplitudes across our regions of interest.** From top left going clockwise to bottom left, the plots show histograms of RMS alpha in the left primary visual cortex, RMS alpha in the right primary visual cortex, RMS mu in the left primary motor cortex, and RMS mu in the right primary motor cortex. 104

List of Tables

| | | |
|-----|--|----|
| 1.1 | PICO criteria of the review | 26 |
| 2.1 | Raw data columns available in a single row of the data set analysed in this study. | 50 |
| 2.2 | Linear regressions of time delta in days between 1st and 95th match on average GPM and KDA between the 96th and 100th match. | 55 |
| 2.3 | Joint frequencies of players in spacing group as defined by k-means cluster (rows) versus time in days delta between 1st and 95th match (columns). | 59 |
| 3.1 | Table of PCA loadings. Column headers display the number of each component ordered by descending amount of variance explained. Cells show the correlations past 0.20 between variables (labelled in the row headers) and the respective components. | 72 |
| 3.2 | Linear regressions of overall score in final game on mean <i>disarray</i> , <i>well preparation</i> , <i>action inefficiency</i> , and <i>decision-action latency</i> in the first game. | 79 |
| 3.3 | Mean <i>disarray</i> , <i>well preparation</i> , <i>action inefficiency</i> , and <i>decision-action latency</i> in each latent state. | 83 |

Acknowledgements

It is due to many wonderful people that the proverbial sweat and tears accompanying this PhD were minimised. Although you all deserve much more, I would like to express my eternal gratitude in these short pages.

First of all, I would like to extend a profound thank you to my supervisors. Tom, thank you for your patience and wisdom, notably in those early years as I struggled to find my footing. Without you I never would have arrived at IGGI, so thank you. Anders, I risk sounding immensely reductive by acknowledging that you are a rock (hey, your words, not mine) in the best of ways. Thanks for being a nigh indestructible totem of support throughout the years. Alex, you are a miracle worker who has injected me with the purest excitement for science and opened doors that I did not know existed. Thank you for being so generous with your time and extensive knowledge, for showing me how to be pragmatic and creative, and for bearing with me. You're an inspiration, and I feel very fortunate to have had you as a supervisor.

Doing a PhD with IGGI was unique in that it connected me with so many wonderful people. I'd like to thank everyone at the helm and EPSRC for making this possible. Jo Maltby, you deserve a special thank you for relentlessly taking care of business, we all miss you. I'd also like to thank Paul Cairns for being a constant source of reassurance and for asking me the fundamental questions.

Thank you Sally Quinn and Dan Baker for allowing me to teach psychology students the joys of statistical programming. Thanks also to everyone at YNiC, in particular to Richard Averyard and Rebecca Lowndes for showing me the (liquid-helium cooled) ropes and for enabling a much privileged final year of research.

To my friends at IGGI, none of this would have been possible without you. Thank you Evelyn, Ryan, Adam, Oliver, Matt, Nathan, and Adrian for being a generally amazing cohort. Thanks especially to Kevin for the constant companionship, Dani for the late night jam sessions, Lisa for your generous nature, Joe for the heart-to-hearts, and Pete for the long discussions and for laughing about Nothing. Thank you Nasos for all the advice and for the evenly matched Tekken. Laughing with you has almost resulted in lung collapse on more than one occasion, so I am not sure how grateful I should be. Thanks also to Valerio for helping me navigate

the trenches of ML and for the tagliatelle. Myat, you deserve a special thank you for always lending me an ear and for reigniting my passion for music.

I would like to thank my friends everywhere else. Thank you Samy, Nikolai, and Jeff, your friendships are invaluable to me. I couldn't ask for a better group of people to go on this adventure with. Thank you Marah for the thoughtful talks that have seasoned my life over the years. I am truly blessed to have you in my life. To Ebru, kanka what can I say yani. Thanks for spontaneously materialising wherever I am in the world. Thanks also to Dariusz, Alexis, Moni, Timur, and Ashleigh for all the sass and for creating such a unique community during the pandemic. Finally, thank you Lorenz and Joshua for taking me under your wing those many years ago and for staying with me.

Next, I'd like to thank my family. Thank you Mum for always believing in me. I'll always be your number one supporter, just as you were mine. Thank you Dad for your stoic advice and for your unwavering help whenever it was needed. Thank you Bora for letting me draw on you for strength. Thanks also to Gada, Gulcin, Gulsen, Gulhayat, Nurhayat, Alpay, for your kindness and for always giving me a second home. I love you all deeply.

Finally, thank you Damla. You took most of the shrapnel from the PhD, and you did so gracefully. Thank you for your warmth and for your patience. Thank you for your love and for keeping me grounded.

Author's declaration

I, Ozan Vardal, declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Parts of the works enclosed herein were performed collaboratively with colleagues and supervisors. These individual contributions are detailed below.

The data set analysed in Chapter 2 was provided by Riot Games through collaboration with Anders Drachen, Alex Wade, and DC Labs. Code for data preprocessing and rule-based clustering analysis written by the author. Code for the neural network and manifold clustering in Chapter 2 was written by Valerio Bonometti and the author. Interpretation of analysis results was performed by the author, Valerio Bonometti, Alex Wade, Tom Stafford, and Anders Drachen.

The archival data set analysed in Chapter 3 was provided and originally collected by Wayne Gray and John Lindstedt Rensselaer Polytechnic Institute. Guidance for application of Hidden Markov modelling to behavioural and in Chapters 3 and 4 was provided Theodoros Karapanagiotidis. Code for the analysis of this data set was written by the author, and interpretation of the results was performed by the author, Alex Wade, and Anders Drachen.

Operation of the Magnetoencephalography (MEG) scanner for data collection of Chapter 4 was conducted by Richard Aveyard at the York Neuroimaging Centre, with assistance from the primary author. Guidance for preprocessing of MEG data was provided by Theodoros Karapanagiotidis and Alex Wade. Code for the original Python implementation of Tetris was provided by Wayne Gray and John Lindstedt. Additional code for the Tetris task facilitating synchronisation with the MEG scanner was written by the author. Code for preprocessing and analysis of behavioural and MEG data in Chapter 4 were written by the author. Interpretation of analysis results was performed by the primary author, Alex Wade, and Theodoros Karapanagiotidis.

Chapter 2 of this thesis has previously been published as a journal paper:

Vardal, O., Bonometti, V., Drachen, A., Wade, A., & Stafford, T. (2022). Mind the gap: Distributed practice enhances performance in a MOBA game. *PLOS ONE*.

Introduction

The study of skill acquisition and expertise (i.e., domain-specific superior performance) has been a central topic in psychology for over a century. Presumably one reason for this is the assertion that if we understand processes of skill acquisition, we can manipulate them to our advantage. In line with this aspiration, a primary focus of this work is to further our understanding of the acquisition and performance of psychomotor skills. By psychomotor skills, we refer to learned behavioural patterns that involve the coordination of motor and cognitive functions (Fitts & Posner, 1967). By performance, we refer to the observed execution of these behavioural patterns.

This investigation is situated in the domain of digital games. In recent years digital games have received increased attention from cognitive scientists interested in the study of learning, owing in part to the availability of large, rich data sets that enable unobtrusive interrogation of human behaviour. Game players develop profound skill over years of play and practice, generating reservoirs of ecologically valid data about practice behaviour and skill development that can be unobtrusively recorded. This thesis will demonstrate that, analysed in detail, such data help us to understand processes surrounding complex skill acquisition and performance.

We anchor this investigation of learning to digital games for two main reasons. The first is to take advantage of the richness and vastness of digital game data, which enables us to study behaviour and cognition at a scale and level of detail that was inaccessible prior to the big data age. The second reason is to conduct research in a naturalistic domain. By extending cognitive science to digital games, we may test whether extant findings survive outside the boundaries of laboratory tasks, and in a real-world setting with millions of stakeholders.

In this chapter, we briefly introduce influential work on human expertise and skill acquisition. This is followed by a literature review of studies that have used digital games as the experimental paradigm to investigate how humans acquire complex skills, and what separates elite players from average and novice players. We conclude with a discussion of respective research gaps and the direction taken in this thesis.

0.1 Psychological Studies of Expertise and Skill Acquisition

In his thesis on the development of perceptual-motor expertise, Ward (2002) traced inquiry into human excellence back some 2000 years, citing early insights by Aristotle and Seneca. According to the former "We are what we repeatedly do. Excellence then is not an act but a habit". Seneca did not view the effects of ordinary, extended experience with such generosity, as evident in his correspondence with his contemporaries: "Toil to make yourself remarkable by some talent or other".

Although the benefits of specialist research methods and knowledge did not exist in the classical era, the intuitions of these thinkers have had enough plausibility to endure until present times. That is, whether expert skill is incidental to experience, the product of some genetic lottery, or the fruit of prolonged effort, has been a centre of much academic debate in recent decades (e.g., Howe et al., 1998; Simonton, 1999). While the issue is not fully untangled, the consensus view in the behavioural sciences appears to acknowledge the interacting roles of innate endowments, environmental factors, and practice (e.g., Campitelli, Connors, Bilalic, and Hambrick, 2015; Den Hartigh, Van Dijk, Steenbeek, and Van Geert, 2016). In order to justify the present focus on practice behaviours alone, it is important to briefly explore evidence pertaining to the contribution of talent.

0.1.1 The role of talent

Most reviews of research on expertise begin with Galton's 1869 seminal work *Hereditary Genius*. Analysing the characteristics and achievements of eminent British individuals across various domains, Galton proposed that excellence is determined by an individual's intellectual ability and personal motivation. Observing that these individuals descended from a small number of families (at a frequency much higher than chance) Galton further proposed that these qualities must be heritable. Thus in this view, individuals are born with innate properties or gifts that are key determinants of eventual excellence. Typically one refers to this configuration of innate properties as *talent*, more formally conceptualised as "any innate capacity that enables individuals to display exceptionally high performance in a domain that requires special skills and training" (Simonton, 1999, p. 436).

Notwithstanding Galton's acknowledgement of the importance of training, the hypothesis that talent is a key determinant of eventual excellence has received serious attention from researchers. To evaluate evidence in support of this position, Howe (1998) systematically reviewed a corpus of correlational research linking var-

0.1 Psychological Studies of Expertise and Skill Acquisition

ious biological factors and measures of cognitive ability with performance across domains ranging from music to sport. Howe concluded that individual differences in motivation, environment, early experiences, and practice were more likely to be determinants of excellence than talent. Noted were recurring issues underlying studies in favour of the talent perspective, such as a reliance on anecdotal evidence, failure to account for practice and training opportunities in retrospective studies, and restriction of range in correlational analyses.

In more recent efforts, Johnston et al. (2018) conducted a systematic review of talent identification research in sport. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher, Liberati, Tetzlaff, and Altman, 2009) statement guidelines, studies were included for review provided they contained an elite sample of athletes, a longitudinal or retrospective design, a comparison between skill brackets, and a removal of "grey-area" topics such as genetic predispositions, birthplace effects, and handedness, resulting in a final selection of 20 studies conducted between the years 1990 and 2015. Overall the authors found this body of work to be inconsistent in its ability to suggest reliable predictors of success in sport. While a fraction of studies identified some variables (e.g., sprint abilities, agility drills) that could differentiate between skill brackets, others did not. In particular, anthropometric measures such as height and weight were inconsistent in their predictive efficacy across different sports. Taken together the results of Johnston and colleagues (2018) are valuable in sketching current conceptualisations of talent in sport, but less so in supporting Galton's original postulation that *heritable* factors are an essential determinant of excellence.

A weakness of this review was the exclusion of studies pertaining to genetic predispositions associated with skill. A starting point to examine this angle is to consider the necessity of certain physiological traits for athletic success. For instance, ballet dancers' ability to turn out their feet appears to be genetically pre-determined, and attempts to force the necessary hip turnout (through practice) without the appropriate genetic makeup can result in injury (Hamilton, 1986). Similarly, it is difficult to argue against the importance of height in basketball. Besides such observable examples, decades of twin studies and advances in genomics have yielded a strong bed of evidence regarding the influence of genes on athletic ability (see Brutsaert and Parra, 2016; Georgiades, Klissouras, Baulch, Wang, and Pitsiladis, 2017). However, these studies also reveal that genes do not account for 100% of the variance in this arena.

Beyond physiological traits, research on the relationship between innate cognitive capacities and performance has typically concentrated on the role of intelligence, most notably in chess. These studies have revealed a relationship between intelligence and chess performance that cannot be easily dismissed (see Grabner, 2014; Burgoyne, Sala, Gobet, Macnamara, Campitelli, and Hambrick, 2016 for a full review). Two main approaches have been adopted in this regard. The first is to

correlate psychometric measures of intelligence with ELO (a standardised measure of chess skill) and other measures of chess performance in players of varying skill. A second approach is to compare expert chess players with novices on measures of performance and intelligence. The combined weight of these studies has shown that intelligence is generally related to chess skill, and that expert chess players almost consistently have an IQ higher than that of the population mean. According to a recent meta-analysis (Burgoyne et al., 2016), correlations between cognitive ability and chess skill are small to medium in effect size ($\bar{r} = 0.24$). Moreover, studies appear to be more consistent in correlating IQ with chess skill in younger players or those in the early stages of chess skill.

Based on the evidence summarised above it would appear that the role of individual differences in (heritable) cognitive or physiological capacities cannot be discarded from models of expertise and its development in several domains including. In sum studies indicate that innate capacities do not comprehensively account for individuals' eventual skill level, regardless of the achievement domain in question, nor do they permit statements regarding mechanisms of expertise development. What remains debatable is the emphasis that should be placed on nature versus nurture. In this regard, Ackerman (2014) asserts that extreme positions are "silly" and that both need to be taken into account. Current domain-general models of excellence typically adopt a middle ground position, factoring in the interrelationship between innate and environmental variables (Den Hartigh et al., 2016; Gagné, 2004; Simonton, 2014, e.g.).

0.1.2 The role of training

In a classic study of practice and performance in Morse code operators, Bryan and Harter (1897; 1899) measured the performance of novice, average, and experienced telegraphers in the sending and receiving of messages for over a year. The authors plotted performance trajectories over this time in weekly tests, where telegraphers were required to decode messages as they were being received, and asked telegraphers task-specific questions pertaining to their attention and thinking during performance. Periods of improvement as a result of repeated message encoding were followed by plateaus in skill acquisition. These plateaus were eventually overcome following intensive efforts to improve and reorganize skill, resulting in qualitative differences in processing. That is, telegraphers consciously attended to more complex units of information (i.e., letters, syllables, then words) as the perception of lower order units became automatic. This study has influenced the perception of some psychologists that mere repetition of behaviours is insufficient to attain maximal levels of performance. For instance, Thorndike (1921) made the general observation that adults perform at a suboptimal level even for tasks that have been repeated numerous times, citing adults' (and clerks') tendency to write slower and more illegibly than they are maximally capable.

0.1 Psychological Studies of Expertise and Skill Acquisition

Ericsson and colleagues (1993; 1996) have arguably conducted the most well known studies on this subject, asserting that a particular kind of sustained training, that they term *deliberate practice*, plays a critical role in the acquisition of elite levels of skill. Reviewing a range of literature on human performance and expertise, the authors made several foundational arguments. Firstly, advancements in human performance in various domains (e.g., Olympic Games events, musical performance, typing speed) over the course of the past century can be partially attributed to improvements in duration, intensity, and structure of training. Secondly, citing work involving the quantity of deliberate efforts required to attain expertise (Chase and Simon, 1973; de Groot, 1978), the authors suggested that up to 10 years of deliberate practice is required to reach maximal levels of performance in any given domain. Deliberate practice, defined as "a highly structured activity, the explicit goal of which is to improve performance" (Ericsson et al., 1993, p. 369) was thus proposed as a framework for characterising those activities most effective for improving performance. In its original conception the framework was constrained in three ways: 1) according to the "resource constraint", 10 years of deliberate practice with access to relevant adequate teachers, training materials, and facilities is required, 2) according to the "motivation constraint", deliberate practice is not inherently enjoyable, and 3) according to the "effort constraint", deliberate practice can only be sustained for a limited time each day.

The authors tested these predictions by collecting quantitative data on practice behaviours from three groups of violinists studying at an elite music academy. Each group differed in level of performance, based on ratings of music professors. Data were collected by asking participants to record times and durations of all practice activities over the course of a week. In an extended biographical interview, each participant also provided retrospective estimates of practice engaged in over each year of their life. Results confirmed the authors' predictions regarding quantity of practice required. Specifically, analyses revealed that the two best groups of violinists practiced almost three times longer than violinists from the third group. According to retrospective estimates, participants in the best group had accumulated approximately 2000 more practice hours than participants in the second best group. The authors also validated the effort constraint, that is, participants rated many practice activities such as practice alone, taking lessons, and solo and group performance as significantly more effortful than the grand mean of all activities. However, results failed to validate the motivation constraint, in that ratings of practice activities were not significantly lower than the grand mean of enjoyment.

Though the theoretical framework of deliberate practice was initially tested in the domain of music, Ericsson & Lehmann (1996) suggested that it can provide a sufficient account of expertise in virtually any performance domain. More specifically, they argued that expert performance is better explained as a result of deliberate practice sustained for a minimum of 10 years (or 10,000 hours), than

it is by talent. As of this year, the original study on deliberate practice (Ericsson et al., 1993) has been cited over 14000 times, and the authors' egalitarian view of excellence has captivated lay audiences (e.g. Gladwell, 2008) and lead to many tests of the framework in different domains, including sports (e.g., Helsen, Starkes, and Hodges, 1998; Baker, Côté, and Abernethy, 2003; Hodges, Starkes, Nananidou, Kerr, and Weir, 2004) and chess (e.g., Charness, Tuffiash, Krampe, Reingold, and Vasyukova, 2005; Gobet and Campitelli, 2007; Campitelli and Gobet, 2011). Adopting diary methods similar to the original work, these studies have generally uncovered quantitative differences in practice between elite, sub-elite, and novice practitioners, that are in line with the original tenets of the deliberate practice framework. Some observations have also echoed the original supposition that deliberate practice activities are "invented to overcome weaknesses" (Ericsson et al., 1993, p. 368). For example, in a study of figure skating (Deakin & Copley, 2003) elite skaters were reported to have spent 20% more time practicing jumps and 30% less time resting on the rink than sub-elite skaters. It was also found that elite skaters fell down more often than sub-elite skaters while practicing specialised jumps. This finding is attributed to the tendency of elite performers to work on techniques at the periphery of their current ability, rather than refining skills that they are already capable of performing well.

The deliberate practice framework has come under scrutiny primarily because the original claim (i.e., that accumulated practice is sufficient to explain expertise) exceeded the bounds of what was observed in the foundational studies. Several reviews have since been conducted across domains in which this construct has been studied (e.g., Baker and Young, 2014; Campitelli and Gobet, 2011; Hambrick, Oswald, Altmann, Mainz, Gobet, and Campitelli, 2014b; Hambrick, Altmann, Oswald, Mainz, and Gobet, 2014a) have indicated that the importance of deliberate practice was largely overstated. A meta-analysis conducted by (Macnamara et al., 2014) has revealed that deliberate practice accounts for 26% of the variance in games (i.e., chess), 21% in music, and 18% in sports respectively. Claims regarding necessary amount of practice have also been called into question. In a study of deliberate practice in chess, Gobet and Campitelli (2007) found that experts exhibited substantial variance in accumulated number of practice hours, ranging from 3000 to 23000 hours. A later study in the same domain found that a minimum of only 3000 hours of practice appears to be necessary (Campitelli & Gobet, 2011) for the attainment of expertise in chess, far from the 10000 hours originally hypothesized by Ericsson and colleagues.

More recently, Macnamara and Maitra (2019) sought to replicate the original study of deliberate practice conducted in violinists Ericsson et al. (1993). The authors improved upon the original design, collecting data from a larger sample than in the original study, employing a double-blind protocol, and running non-parametric tests to account for the small sample size (whereas Ericsson and colleagues ran parametric tests). Interestingly, the original finding was not replicated - although the top two groups of violinists ("best" and "good") differed

0.1 Psychological Studies of Expertise and Skill Acquisition

significantly in accumulated practice than the lowest skill bracket, the best violinists had accumulated *less* practice than the good violinists. Further, practice alone explained only 26% of the variance in performance level, which is in line with the meta-analytic average amount of performance variance (23%) reported by Macnamara and colleagues (2014).

Taken together, previous studies show that effortful practice sustained over many years is an important determinant of human excellence. However, just like the extreme talent perspective, current literature shows that the deliberate practice framework is not sufficient in accounting for the totality of variables that influence the acquisition of elite skill. Moreover findings pertaining to deliberate practice and expertise are often inconsistent. These difficulties may arise due to various reasons. As noted by Macnamara & Maitra (2019), one potential issue is Ericsson and colleagues' inconsistent definition of deliberate practice. In addition to resulting in inconsistent recordings of practice, this is likely to exacerbate the influence of existing differences in the domain-specific skills and respective training protocols that learners engage with. Inconsistency may also relate to high variability in reported practice hours arising from inaccuracy in participants' retrospective estimates of accumulated practice.

Although the deliberate practice framework alone may be ill-suited to provide a comprehensive account of skill acquisition and expertise, some combination of individual differences and behaviour must clearly be examined. In this thesis, we concentrate on the latter for several reasons. Firstly, as noted by Ward (2002) a focus on individual differences and innate abilities precludes an understanding of how the acquisition of expertise develops over time. As these are fundamentally trait-like, performance and behaviour may be better suited to tracing moment-to-moment changes in performance over the long-term. More importantly, in terms of research impact, we argue that it is in the interests of any population interested in accelerating skill acquisition to gain a better understanding of factors that are maximally controllable, such as practice behaviours and other environmental factors. However, advance this research focus in ecologically valid domains such as sports or chess, it may be prudent to move beyond imprecise, wholesale methods of measuring practice, such as retrospective reports of total practice or diary logs of concurrent practice.

Multiple research groups have pointed to digital games task environments to study skill acquisition (e.g., Boot, 2015; Chabris, 2017; Charness, 2017; Allen et al., 2023). This approach is unique in its capability to bypass the problems we have highlighted here, because as computerised tasks, digital games can leave behind accurate records of behaviour, such as what game mode a player played, when they played it, and how well they performed. Certain populations of players, such as those that play so-called "esports" (i.e., competitive games played for spectators) play these games regularly for years on end, developing expertise without the influence of systematic training environments present in more devel-

oped domains. Digital games can thus act as model environments for studying self-guided learning, and have even been compared to chess in terms of their potential to advance our understanding of expertise (Pluss et al., 2019). It has also been suggested digital games are "manageably complex", and let us track multiple metrics associated with complex performance in parallel (Gray, 2017). Combined with state-of-the-art analytic methods, data sets generated by these tasks can assist researchers identifying plateaus in skill trajectories, and potentially pinpointing the behavioural strategies that allow learners to push past these plateaus (Gray & Lindstedt, 2017). As such, we review in the following chapter recent studies of skill acquisition in expertise that have used digital games as experimental paradigms for research.

1 Literature Review

1.1 Previous reviews

In keeping with growing academic interest in digital games and esports, several reviews examining the intersection of games and human psychology have already been published. Although these reviews provide a valuable synthesis of the broader research domain, they each capture areas of research that, while adjacent, are not in immediate alignment with the aims of this thesis. We summarise these reviews here with the aim of contrasting their contents and methods with those of the current review.

With a view toward magnifying research attention on gaming activity, Bányai, Griffiths, Király and Demetrovics (2019) reviewed empirical studies on the psychology of professional esports players. The authors identified 8 empirical studies published between the years 2000 and 2017 concerning the characteristics and development pathways of esports players, as well as the motivations of esports spectators. Following a detailed content summary of each study, the authors concluded that the process of becoming a professional esports player bears similarities to the process of becoming a professional athlete, such as the requisite practice and dedication exemplified in both pathways. The authors additionally discussed parallels between excessive gaming and problem gambling, issuing a call for future research to explore these similarities, as well as to further examine the sport-status of esports and provide further empirical data on esports player psychology.

Mora-cantallops and Sicilia (2018) reviewed any studies involving MOBA games that were published since the year 2011, identifying 23 in total. Due to the general nature of their literature search, these studies comprised a broader range of topics, including player behaviour and motivation, player churn, as well as team dynamics and gender studies. The review consists of a high-level summary of each article, overview of the state of each research topic, and closing suggestions for researchers investigating MOBA games. It is concluded that MOBA games, despite their growing popularity, suffer from underexploration and inconsistent terminology as an area of research. Nevertheless, the authors encourage researchers to make use of opportunities presented by MOBA games in the form of large playerbases and accessible APIs.

1 Literature Review

Pedraza-Ramirez, Musculus, Raab, and Laborde (2020) reviewed studies investigating psychological aspects of esports performance, aiming to summarise available empirical findings in this domain as well as to integrate them into the field of sport psychology. As a result of a rigorously reported search strategy, the authors identified 52 relevant quantitative studies, published between the years 1994 and 2018. These studies were all quantitative in design. Studies were classified as concerning either cognitive (e.g., working memory, inhibitory control) or in-game outcomes (e.g., position in ranking system, game performance) associated with playing esports games, although many concerned both. Studies were further categorised in terms of their focus on either expert-novice differences or (sustained) participation in esports games as the key variable affecting these outcomes. In addition to providing a content summary as in previous reviews, the authors evaluated methodological differences in studies (where such differences may have contributed to differences in results), and postulated a heuristic model of esports performance psychology.

More recently, Allen et al. (2023) discussed the use of games as tools for advancing psychological and cognitive science research, reviewing a range of games and describing how different design features present different advantages and disadvantages for research. The authors propose that games can be used to test and scale theories, highlighting their ability to reveal interactions between cognitive processes, including complex phenomena such as tool use, relational structures, and social behaviours. They also note that games have been instrumental in improving artificial systems' capabilities and have potential for understanding natural intelligence. Finally, it is suggested that researchers can use existing games or create their own, while carefully considering factors like game rewards, complexity, and progression.

Literature reviews adjacent to the current document focus overwhelmingly on esports games as the domain of interest. Although there is overlap between their contents and the objectives of this thesis, it is worth noting that parallel studies relating psychology to digital games, including games that do not meet accepted definitions of "esport", may also offer contributions relevant to the literature. For instance, studies of Space Fortress, a digital game paradigm developed by DARPA for the study of complex skill acquisition, have demonstrated differences in learning outcomes arising from differences in sustained attention across successive gameplay sessions (Donchin, 1995; Mane & Donchin, 1989; Lim & Yen, 2004). Such studies were excluded from these reviews due to their use of search queries that isolated papers involving esports at the expense of non-esport task environments. Relatedly, despite aiming to capture the totality of literature, several reviews covered a relatively small amount of studies, with Bányai and colleagues (2019) identifying only 8 and Mora-Cantalops and colleagues (2018) 23 respectively.

An additional limitation of existing reviews relates to the method of review it-

self. Besides some discussion surrounding the sport status of esports, overview of existing research has previously been provided by collating central findings of reviewed papers. Authors have also recommended some directions for future research and highlighted potential challenges. In particular, Pedraza-Ramirez and colleagues (2020) made several valuable observations, writing on the role of deliberate practice (Ericsson et al., 1993) in esports and the dearth of related empirical literature, as well as on correct interpretation of in-game performance metrics and proper methodological design for research involving games (Dale and Green, 2017). Notwithstanding their contributions, what is lacking from these reviews is consistent consideration of the methods employed in each study and discussion concerning the interpretation and quality of resultant findings. Without this critical step, it is difficult to establish which lines of inquiry are promising, which are problematic, and ultimately what might be most beneficial for the future of this research area.

1.2 Aims and rationale

Taken together, existing literature reviews do not provide a comprehensive summary of studies that have used digital games as a paradigm to shed light on expertise and skill acquisition. The aim of this literature review is to address this gap by producing a synthesis and analysis of current knowledge in this area. Specifically, the review is concentrated on empirical research that:

- i) has produced observations of human performance, skill level, and/or cognitive factors related to digital games
- ii) has related any combination of these variables to longitudinal performance, cross-sectional performance at a different time point, and/or skill level (e.g., in-game rating)

within the same digital games from which data were collected. As such, the present focus excludes related but tangential research domains such as gamification and the transfer of skill obtained via game playing to other contexts outside the respective game environment (e.g., Green & Bavelier, 2015) In contrast to previous reviews, the present scope is not limited to studies that adopt a purely quantitative approach. Rather, studies that adopt qualitative methods are also considered for eligibility, given the potential for such approaches to yield insights that quantitative designs may overlook (Salmon, 2003).

1.3 Search strategy

To capture the state of the art, the search focused on the past decade of published work, ranging from the year 2009 to 2020. Only studies published in peer-reviewed journals and conferences within this time frame were considered. As in Pedraza-Ramirez and colleagues (2020), the Population, Intervention, Comparator, and Outcomes study design model was adopted for formulating the inclusion criteria (Schardt et al., 2007). These criteria delineate the characteristics of a study considered to be a requirement for addressing the aims of this review, and thus be eligible for inclusion (Table 1.1).

| Population | Intervention/ Phenomena | Comparators | Outcomes |
|---|---|--|---|
| Digital games, healthy human population research, excluding AI agents | Psychological and/or behavioural aspects of performance, in particular practice patterns and training interventions | <ol style="list-style-type: none"> 1. Level of practice/training intervention 2. Individual differences in cognitive factors 3. Expertise or in-game rating | <ol style="list-style-type: none"> a. Performance trajectory over time b. Performance at a given time point c. Expertise or in-game rating |

Table 1.1: PICO criteria of the review

The search was executed across multiple electronic databases to capture literature from psychology as well as related interdisciplinary fields: PsycINFO, Web of Science, and SCOPUS. Bibliographies of studies included in the final stage of screening were additionally hand-searched to identify studies that may have evaded our initial electronic query.

The following query was used to search for peer-reviewed studies published in journals and conferences between the years 2009 and 2020: (learning OR skill OR practice OR training OR performance OR expertise) AND ("digital game*" OR "video game*" OR "complex game*" OR "online game*" OR "action game*" OR "space fortress"). The search was executed across multiple databases to capture literature from psychology as well as related interdisciplinary fields: PsycINFO, Web of Science, and SCOPUS. Further, any results that contained the terms "exergame*", "game-based", or "gamification" were eliminated. This was done

to filter out studies investigating the effects of game playing on factors outside the game environment, such as general health and cognition. Additional studies that fit these criteria, but fall outside the original time range of the query, have also been included in this literature review to account for recent developments since its conception.

1.4 Laboratory studies of skill acquisition

1.4.1 Studies using Space Fortress

A large body of work has used digital games to investigate behavioural or cognitive factors that influence skill acquisition. Many of these studies have focused on the effects of different training strategies on the rate at which individuals learn to play Space Fortress (SF; Donchin 1995). SF is a game in which players maneuver a spaceship in a frictionless 2D playfield with the aim of shooting missiles at a central space fortress to amass points and ultimately destroy it. This must be accomplished whilst evading an intermittently appearing mine and remaining within a specific boundary of the playfield. In the original version of SF, players controlled the spaceship (i.e., rotation, acceleration, missiles) using a joystick, although a more recent pygame implementation allows players to play the game using a computer keyboard (Destefano & Gray, 2008, 2016). Originally designed to study the acquisition of psychomotor and cognitive skills in complex multi-tasking environments, SF is particularly conducive to such inquiry as achieving the overarching goal of destroying the fortress requires players to successfully progress towards several subgoals, previously codified in a cognitive task analysis of SF (Wang et al., 2010). Progress towards each of these subgoals, namely dealing with mines, protecting the ship, maneuvering, and managing resources (e.g., missiles) is logged by the game in the form of related subscores that are made visible to the player together with overall score (see Figure 1.1 for an overview of the game display and presented subscores). This logging system allows researchers to measure players' acquisition of component skills and overall skills and investigate how overall skill acquisition might be optimised.

Wang et al. (2010) investigated the effects of two different training protocols on skill acquisition. Participants randomly assigned to a Fixed Priority (FP) or Variable Priority (VP) condition were instructed to either give equal emphasis to each of the four subscores (points, control, speed, or velocity), or to emphasise a different subscore during each block of practice over a training protocol spanning 10 days. Participants engaged in 7 practice blocks each day, beginning and ending with one test block for which they were instructed to emphasise total scores. Comparisons of test block score trajectories between the two training groups revealed a superior rate of skill acquisition for participants in the VP

1 Literature Review

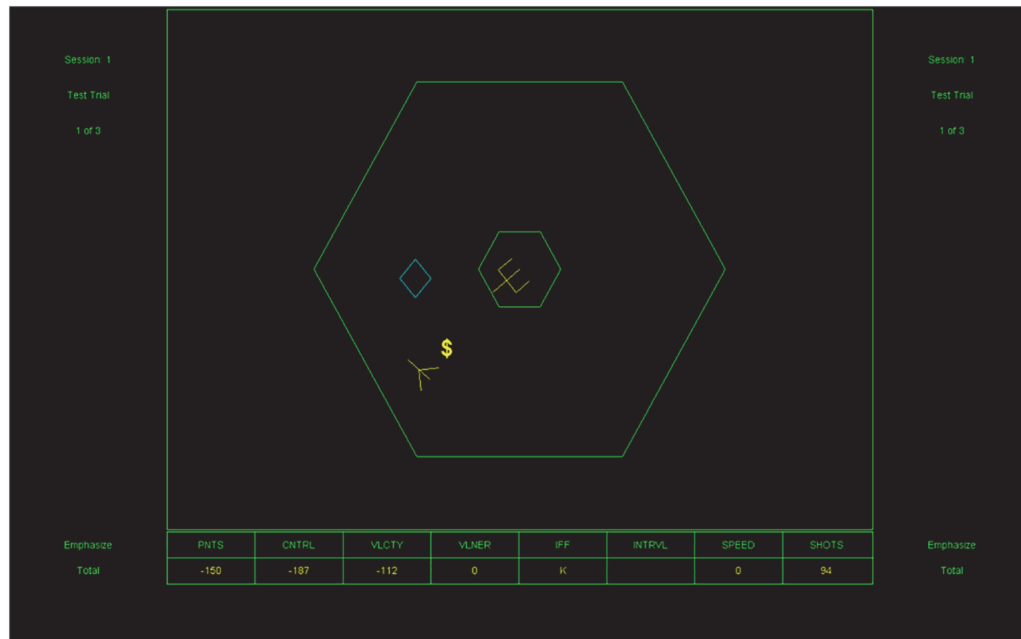


Figure 1.1: Screenshot of Space Fortress game (from Lee et al., 2015).

group. Moreover, differences between VP and FP training were significantly more pronounced for participants that initiated at a lower skill level.

These results extend early work on training strategies in SF, and more broadly extend a skill acquisition literature concerned with the effects and applications of attentional control across different training protocols (e.g., Gopher et al., 1989; Swanson & Law, 1993; Kurtz & Lee, 2003; Wickens et al., 2013; Frerejean et al., 2019), including part- (i.e., decomposing a complex skill and practicing subtasks), whole-task (i.e., practicing all components of a complex skill simultaneously), and part-whole training (i.e., practicing subtasks first followed by the whole task). In a later study, Lee et al. (2015) investigated why low initial performance may have a different effect on opposing training strategies. Comparing two groups of participants in training protocols similar to Wang et al. (2010), they demonstrated that the effects of training are moderated by fluid intelligence, with skill acquisition in a Full Emphasis Training (FET) condition correlating with fluid intelligence, while skill acquisition in a Hybrid Variable-Priority training (HVT; a combination of part-task and VP training) did not correlate with skill acquisition. Blumen et al. (2010) conducted a study of skill acquisition in SF with similar training protocols, but this time in ageing participants that were trained over the course of 36 training sessions spread out over 12 weeks. In addition, participants in the VP condition were instructed to emphasise one subscore each in training session, rather than in each training block as previous studies had done. Possibly related to this altered scheduling of practice, results of this experiment indicated the opposite effect: training with emphasis on total score resulted in

1.4 Laboratory studies of skill acquisition

superior performance as compared to shifting focus on a different subscore each session.

Other studies have used a similar experimental design to investigate how the effects of training strategies on skill acquisition relate to underlying cognitive or neural factors. Erickson et al. (2010) were able to replicate the effects of VP and FP training on naive learners over an analogous 10 session training schedule, which resulted in superior performance for participants in the VP condition. Additionally, pre- and post-training MRI scans revealed areas of the brain that were predictive of skill acquisition, with nucleus accumbens volume predicting performance improvement early on in the skill acquisition trajectory and dorsal striatal volume predicting performance for the VP, but not the FP protocol. Analysing the same data, Vo et al. (2011) found that activity in the dorsal striatum recorded before the onset of training was highly predictive of skill acquisition, highlighting the differences in neuroanatomy that may predict to what extent individuals will benefit from training in a psychomotor skill. Several studies have similarly combined neuroimaging techniques with training interventions in SF to study the neural mechanisms of acquiring complex skills. Prakash et al. (2012) also demonstrated that HVT training results in faster skill acquisition than FP training. Relative to a control group that received less contact with SF and no training instructions, participants in both training groups showed less post-training neural activity in brain regions implicated in attentional control, suggesting increased automaticity of the acquired skills (Poldrack et al., 2005). Parallel work has identified other brain regions associated with visuo-spatial attention and motor control that differ between HVT and FP participants following training Lee et al. (2012), as well as differences in functional connectivity between participants who receive VP versus FP training Voss et al. (2012).

These studies of skill acquisition allow several inferences to be made, the most obvious of them being that practicing a complex psychomotor skill improves performance over time. More importantly, the evidence suggests that the manner in which complex skills are practiced has a significant impact on outcome. In most cases, training protocols that shifted the learner's attention between different components of a multifaceted skill produced superior results compared to a practice regime involving exclusive focus on overall performance. Furthermore, the relationship between training and skill acquisition appears to be moderated by individual differences in cognition, an effect that has also been demonstrated in longitudinal performance data from the domain of chess (Vaci et al., 2019). This suggests that tailoring training protocols to an individual's abilities can maximise the benefits of practice. While this idea is certainly not new (e.g., Vygotsky, 1978), these studies demonstrate how combining complex behavioural tasks with psychometric assessments (and neuroimaging techniques) can produce the high-density data sets required to study how training protocols may possibly be optimised, and to understand the cognitive and neural mechanisms surrounding these processes.

1.4.2 Model-based studies of Space Fortress

A common feature of many studies we have reviewed thus far is the assumption that complex tasks can be broken down into simpler components that, when identified and studied separately, can help us understand how complex tasks are performed and represented in the brain. Referring to this position as the Decomposition Thesis, Anderson et al. 2002; 2011 have proposed two challenges that researchers should address to explain inconsistencies in earlier work: Either the Decomposition Thesis is erroneous on account of the possibility that complex tasks are greater than the sum of their parts, or inconsistent findings are related to an informal decomposition of the task. In line with this proposition, the authors conducted several studies of the Decomposition Thesis using SF (Anderson et al., 2011, 2016, 2019), but with a stricter decomposition of the task.

A key difference in their work is the adoption of a model-based approach. The authors used the Adaptive Control of Thought-Rational (ACT-R; Anderson et al., 2004; Anderson, 2007) model as a computational framework for explaining and predicting the cognitive processes involved in learning SF. ACT-R aims to provide an integrated model of human cognition by describing cognitive processes and their interactions through a set of symbolic representations and production rules that dictate when and how these processes unfold. In the context of SF, ACT-R assumes that the player's cognitive processes can be represented by different modules (Anderson et al., 2011), such as a visual module (e.g., encodes on-screen symbols representing the mine), an imaginal module (e.g., determines how to aim at the mine), and a goal module (e.g., dictates switching of focus from fortress to mine). These modules respond to inputs (e.g., from SF) and coordinate with each other based on guidance from a procedural module (i.e., the aforementioned production rules). Thus, researchers use ACT-R to test predictions about cognition in different environments (e.g. Borst & Anderson, 2013; Van Rij et al., 2010; Liang et al., 2016), making revisions to the model and theory in response to simulations and human experiments.

Anderson et al. (2011) used SF with simultaneous neuroimaging to investigate the brain regions involved in part- and whole-task practice, testing the predictions of ACT-R with behavioural and neural data. While previous studies manipulated the focus of participants' practice by giving them different training instructions, Anderson et al. (2011) manipulated focus on different subtasks directly by modifying the programming of SF for different training conditions. Participants played in one of four different conditions in a 2x2 design where the presence of both the Fortress and Mines was controlled for (present versus not present), resulting in a set of training conditions that necessarily constrained attention to specific subtasks. Although the ACT-R model was fitted only to behavioural data from the Fortress-only and Mine-only (i.e., part-task) conditions, the authors found evidence to support predictions of both behavioural and neural activity across the whole-task conditions. Model predictions for neural activity across predetermined

1.4 Laboratory studies of skill acquisition

brain regions showed a high correspondence with observed activity, suggesting that extant intuitions about the neural underpinnings of complex performance, at least in the context of SF, may be accurate.

In a subsequent study, Anderson et al. (2016) conducted a study of SF performance with similar experimental conditions, but this time aiming to predict skill acquisition based on the sequential activation of participants' mental states during gameplay. These states were identified using a classifier trained on observed behavioural states (e.g., navigating the ship, shooting at mines) and neural data, and ACT-R predictions. The authors found classifying brain activity into discrete states and modifying the sequential activation of these states, was predictive of individual differences in performance and skill acquisition.

1.4.3 Studies using commercial games

SF is one exemplar of how games can be utilised as experimental tasks to generate such data, but researchers have used many other game paradigms to study variables associated with skill acquisition. A promising direction is the use of existing commercially successful games to study skill acquisition. By situating research on behaviour and cognition in a commercial game environment, researchers approach the study of phenomena that have immediate bearing on stakeholders in real world environments, such as casual and competitive players.

Basak et al. (2011) extended work on the neural correlates of skill acquisition previously studied in SF by acquiring high-resolution MRI scans of individuals before and after completing 20 hours of training in *Rise of Nations*, a complex real-time strategy (RTS) game. In contrast to SF, *Rise of Nations* loads less on fine motor control and more on rapid situation assessment and decision-making, as players are required to govern a growing civilisation composed of multiple entities that accept individual command inputs from the player. Regional differences in volume, particularly in the medial prefrontal cortex and anterior cingulate cortex (areas important for tasks associated with motor control and cognitive control respectively), were found to be significantly predictive of improvement in the game. These findings further our understanding of the neuroanatomical correlates of skill acquisition. Other studies have used modified versions of existing first-person shooter (FPS) games, a popular genre of action game, to examine how task difficulty can be adapted to the learner, and how such processes are impacted by differences in cognitive ability and personality (Hughes et al., 2013; Bauer et al., 2012; Richels et al., 2020).

Digital games, in particular commercially successful ones, also present the opportunity of recruiting large samples of willing online participants. Johanson & Mandryk (2016) improved skill acquisition and skill retention in an online sam-

ple of FPS players by providing augmented feedback that taught novice players how to identify opponent location via audio cues. In another experiment, Johanson et al. (2019) tested the generalisability of the distributed practice effect: the phenomenon whereby gaps taken between individual practice episodes can produce superior acquisition and retention skill as opposed to schedules in which all practice episodes are clustered close together in time. The researchers created a copy of Super Hexagon, a casual action game, to collect and analyse data from another online sample of players, demonstrating that performance and skill acquisition can be improved by taking short breaks between sittings and extending the distributed practice effect to an ecologically valid setting.

1.5 Observational studies of skill acquisition and expert performance

1.5.1 Longitudinal studies using telemetry data

Although more detailed review of research on the distributed effect is reserved for Chapter 2, we devote some space to these studies in the present section due to their relevance to the topic of games as research methods. Notably, several research groups have recognised that online games can generate data sets of unprecedented size via remote logging of user behaviour. Stafford & Dewar (2014) were the first to analyse skill acquisition trajectories of a huge player sample obtained through telemetry, collecting data from 854,064 players in a casual online game called *Axon*, designed and deployed in collaboration with professional game developers. These researchers extended previously known effects to this novel setting, confirming that practice amount is a strong predictor of ultimate performance, and that distributing practice over extended time windows leads to better performance than massed practice.

Other analyses of this data set have explored factors influencing the benefits of distributed practice, as well as reasons why players discontinue sessions of play. By comparing subsamples of players who took breaks during time periods likely to contain sleep versus those that did not, Stafford & Haasnoot (2017) were able to test whether sleep consolidation contributed to skill acquisition (but failed to find an effect). Agarwal et al. (2017) reproduced the practice spacing effects found by Stafford & Dewar (2014) using an alternative analysis, and additionally found that the probability of aborting a series of games is higher if the last game ended with much lower performance than previous games. Individuals that persisted in the face of low scores were ultimately able to achieve higher scores than those who quit, which authors related to the psychological construct of grit.

1.5 Observational studies of skill acquisition and expert performance

Work using *Axon* shows what is achievable with telemetry data even when measures are relatively simple (e.g., machine location, game timestamps, game scores), and similar work has also been conducted using commercial FPS games such as *Destiny* (Stafford et al., 2017) and *Halo Reach* (Huang et al., 2017). Researchers can study changes in skill acquisition with large sample sizes and over lengths of time that would be prohibitive to study in a laboratory setting. Moreover, researchers can not only test for effects that are difficult to arrange in controlled experiments but also compare the relative sizes of effects with one another. For instance, (Stafford & Dewar, 2014) showed that the benefit of distributing play sessions in *Axon* over 24 hours was almost equivalent to the experience benefits of playing an additional 5 games in the early stages of skill acquisition. Thus, although these observational studies are unable to test for causal relationships, they can test and scale existing theories in ecologically valid environments, and assist in the identification of research directions that are maximally impactful.

Several studies have analysed observational data from online games that are multiplayer, highlighting another affordance of such data: the ability to study the interaction between social behaviours and skill acquisition. Landfried et al. (2019) studied skill acquisition in *Conquer Club*, a turn-based game of diplomacy inspired by the board game *RISK*, where players compete either individually or in teams in the pursuit of geopolitical objectives. Comparing skill trajectories of individuals with different patterns of sustained social behaviour, the authors found that players with team-oriented play strategies achieved superior skill levels in the long-run as compared to players that mostly played as individuals. Moreover, the authors observed a "loyalty" effect whereby sustained teamwork with the same players over successive games resulted in a significant short-run boost to skill acquisition. Sapienza et al. (2018) examined performance of individuals playing in temporary teams in *League of Legends*, a game from the Multiplayer Online Battle Arena genre (MOBA games; we take a closer look at this particular MOBA in Chapter 2 as we analyse our own data set). Similar to findings from (Agarwal et al., 2017), the authors observed a tendency in individuals to stop playing after performance started declining in successive matches, although the consequences of this behaviour is unclear in their analysis.

1.5.2 Cross-sectional studies using telemetry data

In the previous section we reviewed studies that have used observational methods to trace skill acquisition over the long term. In contrast, many studies have focused on cataloguing differences in performance between players of different skill levels at a single moment in time. Drachen et al. (2014) compared players across four skill tiers in *Dota 2*, a MOBA game in which each player controls the movements and abilities of a single "hero" entity as they fight for control of a 2.5D

1 Literature Review

game arena in team-based five versus five matches. Analysing spatio-temporal data describing moment-to-moment movements of player-controlled "heroes" in *Dota 2*, another MOBA game, the authors found that teams of higher skill level initiated more zone changes across the game arena and exhibited less intra-team distance than lower teams from lower skill levels. Moll et al. (2020) analysed telemetry data from *Fortnite*, a popular online game in which 100 players land their avatars on a remote island and battle until a single player remains. Data visualisation suggested that experienced players made choices regarding initial start position that differed from the choices of beginners. This was weakly supported by inferential analysis that showed a modest correlation between start position and end-of-match placement. In another exploratory analysis of 6615 *League of Legends* matches, Sangster et al. (2016) observed trends suggestive of a link between intra-team familiarity and team performance, although correlations failed to reach statistical significance.

Thompson et al. (2013, 2014, 2017) conducted a series of studies on expertise using a data set of 3360 *Starcraft 2* players, a popular RTS game formerly with a large competitive following. Asserting that many studies of expert-novice differences assume predictors of performance to be uniformly important across the skill curve, Thompson et al. (2013) performed classification analyses to rank the predictive importance of performance variables across 7 different skill levels. Results showed that the predictive importance of many performance variables (e.g., action latencies, use of hotkeys to select units, actions per minute) changed across skill levels, suggesting that assuming the static importance of performance variables, particularly in the context of expert-novice differences, may lead to an erroneous understanding of expertise and the factors that relate to its acquisition. More generally, the authors interrogate a rich set of variables including action latencies and perception-action cycles (Fuster, 2004) in an environment that requires persistent switching between multiple complex tasks. This further demonstrates how game replays from the RTS genre can serve as a testbed for cognitive science in environments analogous to traditional laboratory paradigms.

Further analysis of this data set has produced results relating to cognitive ageing and motor chunking. Regression analyses of "looking-doing latency" (i.e., the time to first action immediately following a switch in position on the in-game map) showed that reaction times in this environment increase with age, starting most prominently at the age of 24 (Thompson et al., 2014). Results also indicated that this decline in performance was not ameliorated by expertise, providing evidence that experience cannot compensate for age-related deficits in this kind of cognitive task. In a third study, the authors extended the study of motor chunking to this domain, that is, the notion that individuals "chunk" sequences of familiar actions in a motor skill and reproduce these individual units of motor memory in relevant situations. Results suggested that motor chunking is a valid model of action preparation as, in line with previous laboratory experiments of

motor skill, the onset of the first action in action sequences was delayed relative to other actions in the same sequence.

1.6 Cross-sectional studies of expertise and performance

1.6.1 Skill-based differences in cognition and behaviour

In this section, we provide an overview of recent studies that have generated primary data sets describing snapshots of individuals at different levels of skill. Reminiscent of early expert-novice differences studies (Chi et al., 1981; Murphy & Wright, 1984), these studies provide cross-sectional comparisons of individuals on salient cognitive and/or behavioural factors relating to complex skill in digital games.

Tanaka et al. (2013) found that expert *Guilty Gear* players, a 2D fighting game, performed significantly better on a visual working memory task than novices with negligible experience in digital games. Additionally, analyses of structural MRI images revealed that experts had larger gray matter volume in the right inferior parietal cortex than novices. In a similar study, Zhang et al. (2015) used diffusion tensor imaging to compare white matter integrity of visual and motor pathways between individuals with considerable digital gaming experience and those without. Several clusters in visual pathways were found to have higher white matter integrity in experienced game players versus non-game playing control participants. Furthermore, experienced players performed better in a visual attention task, and faster reaction times in this task correlated with higher connectivity in the left corticospinal tract.

Kokkinakis et al. (2017) related expertise in digital games to fluid intelligence and age in two laboratory experiments. Collectively, the authors found a significant correlation between scores on the Wechsler Abbreviated Scale of Intelligence (Second Edition, Matrix Reasoning Subtest) and in-game rank in *League of Legends*, a weak correlation between rank and tests of visuospatial working memory, but no relationship between rank and theory of mind as measured by the "Reading the Mind in the Eyes Test" (MITE). Similar to Thompson et al. (2014), it was also found that performance related to age, with performance in two separate samples of FPS players exhibiting an early peak following decline in the late 20s, while two samples of MOBA players exhibited peak performance in the mid-20s.

Many other studies have also focused on MOBA game players as the population of choice for studies relating individual differences to game expertise. Large et al.

1 Literature Review

(2019) conducted a correlational study by deploying a battery of cognitive tasks to a large sample of *League of Legends* players, albeit in an online setting (Large et al., 2019). In-game rank was found to significantly (but weakly) relate to cognitive control, speed of processing, and a statistical learning task (measured using an exploitation-exploration choice paradigm), but not to a deductive reasoning task. Furlough & Gillan (2018) investigated whether and how individuals' mental models of *League of Legends* changed with experience by comparing relatedness ratings of in-game concepts measured using a questionnaire. Results were suggestive of broad structural differences in mental models between low, medium, and high expertise groups, such as high expertise groups exhibiting more conceptual abstractions than less experienced groups. These results are tempered, however, by the fact that expertise was measured through retrospective self-reports of experience.

Rohlcke et al. (2018) failed to detect a relationship between working memory capacity or fluid intelligence (also measured using a matrix reasoning task) and rank in an online sample of *Dota 2* players, in conflict with previous results. However, rank was positively correlated with number of games played as well as psychological "grit", measured using the Short Grit Scale. Nagorsky & Wiemeyer (2020) found similar evidence in support of higher skill levels across competitive digital games relating to differences in practice quantity or structure, suggesting that competitive gamers may naturally titrate their training with a view towards accelerated skill acquisition. Other behavioural work has investigated whether expert and novice *Dota 2* differ in their gaze behaviour. Castaneda et al. (2016) found weak evidence for differences in fixation and scan patterns between experts and novices during gameplay, but found posited that experts' verbal reports of gaze behaviours were qualitatively more abstract than those of novices.

More recently, top ranking and averagely ranked *League of Legends* players were found to significantly differ on measures of executive functioning, including task switching and impulse control (Li et al., 2020). Bonny et al. (2020) measured both cognitive ability and personality in *Dota 2* players, finding that fluid intelligence, theory of mind, and neuroticism, in addition to experience were significantly predictive of in-game rank. An adjacent study of personality and rank in *League of Legends* produced conflicting results, finding that higher ranked players scored lower in agreeableness and extraversion, but higher in openness to experience (Matuszewski et al., 2020). Neuroticism was not found to differ between in-game ranks as in Bonny et al. (2020), suggesting that the relationship between personality and game expertise may not be as clear as it is for cognitive ability.

1.6.2 Detailed behavioural analyses of individual players

Seminal work on expertise by Ericsson & Smith (1991) argued that uncovering the mechanisms in domains of expertise should involve detailed analyses of expert performance, observed and traced in laboratory environments. Asserting that information loss arising from the averaging of performance data may impede the identification of the mechanisms underpinning skill acquisition, several groups have suggested that this "individual-analysis technique" can be extended to complex laboratory tasks such as SF as well (Towne et al., 2016; Boot et al., 2017; Harwell et al., 2018).

Towne et al. (2016) analysed archival performance data of the most skilled SF players from an earlier experiment (Boot et al., 2010), and contrasted their behavioural patterns to those of other players. Qualitative and quantitative interrogation of the time series of SF subscores across games, compared between players, illustrated diversity in behavioural strategies within the sample. For instance, while the top scoring player maintained a careful orbit around the fortress (in line with experimental instructions) with relatively low movement variance, another high performing participant was found to have exploited an alternative but viable flight strategy, in violation of experimental instructions. Among other examples of individual nuances in behavioural patterns that could only be made visible with concentrated analyses of individual players, the authors caution against research approaches in this context that rely exclusively on group-level comparisons.

Extending this work, Boot et al. (2017) conducted an experiment in which a single participant was trained to play SF in laboratory conditions over a period of 20 hours (360 games). This participant was selected through a sampling process in which individuals were screened for SF aptitude through a test battery including a measure of fluid intelligence as well as an SF aiming task. To permit some insight into ongoing cognition during the training process, the authors collected concurrent verbal reports from the participant for every three games out of each block of 10. The authors then analysed these data using the behavioural results surrounding the top performing participant from Towne et al. (2016) as a guide (referred to as "Participant 17") Their participant performed comparably to Participant 17 on all but one subscore, and was found to perform over 50% better than the sample average from Boot et al. (2010), thus supporting their selection procedure. Analyses of verbal protocols allowed the researchers to identify the sources of errors in performance, as well as providing rich descriptions of strategies surrounding subgoals throughout performance. Although this data set showed that performance was clearly disrupted during games with concurrent verbal reports, it is revealing of possible advantages of collecting verbal data as a supplement for rich behavioural time series.

Harwell et al. (2018) also adopted a fine-grained behavioural analysis approach

to examine sex differences in SF performance, reanalysing archival data from Lee et al. (2012) and situating their work in the context of sex differences in visuospatial ability. By individually analysing flight control patterns in low and high performing women, the authors indicate that low performing women rely on sub-optimal behavioural patterns that could be corrected by training. Controlling for this subsample of women almost entirely eliminated sex difference in video game performance in this sample. Supported by these results, the authors highlight that using summary scores of task performance to compare men and women on visuospatial ability may mask important differences underlying behavioural patterns that are correctable through training.

1.7 Future research

The current literature shows that use of digital games to study skill acquisition and performance has received growing attention in recent years. This use is diverse both in terms of methodological approach and of the research questions that have been studied. We summarise the implications of this literature for future work here, concluding with a summary of the rationale for this thesis.

Broadly speaking, research on psychomotor skills in the context of digital games can be characterised as taking two approaches. The first approach tracks measures of performance through time (e.g., days, weeks, months), and relates differences in these performance trajectories to sustained differences in behaviour, or individual differences such as cognition, gender, or personality (e.g., Lee et al., 2015; Wang et al., 2010; Anderson et al., 2011). The second approach takes snapshots of individuals' performance at singular points in the skill curve, typically groups of individuals at the beginning (i.e., novices) and the end (i.e., experts) of the curve. These groups are then compared on variables considered in the former approach, such as behavioural strategies or cognitive ability (e.g., Thompson et al., 2014; Tanaka et al., 2013; Kokkinakis et al., 2017; Bonny et al., 2020). Both of these approaches appear to be enhanced when behavioural and cognitive variables are measured simultaneously, due to the evident interrelationship between cognitive ability and behaviour in the context of skill acquisition (e.g., Prakash et al., 2012; Erickson et al., 2010; Vaci et al., 2019). In particular, combining neural data with detailed records of in-game behaviour has allowed researchers to identify neural signatures associated with particular aspects of psychomotor skill (e.g., Erickson et al., 2010; Vo et al., 2011; Basak et al., 2011), and to test theories regarding the mechanisms of skill acquisition (e.g., Anderson et al., 2011, 2016, 2019). It remains to be seen whether these findings are restricted to particular games such as SF, or whether some findings may generalise to other complex games that also tap multi-tasking. Likely this can only be ascertained through a more formal classification of the skills involved in specific tasks.

With the advent of online gaming and resultant large observational data sets, it is now possible to test the validity of laboratory-born theories of skill acquisition in real world settings where "participants" naturally generate data sets describing the trajectories of their skill development. Researchers adopting this approach benefit from ecological validity, large sample sizes, and potentially even opportunities to collaborate with game developers and other stakeholders involved with competitive games, who may have an interest in the science underlying how players improve their skills. Given the demonstrated link between fluid intelligence and in-game scores, some studies have suggested that telemetry data from games may also be used for purposes of "cognitive epidemiology", that is, the tracking of population-level cognitive health on the basis of fluctuations in online game scores (e.g., Kokkinakis et al., 2017).

Most importantly, we propose that each of these approaches would benefit by borrowing from the strengths of the other of the other. For instance, many laboratory studies of expert-novice differences measure expertise using self-reports of previous experience. Inaccuracy relating to this method of measurement can be mitigated using telemetry data tracking actual hours of experience or algorithmic measures of skill rank. On the other hand, longitudinal studies may benefit from combined laboratory/observational approach whereby online participants are brought to the lab for a one-off measurement of demographic and cognitive variables before their performance is tracked through telemetry methods. Perhaps the most fruitful direction would be to combine commercial games with laboratory methods. Adapting successful commercial games for use in the laboratory can allow researchers to study ecologically valid tasks in controlled conditions, and record detailed behavioural data that may be non-trivial to acquire through developer APIs. Without the ability to track metrics describing progress in individual subskills, such as in SF, commercial games may be ill-suited as paradigms to study more difficult questions, such as how training may be optimised when multi-tasking is involved, and how plateaus in the acquisition of parts of a skill affect development of the whole skill (see Gray & Lindstedt, 2017). This approach also allows researchers to modify the task based on the demands of particular research questions and experimental requirements, making it more flexible than using commercial games alone. Finally, deploying adaptations of commercial games to online experiments could allow for the scaling up of findings to larger samples that are truly representative of populations of interest.

This thesis is an attempt to extend previous research on the acquisition and performance of complex skills that have used digital games as experimental paradigms. The first chapter is an observational study that iterated on research into the distributed practice effect. We used a large telemetry data set of *League of Legends* players to investigate the effects of distributed versus clumped practice schedules on skill acquisition, and used time series clustering techniques to address potential weaknesses related to previous operationalisations of practice spacing. In addition to confirming the relevance of distributed practice in a com-

1 Literature Review

petitive online game played by millions of people, this chapter made apparent difficulties related to the use of telemetry data taken from commercial games. The second chapter aimed to address these limitations by exploring the use of a commercial game, *Tetris*, that has previously been adapted for the laboratory for use as an experimental paradigm. This was done by analysing an archival data set of Tetris players collected by Lindstedt & Gray (2015). We investigated how rich behavioural data with high temporal resolution can be used to characterise different aspects of performance in Tetris, distinguish between players of different skill level, and build models that can detect shifts in players' internal state during gameplay. In the third chapter, we investigated the validity of our approach to identifying players' internal states by collecting simultaneous behavioural and neural data as players played Tetris in an MEG scanner. After fitting a Hidden Markov model to players' behavioural data, we compared the amplitudes of occipital alpha, an index of visuospatial attention, to examine whether players switch between neurally distinct states during gameplay.

2 Effects of Practice Distribution on Acquisition

2.1 Introduction

Among the many determinants of expertise in skilled human endeavour, the accumulation of experience is one over which the aspiring expert has significant control. Research on skill acquisition and expertise, in particular the framework of "deliberate practice" (Ericsson et al., 1993; Ericsson & Lehmann, 1996), has demonstrated that the quantity and quality of sustained engagement within a domain of skill is an important driver of ultimate performance. The relationship between practice, performance, and expertise has been subjected to much scientific inquiry (e.g., Baker et al., 2003; Hodges et al., 2004; Ward et al., 2007; Tenison & Anderson, 2017; Macnamara & Maitra, 2019), and despite much debate surrounding its importance in relation to other factors, the effect of practice is widely accepted to be substantial (Baker & Young, 2014; Hambrick et al., 2014c). Researchers seeking to understand and accelerate skill acquisition have adopted a mixture of approaches, including the measurement and comparison of expert and novice performance (e.g. Shapiro & Raymond, 1989; Wiggins et al., 2002), the tracing of expert thought during practice (e.g., Gegenfurtner & Sepänen, 2013; Eccles & Aarsal, 2017; Samson et al., 2017), and use of interview methods to elicit expert knowledge (e.g., McAndrew & Gore, 2013; Den Hartigh et al., 2014). Unfortunately these methods share several difficulties - notably the expenses of recruiting human (expert) samples, the detailed tracking of cognition and behaviour over periods of training, as well as the use of laboratory tasks that may fail to generalise to the real world.

As introduced in Chapter 1, researchers have recently proposed the use of games as a solution to some aspects of these problems (e.g., Boot, 2015; Charness, 2017; Gray, 2017). The competitive and immersive nature of many games encourages players to develop profound skill over hours, days and even years of practice. Because most actions taken during a game are recorded on a computer, many competitive online games generate huge reservoirs of ecologically valid performance data that can be requested and interrogated by the curious analyst. Due to their size and richness, naturally occurring data sets from online games afford both statistical power and the ability to extract and examine "participants"

2 Effects of Practice Distribution on Acquisition

that exhibit features of interest to the researcher - features that would usually require experimental manipulation to permit empirical investigation (Goldstone & Lupyan, 2016). In the present study we analysed the relationship between skill acquisition and the distribution of practice across time using a data set drawn from *League of Legends*, an immensely popular online game that has previously been estimated to generate over one billion hours of game play per month (Kenreck, 2012), with a current tournament viewership of over four million spectators (Esguerra, 2021). In doing so we generalised a known effect in the psychological literature to a real-world context comprising millions of stakeholders, and extended previous methodological approaches in this space by using clustering techniques to interrogate how learners space their practice sessions across time.

2.1.1 Effects of practice distribution on learning

One aspect of practice that has received considerable attention from researchers is its distribution across time. In the literature on learning and skill acquisition, the effect of distributed practice refers to the tendency of learners to exhibit superior performance following a practice schedule containing rest periods between practice sessions (i.e., distributed practice), compared to a practice schedule containing shorter or no rest periods (i.e., massed practice). The terms distributed and massed practice lack strict definitions - researchers distinguish between the two in terms of the relative amounts of rest time between sessions in different practice schedules (Magill & Anderson, 2017). While there is some consensus that distributed practice leads to better learning than massed practice (e.g., Lee & Genovese, 1988; Donovan & Radosevich, 1999; Benjamin & Tullis, 2010; Smolen et al., 2016), it is important to examine what is meant by "learning" in this context, and to consider factors that have been shown to moderate this effect.

The study of distributed practice can be traced back to early studies on the recall of verbal material by Ebbinghaus (1885/1964), and so a significant amount of related work has been conducted on the effects of spaced studying on verbal memory, which we will not consider here. However, the effect has also been demonstrated in psychomotor learning (Adams, 1987). In its simplest form, a study of distributed practice in this context involves participants practicing some motor task (e.g., mirror tracing, rotary pursuit) over a block of practice trials. The amount of rest time between a block of trials (i.e., the "intertrial" or "inter-session" interval) in a distributed practice condition is greater than in a massed practice condition, but the spacing between individual trials within each block is kept constant. The researcher then compares performance on a final "test" trial between the two groups. Because learning is said to have occurred when changes in performance are relatively stable (Fitts & Posner, 1967), more involved designs include a final trial or block of trials separated from the practice block by a non-trivial amount of time (≥ 24 hours). By comparing performance

in the "retention" block and the practice block, it can be judged how well acquired performance is retained following a period of no practice. Donovan and Radosevich (Donovan & Radosevich, 1999) use the terms acquisition performance (performance in the last trial of the practice block), and retention performance (performance in the first trial of the retention block) to denote this distinction.

Overall, distributed practice appears to have a moderate to large positive effect on motor learning. For example, in a meta-analysis of 47 psychomotor studies, Lee and Genovese (Lee & Genovese, 1988) reported a large weighted average effect size of .91 for acquisition, and a moderate average effect size of .49 for retention, although the spread on these effect sizes was large. A later meta-analytic review of 61 studies (Donovan & Radosevich, 1999) yielded a smaller mean weighted effect size of .46, with a 95% confidence interval ranging from .42 to .50. The authors computed separate averages for effects sizes describing acquisition performance (.45) and retention performance (.51). Noting the importance of the type of task trained in these studies, the authors conducted additional moderator analyses to estimate how task type may influence the magnitude of the distributed practice effect. Ratings of task complexity were collected from 95 graduate and undergraduate students across three dimensions (overall complexity, physical requirements, mental requirements) for all 28 tasks examined in these studies. A cluster analysis resulted in four clusters of task complexity, optimised for maximal within-group homogeneity with meaningful between-group heterogeneity. Correlating between task complexity and effect size suggested that the distributed practice effect is diminished with increasing overall complexity (Pearson's $r = -.25$, $p < 0.05$), while mental and physical requirements were not significantly correlated with the effect. Moreover, bucketing studies into four different levels of intertrial interval, the authors considered the relationship between intertrial interval and task complexity by examining a 4 x 4 matrix of effect sizes. While it was noted that tasks of different complexity may have a different "optimal" intertrial spacing, the observation is caveated by a small number of effect sizes per cell.

2.1.2 Distributed practice in digital games

As mentioned previously, one approach to mitigating difficulties associated with laboratory-based experimentation is through the use of digital games. In cognitive science, a growing body of researchers have advocated for the use of games as an environment for the study of skill learning (e.g., Boot, 2015; Charness, 2017; Gray, 2017) noting several advantages afforded by games that allow researchers to bypass limitations of experimentation. These include large observational data sets (affording statistical power and ecological validity), participants that are intrinsically motivated to engage with the task, and a level of task complexity resembling that of real-world tasks. We review here studies that have used digital

2 Effects of Practice Distribution on Acquisition

games to investigate the spacing effect of practice, in order to provide background on the current work.

Three observational studies of skill acquisition examined the relationship between practice and performance in *Axon*, a casual computer game where players click on periodically generated targets with a mouse to maximise growth of an axon. Performance is measured by a single game score - the final length of the axon. In a first study, Stafford & Dewar (2014) analysed digital records of $\sim 850,000$ *Axon* players to test the impact of spacing on acquisition. Players were identified heuristically as having distributed (versus massed) their practice if their first 10 plays took place in a >24 hour window (versus <24 hour window). Defined this way, distributed practice had a small but significant effect on subsequent performance (highest score on plays 11 to 15; $d = 0.11$, $p < 0.00001$). Further analysis showed that the association between spacing and acquisition remained after testing separately on subsamples of players with comparable initial ability.

Stafford and Haasnoot (Stafford & Haasnoot, 2017) extended this work by investigating whether the presence of sleep could explain the effect of distributed practice, and by examining the magnitude of the effect at different levels of spacing. Players in the aforementioned distributed practice category were categorised into a "sleep" or "wake" group based on the timing of their breaks, accounting for geographical location. Comparing average scores between these groups showed no additional benefit of sleep (in fact, players in the wake group had slightly higher scores than their counterparts). To examine how different rest intervals affected acquisition, the authors plotted average scores of players on plays 11 to 15 against amount of time elapsed between games 1 and 10 - an amount ranging from 0 to 60 minutes, discretised into 16 bins. The resulting curve suggested that the relationship between practice distribution and acquisition can be described by a non-monotonic function, where optimal spacing between games lies in the middle of this range.

Agarwal, Burghardt, and Lerman (Agarwal et al., 2017) also investigated the relationship between practice and performance by revisiting the *Axon* data set. After segmenting the players' games into sessions (defined as a sequence of games with no longer than 2 hours between consecutive games), they plotted aggregated performance trajectories for sessions of different length (ranging from 4 to 15 games per session), observing that players scored abnormally high on the last game of a session, regardless of session length. Consequently, the authors suggested that the spacing related performance boost observed by Stafford & Dewar (2014) could be attributed to this score spike at the last game of a session. The accuracy of this claim is difficult to assess, however, as the two groups of researchers had different quantifications of rest interval, and Agarwal and colleagues did not report any statistics to support this particular observation.

Two studies investigated the effect of distributed practice on acquisition in first-

person shooters (FPS), a genre of action video game characterised by fast-paced weapon-based combat in a three-dimensional environment. Importantly, these games are considerably more complex than *Axon* (and many motor tasks employed in the study of distributed practice), seeing as they are played against human or AI opponents, load on bimanual dexterity, and involve communication with other players on the same team. Huang and colleagues (2013; 2017) reported on the effects of play frequency and breaks between play on performance in *Halo Reach* using a longitudinal data set comprising performance of 3.2 million players over a 7 month period. Players were subsampled by play frequency (operationalised as number of matches played per week), and average performance of each group was plotted first against match, then against time. This produced two perspectives. Players who played a relatively small number of matches per week (4 - 8) had the fastest acquisition *per match*, while those who played more frequently (>64 matches per week) had the fastest acquisition *over time*. Despite starting lower on initial performance, these players had the highest performance by the end of the 7 month period. These findings show some agreement with the literature on deliberate practice, and illustrate a trade-off inherent to spacing - taking breaks between practice sessions results in greater learning per unit of time invested into practice, but massing of practice can result in the fastest acquisition within a given time period. Additionally, the authors reported a reduction in skill rating following a break from the game longer than a day. However, the magnitude of this reduction grew smaller with an increase in gap size, and in most cases players regained their pre-break skill level after several hours of play. In contrast to Agarwal et al. (2017), the *Halo Reach* data suggested that players terminate a session of play after a decline in performance rating (associated with a loss) as opposed to after an atypically strong performance.

Stafford et al. (2017) obtained similar results by observing the performance of players in *Destiny*, another FPS game. Performance was measured by a proprietary "Combat Rating", a Bayesian skill rating system comparable to TrueSkill and Elo (Herbrich et al., 2007), systems fundamentally based on a player's win/loss ratio. The authors reported a small but significant positive correlation between performance and distribution of practice ($r = 0.18$, 99% CI [0.14, 0.22]), operationalised as the time range over which players recorded their first 25 days of play. In contrast to results from Huang and colleagues (2013; 2017), players who spaced their practice started slightly lower on initial ability (Pearson's $r = -0.09$, 99% CI [-0.14, -0.05]). Additionally, performance over the first 50 matches were plotted for players in the top and bottom quartiles of spacing, defined as the time gap between the 1st and 25th match. Players who distributed their first 25 matches over a greater time range had higher performance in their subsequent 25 matches. However, this difference was not tested for statistical significance.

Johanson and colleagues (Johanson et al., 2019) are the first group, to our knowledge, to have procured experimental data on distributed practice in digital games. In an online experiment participants played *Super Hexagon*, a minimal action

2 Effects of Practice Distribution on Acquisition

game where players must rotate a triangle inside a hexagon with the aim of avoiding incoming obstacles. Players control the triangle using left and right arrow keys on a keyboard and performance is measured as time until failure. Participants played the game for 5 trials of duration 5 minutes, separated by a rest interval of varying length (5 conditions, ranging from 3 seconds of rest to 1 day). The last trial was a retention test, separated from the preceding trial by one day across all conditions. Analyses revealed a small but significant overall effect of distributing practice on acquisition performance ($\eta^2 = .127$, $p < .001$) and a marginally significant effect on retention performance ($\eta^2 = .108$, $p = 0.44$). Additional pairwise comparisons showed that practice with no gap resulted in significantly inferior acquisition compared to most conditions. However, the effect on acquisition did not differ significantly between groups with rest intervals, and pairwise differences in retention were not significant at all.

Expanding on this work, Piller et al. (2020) tested whether the effects of spaced practice are present in a game more complex than *Super Hexagon*, as well as to test differences in acquisition arising from types of break taken. The researchers developed a 2D side-scrolling platformer called *SpeedRunners*, in which players controlled an avatar with the ability to run, jump, and swing with a grappling hook to run laps around a circular obstacle course. Performance was measured as average lap time as well as total distance travelled. Participants played 20 minutes of *SpeedRunners* split into four 5-minute sessions. Participants in a spaced practice group had breaks of 2 minutes in between sessions, while those in the continuous practice group had 3-second breaks. Participants also returned for a 5-minute test of retention one week after the 20-minute training block. Analyses did not support a positive overall affect of spaced practice on acquisition, but did reveal a small effect of spaced practice on retention performance ($\eta^2 = 0.093$, $p = 0.042$ for average lap time; $\eta^2 = 0.087$, $p = 0.046$ for distance travelled).

2.1.3 Contributions of studies using behavioural telemetry from action games

What do these studies of skill learning in digital games reveal about distributed practice? The reported data are generally in line with previous experiments showing that the cramming of practice into relatively short time frames tends to produce depressed performance following a training period. More specifically, players who distributed their game play sessions over longer time windows exhibited higher performance in subsequent game play sessions, and in some cases during the "training" period itself. In sum, this body of work answers the question as to whether or not practice spacing affects performance, and perhaps learning, in games. Unsurprisingly, it does. Unfortunately, comparing it to previous laboratory experiments of psychomotor tasks is difficult for several reasons.

For one, the majority of these studies were observational in nature, and operationalisations of practice distribution consequently diverged from previous (experimental) approaches. Where in earlier studies practice distribution referred to the amount of time elapsed between individual practice trials or sessions, working definitions in the present studies included the time gap between first and last recorded game instance (Stafford & Dewar, 2014) or game session (Stafford et al., 2017), as well as the number of game instances recorded within a week (Huang et al., 2013, 2017). Thus, the possible conflation of practice distribution with practice frequency is a concern. In some cases, data visualisation lacked supporting inferential statistics, making the interpretation of effect significance and size impossible (Agarwal et al., 2017; Huang et al., 2013, 2017). Finally, interpreting players' performance dynamics in commercial games is less straightforward than in laboratory tasks, as performance in the former is typically described by proprietary scoring systems. Taken together, while evincing that the effects of practice distribution persist in complex psychomotor tasks such as action games, the difficulties described above prohibit additional commentary, for instance on the conditions under which the effects might be strongest.

Despite these drawbacks, the studies summarised above highlight several advantages associated with the interrogation of longitudinal, observational data sets. Traditional laboratory experiments of skill acquisition are difficult: Although an observational approach sacrifices experimental control, a sufficiently large data set permits the subsampling of "participants" that meet multiple conditions of interest (e.g., practice at various levels of spacing), and enables the study of skill acquisition over far longer periods than is ordinarily practical (e.g., months). Such data also make it possible to compare the relative impacts of different factors on the dependent variable of interest. For example, Stafford and Haasnoot Stafford & Haasnoot (2017) made an argument for the relevance of distributed practice by demonstrating that the effect of spacing was comparable to tripling the practice amount. In light of these features, the capacity to test theory-led hypotheses using large observational data sets of game performance seems promising.

2.1.4 Aims of the present work

In the current study we extended this line of enquiry to a popular commercial action game, with the aim of generalising work on distributed practice that has been conducted using artificial tasks created by researchers, to a non-artificial, ecologically valid environment with which researchers have not interfered. We analysed a large body of observational performance data to investigate the effects of distributed practice on performance, mirroring operationalisations of practice distribution adopted in recent studies, and extending previous work by using machine learning techniques to investigate how the timing of breaks influences performance gains. In conducting iterative work of this nature, we tested the generalisability

2 Effects of Practice Distribution on Acquisition

of the distributed practice effect in a non-laboratory context comprising millions of stakeholders (e.g., amateur to professional action game players) with a vested interest in fast and efficient acquisition of skill.

2.2 Materials and Methods

We used a Python 3.8 (Van Rossum & Drake, 2009) environment to preprocess and analyse data, with additional packages for data munging, analysis, and visualisation including Pandas (pandas development team, 2020; Wes McKinney, 2010), NumPy (Van Der Walt et al., 2011), and SciPy (Virtanen et al., 2020). We used the Pingouin (Vallat, 2018) and statsmodels (Seabold & Perktold, 2010) packages for all statistical analyses. All analysis code are publicly available at (https://github.com/ozvar/lol_practice_distribution), together with additional documentation detailing all required software dependencies.

2.2.1 Task environment

Our study focuses on *League of Legends*, a subgenre of action game referred to as Multiplayer Online Battle Arenas (MOBAs). *League of Legends* is one of the most popular competitive online games (esports) in the world, having previously recorded a monthly player base of 67 million players, many of which participate annually in international tournaments (Segal, 2014). Like other MOBAs, *League of Legends* is a team-based invasion game that involves a high degree of team coordination and fast-paced action as two teams of five seek to destroy the opposing team's headquarters entity, located on the opposite corner of a 2.5D game arena. Each player uses a keyboard and mouse to control a single game entity (a "champion") selected at the start of each game out of a pool of 150, each with a different set of synergistic combat abilities (e.g., boosting the attributes of friendly champions, immobilising opposing champions). Players must use their abilities to eliminate opponent champions (reanimated after a scaling delay) and computer-controlled entities, as well as to support teammates, in order to reach the win condition of destroying the opposing team's "Nexus". Over the course of the game, each player accumulates "gold" and "experience points" (XP) in proportion to their successful participation in combat with enemies and contest over intermediary map objectives. These resources can be used to strategically modify the abilities and attributes of champions as the game progresses, in order to best adapt to the current game state. The combination of decision making involved in champion selection, modification, and combat, together with the fine motor skills necessary to effectively control champions, makes *League of Legends* a complex game that is hard to master.

Previous studies have used *League of Legends* as an environment to study longitudinal skill acquisition (Aung et al., 2018), model the relationship between engagement and individual performance in team-based games (Sapienza et al., 2018), and investigate teamwork at different temporal resolutions (Kim et al., 2016, 2017; Sangster et al., 2016). Moreover, as the participation of many players in esports is driven partly by a commitment to skill mastery (Seo, 2016), we anticipate these results to be of interest to relevant stakeholders such as players and professional esports teams, in addition to researchers interested in skill acquisition.

2.2.2 Measures

Whenever players queue online for a match, Riot's servers attempt to balance the teams to ensure a fair game. This balancing is strongly weighted by each player's Match Making Rating (MMR), a relative skill score calculated using a method broadly similar to those used in *Destiny* and *Halo Reach*. That is, a player's rating updates following each match based on the relative skill level of opponents, with wins resulting in an increase and losses a decrease (Laserface, 2022). While MMR is kept hidden from players, it is used to predict a player's ranking in different public tiers and divisions. A player's ranking is visible to other players and determines the skill bracket within which they may play, as well as tournaments that they may qualify for. Thus, while MMR is reflective of skill, individual changes in MMR from match to match may not directly reflect on the performance of any individual player, as MMR is primarily governed by the ratio of wins to losses (Laserface, 2022; Rio, 2021), and the likelihood of a win is dependent on more than the contribution of any single player (e.g., performance of teammates and opponents). For this reason, we concentrated our analyses on post-match statistics that describe the performance of an individual at each match. These included the the amount of gold per minute (GPM) earned in a match, and the ratio of kills and assists scored against opposing champions to the number of deaths experienced by the player's own champion (KDA), calculated using the formula $(\text{kills} + \text{assists}) / \max(1, \text{deaths})$. While metrics like this can be impacted by the role that their chosen champion may fill (e.g., Demediuk et al., 2019) (e.g., support roles typically earn less gold than the "carry" role), we judged these to be the best available to work with, and had no expectation of systematic bias as players play a variety of roles across their trajectory. As League of Legends developer Riot Games keeps the MMR algorithm confidential, we normalised all values of MMR across the data and analyses reported here.

2 Effects of Practice Distribution on Acquisition

Table 2.1: Raw data columns available in a single row of the data set analysed in this study.

| Column | Description |
|---------------|---|
| Account ID | Unique anonymised numeric identifier of player account |
| Platform ID | Identifier of server the match was played on |
| Game ID | Unique numeric identifier of match |
| Neutral Creep | Number of neutral AI entities killed |
| Enemy Creep | Number of opponent AI entities killed |
| Win | Boolean indicator of match result |
| Timestamp | Unix timestamp indicating when the match was logged |
| Date | Date on which the match was played |
| Hour | Hour at which the match was played |
| Gold Earned | Total amount of Gold earned by the player |
| Damage Dealt | Total damage dealt by the player to opponents |
| Time Dead | Total time in seconds the player champion spent dead |
| Time Played | Total time in seconds played in the match |
| Kills | Total kills scored on opponent champions |
| Deaths | Total number of times the player champion was killed |
| Assists | Total number of times the player assisted in scoring a kill |
| Rating | Normalised MMR of the player before the match |
| Position | Role of the player champion |

2.2.3 Data and preprocessing

League of Legends developers Riot Games digitally log all match events and summary statistics, and make a subset of all global game logs available to access through a public Application Programming Interface (API). Presently, we anal-

use a large data set of game logs describing the longitudinal performance trajectories of *League of Legends* players across matches. Our data closely resemble that which is available through the API, but were provided to us by Riot Games and therefore differ in that they additionally contain a record of player MMR at each match, which is ordinarily not publicly available. The data comprise all ranked matches played by a random sample of 482,415 new *League of Legends* accounts over the course of a competitive season, dating from 21 January 2016 to November 2016. All analyses were in compliance with the terms and conditions for data usage made clear to us by Riot Games. All matches correspond to the default "Solo/Duo Queue" ranked mode of play, with five players on each team. Each row in the data lists a single match for a single given player, containing a unique player identification number, unix timestamp, and various performance and outcome variables (see Table 1 for an overview of the raw data). Importantly, these were newly created accounts that had not previously been registered with any competitive *League of Legends* play prior to the start of this season. New player accounts are initialised at the same MMR value when they first enter ranked play, and therefore nominally appear to be of equal skill at the start of their trajectories. However, as the data set lacks records of unranked matches that may have been played in order to unlock the ranked game mode, we are limited in our knowledge of differences arising from prior experience. Additionally, as all account IDs are anonymised, we cannot associate each ID with a single unique player, and acknowledge hereby another source of potential bias, although we do not expect it to be systematic.

We took several steps to ensure the quality of the data prior to analysis. These preprocessing steps were focused on ensuring data quality for an initial window of 100 matches, as visualisation indicated that this was the period in which most players appeared to reach asymptotic performance. We first dropped all players who had not played a minimum of 100 games over the course of the season, and any players with missing values in any of their first 100 match records. We dropped any players who had a non-default initial MMR value, as well as players with records in multiple servers, as these observations violate our assumption of equal starting experience. These inconsistencies can occur when a player migrates from one server to another, and would have confounded our assumption that all accounts in the sample started with similar experience. We also dropped any players with matches that lasted less than 900 seconds within the first 100 matches we sampled, as this is indicative of a match which has been abandoned by one or more players, and thus does not reflect a match experience that is on equal terms with all others in the sample. Finally, we removed any players with games in which they were likely completely inactive (i.e., matches in which they scored 0 Kills, Assists, Deaths, and Creep Kills). In addition to dropping players that did not meet analysis requirements, we performed several linear combinations of columns from the raw data to generate additional variables of interest: KDA, GPM, and the time gap between the end of one match and the start of the next.

2 Effects of Practice Distribution on Acquisition

We retained a total of 162,417 players following preprocessing and a corresponding 16,241,700 rows worth of data (at 100 matches per player).

2.3 Results

To assess general changes in performance as a function of experience, we first plotted the trajectory of GPM and KDA against matches played for all players in the sample (Fig 2.1). The trajectories of average GPM and KDA per match displayed a sharp initial climb with decelerating gains. This is in line with previous studies that have found good fit between the power or exponential function and averaged performance, demonstrating the diminishing returns of sustained experience on performance across a range of domains (e.g., Gaschler et al., 2014; Heathcote et al., 2000; Haider & Frensch, 2002). We also plotted the averaged MMR trajectory of all players in the sample which, in contrast, sharply decreased before showing a gradual rise towards later matches (S1 Fig). We attributed this initial rating drop to our sample being composed exclusively of new accounts. Specifically, we expected new players to suffer more losses against the relatively more experienced majority (unobserved in the sample) towards the start of the season, where the matchmaking algorithm has begun to calibrate for fair matches. This intuition is supported by the trajectory of loss percentage, which descends to 50% as the average rating of the sample stabilises (plotted together with MMR).

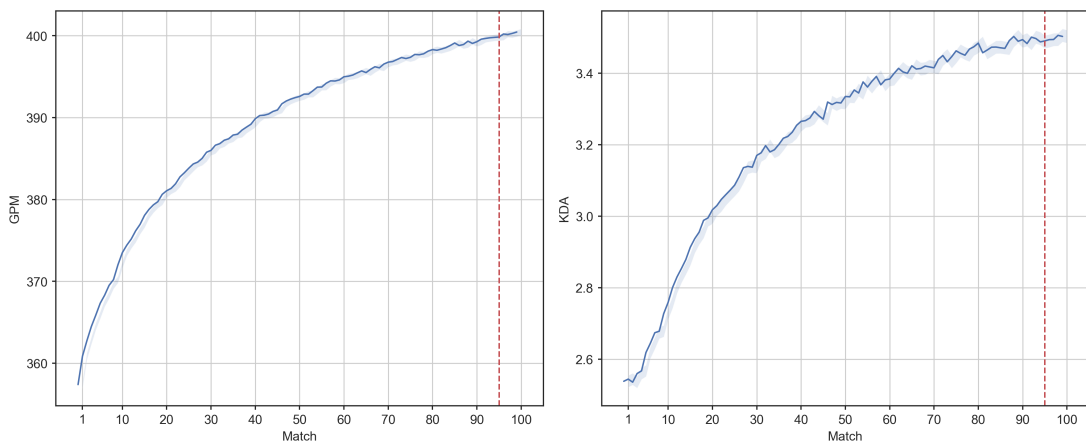


Figure 2.1: Trajectories of mean GPM (left panel), and KDA (right panel) of all players against match. Shaded regions indicate 95% confidence intervals.

We assessed the effects of spacing on acquisition performance first by subsampling and comparing groups of players with different patterns of spacing. We concentrated these analyses on the first 100 matches, as player performance appeared to asymptote towards the end of this window, and we were predominantly interested in acquisition effects. Similar to Stafford et al. (2017), spacing was operationalised as the gap in days between the 1st and 95th game. After visualising the frequency

distribution of time in days elapsed between the first and 95 match for each player (S2 Fig), we subsampled three groups of players that were sufficiently discrete in terms of their break schedules, and that were adequately sized for statistical analysis: players that took between 136-150 days, 76-90 days, or up to 15 days to play their first 95 matches. Visualising the impact of gap size on mean performance over the final five (96th to 100th) matches, we initially observed that while players who spaced their first 95 matches over a greater range had higher acquisition, players who massed their matches in a shorter range initialised at much higher initial performance (close to the maximum observed performance). Due to the negative correlation between this time range and initial GPM (Pearson's $r = -0.295$, 95% CI [-0.30, -0.29]), we suspected our spacing measure to be confounded by initial performance, potentially explained by a combination of play intensity and other factors related to ability.

In order to control for initial levels of absolute performance, we subsampled players who scored a mean GPM of between 315 and 385 (an interval centered on the median of mean initial GPM; 350 ± 25) over their first five matches, resulting in a subsample of size $n = 52,440$. Analogously, we replotted KDA trajectories after subsampling players with a mean KDA of between 1.64 and 2.24 (median initial KDA 1.94 ± 0.19), resulting in a subsample of size $n = 17125$. Figure 2.2 shows the mean GPM and KDA trajectories of players who took between 136-150 days, 76-90 days, or up to 15 days respectively to play their first 95 rated matches. Players who clustered their matches the most exhibited a faster initial climb in initial, but lower performance overall by the end of their trajectory. Although we produced an analogous plot for mean trajectories of MMR (S3 Fig), we neglected to conduct further (statistical) analyses of this metric due to the aforementioned opaqueness of the MMR algorithm and the ubiquitous downward trend in MMR across our entire sample, which we believe lent itself poorly to a study of learning.

Players with the largest time range between their 1st and 95th match achieved an average GPM in their final five matches that was 6.91 points higher (95% CI [3.74, 10.07], $n = 1236$, $M = 399.71$, $SD = 49.65$) compared to those with the smallest time range (i.e., 1-15 days; $n = 2790$, $M = 392.81$, $SD = 46.18$). This was statistically significant following a t-test at $t(4024) = 4.28$, $p < 0.001$, albeit for a small effect size (Cohen's $d = 0.146$). For the subsample matched on initial-KDA, players in the former ($n = 373$, $M = 3.76$, $SD = 2.18$) achieved a KDA 0.49 points higher (95% CI [0.28, 0.71]) points higher than those in the latter spacing group ($n = 1159$, $M = 3.27$, $SD = 1.74$) This difference was also statistically significant [$t(1530) = 4.45$, $p < 0.001$, $d = 0.265$].

By binning players using our spacing measure, we produced a snapshot of the effects of practice distribution on performance. To produce a fuller account of this relationship using the entire range of our practice distribution variable, we linearly regressed spacing on both GPM and KDA (Fig 2.3). We report regression

2 Effects of Practice Distribution on Acquisition

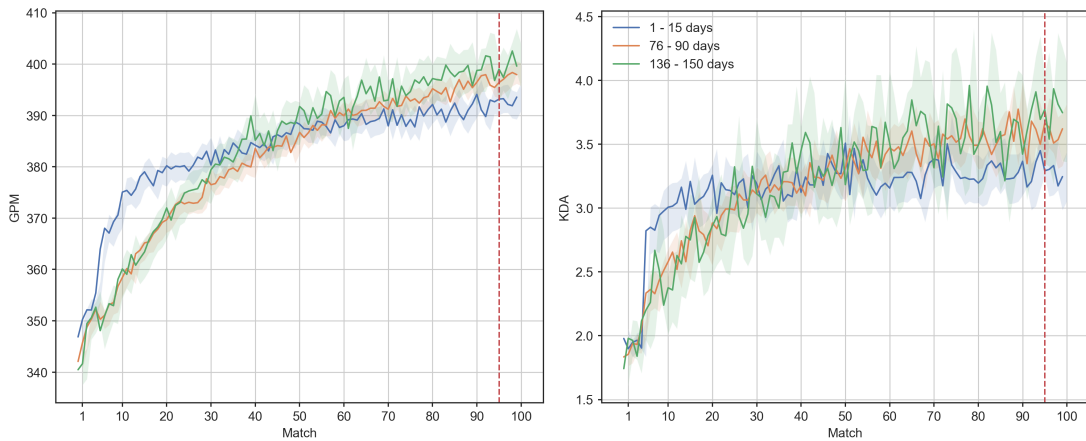


Figure 2.2: Trajectories of mean GPM (left panel) and KDA (right panel) against match for players with different patterns of match spacing. Data in the figure are a subsample of players who initiate at a similar range of GPM and KDA (approximately surrounding the original sample median). Shaded regions indicate 95% confidence intervals.

slopes and supporting statistics for both variables in Table 2. We report White's heteroscedasticity-consistent standard errors (White, 1980) due to nonconstant variance in our residuals.

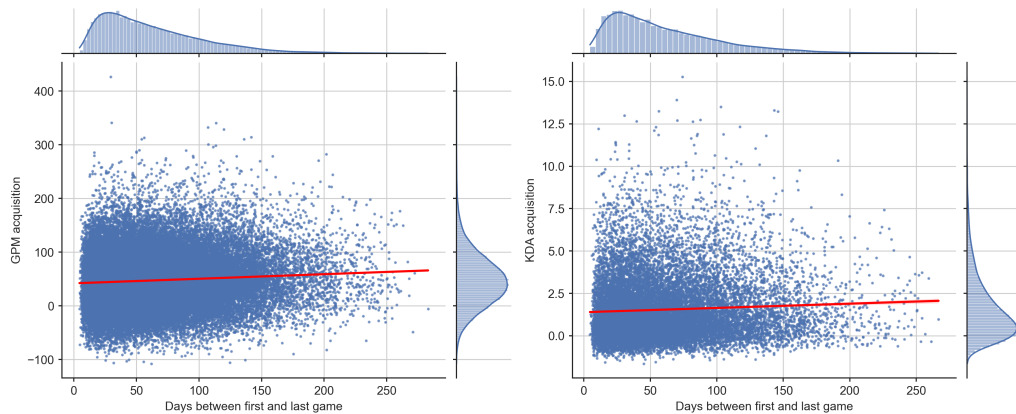


Figure 2.3: Scatter plots of GPM and KDA against time range in days between first and 95th game respectively, with line of best fit. Axis plots show distributions of respective axis variables.

2.3.1 Time Gap Clustering

One issue with operationalising practice distribution as the time range between two matches, is that different schedules of practice may coexist within identical time ranges. For instance, a player with a consistent schedule of 1-2 matches per

Table 2.2: Linear regressions of time delta in days between 1st and 95th match on average GPM and KDA between the 96th and 100th match.

| | B | Std. Err. | β | T | p | R ² | 95% CI |
|-----|--------|-----------|---------|--------|--------|----------------|----------------|
| GPM | 0.0849 | 0.005 | 0.0727 | 16.045 | <0.001 | 0.005 | [0.075, 0.095] |
| KDA | 0.0025 | <0.001 | 0.0568 | 7.396 | <0.001 | 0.003 | [0.002, 0.003] |

day could be grouped with a player who played 10 matches per day followed by a handful of matches after a 10 week break. To explore whether our spacing groups reflected the differences in practice distribution that we were interested in, as opposed to some other systematic and unanticipated differences in play schedules, we conducted an alternative analysis to the rule-based slicing performed above by clustering our original sample of 162,417 players by their time series of time gaps between matches. First, we leveraged the Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018) to perform a visual inspection of how different players distributed their matches over time. The UMAP algorithm is a non-linear dimensionality reduction technique based on manifold learning. Given a high dimensional data-set, UMAP first infers its topological structure and then using stochastic gradient descent attempts to structurally reproduce it in a lower dimensional space (two or three for visualization purposes). In our case, the original data-set was represented by an $N \times T$ matrix of between matches time gaps, with $N = 162,417$ being the number of considered players and $T = 95$ the number of matches in the observation period. We chose this range to align with the previous step of our analysis, allowing a window of five final matches with which to analyse the effects of different spacing patterns on final performance. The transformation performed by UMAP generated an $N \times D$ matrix with $D = 2$ being the number of target dimensions. In this 2D representation, players with a similar pattern of inter-matches temporal gaps were represented closer in space while players with a dissimilar spacing profile were represented as far apart. The topological structure of the original data-set was inferred by computing the euclidean distance in a local neighborhood of 1000 points, while the dimensionality reduction was achieved by running the optimization part of the algorithm for 1000 iterations. The remaining parameters were left at their default value as provided by the python library used for our analysis (i.e., UMAP-learn (McInnes et al., 2018)). The generated 2D representation can be observed in Figure 2.4.

As we can observe from Figure 2.4, a number of naturally occurring groups appear to emerge (i.e., the areas where the density of dots increases), suggesting the existence of different profiles of play distribution. In order to formally evaluate whether differences in naturally occurring spacing patterns truly exist, we decided to run a clustering analysis. This was done to test the consistency of the individuated profiles arising from clustering. We describe and report here

2 Effects of Practice Distribution on Acquisition

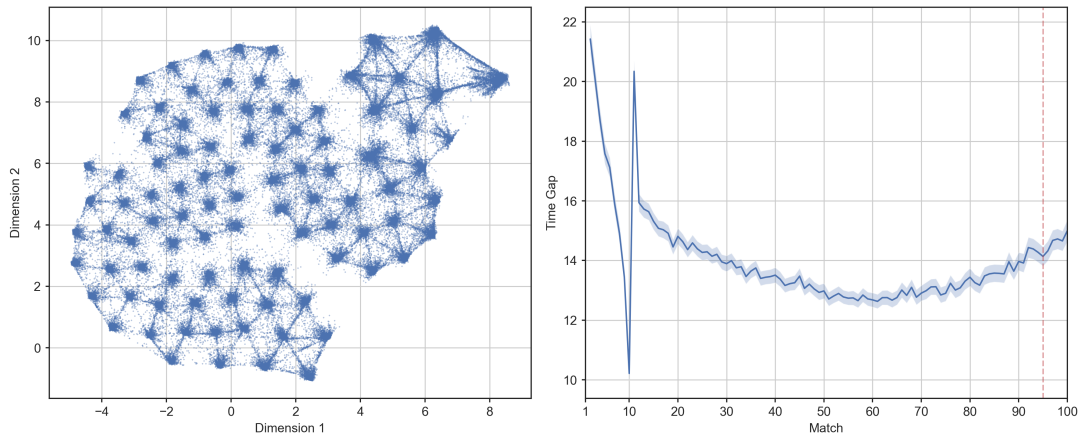


Figure 2.4: The left panel shows the two-dimensional projection of the observed 95 inter-match gaps in hours as generated by UMAP for the entire sample. The y and x axes represent the two dimensions individuated by UMAP. As opposed to Principal Component Analysis their associated values should be interpreted as coordinates on a plane rather than indicators of the magnitude of the two components. Each dot represents the history of inter-matches gaps for a single player while distance between dots indicates the degree of similarity between different patterns of spacing. **The right panel shows the average evolution of inter-match gap in hours for the entire sample.** The y axis indicates the time in hours elapsed since the previous match while the x axis indicates the order of the match. The solid line indicates the mean value while the shaded region shows the 95% confidence interval. The dotted red line separates the observation period (i.e., the first 95 matches) from the evaluation period (i.e., the last 5 matches).

the results derived from a combination of recurrent autoencoder and mini-batch K-means.

Recurrent Autoencoder and K-Means Autoencoders are a specific type of artificial neural network (ANN), which given an input x attempt to produce a copy of the same (Goodfellow et al., 2016). This is done by simultaneously learning the parameters of a function $h = f(x)$ (called encoder), mapping the original input to a latent representation h , and of a second function $\hat{x} = g(h)$ (called decoder), generating a copy \hat{x} from the same latent representation (Goodfellow et al., 2016). Learning occurs through stochastic gradient descent, minimizing a reconstruction loss that measures the mismatch between x and \hat{x} . Once the training process is terminated the latent representation h can be extracted, and should carry meaningful properties of the original input. In this sense, the operations performed by the encoder function can be seen as a form of automatic feature extraction.

In order to force the autoencoder to produce an h with interesting characteristics, a series of constraints are usually applied during the learning process. In our case we adopted a combination of denoising and undercompleteness strategies. The first corrupts the input (usually through random gaussian noise) forcing the autoencoder to learn a representation capable of undoing the noise, while the second requires the dimensionality of h to be much smaller than that of the original input (Goodfellow et al., 2016). Since we were dealing with time-series data, we parameterized the encoder and decoder functions using two recurrent neural networks (RNN), a specific type of ANN able to capture temporal dynamics (Goodfellow et al., 2016). The first RNN tasked to generate h , was composed of two Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) layers respectively with 60 and 30 hidden units. The second RNN, used to reconstruct the corrupted input was a single LSTM layer with 60 hidden units. The autoencoder minimized the Mean Absolute Error (MAE) between the reconstructed and original inputs and used the Adaptive Moment Estimation (Adam) optimizer (Kingma & Ba, 2014) for gradient descent. Training was carried out by passing random batches of 512 inputs and monitoring the reduction in MAE on a 20% held-out subset of the original data. Training was terminated once the reconstruction loss stopped decreasing in the held-out subset by a minimum of $\delta = 0.0001$ for more than 15 consecutive epochs. At this point, we proceeded to generate features from the original input passing a $N \times T \times 1$ tensor of between matches gaps through all the operations carried out by the encoder function. This generated an $N \times h$ matrix, with $h = 30$ being the dimensionality of the last layer of the encoder, which other than offering a more compact representation of the original input (making it easier to perform a cluster analysis) should have also distilled its most salient characteristics.

Finally, in order to obtain different spacing profiles we applied Mini Batch K-Means (a more scalable version of K-Means) (Sculley, 2010) to the representation generated by the encoder. We selected the number of centroids k by generating an elbow plot after running the algorithm for a range of 2 to 10 k , with 2000 random initializations for a maximum of 3000 iterations each, passing the inputs in random batches of 512 elements. Following the methodology proposed by Satopa et al. (Satopaa et al., 2011), the optimal $k = 4$ was found by individuating the point of maximum curvature in the aforementioned elbow plot (S4 Fig). In order to derive interpretable profiles from the individuated cluster, we averaged the time series of between-match time gaps (along with GPM and KDA) over the labels provided by the Mini Batch K-Means. The autoencoder was realized using tensorflow's high level API keras (Abadi et al., 2015; Chollet et al., 2015), while the Mini Batch K-Means implementation we employed was the one provided by the library scikit-learn (Pedregosa et al., 2011). Results of this clustering analysis can be seen in Figure 2.5 and Figure 2.6.

Looking at Figure 2.5 we can see how the location and extension of the clusters on the 2D reduction provided by UMAP tells us when, for how long and how

2 Effects of Practice Distribution on Acquisition

intensely the players in those clusters spaced their matches on average. Interestingly, the areas of high density in this representation seem to identify groups of players taking a single long break at specific points during our observation period. With the exception of a single period characterised by longer breaks (more hours) between matches, players appear to maintain a consistent play schedule. Following the representation in the right panel of Figure 2.5 we can see that clusters 1 and 3 represent the extremes of a continuum going from a relatively early versus late rest period. Clusters also differed on the intensity of this rest period, with spacing cluster 3 exhibiting the longest breaks during the shortest rest period, followed by the most consistent streak of play.

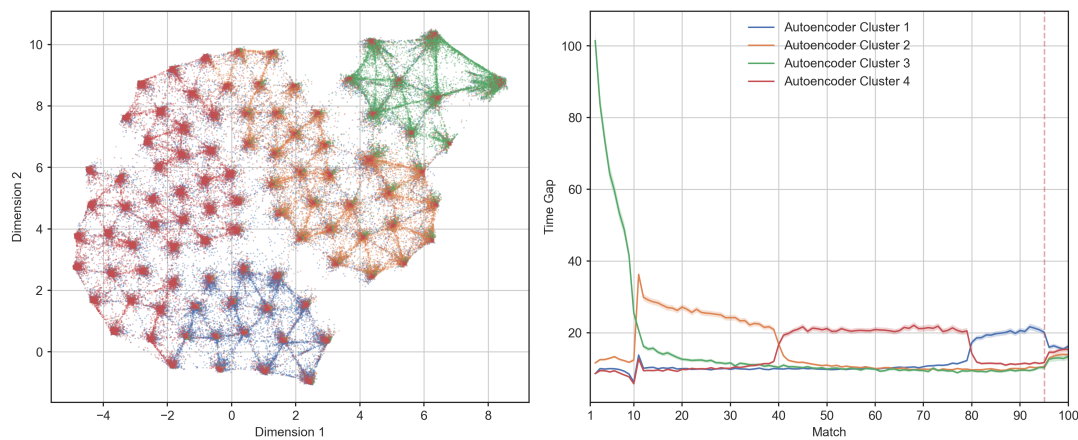


Figure 2.5: The left panel shows the two-dimensional projection of the observed 95 inter-match gaps in hours as generated by UMAP for each spacing cluster across the entire sample. The y and x axes represent the two dimensions individuated by UMAP. As opposed to Principal Component Analysis their associated values should be interpreted as coordinates on a plane rather than indicators of the magnitude of the two components. Each dot represents the history of inter-match gaps in hours for a single player while distance between dots indicates the degree of similarity between different patterns of spacing. **The right panel shows the average evolution of inter-match gap in hours for players in each spacing cluster.** The y axis indicates the time in hours elapsed since the previous match while the x axis indicates the order of the match. The solid line indicates the mean value while the shaded region shows the 95% confidence interval. The dotted red line separates the observation period (i.e., the first 95 matches) from the evaluation period (i.e., the last 5 matches).

Tabulating the joint frequencies of players across each of the clusters and original categories (Table 3) showed that players in a given spacing category do not display uniform membership to a single spacing cluster, supporting our intuition that operationalising practice distribution as a time range may mask differences in underlying play schedules.

Table 2.3: Joint frequencies of players in spacing group as defined by k-means cluster (rows) versus time in days delta between 1st and 95th match (columns).

| | 1-15 Days | 76-90 Days | 136-150 Days |
|-------------------|-----------|------------|--------------|
| K-means cluster 1 | 7701 | 3477 | 914 |
| K-means cluster 2 | 1644 | 4120 | 1102 |
| K-means cluster 3 | 583 | 2093 | 523 |
| K-means cluster 4 | 2014 | 3430 | 1006 |

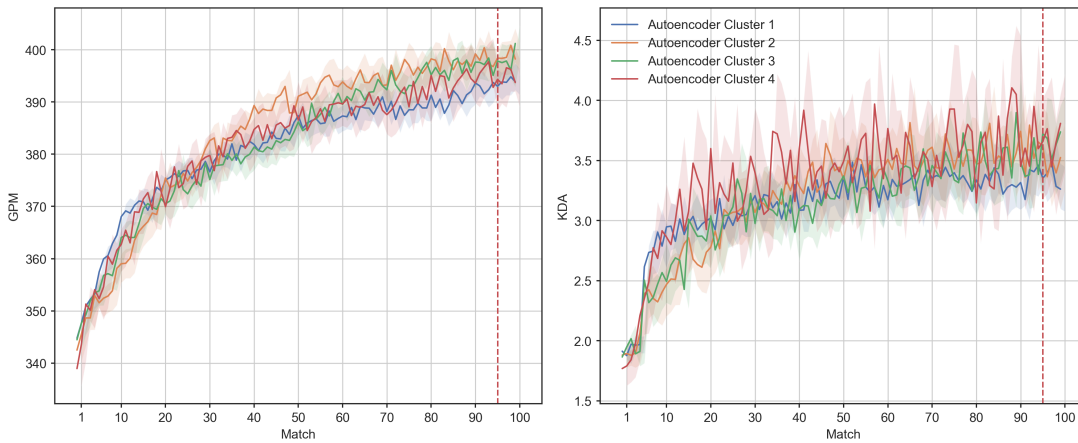


Figure 2.6: Trajectories of mean GPM (left panel) and KDA (Right panel) against match for players in our 4 autoencoder clusters. Data in the figure are a subsample of players who initiate at a similar range of GPM and KDA (approximately surrounding the original sample median). Shaded regions indicate 95% confidence intervals.

Figure 2.6 displays the typical averaged trajectories of GPM and KDA for players in each cluster. Compared to the analysis of groups sliced by time range, there were no large differences between spacing clusters in final GPM or KDA. We conducted one-way ANOVAs to test these differences in mean final performance (average GPM and KDA over the last 5 matches). This was significant for both GPM [$F(3, 162413) = 517.93, p < 0.001$] and KDA, [$F(3, 162413) = 439.87, p < 0.001$] but for negligible effect sizes ($\eta^2 < 0.01$). Additionally, we conducted pairwise comparisons (Holm-Bonferroni corrected t-tests) in GPM and KDA between each pair of clusters. We identified significant differences in GPM between clusters 1 and 2 [$t(31839) = 5.45, p < 0.001, d = 0.061$], clusters 1 and 3 [$t(24758) = 3.65, p < 0.001, d = 0.051$], as well as clusters 3 and 4 [$t(30911) =$

2 Effects of Practice Distribution on Acquisition

4.23, $p < 0.001$, $d = 0.049$], but only negligible effect sizes. Clusters 1 and 3 were also significantly difference in mean final KDA [$t(8195) = 3.05$, $p < 0.001$, $d = 0.075$], but again with a negligible effect size.

2.4 Discussion

Analysing a large data set drawn from *League of Legends* - one of the world's most popular competitive online games - we extended recent work on the distributed practice effect (Stafford & Dewar, 2014; Stafford et al., 2017; Stafford & Haasnoot, 2017; Huang et al., 2013, 2017) in an ecologically valid and complex perceptual-motor skill environment. Players in our data set showed monotonic gains in measures of absolute performance (GPM, KDA), which tapered off after approximately 100 matches. After matching players on initial ability and subsampling groups defined by the amount of time elapsed between their 1st and 95th game, we found that players who spaced practice the most showed initially depressed gains but superior final performance, albeit for a small effect size, and only for a large time range of spacing. These effect sizes were in line with those previously reported in action video games (Stafford & Dewar, 2014; Stafford et al., 2017; Johanson et al., 2019). In a second analysis, we applied clustering techniques to identify and analyse differences in the *timing* of practice spacing in our data set, and tested whether the "when" of practice distribution has an effect on performance. Our analyses indicated that, for this task environment, only the total amount of rest is what matters, and not the timing of these rest periods. Practically speaking, our results suggest that by their 100th match, a player who maximised spacing would be earning on average 228 gold more and scoring a KDA of 0.49 higher per match than a player who crammed their matches, given the typical match lasted roughly 33 minutes in our sample. Although highly significant, our effects are limited by large spread around our group means. This observation echoes concerns raised in recent research, namely, that analyses of aggregated data sacrifice the ability to accurately describe dynamics of the individual (Charness, 2017).

For the sake of completeness, we also reported players' trajectories of MMR, a relative measure of performance calculated by a proprietary algorithm that is heavily weighted by match outcome (i.e., win versus loss; (Herbrich et al., 2007; Laserface, 2022; Rio, 2021)). A full description of the MMR algorithm is kept hidden from the public, making MMR significantly more opaque than GPM or KDA as a measure of performance. Moreover, although MMR is partly dependent on match outcome, the probability of winning a match is dependent on many factors (including the behaviour of teammates and opponents), and is itself the subject of many efforts in prediction. For these reasons we neglected to conduct further statistical analyses of MMR, and instead concentrated our

efforts on GPM and KDA, which we believe to provide a clearer perspective on individual performance from match to match.

The size of our effects (Cohen's $d = 0.146$ for GPM; 0.265 for KDA) are in keeping with other studies of digital games that reported on the distributed practice effect. For example, Stafford and colleagues reported a small effect size of distributed practice on subsequent performance in *Axon* (Cohen's $d = 0.11$; (Stafford & Dewar, 2014)), a small correlation of distributed practice on the slope of performance in *Destiny* (Pearson's $r = 0.18$; (Stafford et al., 2017)), while Johanson and colleagues (Johanson et al., 2019) reported a small effect of distributed practice on acquisition ($\eta^2 = .127$, $p < .001$) as well as a marginally significant effect on retention ($\eta^2 = .108$, $p = 0.44$). Importantly, it is also consistent with early meta-analytic work that observed smaller effect sizes in studies involving motor tasks of lower overall complexity (Donovan & Radosevich, 1999). Despite efforts to mimic related work, we are cautious to make direct comparisons between the effects reported here and similar studies due to differences in elements of study design, such as the length of our training window and our operationalisation of practice distribution, as well as the exploratory nature of our design. An explanation as to why practice distribution is less beneficial for more complex tasks presumably depends on a fuller understanding of the mechanisms underlying memory consolidation and the effects of extended inactivity on subsequent recall. Ultimately this is a question for future experimental work that investigates the effects of distributed practice while directly manipulating levels of task complexity.

Our initial results appeared to be confounded by pre-existing differences in gameplay habits. Similar to Stafford and colleagues (Stafford et al., 2017), the distribution of practice was significantly related to the intercept of performance in our sample, but to a more extreme degree. Specifically, players who clustered their matches in relatively shorter time windows initiated at much higher levels of absolute performance. Plausibly, we were observing in our "groupers" a category of player characterised by intense, frequent play. Such players may be more motivated to engage with the game, and would potentially have accrued a commensurately higher amount of experience during the early initiation period of the game where only unranked matches can be played. We attempted to control for this confound by running our analysis on a subsample of players matched on initial performance, but acknowledge that lingering effects of this confound may nonetheless impact our reported statistics.

Similarly, as our sample consisted only of ranked matches, we were agnostic to any experience that players may have acquired in unranked matches that were played between the ranked matches recorded in our data set. A related concern is that players we found to have spaced their matches the most may have played more matches generally than players in our massed practice group, having had more opportunities to play unranked matches during breaks from the ranked game mode. However, we contend that our observations are inconsistent with this

2 Effects of Practice Distribution on Acquisition

hypothesis, as we would then have expected the players that spaced their matches the most to have a more accelerated learning trajectory than what is observed in Figure 2.2, reflecting the additional practice hours that they accumulated. Nevertheless, we suggest it is important for related future work to eliminate any such ambiguity by ensuring that the entire history of player experience is visible when curating the data. In this regard, it may be also be worthwhile to record players' past experience with other digital games. In their analysis of gameplay patterns in *Halo Reach*, Huang et al. (2013) reported separate rating trajectories for players that had previous experience in various related games, such as previous iterations of the Halo series, or other FPS games. This showed that differences in prior experience resulted in differences in current rating. Thus, we suggest that future work could deliver more precise results by capturing pre-existing differences in game experience, for instance through an additional survey component.

Previous work that has leveraged game telemetry data to study distributed practice in games has made use of data slicing techniques to isolate play schedules of interest (Stafford & Dewar, 2014; Stafford et al., 2017; Stafford & Haasnoot, 2017; Agarwal et al., 2017). As an extension to this approach, we used machine learning to cluster players by their time series of gap between matches. In doing so we aimed to reveal naturally occurring play schedules in our data set and investigate whether these underlying patterns have any bearing on effects arising from our data slicing procedures. Our results showed that players in the same spacing group, defined by the time delta between two matches, may diverge considerably in their underlying play schedules, as identified by our time series clustering. Specifically, players across different spacing clusters differed in the timing of an extended "rest" period, characterised by less frequent gameplay. This suggests that operationalising practice distribution as a time delta between two matches may not be as straightforward an analogue to classical operationalisations as one would have hoped. Nonetheless, players across these spacing clusters did not differ significantly in their final performance, suggesting that it is indeed the amount of time spent on breaks is positively correlated with acquisition, but not necessarily the timing of these breaks.

By identifying and attempting to control for confounds in our data, we highlight both a weakness and a corresponding strength of telemetry-based big data analysis. The use of observational data in behavioural science sacrifices total control of participant behaviour. In our case, the absence of experimental control restricted our ability to compare groups of players with homogenous time gaps between each of their play sessions, as has been done in laboratory studies of distributed practice (Magill & Anderson, 2017). Our solution, similar to other studies that have used game telemetry (e.g., Stafford & Dewar, 2014; Stafford et al., 2017; Stafford & Haasnoot, 2017) was to use a proxy for intersession time interval, namely the time gap between the first and last match. Although time between first and last match is likely related to time between individual trials,

we acknowledge that use of this alternative operationalisation limits our ability to generalise from laboratory work to a non-artificial environment.

An additional consequence of using observational data is susceptibility to the effects of both known and unknown nuisance variables that may systematically skew results in unpredictable ways. Presently we attempted to filter out potential confounds, such as players that migrated server (accumulating additional "hidden" experience), or players whose records contained matches with abnormal participation (i.e., complete inactivity). In doing so we dropped approximately two thirds of our data, but were nonetheless left with a sample size that offered ample statistical power. However, despite our attempts to isolate our variables of interest, we remain cognizant of the potential for additional confounding variables. These may include the presence of multiple players using the same *League of Legends* account, or the existence of highly experienced players who create new accounts to enjoy lower levels of ranked play ("smurfs").

2.4.1 Optimising training for MOBAs

Additional work is required before the relevance of our work for competitive players can be made clear. For a training window of 100 matches, we observed significant gains in GPM and KDA for players that distributed their practice over a range of 136-150 days (a rate lower than a match a day). Regression analyses further suggested modest improvements to these performance measures by the end of the 100 match window for each day added to the time window of play. While these gains appear large enough to make a impact in a match, the measures taken to achieve may be impractical for some. A player invested in consistent improvement of performance will likely require a high volume of daily practice comprising different training foci. Practical considerations notwithstanding, there is also the matter of preference. Clearly, many League of Legends players are unlikely to desire a temporary reduction in play frequency to maximise performance at some future time point. Indeed, in a classical study of distributed practice with postal workers, Baddeley and Longman reported lower preference for spaced than clustered practice (Baddeley & Longman, 1978).

As opposed to the mere *spacing* of practice, what is likely to be of practical relevance is the *interleaving* of practice of different skill components. Research on skill acquisition distinguishes between blocked practice, that is, a sequence of practice where trials on one task are done consecutively before moving onto the next task, from random practice, in which practice of different tasks are interleaved. Studies of motor learning have shown the latter to produce superior learning in certain contexts (Schmidt & Lee, 2019, Chapter 10). Random practice can be viewed as the distributed practice of multiple skill components in a training schedule and is thus intrinsically related to practice distribution as has been

2 Effects of Practice Distribution on Acquisition

presently examined. Although an investigation of such interleaved practice was not within the scope of the present study, future work on skill acquisition in digital games may uncover applicable insights into practice and performance by turning its attention to the interleaving of skill components within a training schedule.

2.4.2 Conclusion

Research on motor learning has demonstrated that taking breaks between practice sessions, as opposed to massing them in relatively short time windows, benefits ultimate performance (Lee & Genovese, 1988; Donovan & Radosevich, 1999). By analysing an observational, longitudinal data set describing player performance in a massive, commercially successful video game, we showed that the distributed practice effect is relevant in an ecologically valid context comprising stakeholders with a vested interest in improving their skills. Although data sets such as ours afford strong statistical power and the ability to filter through observations that meet desired experimental conditions, they are also complicated by noise and potential confounds. As a solution, we propose that researchers seeking to use telemetry data adopt a hybrid approach, collecting demographic information on players before tracking their play records through game APIs. In doing so, interested researchers may control for variables related to initial performance, such as age or cognitive ability (Kokkinakis et al., 2017), and account for sources of data pollution such as players generating data on multiple accounts.

3 Tetris as an experimental platform for cognitive science

3.1 Introduction

In the previous chapter we demonstrated the capability of using a commercial game to test theories of cognition in a naturalistic setting. Using a large, observational data set drawn from the world's most popular commercial online game, we showed that distributing play sessions over longer time spans results in greater net performance gains as opposed to clustering the same number of sessions in a shorter span of time. These effects were in line with an established literature on the distributed practice effect (Lee & Genovese, 1988; Donovan & Radosevich, 1999).

Although we successfully extended previous findings to this setting, adopting this approach also exposed several restrictions in the use of telemetry data obtained from a commercial digital game. Firstly, game developers are not as interested in generating data sets conducive to cognitive science as cognitive scientists are. Hence, the measures available through data telemetry can be superficial. In our case, we had access to gross measures describing the overall performance of individuals at the end of each game, but no measures describing granular aspects of performance, such as control inputs or communication with other players at each minute or second. Thus, although this data set had the capacity to summarise average longitudinal trends in performance, it was less capable of yielding insights into the dynamics of performance *during* gameplay.

An additional caveat of this game was related to its complexity. Although this was covered in the previous chapter, we reiterate that League of Legends is a game played by two competing teams of five players, where each player controls a single avatar selected from a roster of over a hundred possible avatars, each differing in control inputs, as well as tactical and strategic possibilities. These elements introduce a high degree of variance from session to session, making it harder to detect signals relating to underlying behavioural or cognitive variables of interest to this thesis. In this chapter, we focus on addressing the limitations to research imposed by methodological approaches that rely on commercial games and telemetry data. We investigate the viability of using a particular commercial

game, in this case Tetris, that has been adapted for the laboratory, thus possessing both the desirable qualities of a commercial game and some of the essential qualities of a laboratory task.

3.1.1 What makes a good research game?

As discussed in Chapter 1, games of many different varieties have previously been used for research purposes. These purposes have included both games as a treatment (e.g., investigating the effects of playing games on cognition), and games as a lens into the human mind. In line with the overarching aim of this thesis, we concentrate in this section on the latter. That is, we consider the question, what makes a game suitable for studying cognition in relation to complex skills?

One aspect of games that makes them interesting research tools, is that they allow cognitive scientists to test theories in domains that simulate the complexity of real world tasks. While games can load on many cognitive factors simultaneously, many laboratory tasks typically isolate singular constructs of interest. For instance, the Towers of London task has been used extensively to study problem-solving (Berg & Byrd, 2002; Bull et al., 2004; Ward & Allport, 1997), while simple and choice response time tasks have been used to study reaction times (Deary et al., 2010; Heitz, 2014). In contrast, many competitive online games such as MOBAs tap into several such constructs simultaneously. For example, success in games such as League of Legends requires both sound strategic decision-making in addition to motor skill execution. Such a combination of multiple and potentially interacting variables in a single task environment can allow researchers to investigate an expanded territory of questions, such as how plateaus in one skill component can bottleneck progress in overall performance, and how learners may optimally schedule practice of individual skill components to accelerate learning.

Naturally, complexity is not always a desirable quality in a laboratory task. As illustrated in the previous chapter, too much complexity may reduce signal-to-noise ratio. Additionally, complexity is only beneficial to researchers insofar as task elements that contribute to said complexity are appropriately measured. Thus, while League of Legends may serve as an example of a complex game, it may not serve as an example of an appropriately complex research task, given that proxy measures of motor execution or planning (e.g., moment-to-moment control inputs and communication) are not always readily accessible to researchers. We argue that the complexity of games can emulate the behavioural richness of real-world tasks, but too much complexity may be detrimental for the research. Moreover, elements that make a task complex must be logged at a level of granularity sufficient to study pertinent research questions.

Another advantageous quality of games is that they enable the study of human behaviour in an intrinsically motivating environment (Baldassarre et al., 2014). Recruiting willing participants for laboratory experiments is not a trivial endeavour, and much less so if the same participants are to be studied over multiple sessions. Moreover, classical experimental paradigms often rely on unengaging tasks that require participants to be briefed with detailed instructions. Games bypass these issues entirely as they are intuitive and engaging by design, prompting players to return to the same game time and time again. As has been demonstrated by our own and by previous research (e.g., Sangster et al., 2016; Huang et al., 2013, 2017; Stafford & Dewar, 2014; Stafford et al., 2017), this can result in longitudinal data sets with sample sizes that are unprecedented in cognitive research.

3.1.2 Previous games for research

One way of profiting from the desirable qualities of games in research while avoiding some of the pitfalls of relying on data from commercial games, is for researchers to develop their own games. We have previously discussed Space Fortress as an example of a game developed explicitly for the study of learning and training strategies. While Space Fortress satisfies the criteria of complexity and granular logging of behavioural measures, its original version was not introduced online nor widely received, limiting its potential as a generator of large data sets as in other examples.

An alternative approach is for researchers to collaborate with developers to create or modify games that are designed to be engaging while also being capable of addressing relevant research objectives. For example, working with the developer of *Little Alchemy 2*, a simple game in which players combine various inventory elements to explore a sprawling tree of object combinations, Braendle and colleagues (Brändle et al., 2021, 2022) collected a data set of 29,493 players to study individuals' exploration strategies. After concluding that players explore with the aim of maximising additional object combinations, they moved to a laboratory setting by creating two versions of the game: one copy and one version with a simpler semantic structure. In doing so, the authors were able to replicate and extend their findings, showing that players resort to simpler strategies in the absence of rich semantic information.

Another example is Sea Hero Quest: a mobile game resulting from the collaboration between an independent game company and Alzheimer Research UK, that was developed with the purpose of collecting data to study human navigation. The launch of Sea Hero Quest resulted in a data set comprising virtual-navigation behaviour from 4 million participants across 195 countries. These data have been used to study various aspects of spatial navigation, including the impact of age

and Alzheimer’s disease on navigation ability (Coughlan et al., 2019), the relationship between virtual navigation and real-world navigation (Coutrot et al., 2022), and the differences in navigation ability between different populations and cultures (Coutrot et al., 2018). These studies show the capacity of tailor-made online games to produce insights into cognition as well as benchmark specific elements of cognitive function.

3.1.3 Aims of the present work

Where opportunities are not present for researchers to work with game designers, it remains possible to convert existing commercial games to a laboratory format by programming clones that log actions and game events. Recent examples include a replica of Super Hexagon created using the Unity game engine to study practice scheduling effects (Johanson et al., 2019), and generalisation of tic-tac-toe (where participants must connect four tokens in a row instead of three) created to study problem-solving (van Opheusden et al., 2021).

The aim of this chapter is to explore the viability of one such commercial game adapted for the laboratory as a platform for cognitive research: Tetris. Tetris is a popular game that is easy to comprehend, but deceptively hard to master. It demands multiple cognitive faculties to work in synchrony, including planning, decision-making, mental rotation, and motor control, satisfying our criterion of task complexity. Moreover, recent development of a laboratory adaptation of Tetris has demonstrated the feasibility of the granular, real-time logging of game events and participant behaviour throughout gameplay. Using a shared data set from an experiment conducted using this task, we conducted exploratory analyses with the following aims:

1. To identify measures that capture cognitive variables germane to the study of psychomotor skills
2. To test whether these measures are capable of distinguishing between good and bad performances (or players)
3. To test the capability of these measures to characterise the moment-to-moment performance within sessions of gameplay

3.2 Methods

As in the previous chapter, we used a Python 3.8 (Van Rossum & Drake, 2009) environment for all preprocessing and analysis. Data munging and preprocess-

ing were performed using pandas (pandas development team, 2020) and (Van Der Walt et al., 2011) and we used matplotlib and seaborn for visualisation. Additional packages for corresponding analysis techniques are detailed below under Results. All analysis pipelines and supporting software are publicly available at https://github.com/ozvar/tetris_osf_exploration, together with details regarding requisite software dependencies.

3.2.1 Task environment

Although it will certainly be familiar to most readers, Tetris is a classic tile-matching puzzle video game in which players manipulate an infinite succession of falling geometric shapes called "tetrominoes" (or "zoids") to fit them together in a rectangular playfield. The objective of the game is to create as many complete horizontal lines as possible. Lines without gaps are cleared from the playfield, scoring points and clearing up space for more blocks to be placed. Every 10 line clears increases the game difficulty, increasing the speed at which the blocks fall and making it more challenging to maintain a solvable pile structure. Importantly, the game ends once the tetrominoes are stacked to the top of the playfield and cannot be stacked any higher.

Players control the drop destination of tetrominoes by moving them laterally (translations), by rotating them, or by accelerating their downward movement speed. Players also have information about the next incoming tetromino, allowing them to plan their pile structure as they manipulate the current tetromino. Players can score bonus points for clearing more than one line simultaneously, with maximum bonus points awarded for the "tetris" maneuver, requiring four simultaneous line clears. Executing these moves under increasing time pressure requires a combination of forward planning, rapid decision-making and efficient execution of chained control inputs.

Tetris is a popular game that is played internationally in tournaments that are viewed by thousands of people (Sweet, 2021). While Tetris has been used in previous cognitive research, these studies have typically used the game as a treatment condition, or as a task for studying individual differences in spatial ability. In recent years, the development of *Meta-T* (Lindstedt & Gray, 2015), a laboratory adaptation of Tetris that collects high fidelity behavioural data, has demonstrated that Tetris is capable of much more as an experimental paradigm. It is designed specifically for use in cognitive science experiments, and it includes features that allow researchers to control and manipulate various aspects of the game.

Firstly, *Meta-T* allows researchers to configure features of the task to meet bespoke experimental criteria, such as game difficulty, visual cues, or the size and shape of blocks. These modifications make it possible to investigate different as-

3 Tetris as an experimental platform for cognitive science

pects of cognition, such as attention, working memory, and visuospatial ability, under controlled laboratory conditions. Secondly, Meta-T includes features that allow for the collection of data on player performance, such as the number of lines completed, the number of blocks placed, and the time taken to complete the game. These data can be analysed to investigate the relationship between cognitive factors and game performance. Theoretically, the fidelity of game logs in Meta-T enable full recreation of a participant's interaction with Meta-T within a full session of gameplay (Lindstedt & Gray, 2019). These factors, together with its use in several studies of cognition and expertise, demonstrate its potential as a tool for researchers interested in conducting cognitive research using a task that is both engaging and, given its popularity, representative of psychomotor performance in real-world settings.

3.2.2 Measures

We used a secondary data set made public by Lindstedt and colleagues through the Open Science Framework (<https://osf.io/78ebg/>). We describe the data set here following the original experimenters' (Lindstedt & Gray, 2019) reports as well as our own examination of its contents. The data comprise one hour of Meta-T performance from each of 240 participants, collected under laboratory conditions. Meta-T was configured to resemble "Classic Tetris" (and verified as a faithful representation thereof by a software expert affiliated with the Classic Tetris World Championship). Each participant was seated in front of a computer and instructed to play 50 minutes of Meta-T with a provided Nintendo Entertainment System controller. Players repeatedly played successive games of Tetris until 50 minutes had elapsed, restarting games upon failure.

The data set comprises three log files, each detailing all 240 participants' task engagement at three different levels of time: one describing behaviour at the time of each button input, one describing behaviour in the time spanning the appearance to dropping of each tetromino, and one summarising behaviour at the level of the entire match. We concentrated our analyses on logs of tetromino drops at each match, as these provided the highest density of measures across all log files. As mentioned previously, Meta-T captures a wealth of information throughout gameplay. Each row in the episodic log file details over 60 variables for the current tetromino drop, including:

1. Features summarising the session (e.g., participant ID, game number, timestamp),
2. Game state features relating to the tetromino (current and next tetromino, current tetromino position),

3. Game state features describing the pile (e.g., height, circumference, number of unplayable cells)
4. Motor execution features (e.g., number of control inputs, latency before and between actions),
5. Features describing tetromino placement (e.g., number of lines cleared, landing height).

We refer the reader to Lindstedt and colleagues (Lindstedt & Gray, 2019) for an exhaustive list of all variables that are logged by the Meta-T. In the next section, we describe the variables we chose to concentrate our analysis on, which were principally those describing the pile structure and those describing motor execution.

3.3 Results

3.3.1 Principal component analysis

Our first aim was to reduce the data set into a subset of variables relating to orthogonal aspects of performance in Tetris. To this end, we performed an exploratory principal component analysis (PCA) using the sklearn package (Pedregosa et al., 2011). We first inspected the data for extreme outliers and other anomalies, removing two players from the data set who never made it past level 0 during gameplay. We then subset the data, removing all variables describing the session (e.g., participant ID, time stamp) or game-state (e.g., current tetromino, tetromino orientation), effectively retaining only those variables relating to performance for dimensionality reduction (see Table 6.1 for a list of variables retained for PCA). PCA was then performed on this trimmed data set, initially with an unconstrained number of components. To identify the optimal number of components to retain, we produced a line plot of the amount of variance explained by each successive component (Figure 3.1). Identifying the point of maximum curvature in this plot indicated four components as being the optimal number to retain, as adding more would have had a limited impact on the explained variance.

The four principal components that we chose to retain explained a combined total of 53% of variance in the data set. Table (3.1) displays the PCA loadings, describing the correlation between each variable and principal component (only correlations past 0.20 are displayed). We take a moment here to interpret each component in detail.

3 Tetris as an experimental platform for cognitive science

| | 1 | 2 | 3 | 4 |
|----------------|--------|---------|---------|---------|
| mean_ht | 0.2960 | | | |
| row_trans | 0.2890 | | | |
| weighted_cells | 0.2890 | | | |
| max_ht | 0.2870 | | | |
| landing_height | 0.2850 | | | |
| pit_rows | 0.2790 | | | |
| pit_depth | 0.2730 | | | |
| col_trans | 0.2710 | | | |
| pits | 0.2710 | | | |
| lumped_pits | 0.2670 | | | |
| min_ht | 0.2530 | -0.2080 | | |
| wells | | 0.3810 | | |
| deep_wells | | 0.3800 | | |
| max_well | | 0.3720 | | |
| cuml_wells | | 0.3690 | | |
| jaggedness | | 0.3220 | | |
| max_diffs | | 0.2770 | | |
| cd_1 | | 0.2110 | | |
| cd_9 | | -0.2090 | | |
| min_rots_diff | | | 0.4440 | |
| rots | | | 0.4420 | -0.2230 |
| min_trans_diff | | | 0.3980 | |
| drop_lat | | | 0.3730 | |
| trans | | | 0.3670 | -0.2460 |
| prop_u_drops | | | -0.2590 | |
| avg_lat | | | | 0.5070 |
| initial_lat | | | | 0.4560 |
| d_max_ht | | | | 0.3550 |
| matches | | | | -0.3690 |
| cd_2 | | | | |
| cd_3 | | | | |
| cd_4 | | | | |
| cd_5 | | | | |
| cd_6 | | | | |
| cd_7 | | | | |
| cd_8 | | | | |
| pattern_div | | | | |
| resp_lat | | | | |

Table 3.1: Table of PCA loadings. Column headers display the number of each component ordered by descending amount of variance explained. Cells show the correlations past 0.20 between variables (labelled in the row headers) and the respective components.

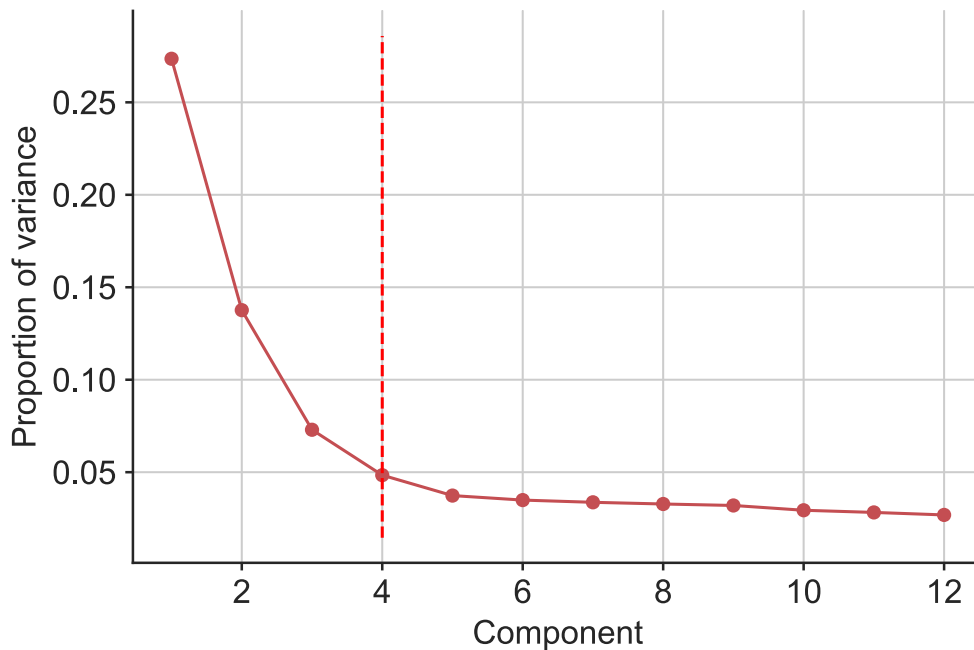


Figure 3.1: Scree plot of PCA. Points in the line show the variance explained by each successive principal component, where components are ordered from most to least variance explained. The vertical dashed line indicates the point of maximum curvature in the line, corresponding to our choice of number of components to be generated for our analysis.

The first component was predominantly positively correlated with variables describing the height of the player's pile, including the mean, max, and minimum height of the pile across all ten columns. It was also correlated with variables describing the number and depth of rows with pits (i.e., empty cells that are rendered unplayable by obstructing tetrominoes). A tall pile in general may indicate a player preparing for simultaneous or sequential line clears for combo points. However, a combination of unplayable cells and a tall pile is more indicative of a disorderly tetris pile that prohibits line clears on rows where pits are present. We refer to this component henceforth as *disarray*, viewed as an index of pile difficulty and extant pressure in the game environment.

In contrast to *disarray*, the second component was negatively correlated with the minimum height of the pile, and positively correlated with wells (i.e., low-height columns flanked on one or both sides by higher columns). Wells provide options for fitting particular shapes of tetrominoes, and deeper wells provide options for clearing multiple lines at once, and in particular enable the coveted "tetris" maneuver whereby four lines are cleared at once by inserting a line block into a well that is four cells deep. This component was positively correlated with the number and depth of wells, as well as other variables indicating large differences

3 Tetris as an experimental platform for cognitive science

between the tallest and shortest column (e.g., "jaggedness", "max_diffs"), which is suggestive of players planning simultaneous line clears for bonus points. We refer to this second component as *well preparation* to reflect this strategic decision-making and planning element in players' behaviour.

The third component appeared to be an index of the efficiency of participants' control inputs. For each tetromino, there is a minimum number of translations and rotations required for the tetromino to reach its final destination. Meta-T records unnecessary translation and rotation inputs in excess of these minima using the "min_rots_diff" and "min_trans_diff" variables. Both of these variables, as well as those tracking total number of translations and rotations, and finally "drop_lat", which tracks the total time elapsed between tetromino appearance and drop, were positively correlated with this third component. To reflect this inefficiency in motor execution, we termed this component *action inefficiency*.

The fourth and final component appeared to relate to lagged and sloppy tetromino drops, being positively correlated with the latency to first input, average latency across all inputs for the current drop, and correlated with indices of poor fit between the tetromino and the pile. We viewed this variable as tapping into executive function such as short-term planning and working memory, and refer to it as *decision-action latency* henceforth.

Taken together, our PCA captured four components that we believe captures distinct aspects of Tetris, each of which may act as proxy measures of cognitive-behavioural aspects of psychomotor performance. Our results also resemble dimensionality reduction performed by Lindstedt and colleagues, who retained three orthogonal components in their analysis using the same data set (Lindstedt & Gray, 2019). We retained one component describing (sub)optimality of the game state, one component describing strategic planning in relation to pile structure, and two components that potentially tap into aspects of executive function, such as working memory and planning.

To put the results of our PCA into the context of task performance, we visualised the averaged trajectories of each of these components across the first 50 tetromino drops of two example games from the sample (Figure 3.2). We chose to visualise 50 drops as over half the sample had played at least this many tetrominoes in every one of their games. Evident in this visualisation is that *disarray* appears to monotonically increase across both games, likely relating to increase in pile complexity as the game goes on. *Well preparation* also appears to trend upwards, as would be expected from players attempting to play strategically in the face of incoming tetrominoes. The final two components appear to be statistically stationary, fluctuating without any visible periodicity within these first 50 episodes.

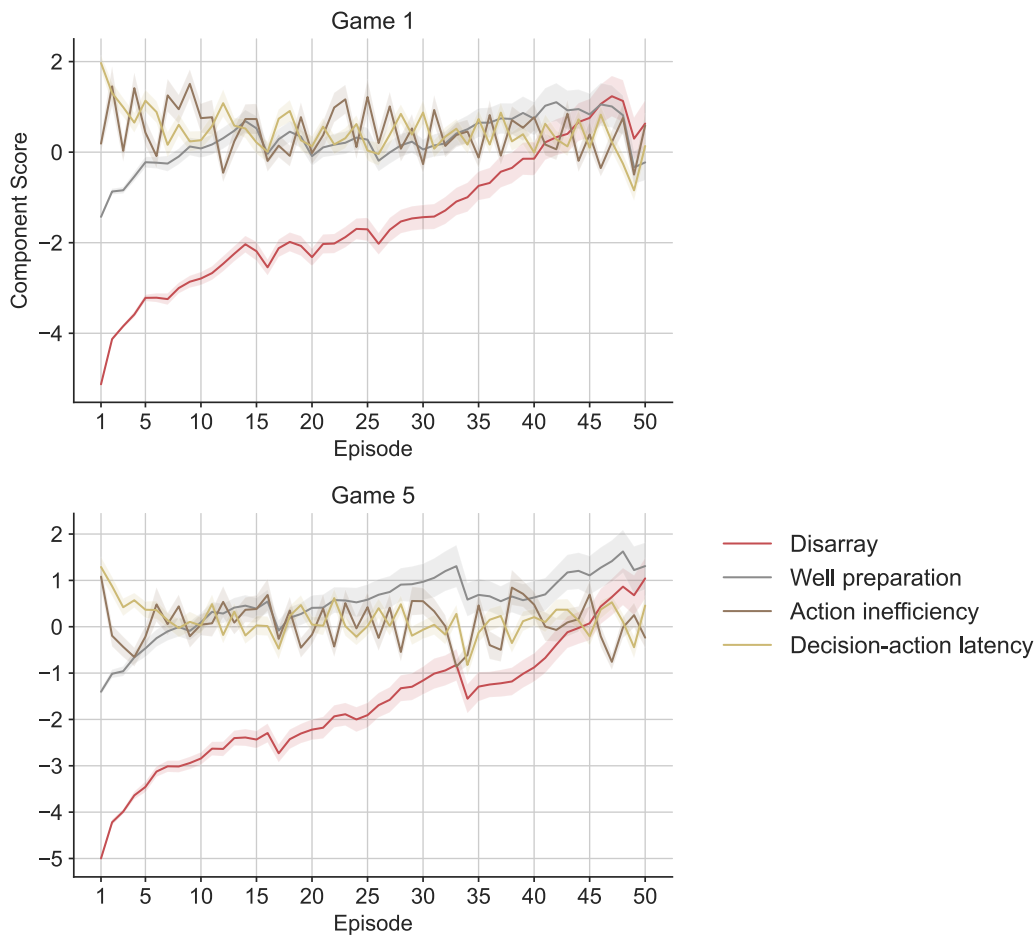


Figure 3.2: Averaged component score trajectories across two exemplar games. The top and bottom panels show mean *disarray*, *well preparation*, *action efficiency*, and *decision-action latency* from the first and fifth game across the entire sample (including only those participants who played at least five games) respectively. Shaded regions show 95% confidence intervals of the mean.

3.3.2 Distinguishing top from bottom scorers

If these measures are meaningful, we would expect them to distinguish between good and bad performances, or between heterogeneous groups that exist within our sample, such as top scoring players versus bottom scoring players. To the latter idea, we first split players into a top and bottom scoring group by sorting players based on their average score on their first few games, and then taking the top and bottom quintiles respectively for each group. We then visualised, for each group separately, the averaged trajectory of each component over the first 50 episodes of the first game played (Figure 3.3).

3 Tetris as an experimental platform for cognitive science

It is evident across every panel that top scoring players differ significantly in mean respective component score across time. As a reminder, participants are exposed to identical tetromino sequences for each successive game that is played. Thus, it is striking that while the peaks and troughs in *action inefficiency* and *decision-action latency* appear similar between top and bottom scorers, the top scorers appear to be more efficient and quick in their gameplay at almost every tetromino drop in the game. Moreover, while *disarray* in both groups appears to trend upwards, the upward trend is much more aggressive in bottom scorers than in the top scoring group. Conversely, bottom scoring appears trend downwards in their *well preparation* while top scoring players appear to trend upwards.

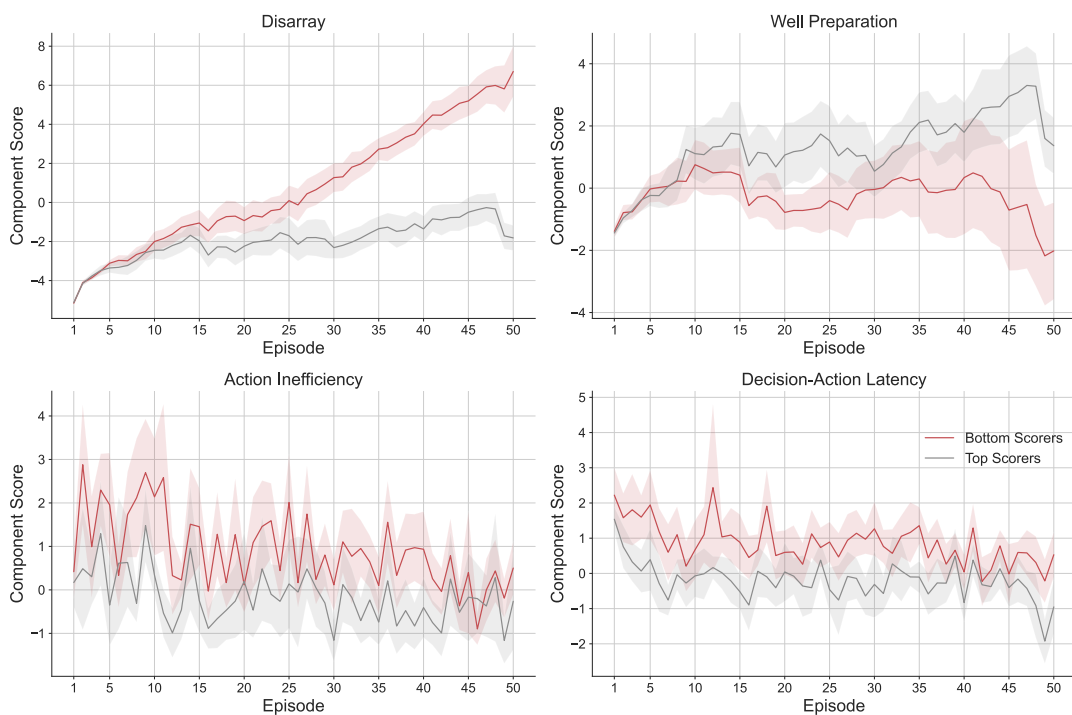


Figure 3.3: Comparison of moment-to-moment performance between top and bottom scorers. Each panel depicts the mean trajectory of the respective performance component for bottom (red line) versus top scorers (grey line) for the first 50 tetromino drops from game 1. From top left going clockwise to the bottom left panel, the subplots show trajectories for mean *disarray*, *well preparation*, *decision-action latency* and *action inefficiency* respectively. Shaded regions depict 95% confidence intervals of the mean.

We tested the statistical significance of visible trends in our plots by conducting a mixed ANOVA of each performance component, with scoring decile as the between subjects factor, and tetromino drop as the within-subjects factor. Between-subjects effects for *disarray* [$F(1, 34) = 81.93, p < 0.001, \text{partial } \eta^2 = 0.71$], *well preparation* [$F(1, 34) = 10.90, p = 0.002, \text{partial } \eta^2 = 0.2427$], *action inefficiency* [$F(1, 34) = 26.38, p < 0.001, \text{partial } \eta^2 = 0.44$], and *decision-action*

latency [$F(1, 34) = 21.75, p < 0.001, \text{partial } \eta^2 = 0.39$] were all statistically significant.

Additionally, interaction effects between scoring decile and tetromino drop were statistically significant for *disarray* [$F(49, 1666) = 36.79, p < 0.001, \text{partial } \eta^2 = 0.52$], *well preparation* [$F(49, 1666) = 3.84, p < 0.001, \text{partial } \eta^2 = 0.11$], and *action inefficiency* [$F(49, 1666) = 1.42, p = 0.30, \text{partial } \eta^2 = 0.04$], but not for *decision-action latency* [$F(49, 1666) = 1.09, p = 0.31$], suggesting that between-groups differences in the former three components grew statistically significantly more pronounced as games went on.

3.3.3 Relating skill components to overall score across games

Another question we asked was whether these components relate meaningfully to overall performance, as captured by post-game points score. We also considered whether the significance of this relationship spanned multiple games. As previous research on psychomotor performance using digital games has suggested that performance windows spanning as few as a player's first 10 games can be predictive of a player's ultimate skill level (Aung et al., 2018), we decided to test whether component performance in the sample's first game is predictive of overall score in the final game. To do so, we first performed a linear regression of final score each on mean component score in the first game. As scatter plots of the residuals indicated non-constant mean and variance, we repeated these regression analyses after box-cox transforming our dependent variable in order to satisfy the assumptions of OLS regression (i.e., score in the final game, which had a heavily right-skewed distribution). The results of these regression analyses are visualised in Figure 3.4, and statistical results are presented in Table 3.2.

All of these regressions were statistically significant. Unsurprisingly, mean *well preparation* in the first game was the only variable to be positively predictive of score in the final game, with all other three components having a negative relationship with the final score. In particular, *disarray* and *action inefficiency* had the strongest linear relationship with final score, each accounting for 8% of the total variance in our dependent variable. In sum, our exploratory analysis of skill components in Meta-T confirmed that our principal components meaningfully related to performance in Tetris, both being capable of distinguishing between expert and novice performance within isolated performance sessions, as well as of explaining a significant portion of statistical variance in participants' final overall tetris scores.

3 Tetris as an experimental platform for cognitive science

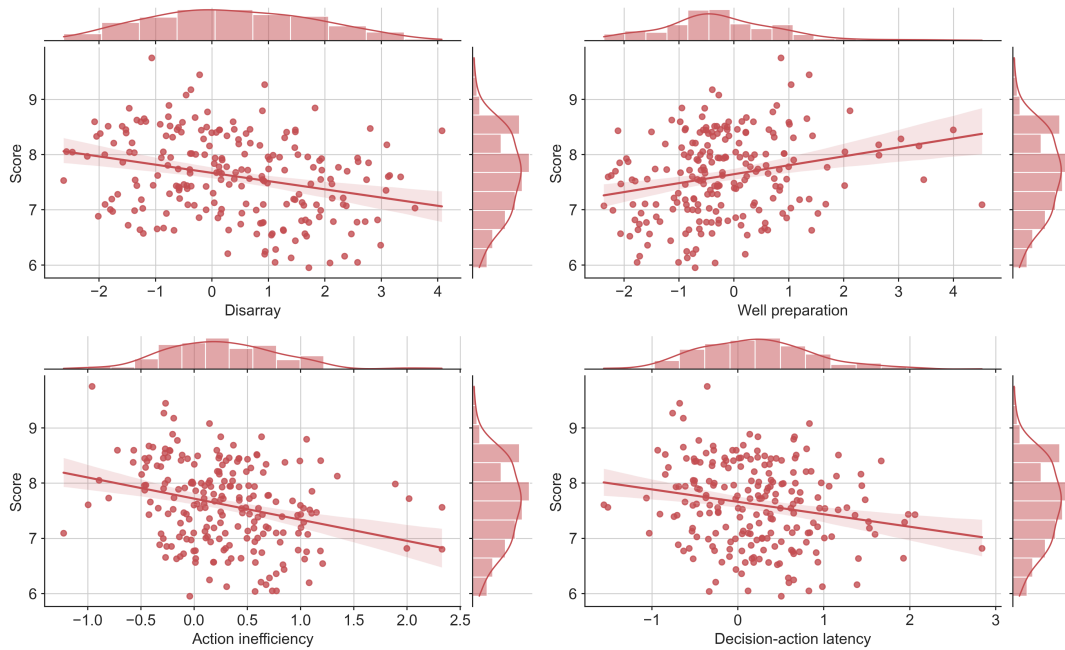


Figure 3.4: Scatter plots of OLS regressions of score on performance component.

3.3.4 Hidden Markov models of skill components

Our results thus far indicate that our decomposed skill components provide a meaningful window into aspects of Tetris performance. However, although participants within respective scoring groups appeared to possess similar skill levels, comparing the trajectories of *individuals'* component scores with each other across games indicated marked differences in how they performed the same task. One approach to making sense of such variations in performance is to model patterns of behaviour as influenced by unobservable psychological states, such as periods of extreme concentration, "flow" (Csikszentmihalyi, 1991), or conversely, periods of inattention. Can we identify such states in Tetris gameplay, such as states of duress characterised by erratic button pressing, or a performant state where participants plan effectively and input controls efficiently? We explored an approach to modelling this possibility in an unsupervised fashion by fitting a hidden Markov model using the `hmm_learn` (HMMLEARN., 2022) package in Python.

Hidden markov models represent sequences of observations in terms of an underlying "hidden" layer of discrete states. The model is used to infer the most likely sequence of states given the observations, and depending on the use case may also serve to generate sequences of observations given the inferred states, with a view towards predicting how the modelled system might behave next. The analytic

| | B | Std. Err. | T | p | R ² | 95% CI |
|-------------------------|-------|-----------|-------|--------|----------------|----------------|
| Disarray | -0.15 | 0.04 | -4.29 | <0.001 | 0.08 | [-0.22, -0.08] |
| Well preparation | 0.16 | 0.04 | 3.72 | <0.001 | 0.06 | [0.08, 0.25] |
| Action inefficiency | -0.38 | 0.09 | -4.37 | <0.001 | 0.08 | [-0.56, -0.21] |
| Decision-action latency | -0.23 | 0.07 | -3.04 | <0.005 | 0.04 | [-0.37, -0.08] |

Table 3.2: Linear regressions of overall score in final game on mean *disarray*, *well preparation*, *action inefficiency*, and *decision-action latency* in the first game.

pipeline for using a hidden Markov model typically involves three main steps: model definition, parameter estimation, and decoding.

1. In the model definition step, the structure of the HMM is defined, including the number of hidden states, the transition probabilities between states, and the observation probabilities for each state.
2. Parameters of the model are then estimated from a training data set. This involves application of the Expectation-Maximization (EM) algorithm, which works by alternating between estimating the expected values of the hidden states given the observations (the expectation step) and updating the parameters of the model to maximize the likelihood of the observed data (the maximization step).
3. Estimated parameters are then used to infer the most likely sequence of hidden states given the input data. This is typically done using the forward-backward algorithm, which works by computing the forward probabilities of the model (the probability of each state given the observed data up to a certain point) and the backward probabilities (the probability of the observed data after a certain point given the state). These two sets of probabilities are then combined to produce a probability distribution of states at any point in the observation sequence.

We hypothesised the presence of three latent states that could influence performance in Meta-T. A default engaged state, where participants exhibited close-to-average performance (as indicated by their component scores). A hyper-engaged state, where participants were expected to perform more effectively and score a greater number of points than in the other states. We anticipated that this

3 Tetris as an experimental platform for cognitive science

state would arise from board configurations that presented opportunities for line clears. And finally, we expected to observe a suboptimal performance state characterised by inefficient and clumsy play, potentially arising from unfocused, inattentive play, as participants wrestled with difficult pile structures close to the top of the playfield. In line with these expectations, we fit a three-state HMM to a combined sequence of our four component scores. At this exploratory stage, we were interested in describing performance at a sample-wide level. For this reason, we fit our model to the concatenated sequence of observations spanning all games and players.

HMMs can model both discrete and continuous observations. Presently, as we sought to identify latent states influencing our observed sequences of performance components at each tetromino drop (i.e., modelling continuous observations), we assumed each component as being described by a normally distributed random variable at each time step. We therefore fit a three-state Gaussian HMM to our four sample-wide sequences of performance components, fitting with a diagonal covariance matrix and stopping the algorithm after 200 iterations. HMMs are generative models and will infer a state sequence no matter what the inputs are. To improve our confidence that our model was producing meaningfully interpretable output, we created a "null" model by fitting an HMM with identical parameters to a randomly shuffled sequence of observations.

We start our presentation of the HMM results by discussing the transition matrix. Figure 3.5 displays the sample-wide probability of participants switching from one state to another, with cells on the diagonal displaying the self-affinity or "stickiness" of states. We can observe that participants are likely to remain in states that they enter (as opposed to switching rapidly between them) given the high probabilities on the diagonal. State 3 is the state least likely to be occupied through consecutive drops at 91% - 5% less likely than the other two states. To further overview the occupancy rates of our states, we decoded the state that was occupied at each tetromino drop using the Viterbi algorithm, and then visualised the amount of time each state was occupied as a fraction of the total amount of time spent across all tetromino drops (Figure 3.6). State 2 accounts for the highest fractional occupancy across the data set $\sim 50\%$, followed by State 3 and State 1 at $\sim 26\%$ and $\sim 24\%$ respectively.

After exploring the temporal dynamics of the states, we sought to describe the behavioural signatures of each state. For each state, we explored the average performance profile of all participants in the sample, as described by the score distributions of each principal component. We computed the mean score of each principal component across all instances of each state (Table 3.3) and visualised the distributions of component scores to provide an overview of both central tendency and spread (Figure 3.7). We describe these trends here.

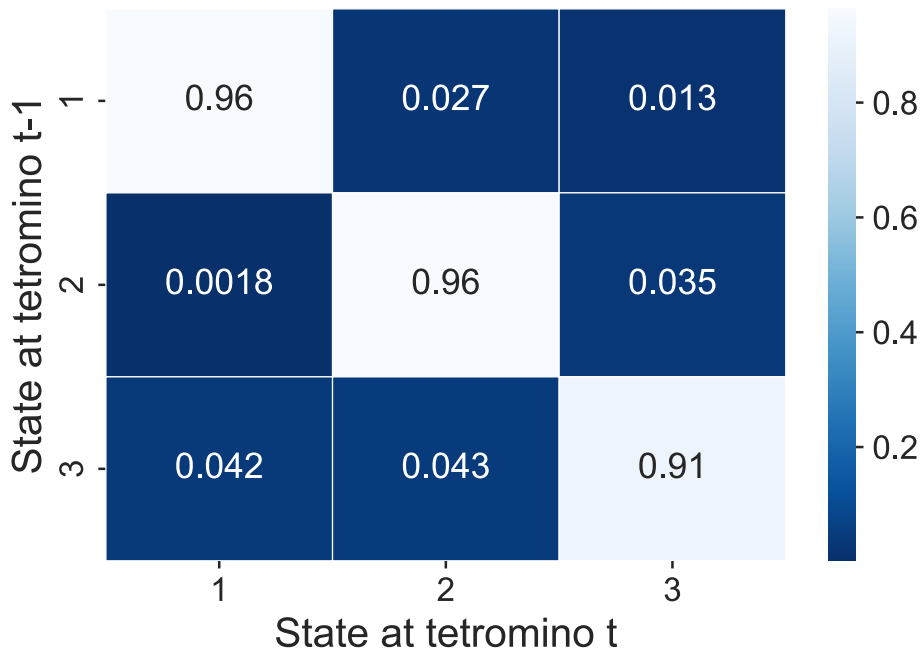


Figure 3.5: HMM state transition matrix. The figure shows the inferred transition matrix of the model, describing the probability of participants switching from one state (y axis) to the next (x axis).

State 1 State 1 appears to be characterised by relatively high *disarray* and *decision-action latency*, while *well preparation* and *action inefficiency* are relatively low. This was also the state with the lowest fractional occupancy, accounting for $\sim 24\%$ of state occupancies across all tetromino drops. We interpret State 1 as an inattentive state that participants sporadically enter as they struggle to keep manage a dangerously tall and messy pile.

State 2 In contrast, State 2 was characterised by low *disarray* and *decision-action latency*, while *well preparation* and *action inefficiency* were close to the average. The fact that participants spent the most amount of time in State 2, with the highest fractional occupancy at $\sim 50\%$. This leads us to view State 2 as the default state of engagement in this data set, characterised by usual, attentive gameplay, as well as uneventful pile structures and relatively short delays in between actions.

State 3 Finally, State 3 was characterised by the highest *well preparation* and *action inefficiency* on average. This state also had relatively low fractional occupancy at $\sim 26\%$, We interpreted this state as the opportunity or point scoring state, in which participants rushed to capitalise on pile structures adequately

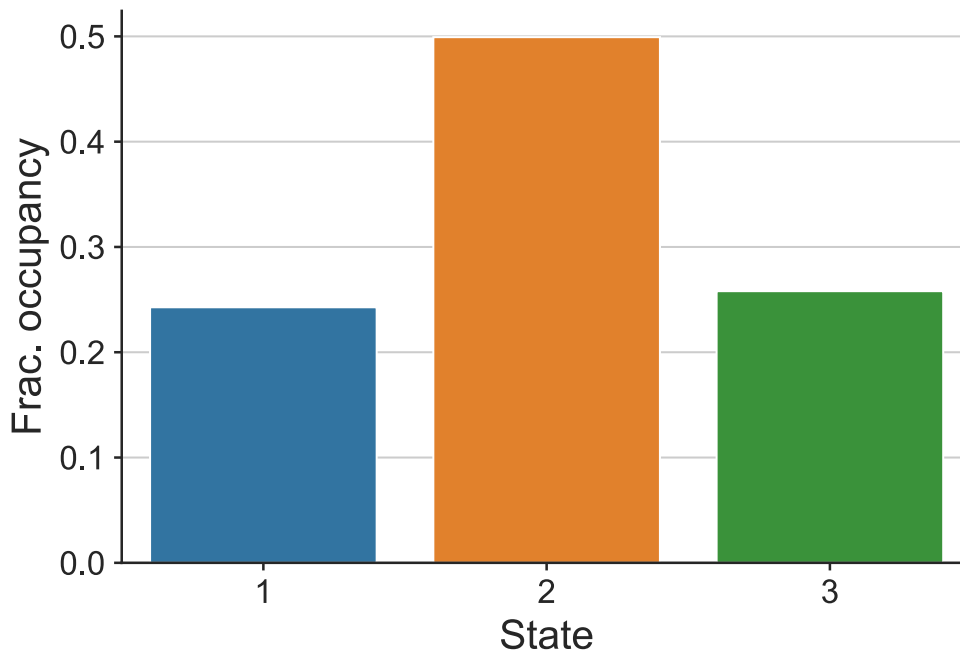


Figure 3.6: Bar chart of state fractional occupancies. Each bar shows the fractional occupancy of the respective state, defined as the fraction of total time on task spent in that state across the entire data set.

prepared for line clears or the Tetris maneuver. However, we interpret this state with some degree of caution due to the large spread in component scores across instances of State 3.

3.4 Discussion

In this chapter, we investigated the viability of using Tetris as a platform for cognitive research by analysing a secondary data set of gameplay logs recorded in Meta-T; a bespoke version of Tetris adapted by Lindstedt and colleagues for the laboratory (Lindstedt & Gray, 2015, 2019). We started with a data set containing over 60 task features comprising second-by-second game-state and behavioural logs of Meta-T gameplay from each of 240 participants. We first performed a PCA to reduce the data to a subset of four principal components, each describing a unique aspect of Tetris performance and explaining a combined total of 53% of the variance in the data set. We then showed that these features provide meaningful insights into Tetris performance, demonstrating that they differentiate between players of different skill levels, and that component scores in participants' first game have a significant linear relationship with overall score in participants' final game. Finally, using HMMs, we explored a novel methodological approach

to analysing moment-to-moment performance by modelling our time series of performance components as observations influenced by unobservable latent states. Taken together, we argue that Meta-T is a resourceful vehicle for the study of psychomotor performance, given breadth and depth at which it logs complex behaviour through time.

| | State 1 | State 2 | State 3 |
|-------------------------|---------|---------|---------|
| Disarray | 3.4389 | -2.4919 | 1.0584 |
| Well preparation | -1.7641 | -0.3752 | 2.7927 |
| Action inefficiency | -0.0942 | -0.0068 | 0.1224 |
| Decision-action latency | 0.0687 | -0.0754 | 0.0727 |

Table 3.3: Mean *disarray*, *well preparation*, *action inefficiency*, and *decision-action latency* in each latent state.

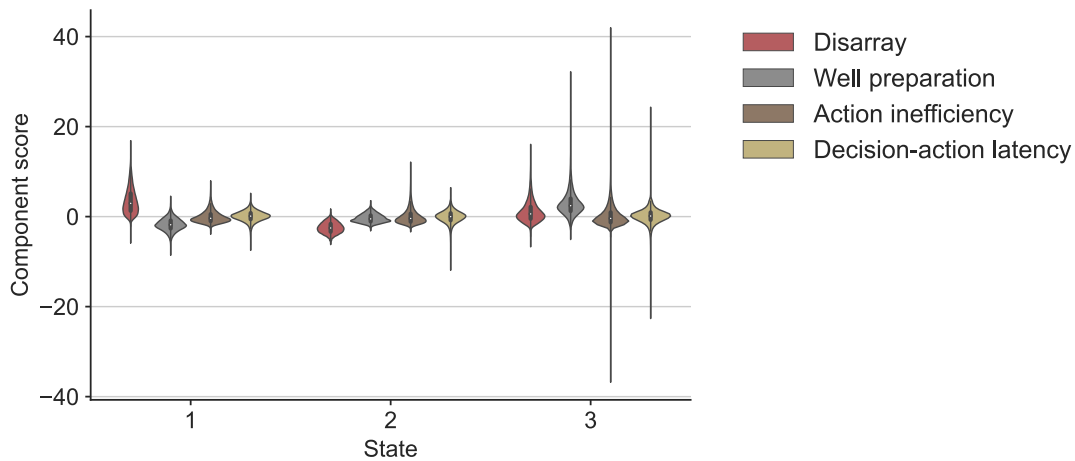


Figure 3.7: Violin plot of principal component distributions across states.

Each "violin" depicts the distribution of the corresponding performance component across all tetromino drops within occurrences of a given state. The black bar in the center of the violin is a box plot, with the center showing the median observed value in the distribution. The coloured portion of the violin shows the kernel density estimate of the distribution. The tails of the violin extending beyond the inner box plot show the range of extreme outliers.

It is reassuring that our own independent attempt at dimensionality reduction echoed original analyses of this data set by Lindstedt and colleagues (2019). Our application of PCA to a trimmed subset of 39 task features yielded four orthogonal components of performance: overall *disarray* of the Tetris pile, preparation of wells in the pile for simultaneous clearing of lines, motor execution inefficiency, and *decision-action latency*. In comparison, Lindstedt and colleagues identified a trio of performance features that captured pile *disarray*, pile preparation, and

3 Tetris as an experimental platform for cognitive science

a joint component capturing both action-latency and motor execution efficiency. As argued by these previous authors, we reiterate that these performance components likely tap into distinct aspects of cognition and behaviour. For instance, the *well preparation* and *disarray* components may be characterised as indices of planning, while *action inefficiency* may capture motor coordination and mental rotation ability. Finally decision action-latency and *disarray* tap into rapid decision-making and situation assessment respectively. In sum, we believe that Tetris is a task that captures distinct aspects of cognition, each of which can be captured through a reliable automated logging system that measures behaviour as the task is engaged with. Future work could seek to verify the relationship between these performance components and purported cognitive factors by correlating each of these measures with validated measures of cognition.

We tested whether the characteristics of these components differ across players of different skill levels. In general, we found that top scoring players exhibit less *disarray* and accrue it at a slower rate than bottom scoring players throughout gameplay. Conversely, top scoring players demonstrated a superior ability to prepare wells for "tetris" maneuvers and simultaneous line clears. They also exhibited less *decision-action latency* and *action inefficiency* on average throughout gameplay as compared to lower scoring players. Given that Meta-T records component scores at each tetromino drop, Meta-T grants researchers the ability to record moment-to-moment dynamics of overall performance, but also *components* of skill that demonstrably distinguish between good and bad performances. We have previously argued that task complexity and richness of data are advantageous qualities of games previously used to produce insights into human cognition, such as Space Fortress (Mane & Donchin, 1989; Donchin, 1995). We show here that Meta-T also possesses these qualities, in addition to being a familiar and engaging game that has the potential to attract large populations of players.

Previous work has shown that performance in a complex video game is sufficient to predict ultimate skill level in a sample of new learners (Aung et al., 2018; Stafford & Dewar, 2014). Although the prior skill level of our participants is unknown, our analyses showed similar trends. By regressing overall score in participants' final game on the average of each component score in participants' first game, we showed that component scores not only statistically significantly relate to overall performance, but are also indices of participants' global skill level. These results are similar to those of Lindstedt and colleagues (2019), who found behavioural components measured in level-1 gameplay to be predictive of participants' overall scores. Because Meta-T granularly measures components of skill that contribute to overall performance, longitudinal experiments into skill learning that involve Meta-T could have the potential to pinpoint factors underlying learning plateaus. For example, having identified points in a learning curve where learning appears to be halted, researchers can proceed to interrogate behavioural data to see whether these plateaus correlate with deficiencies in skill components such as planning or motor execution (e.g., Gray & Lindstedt, 2017). In contrast, conventional

psychomotor tasks that output univariate measures might not allow this level of analysis and hypothesis generation.

In addition to providing evidence for the validity of Meta-T as a suitable research platform, we explored a novel method for segmenting behavioural time series into state-like epochs with distinct behavioural characteristics. We used a three-state unsupervised HMM to model our time series of PCA-derived performance components as a time series of observed behaviour that is influenced by a discrete sequence of unobservable states. In the present data set, each of these states manifested a different average behavioural signature in the sample. One state was characterised by lagged decision-making and disorderly pile structure. Another state was characterised by relatively smooth inputs and decision-making with uneventful pile structures. Finally, the third state was characterised by high *well preparation* (planning the pile for point scoring) and erratic control points. We believe that this approach to analysing multivariate behavioural time series affords researchers several opportunities. Departing from the assumption that behaviourally rich time series are important for the study of skill acquisition, segmenting complex sequential observations in this could allow researchers to investigate the factors that influence variation in performance by pinpointing individual or groups of epochs with particular behavioural signatures, and interrogating the data to generate novel hypotheses. For instance, future research might ask what cognitive or neural dynamics are present during states of suboptimal or optimal performance, although such an approach would require simultaneous recording of gaze behaviour or brain activity using appropriate measurement techniques. Another question one might ask is whether players' fractional occupancy of states changes as they gain experience in the game: as players improve, do they transition to inattentive states less or occupy performant states more often, and do expert players transition between states less often? Future work involving longitudinal measurement and HMMs could address these questions using the approach demonstrated here.

3.4.1 Limitations and future work

A caveat of our initial behavioural analysis is that we restricted our analysis window to the first 50 tetromino drops of games in order to use the entirety of the player sample. However, this limited our scope to easier levels of Tetris, and equally interesting may be analyses that focus on later levels of Tetris that are more cognitively demanding. Furthermore, it is also possible that grouping the entire sample together for our HMM resulted in a less precise model than a separate model for experts and novices. This may be the reason why we observed such high variance in our component scores for State 3. Thus, although variation in participant skill level may have increased the generalisability of our exploratory model, follow-up work could model high versus low skilled players separately, and

3 Tetris as an experimental platform for cognitive science

even model a data set featuring Tetris from a specific difficulty band (e.g., only level-1 or level-9).

Might other games provide a better window into specific aspects of cognition. Tetris is a game that might lend itself well to studying specific cognitive abilities, but it is restricted to these abilities specifically. As a single player game, Tetris cannot allow researchers to investigate teamwork or communication, for which team games such as MOBAs (e.g. League of Legends) are more appropriate, nor can it be used to study game theoretical constructs, for which games like Diplomacy might serve well. Nevertheless, given its popularity and accessible nature, future work could benefit from using Tetris (and in particular Meta-T) for a range of studies that could otherwise be difficult to recruit for (e.g., longitudinal studies). For example, while this study focused on states that arise during gameplay, parallel work could shed light on state transitions immediately before and after gameplay across multiple sessions. These benefits could be compounded by translating Meta-T to a format that is deployable online, permitting the type of mass recruitment that we alluded to previously.

3.4.2 Conclusion

Despite their growing usage in cognitive research, many games produce information-poor data sets that are inadequate for the studying of complex psychomotor skills. In this study, we analysed a behaviourally rich archival data set of Tetris gameplay, obtained via the Meta-T research platform. Using PCA and inferential statistical techniques, we showed that Meta-T has the capacity to produce granular data sets that can describe distinct cognitive-behavioural aspects of performance through time, and that these variables distinguish between players of different skill levels. Further, we took a novel approach to analysing moment-to-moment variations in sequential performance episodes by applying an unsupervised HMM to participants' time series of tetromino drops. In doing so, we demonstrated a way in which researchers can segment complex behavioural time series data into epochs that may open a window into the cognitive dynamics that shape moment-to-moment psychomotor performance. We conclude that Meta-T is more than adequate as a platform for studying cognition, presenting many opportunities for novel research into psychomotor performance and skill learning.

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

4.1 Introduction

Thus far, this thesis has endeavoured to demonstrate that digital games are a promising paradigm for research into human cognition. To reiterate some examples, researchers have recently used telemetry data recorded in commercial games to investigate theories of motor chunking (Thompson et al., 2017, 2019), ageing (Thompson et al., 2014; Kokkinakis et al., 2017), and sleep consolidation (Stafford & Haasnoot, 2017), among some examples of problem domains. On the other end of the methodological spectrum, games that have been tailor-made for laboratory research have been used to further our understanding of neural plasticity (Lee et al., 2012; Voss et al., 2012), skill transfer effects (Anderson et al., 2011; Boot et al., 2010), and have been used to test cognitive architectures that model human learning as a whole (Anderson et al., 2019).

One problem associated with games in research is the use of total or end-game scores to compare the aggregated performance of groups of individuals. We have described numerous studies of game telemetry data, including our own work in Chapter 2, that for various reasons have relied on univariate scores describing performance in the chosen task (e.g., Stafford & Dewar, 2014; Stafford et al., 2017; Huang et al., 2013, 2017; Johanson et al., 2019; Aung et al., 2018). Unfortunately for investigators interested in sterile research environments, games are frequently complex and designed to be engaging, often meaning that players encounter variations of problem spaces across separate interactions with a given game. Although summary scores are sufficient for many types of research questions, exclusive use of such scores may make more detailed analyses of performance difficult, as they may mask underlying factors that can vary across trials and sessions, for instance, changes in player behaviour as a response to novel situations in the game. A proposed solution is that researchers look past total scores by interrogating variables describing *components* of performance, such as

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

patterns of players' control inputs and decisions (Towne et al., 2014; Gobet, 2017; Stafford & Vaci, 2022). This is a particular advantage afforded by digital games as experimental tasks, as their programming often allows researchers to extract high-density behavioural data describing multiple aspects of performance. In the previous Chapter, we demonstrated this affordance by introducing a laboratory version of Tetris (Meta-T; Lindstedt & Gray, 2015) that produces fine-grained behavioural data describing moment-to-moment performance.

A related, broader problem, and one that is potentially exacerbated by the variability in problem spaces inherent in digital games, is that individuals may alternate between periods of good and bad performance within single sessions of play despite proficiency in the game. In behavioural neuroscience, trial-to-trial variability is present even in low-level psychophysics tasks. These confounding behavioural observations have often been the subject of modelling efforts, whereby researchers seek to explain fluctuations in observed behaviour in terms of changes in latent cognitive factors, such as noise in perceptual systems or shifts in attention (van Maanen et al., 2011; Renart & Machens, 2014). In recent years, novel applications of unsupervised learning techniques have allowed researchers to statistically relate trial-to-trial variability to these unobservable cognitive factors, typically modelled as discrete shifts in internal *states* (Chen, 2015; Calhoun et al., 2019; Ashwood et al., 2022). Inspired by these approaches, the aim of this study is to address the two problems described above through the identification of latent perceptual states in the context of human psychomotor performance. More specifically, using simultaneous, high-density behavioural and magnetoencephalography (MEG) recordings of participants playing Meta-T (Lindstedt & Gray, 2015), a laboratory version of Tetris, we show that the dynamics of psychomotor performance can be modelled as observations influenced by latent perceptual states, which we relate to neural markers of attention.

4.1.1 Detecting latent states

A convenience assumption that is made in many cognitive experiments involving sequential measurements, is that data are independent and identically distributed samples from a shared distribution. We can consider an alternative and arguably more realistic perspective by assuming that multiple processes with distinct mechanisms contribute to recurring sequences of observations in a given data set (Kunkel et al., 2020). By treating these temporally recurring patterns in time series as being influenced by distinct, unobservable processes, we can segment continuous data into patterns of observations based on inferences about underlying latent states. For example, in the context of digital games, differences in patterns of performance may arise from players consciously adopting different behavioural strategies (Harwell et al., 2018), or from players shifting attention between different aspects of skill (Lim & Yen, 2004). Modelling the dynamics

of moment-to-moment performance in this way can allow us to make sense of fluctuations in individuals' performance that are typically dismissed as noise, but may actually be important sources of information.

Previous studies adopting this modelling framework have shown that brain states during waking behaviour shape the dynamics of cortical activity, stimulus-response and task performance in different animals and in recent cases have been able to describe behaviour with striking accuracy (Vidaurre et al., 2019; Eyjolfsson et al., 2017; Wiltschko et al., 2015; Calhoun et al., 2019). For instance, researchers investigating the acoustic courtship behaviours of fruit flies were able to precisely predict distinct patterns of song behaviour by statistically inferring latent states from flies' movement data, capturing 84.6% of all remaining song patterning information that previous models lacking a latent state component could not explain (Calhoun et al., 2019). Accurate segmentation of the latent state sequence allowed detailed description of the flies' sensorimotor-strategies corresponding to each state and, following an optogenetics component of the study, identification of the neurons responsible for switching between states. Similarly, latent state models of rodent decision-making can accurately predict choice strategies corresponding to states of optimal engagement versus bias (Roy et al., 2021; Ashwood et al., 2022), permitting reliable detection of blocks of trials with heterogeneous error-rates, as opposed to previous models that would assume errors are scattered throughout all trials in a session with equal probability.

While varying in scope and problem domain, common to some of these studies is their use of hidden Markov Models (HMM), which model observable processes in terms of an underlying sequence of unobservable (i.e., hidden) states that transition with fixed probabilities. Roughly speaking, the typical modelling pipeline involves specification of the number of states that are assumed to influence the process as well as the probabilities of the model initializing at each state, following which the parameters of the model are estimated via maximum likelihood estimation. As described previously, successful validation of HMMs in cognitive task environments allows post-hoc relation of observable behavioural dynamics to underlying brain states, resulting in rich-descriptions of moment-to-moment performance and cognition. These can exist at the group but also the individual level, for instance by analysing how much time individual participants spend in each state and how often they transition between states (Vidaurre et al., 2018).

Depending on the objectives of modelling, the specification of the states can take on different forms. In the examples outlined above, researchers specified a distinct generalized linear model (GLM) for each state that acted as a psychometric function mapping stimulus to sensorimotor response (Calhoun et al., 2019; Ashwood et al., 2022). This approach paired the HMM with a previously tested GLM with proven application in tasks with discrete outputs. A similar usage tested stagelike models of human skill acquisition by pairing each latent state (i.e., stage of learning) with a different speedup function describing participants'

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

response latencies in a novel arithmetic task (Tenison & Anderson, 2016). Other investigations of latent states in humans have included the identification of brain states during wakeful rest or motor task performance by fitting HMMs to electrophysiological time series (Vidaurre et al., 2018, 2019; Karapanagiotidis et al., 2020). As such, studies have demonstrated that HMMs provide a flexible and task-agnostic framework for segmenting behavioural or neural time series into meaningful state-dependent epochs.

4.1.2 Neural correlates of internal states

Thus far, we have highlighted how trial-to-trial variability in psychomotor data obtained through digital games (and in psychological data sets in general) remains unaccounted for in many studies, despite the burgeoning use of digital games as paradigms to investigate psychomotor learning and other aspects of cognition. We have also provided an overview of how sequence classification of high-density behavioural or neural time series, in particular through the use of HMMs, can help researchers to make sense of trial-to-trial variability across various task environments by identifying latent states that subjects shift between as they engage in a task. Accordingly, we continue by considering how latent state identification through this approach may be applied to high-density psychomotor data. In particular, we ask the question: what might states identified through such means represent in terms of their underlying neural and cognitive dynamics?

One aspect of neural activity that can inform us about internal states is endogenous rhythms. These refer to the cyclical patterns of neural oscillations that occur naturally in the brain (independent of external stimuli) as neurons settle into stable firing rates. Endogenous rhythms can be observed at various levels of neural organization, from individual neurons to large-scale networks, and can be characterized by their frequency, amplitude, and phase. Some examples include oscillations in the *alpha* (8-12Hz), *gamma* (25-80Hz), and *theta* (4-8Hz) frequency bands, which have been linked to various cognitive functions such as attention (de Vries et al., 2021; Foster et al., 2017; Bagherzadeh et al., 2020), learning (Popescu et al., 2009; Li et al., 2021), and memory (Osipova et al., 2006; Nyhus & Curran, 2010). While endogenous rhythms are also associated with coordination and communication between different brain regions (Palva & Palva, 2011), we focus presently on the cognitive correlates of rhythms within individual brain regions.

The alpha rhythm was the first human brain rhythm to be detected in the human brain, and is easily measurable across the cortex using electrophysiological methods (Berger, 1929). Despite ongoing conflicts regarding the underlying mechanisms, there is a well-established relationship between occipital alpha and visual spatial attention (Foster & Awh, 2019; Peylo et al., 2021). More specifically, direc-

tion of the attentional "spotlight" from one location in the visual field to another in the absence of eye movement has been shown to correlate with modulations in the amplitude of alpha rhythm in both the parietal and primary occipital cortices (Yamagishi et al., 2003; Sauseng et al., 2005). Alpha can be robustly detected in occipital regions, where it is thought to be associated with visual attention through the suppression of task irrelevant information (Pfurtscheller, 2003; Foxe & Snyder, 2011). Additionally, studies have demonstrated that neural activity in this area is modulated by attention even when visual stimuli are not present (Heinemann et al., 2009; Sundberg et al., 2012).

4.1.3 Aims of the present work

Crucially, the approach described here depends on high-density input data, such as pupillometry and movement in addition to task outcome. In contrast, many studies of human performance that use commercial games are limited to analyses of outcome scores, such as match wins or number of points scored in a round. We bridged this gap by using a previously tested version of a commercial game (Tetris) that has been adapted for use in the laboratory. This version of Tetris records control inputs at every stage of the game, and outputs logs of variables that are of interest to cognitive researchers, such as response latency and motor efficiency. Our aim was to identify latent states (e.g., states of high versus low engagement), to characterise recurring patterns of performance during engagement with our task. Using MEG, we then investigated cortical activity relating to attention in each state.

4.2 Methods

4.2.1 Participants

15 healthy, right-handed participants were recruited through the York Neuroimaging Centre (YNiC) participant pool. All participants provided informed consent, and the study was approved by the York Neuroimaging Centre ethics committee. All participants were familiar with playing Tetris, and provided a self-report of their proficiency on a 5 point Likert scale ($M=3.08$, $SD=1.04$), as well as their proficiency in digital games in general ($M=3.38$, $SD=1.19$). Data from 2 participants were excluded from analysis due to poor MEG data quality, resulting in a final sample of $n=13$ participants (4 female, $M_{\text{age}}=33$, $SD_{\text{age}}=11.31$).

4.2.2 Behavioural data acquisition

Participants played a pygame implementation of "Classic Tetris" called Meta-T (Figure 4.1), developed by Lindstedt and colleagues for the purpose of studying human expertise and learning (Lindstedt & Gray, 2015). Meta-T is a near-identical representation of the original NES Tetris, with the exception of minor visual differences relating to the use of Python as the development language. Importantly, Meta-T possesses several additional features that make it suitable for cognitive science, and has been used as a task environment in several published studies on human and machine expertise (Lindstedt & Gray, 2013; Sibert et al., 2017; Sibert & Gray, 2018; Lindstedt & Gray, 2019). Firstly, Meta-T outputs several data files at the end of each session that, in addition to detailing the participant's ID and other session-specific information, include a log of post-game summary statistics for each game, a log of game-state (e.g., pile structure) and behavioural (e.g., action latencies) information describing performance for each "episode" of play (i.e., the time between a tetromino appearing to the time at which it is placed), as well as a complete log containing key-press information at the millisecond level (See Lindstedt and Gray 2015 for an exhaustive description of logged variables). Secondly, researchers can modify game parameters such as the screen size, game length, or difficulty curve, by editing the default configuration file. In doing so, researchers can constrain participant behaviour to bespoke experimental conditions according to the requirements of their research question.

For the present study, we configured Meta-T to run in a full-screen environment without in-game music. We also fixed the set of numerical seeds determining the tetromino sequence in each game, such that each subsequent game had a different sequence of tetrominoes, but the variations in tetromino sequence were the same across all participants. Games had an indefinite length (i.e., players played each game until loss) and all other options were left at the default values for Classic Tetris. The original stimulus code was further adapted to accommodate a fibre optic response interface (Cambridge Research Systems 905 package) connecting between the stimulus computer and MEG scanner, which allowed participants to use a non-electronic, non-magnetic five button response pad to play Meta-T without adding additional noise to the scans. We configured Meta-T to send triggers to the MEG record via this interface upon the occurrence of salient events. These included button inputs, tetromino appearances and drops, line clears, as well as game start and game end.

4.2.3 Magnetoencephalography (MEG)

MEG scanning was conducted using a 4D Neuroimaging Magnes WH3600 scanner (248 channels + 23 reference channels) at YNiC and data were recorded at a

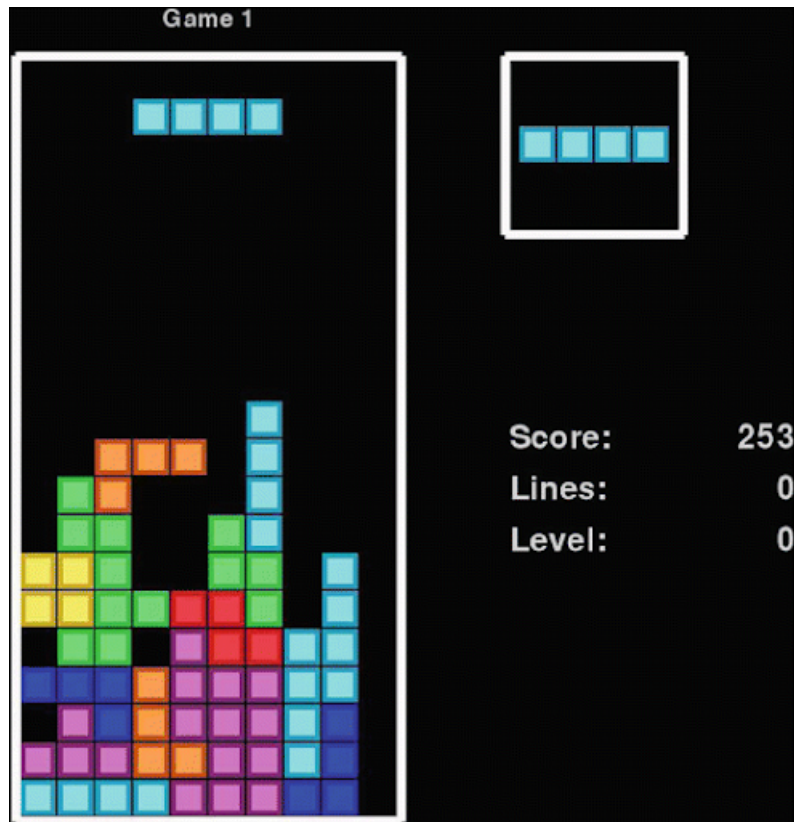


Figure 4.1: Depiction of Meta-T user interface. The left side of the screen shows the current board, including the tetromino that is currently being controlled above it. The top right side of the screen shows the *next* tetromino that will be played after the current one is dropped. The player is also presented with the current game number, their current score, number of cleared lines, and the current level. Image taken from Lindstedt and Gray (2019).

sampling rate of 500 Hz. Prior to scanning, five fiducial head-coils were attached to each participant’s head with hypoallergenic tape. Facial landmarks (nasion, left and right preauricular) and head shape were then recorded using a Polhemus Tastrak 3D digitizer. To assess head movement inside the scanner helmet, we measured the position of the head-coils before and after every scan, and then compared these measurements to the spatial relation between head-coils recorded outside of the scanner. Movement $< 0.5\text{cm}$ was our acceptance threshold for head movement, beyond which we reran our coil-on-head (CoH) scan to confirm any discrepancies in coil position and to subsequently recalculate coil positions using non-displaced coils.

After being briefed and prepared for scanning, each participant was given some time to practice playing Meta-T in the scanner until they reported feeling well-adjusted to the button inputs, during which time the data acquisition software was configured for scanning. Participants played Meta-T in a seated position

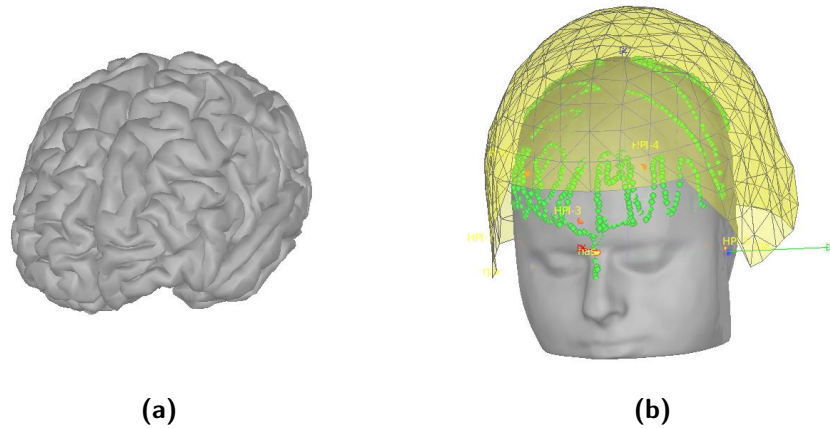


Figure 4.2: Partial depiction of MEG-MRI co-registration process. The left panel shows the 3D rendering of the participant’s cortex following import of their preprocessed MRI data into the Brainstorm analysis environment. The right panel shows the alignment of the participant’s digitized head surface with their 3D head model.

and were instructed to play until they lost, while keeping their head as still as possible. Each scan was initiated five seconds before the start of each game and scan duration varied for each participant depending on their performance across games (i.e., better performance resulted in longer games). As described above, each scan and concurrent game was preceded and followed by a CoH scan, allowing us to assess head movement while the participant took a short break. Each participant typically played two or three games, resulting in an average acquisition duration of $M = 7.84$ minutes per game ($SD = 2.88$) and an average total acquisition duration of $M = 21.55$ minutes per participant ($SD = 4.54$).

4.2.4 Structural magnetic resonance imaging

To estimate the neural sources of our MEG recordings, MEG data were co-registered with high resolution structural MRI images. T1-weighted structural MRI scans were acquired for each participant using a Siemens Prisma 3T MRI scanner, and the Freesurfer pipeline (Dale et al., 1999; Fischl et al., 2004) was used to perform image segmentation and cortical reconstruction. Head surfaces digitized during MEG preparation were then aligned with reconstructed MRI images based on the aforementioned fiducial landmarks. We provide an example of 3D rendering of a participant’s cortex as well as a depiction of the alignment process (Figure 4.2).

4.2.5 Analysis

Dimensionality reduction of behavioural data

We concentrated our behavioural analyses on episodic logs describing behaviour and game-state at the level of each tetromino drop, as these logs provided the greatest breadth of information relating to moment-to-moment performance. For each game that was played, Meta-T produced one such log file as a .tsv spreadsheet. Each row in these files corresponded to one tetromino drop, describing the input behaviours from the moment the tetromino appeared to when it was dropped, changes to the game-state following the tetromino drop, as well as summary variables describing participant and session related information.

We first sought to reduce our behavioural data set into a more manageable and interpretable subset of variables whilst retaining the available information relating to performance. To do so, we referred to the PCA we performed in Chapter 3 (using the sklearn library (Pedregosa et al., 2011)) on an archival data set of 240 Meta-T players collected by Lindstedt and colleagues (Lindstedt & Gray, 2019), which contained analogous logs to those produced by our own protocol. We restricted the PCA to variables that related directly to participants' performance, that is, we excluded variables that describe session and game-state variables describing events independent of participant behaviour (e.g., labels for the current or next tetromino).

As described in Chapter 3, after inspecting an elbow plot of proportion variance explained against the number of components in the model, we reduced our data set to four components based on the point of maximum curvature in our visualisation. We provided meaningful labels to each of these components, similar to Lindstedt and colleagues 2019, according to the unique aspect of Tetris performance captured by each one. Together these components explained up to 53.3% of the variance in Meta-T performance. We describe each of these components below.

1. *Disarray*. Players that fail to clear lines as their tetris pile increases in size are prone to developing an unfavourable tetris pile. Disarray captures this deficiency in pile structure, and is associated with unplayable holes and jaggedness of the pile, as well as overall pile height.
2. *Well preparation*. Achieving a high score in Tetris requires capitalising on opportunities to score bonus points, typically by clearing multiple lines with a single tetromino. Well preparation relates to the forward planning required to achieve multiple line clears, such as by reserving a single, deep well and maintaining a neat pile structure.

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

3. *Action inefficiency.* Action inefficiency captures inputs (e.g., rotations, translations) that are made in excess of the minimum number of inputs required to place a tetromino at its final destination. This may relate to poor motor execution and planning.
4. *Decision-action latency.* This component corresponded to the initial lag and average lag between actions associated with each tetromino placement. It also corresponded to the local quality of placement for each tetromino (i.e., the reduction in pile height caused by placement and amount of contact with tetrominoes in the pile). Taken together we view this component as capturing both the speed and quality of decision-making as it relates to identifying optimal tetromino placement.

We used the PCA performed in Chapter 3 to derive performance components from the behavioural data collected in the current chapter. Specifically, we computed performance components by scaling each row of the current behavioural data by the weights of each component from the previous PCA, producing for each participant an additional time series of four components relating to the aspects of Tetris performance described above. In contrast to Chapter 3, we used the mathematical difference of disarray in lieu of the raw disarray value. The reason for this is that while raw disarray provides a status description of the current board configuration, the difference of disarray is a more direct reflection of participants' moment-to-moment interactions with the game (assuming that a participant is indeed providing inputs to the game), as it is a measure of the net change in disarray at the moment of each tetromino drop. We believe that this measure is therefore more appropriate for an analysis of moment-to-moment performance. After projecting the previous PCA onto the current behavioural data, the scores of each component were then standardised to permit comparison between components with different scales. Figure 4.3 depicts the score distribution of each performance component across all tetromino drops in the sample.

Hidden Markov Models

We used the Python `hmmlearn` package (an open source module with an API similar to `scikit-learn`; `hmmlearn`, 2022), to fit a three state HMM to the time series of PCA-derived performance variables, where each array in the time series describes participant performance at the current tetromino drop. We chose a three state model assuming three modes of engagement with Meta-T: a default state where participants were engaged and attentive, a performant state where participants were both engaged and playing optimally, and a "panic" state involving suboptimal moves and blunders, potentially relating to inattention.

Our model was fit to our data at the group-level (as in Karapanagiotidis et al.,

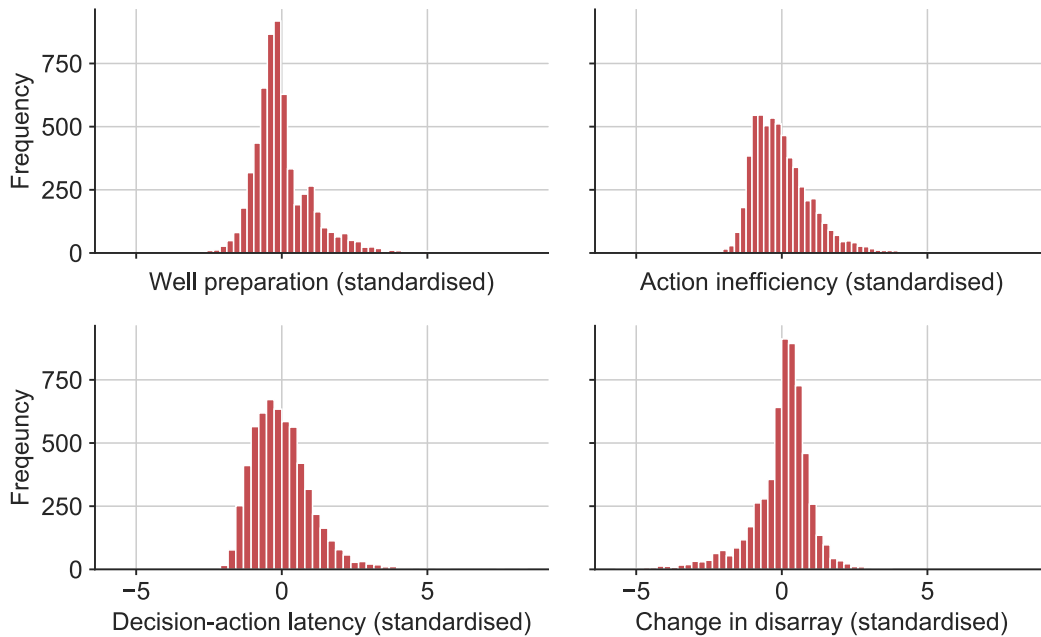


Figure 4.3: Distributions of component score at each tetromino drop for each component. From top left going clockwise to bottom left, the plots show histograms of z-scored well preparation, action inefficiency, decision-action latency, and change in disarray respectively.

2020) by concatenating the data across all of our participants and games. We fit a Gaussian HMM, as our the observations were assumed to be well-described by a Gaussian distribution (see Figure 4.3), which was confirmed using the Shapiro-Wilks test for normality. The model was fit with a diagonal covariance matrix and a 200 iteration upper bound for training, ensuring that the Expectation Maximisation (EM) algorithm stopped either after 200 iterations or on convergence to a maximally likely solution before reaching the iteration limit. As an additional check of model robustness, we compared the log-likelihood of our true model to a randomised chance model that we produced by fitting an HMM with identical parameters to a randomly shuffled time series of our observations. We observed a consistently higher model fit in our true model as compared to our chance model.

MEG data pre-processing

Analysis and pre-processing of MEG was performed in MatLab using Brainstorm (Tadel et al., 2019). Data were first band-pass filtered between 1 and 40 Hz using a finite impulse response filter. We performed an Independent Component Analysis (using the infomax algorithm; Bell & Sejnowski, 1995) to reduce the MEG data and proceeded to identify and reject components capturing periodic

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

physiological artefacts such as blinks and heartbeats. Raw time series for each scan were then inspected manually in epochs of 50 seconds, and any periods contaminated by additional artefacts were manually removed.

After co-registering the MEG data from each scan with the corresponding participant's structural MRI data, we used the minimum-norm imaging function in Brainstorm to estimate the amplitude of sources across the cortical surface via minimum-norm estimation (Hämäläinen & Ilmoniemi, 1994). To extract source time series for each state, we first aligned the time series of the behavioural and MEG recordings, before segmenting the behavioural time series into bins of one second in length. For each bin, we then extracted the time series of squared amplitudes from regions of interest (ROIs) for each bin and state. ROIs were parcellated using the Brodman atlas (Amunts, 2018) as instantiated by Brainstorm. Specifically, we opted to use parcellations of the occipital cortex to facilitate testing of our primary hypothesis regarding occipital alpha, as well as from the motor cortex due to the high amount of motor control present in Tetris gameplay. We maintained a uniform length of one second for each bin to ensure that fourier transforms of each bin, computed with the Fast Fourier Transform algorithm (FFT), were equal in length. Finally, we obtained a measure of the mean amplitude for frequency bands across each state, by computing the root-mean square (RMS) of amplitude across frequencies for all bins for each given state, and each participant. Specifically, we computed the RMS of alpha rhythm (8-12Hz) in V1, and RMS of mu rhythm (8-13Hz) in M1 across each HMM state for each participant.

4.3 Results

4.3.1 SVM decoding

To assess data quality, we checked that we were able to distinguish between neural responses to button presses executed with the left versus right thumb (i.e., the buttons used to translate the tetromino left and right respectively), as this would provide us with confidence that our MEG and behavioural data were both adequately synchronised and contained meaningful information. For each participant, we trained a linear Support Vector Machine using the libsvm library (Chang & Lin, 2011) to decode left versus right translation inputs using the MEG time series extracted from -400ms to 400ms relative to each button press. To improve computational efficiency and signal-to-noise ratio, trials from each class (i.e., left versus right translation) were randomly assigned to 5 folds. Trials in each fold were then subaveraged, yielding a total of 5 subaveraged trials per class. Decoding was then performed on the 5 subaveraged trials following a leave-one-out cross-validation procedure, and the process was iterated 50 times.

Classification accuracy was averaged across the 50 iterations for each millisecond across the trial time range, and plotted for each participant (Figure 4.4). We attained a classification accuracy of 100% for every participant approximately 0.1 seconds the response.

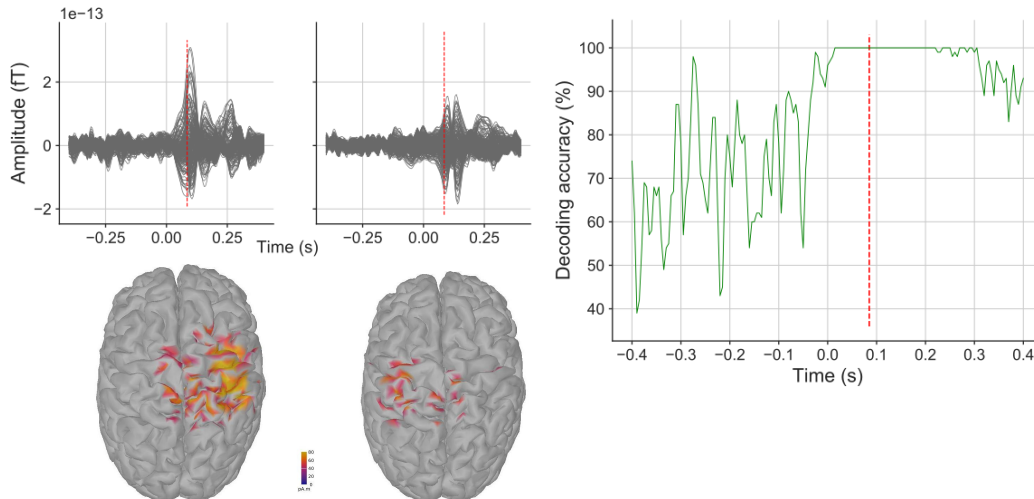


Figure 4.4: SVM decoding of left versus right button inputs for participant R3154. **The top left two panels** show the averaged 244 channel MEG time series from -400ms to 400ms for left and right translation inputs respectively. **The bottom left two panels** show corresponding heat maps of cortical source estimates in picoAmpere at 0.085 seconds after the input was registered by Meta-T. **The right most panel** shows the averaged percentage accuracy of the classifier at each millisecond of the MEG time series. The red vertical dotted line on each line plot shows the time point corresponding to the presented cortical activity.

4.3.2 Hidden Markov Model analysis

State temporal dynamics

We evaluated our three state HMM using a range of metrics describing both the temporal dynamics of the states as well as Meta-T related performance across each state. The central output of the model is the transition matrix, which describes the probability of participants switching between each pair of states from one tetromino drop to the next. Our transition matrix showed that switches between some pairs of states are more probable than others (Figure 4.5). In particular, the probability of switches from State 1 to State 1 and State 3 to State 3 were high (0.69 and 0.79 respectively) showing that participants have an affinity to remain in these states once they enter them. The probability of switching from State 1 to State 2 was also relatively high, while the switches from State 2 to

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

State 2 or State 3 were relatively low (0.2 and 0.13 respectively), suggesting that State 2 was a transient state that participants switched to mostly from State 1 but seldom remained in.

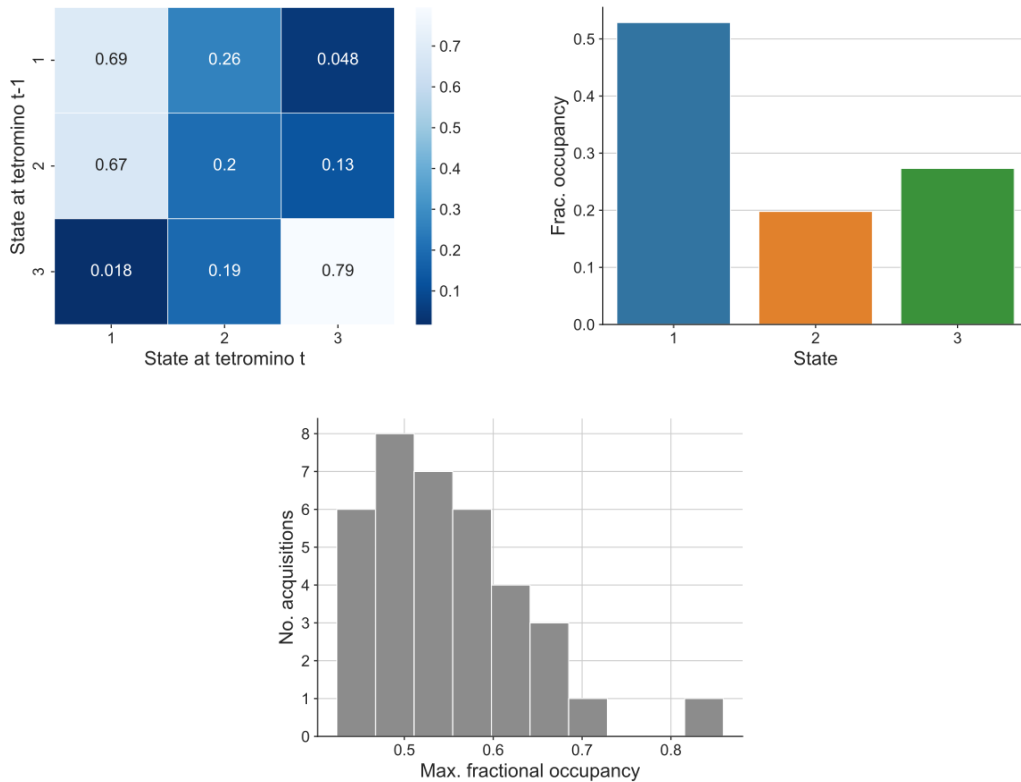


Figure 4.5: Overview of three state HMM temporal dynamics. The top left panel shows the inferred transition matrix of the model, describing the probability of switching between each pair of states. The top right panel shows the fractional occupancy for each state, defined as the fraction of total time spent in that state. The bottom panel shows the distribution of maximum fractional occupancy across acquisitions in the sample. That is, for each data acquisition in the sample, the maximum fractional occupancy represents the fraction of total time spent in the state that the participant occupied for the most amount of time for that acquisition.

To glean further information about the temporal dynamics of our model, we used the Viterbi algorithm to predict the optimal state sequence of our model, and then computed the fractional occupancy of each state, that is, the fraction of total time that is spent by our sample in each state, both in the data set as a whole as well as in each individual game (Figure 4.5). Previous applications of HMMs to the analysis of human brain dynamics have evaluated HMM validity by examining how state occupancy is distributed across participants. An effective HMM would be expected to output state sequences that show participants occupying multiple states without huge discrepancies in state occupancy (suggestive

of single states overwhelming entire participants or recordings). One statistic that reflects this requirement is the maximum fractional occupancy, that is, the fraction of time taken by the state that occupies the most amount of time in a given data acquisition or participant.

To examine this criterion, we visualised our transition matrix together with a bar chart depicting fractional occupancy in each state, as well as a histogram of maximum fractional occupancy across all data acquisitions. In our case, the majority of games had maximum fractional occupancy below 0.6 (mean fractional occupancy was 0.54), demonstrating that our participants' time was shared across all states in our model. Our plot showed that a little over half of all time ($\sim 52\%$) on task was spent in State 1, making this the dominant state throughout task performance. This was followed by State 3, accounting for $\sim 28\%$ of state occupancy, and State 2 with $\sim 20\%$.

State performance dynamics

Together these visualisations inform us about how participants transition between and how frequently they occupy states as they play Meta-T, but they do not tell us how behaviour and cognition varies across states. To investigate the dynamics of our performance components across states we started by visualising the time series of observed performance components for individual participants and games in parallel to the time series of posterior probabilities; a secondary output of our model that describes the probability of each of the three states being active given our observations for any given participant and game (Figure 4.6). By plotting these two time series in parallel, it becomes possible to visually relate patterns of performance to particular states in any given segment of our data. For instance, looking at a game from participant R3154, it is apparent that when well preparation is high, pile disarray is reduced. This pattern appears when the participant enters State 2 which, consistent with our interpretation of the transition matrix, appears to be a transient state with relatively short dwell times. The inverse is the case during State 3, which is associated with low well preparation and increases in pile disarray. It is also apparent that the participant's motor executions are most efficient during State 1, as is evident from dips in action inefficiency following transitions to this state.

State 1. Interpreting performance dynamics across states by visualising individual matches is helpful but not entirely straightforward. To assist in the interpretation of state-performance dynamics across the entire sample, we visualised the distributions of our components across each of our states (Figure 4.7). On average, State 1 is characterised by relatively quick decision-making and efficient motor execution, as well as slightly under-average well preparation and slight in-

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

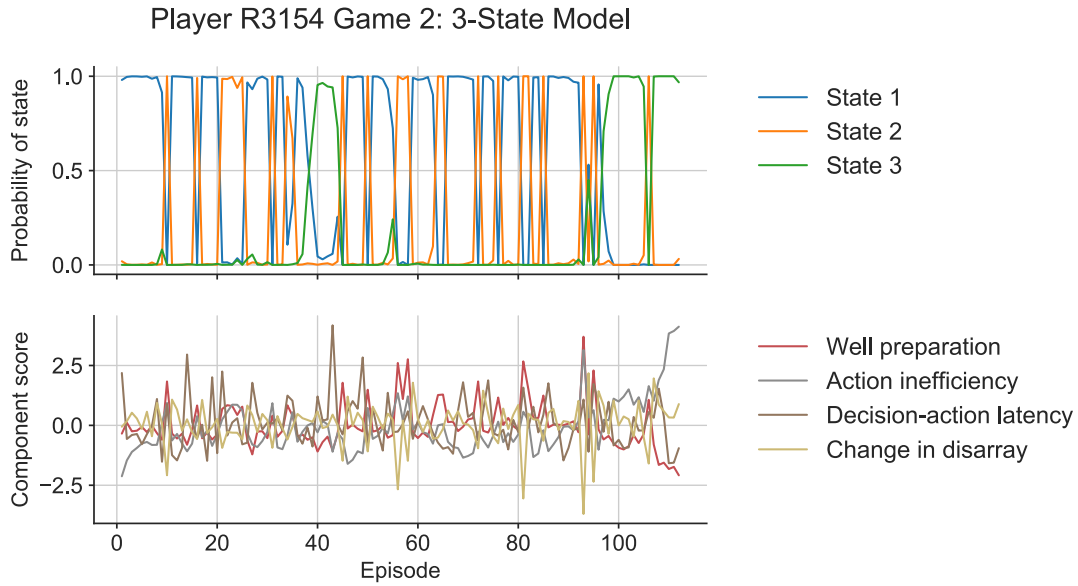


Figure 4.6: Time series of posterior probabilities and observed performance in an example game. The top panel shows the time series of state posterior probabilities, describing the inferred probability of each state being active given the data, across all tetromino drops. The bottom panel shows the time series of all four performance components, displayed as z-scores, across all tetromino drops.

creases to pile disarray at each tetromino drop. In line with State 1 being the most occupied state across the data set, we view State 1 as the default "engaged" state, corresponding to usual, attentive Tetris gameplay.

State 2. On the other hand, State 2 is characterised by high well preparation, reductions to pile disarray, and high motor inefficiency and decision-action latency. Additionally, given that dwell times in State 2 appear to be short, we interpreted State 2 as a transient "opportunity" state, during which the participant is prepared to either score significant points through line clears, or fumble and compromise the established pile structure. We pursued this idea by calculating the percentage of tetromino drops in State 2 that resulted in at least one line clear. This number was 97%, confirming our initial intuition. The remaining 3% of State 2 drops that did not result in a line clear were distributed across 11 players in the sample, indicating that this state does not exclusively capture cleared lines, but rather pile structure conducive to line clears that most players in the sample occasionally failed to take advantage of.

State 3. Finally, and in contrast to State 2, State 3 was characterised by the lowest well preparation, increases to pile disarray, as well as relatively high motor

inefficiency and high but extremely variable decision-action latency. We interpreted State 3 by considering these trends in tandem with aforementioned temporal dynamics. That is, instances of State 3 showed higher dwell times than State 2, and transitions to State 3 were over twice as likely from State 2 than from State 1. Taken together, we interpreted State 3 as a "hyper-engaged" state characterised by poor motor execution and planning, during which participants attempt to resolve difficult pile structures that likely arise from sudden and significant changes to structure that may occur in State 2.

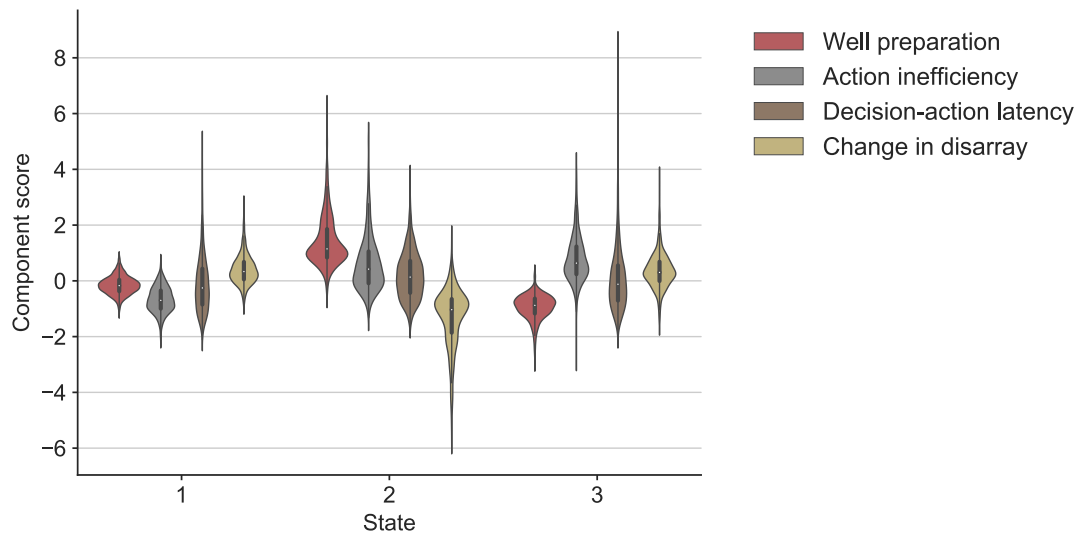


Figure 4.7: Violin plot of performance component distributions across states. Each "violin" depicts the distribution of the corresponding performance component across all tetromino drops within occurrences of a given state. The black bar in the center of the violin is a box plot, with the center showing the median observed value in the distribution. The coloured portion of the violin shows the kernel density estimate of the distribution. The tails of the violin extending beyond the inner box plot show the range of extreme outliers.

Endogenous rhythms across states

We investigated whether our states were neurally distinct by comparing the averaged amplitude of activity within frequency bands and regions of interest between our states. Specifically, after computing the fourier transform of bins across each state, we aggregated neural activity within states for each participant by computing the RMS of frequency bands corresponding to endogenous rhythms of interest. Principal among these rhythms was the occipital alpha rhythm, which has previously been linked to attention. We also compared mu band activity in the primary motor cortex across states.

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

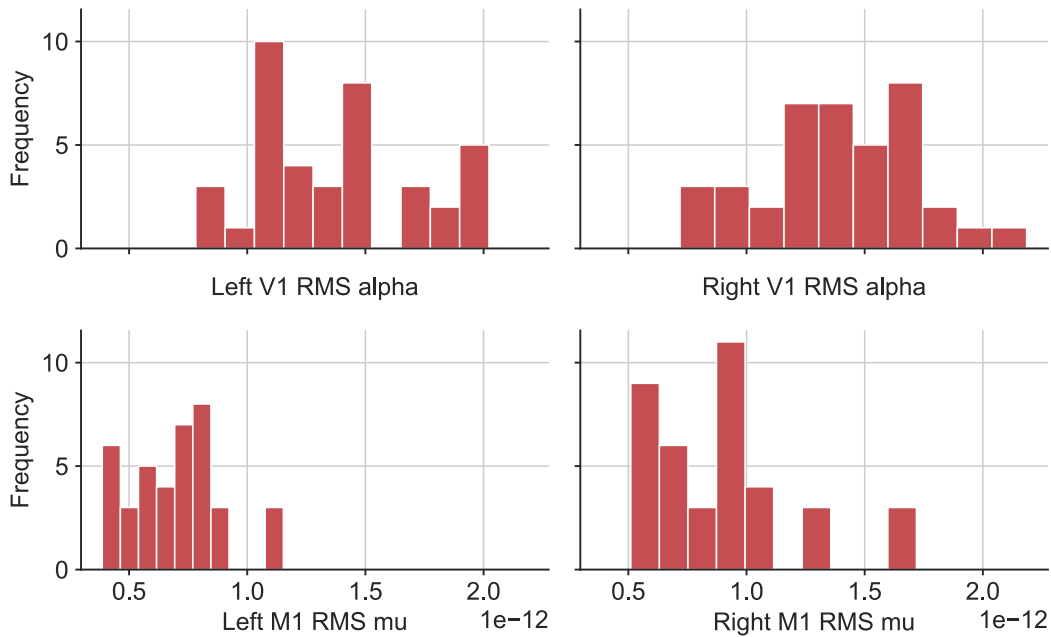


Figure 4.8: Distributions of RMS cortical amplitudes across our regions of interest. From top left going clockwise to bottom left, the plots show histograms of RMS alpha in the left primary visual cortex, RMS alpha in the right primary visual cortex, RMS mu in the left primary motor cortex, and RMS mu in the right primary motor cortex.

We first conducted a one-way repeated measures ANOVA to test for within-participants differences in V1 RMS alpha between states. For each ANOVA, we entered HMM state as the within-group factor, and values of RMS alpha across the aforementioned bins as observations. Interestingly, V1 RMS alpha and M1 RMS mu were different between the left and right hemispheres (see Figure for distributions of the respective observations). For this reason, we conducted separated statistical analyses for each hemisphere of occipital cortex and motor cortex. These tests were significant for both left V1 [$F(2, 24) = 3.6317, p = 0.0419$] and right V1 [$F(2, 24) = 4.2665, p = 0.0260$]. Both tests yielded small effect sizes ($\eta^2 = 0.0024$ and $\eta^2 = 0.0046$ respectively). We also conducted one-way repeated measures ANOVAs to test for within-participants differences in M1 RMS mu between states. Differences in neither left M1 [$F(2, 24) = 0.7357, p = 0.4896$] nor right M1 [$F(2, 24) = 0.8488, p = 0.4404$] were statistically significant.

Post-hoc differences for within-participants differences in occipital alpha across states showed significant differences in alpha activity in the left primary visual cortex between states 1 and 3 ($p = 0.0374$, Cohen's $d = -0.1036$), as well as significant differences in the right primary visual cortex between states 1 and 2 ($p = 0.0194$, Cohen's $d = 0.0835$) and states 2 and 3 ($p = 0.0436$, Cohen's $d = -0.1585$). These results suggest that, in addition to our states displaying distinct

patterns of alpha activity, participants manifest the highest levels of alpha in State 3, followed by State 1.

4.4 Discussion

Drawing on recent advances in behavioural neuroscience, we used an HMM to identify hidden states in multivariate psychomotor data obtained from an ecologically valid task, showing that humans shift between latent states during psychomotor performance that differ in behavioural and neural characteristics. Our task was a laboratory version of Tetris that logs granular performance metrics through time, and was performed in an MEG scanner. We identified three states with unique temporal and behavioural dynamics: 1) a default "engaged" state in which participants spent the most amount of time, characterised by averagely-difficult structure of the Tetris pile, quick decision-making and efficient motor execution, 2) an "opportunity" state with the lowest occupancy and shortest dwell times, characterised by points scoring and high well preparation, but poor motor execution and decision-making speed, and 3) a "hyper-engaged" state with the second highest fractional occupancy, where participants contended with difficult pile structures, often with poor motor execution and decision-making speed. In addition to highlighting differences in performance, we showed that states differ in their neural signatures by comparing the amplitudes of endogenous rhythms between states. Comparisons of neural activity between our three states revealed statistically significant differences in amplitudes of occipital alpha-band activity, a signal associated with attentional state, indicating that differences in cognition across states may relate to attention. Taken together, our findings show that humans switch between behaviourally and neurally distinct states as they engage in complex psychomotor performance. We show that the dynamics of these state transitions can be captured using synchronised behavioural and neural measurements, and subsequently modelled using unsupervised learning techniques to describe the relationship between latent states and performance.

Previous latent state models of behaviour have concentrated on animal behaviour in relatively well-studied task environments, such as courtship behaviours in fruit flies (Calhoun et al., 2019), visual detection in mice (Chen, 2015; Roy et al., 2021; Ashwood et al., 2022), or swim bouts in larval zebrafish (Sharma et al., 2018). These paradigms lend themselves well to models of latent states as the resultant observations are intuitively discretisable. Additionally, many of these studies are high in ecological validity, modelling behaviours that would be natural to observe in an animal's usual behavioural repertoire. In comparison, the application of sequence classification techniques to identify latent states in humans has predominantly involved artificial tasks (e.g., motion coherence task Ashwood et al., 2022) or resting-state fMRI (e.g. Vidaurre et al., 2018, 2019; Karapanagiotidis et al., 2020). In the present work, we used a laboratory adaptation of a highly

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

popular and commercially successful digital game, paralleling growing interest in the use of naturalistic stimuli within the domain of cognitive neuroscience (e.g. Sonkusare et al., 2019; Reggente et al., 2018). Specifically, participants played a laboratory adaptation of Tetris (Lindstedt & Gray, 2013, 2015) that collects numerous cognitive-behavioural variables relating to game state, motor execution, and motor planning. Our analysis included a feature engineering component whereby behavioural measurements were decomposed into four performance components based on data obtained by an independent laboratory using the same task (Lindstedt & Gray, 2019). Thus, using a tried and tested version of a digital game explicitly tailored for laboratory research, we add to a growing body of literature that uses digital games for research in cognitive neuroscience (e.g. Voss et al., 2012; Bavelier et al., 2012; Boot, 2015; Zhang et al., 2015).

Many studies of cognition that use digital games, in particular commercial digital games, analyse univariate measures of performance such as end of match summary metrics (e.g., win/loss, points scored), or time-bound measures of performance (e.g., points scored per minute). We show here that the analysis of multivariate behavioural time series can generate inferences and research questions that may be difficult to access with summary metrics alone. Relatedly, and partly as a consequence of this limitation, studies of digital games that involve repeated measures often aggregate data within and across sessions of engagement. Previous work has advised against this on theoretical (e.g. Towne et al., 2014; Gobet, 2017; Stafford & Vaci, 2022) as well as empirical grounds, demonstrating how certain insights into individual differences (e.g. Harwell et al., 2018) or skill acquisition (e.g. Towne et al., 2016; Rahman & Gray, 2020) can only be achieved after disaggregating data and considering behaviour in a more detailed fashion. Although this study involved detailed analysis of behaviour through time, we are guilty of the sin of aggregation as we too considered our sample as a single homogenous group, despite variation in players' average scores indicating a heterogeneity in skill level.

One related implication for our analysis is that phases of gameplay that are more demanding for less skilled players may place lower demands on best players in the sample. Having concatenated all observations to produce our input time series for model fitting, our model would not have accounted for the potential effects of variation in skill. This is an important consideration, given previous evidence highlighting that variables discriminating between less versus more skilled players are not the same across skill brackets (Thompson et al., 2013). In parallel research involving animal subjects, this issue is either resolved through extensive training, or it is completely bypassed by observing naturally ingrained behaviours. Presently, we made efforts to recruit participants who reported proficiency with Tetris, but we were unable to control for how proficient they were. Additionally, we realised during data collection that many participants were familiar with modern versions of Tetris with nuanced differences that confounded their initial experiences for the game. For instance, our configuration of Meta-T emulates

Classic Tetris and therefore lacks visual guidelines indicating each tetromino's destination, and prohibits rotating tetrominoes at the very edge of the well; both of these are mechanics that some of our more experienced participants reported relying on in their usual recreational gameplay. We acknowledge that these issues are likely to have introduced noise to our model.

Compared to previous latent state models of low-level psychophysical phenomena, we opted for a complex behavioural environment that is high in ecological validity. In doing so, we acknowledge our position in the trade-off between simple behavioural data suitable for predictive modelling versus rich behavioural data that makes prediction much more difficult. Given the nature of our input data (i.e., our time series of performance components), our model infers parameters that describe the temporal dynamics of our states, and generates emission probabilities describing In the case of our Gaussian HMM, the emission probability parameters of each state were the mean and standard deviation parameters describing the Gaussian probability density function of each performance component in the respective state. In short, our model describes the expected distribution of each performance component across states, but this is a far cry from predicting how and where participants drop their tetrominoes at each epoch. This is in contrast to aforementioned alternatives such as the GLM-HMM validated their model, in part, by evaluating the fit of each state's corresponding psychometric curve.

The validity of our model is supported somewhat by the correspondence between the behavioural characterisations of our states and the underlying neural signatures of each state. That is, in a model that failed to distinguish between cognitively meaningful states, we would expect to observe no differences in neural signatures associated with cognition. Instead, comparisons of neural activity across our inferred states revealed statistically significant differences in occipital alpha, a signal that has been previously linked to attention. In particular, post-hoc tests revealed elevated occipital alpha in State 3 as compared to State 1, and higher occipital alpha in State 1 as compared to State 2. Given the existing association between occipital alpha and attention, this result is consistent with our analysis of performance, which was suggestive of increased attentional demand in State 3 due to the presence of difficult pile structures and large variance in participants' decision-making latency. However, although it is likely that our model is detecting shifts in attention, it is difficult to infer precisely what aspects of the task are being attended to across different states, as we did not manipulate attention as previous experiments investigating performance and attention have done (e.g. Cohen & Maunsell, 2009; Mitchell et al., 2009).

It is also possible that attention shifts continuously, and not discretely as assumed by our model. Ashwood and colleagues (2022) found superior model fit in their discrete model as compared to a model with continuous latent states (Roy et al., 2021), albeit in the context of a different task. Additionally, these authors found

4 Identifying latent perceptual states using behavioural and neural recordings during Tetris play

that a two-state discrete model fit human data from a motion coherence task better than a three-state model. We are open to the possibility that models with different assumptions may describe performance in the present context better than our three-state gaussian HMM, for instance a two-state, engaged versus disengaged model. However, this is a question for future work.

4.4.1 Limitations

A limitation of our study is that, in contrast to previous latent state models of behaviour, we did not have adequate time to train our participants on the task. Ours was a complex psychomotor task requiring both rapid perceptual decision-making and skilled motor inputs. Although we recruited participants who all indicated ample prior experience with Tetris, as mentioned before, we are nonetheless conscious of large variation in participant skill, as well as noise arising from unfamiliarity with our specific configuration of Meta-T, the controller, and the scanner environment in which the task was performed. In addition, we note the absence of a "ground truth" model with which to validate our model. Instead, we compared the log-likelihood of our model to a randomised chance model, which indicated superior fit of the true model. However, we acknowledge as a limitation that due to the nature of our input data and the type of HMM that we used, the predictive capacity of our model is restricted.

4.4.2 Conclusion

Recent work has demonstrated that, in the context of repeated trials within sessions of behaviour, animals and humans shift between discrete behavioural strategies during performance. Using simultaneous behavioural and neural recordings of participants playing a laboratory version of Tetris, a popular commercial digital game, we extend previous work by capturing shifts in latent states during performance in an ecologically valid task. Individuals in our sample shifted between three states, each with unique performance characteristics during gameplay. Further, MEG analysis revealed significant differences in occipital alpha across states, suggesting that one neural marker of internal state may be the amplitude of endogenous rhythms, and that in the context of psychomotor performance a cognitive marker of internal state may be attention. Our results show that analysing sessions of data by averaging summary statistics alone may mask a wealth of information describing the dynamics of performance and cognition. We demonstrate how these dynamics can be captured by fitting unsupervised HMMs to high-density time series data.

5 Discussion

The aim of this thesis was to answer the question "*How can digital games be used as experimental paradigms for the study of psychomotor performance and skill acquisition*". Understanding this question would have profound implications for the fields of psychology, neuroscience and game analytics.

We investigated this question first by analysing a large, observational data set drawn from a commercial digital game (Chapter 2). This dataset allowed us to examine the influence of practise spacing on learning. We chose this approach because there is a large existing literature base on learning and spacing but most studies are performed with relatively small numbers of subjects ($n < 100$) and relatively infrequent measurements (typically less than a dozen measurements per person). Here we had access to a dataset with well-defined, objective outcome measures (in-game performance metrics such as KDA and MMR) which were measured at a fine temporal scale (summary metrics were provided for each game). Importantly, we had a very large number of subjects ($n = 162,417$). Our summary data were therefore able to provide excellent approximations to those of the overall population. The results of this analysis are summarised below.

To address the limitations of this 'big data' approach, we then switched our focus to a laboratory adaptation of a different commercial digital game: Tetris. Our goal here was to explicitly identify neuronal and behavioural correlates of cognitive 'state switches' during long periods of gameplay. We first (Chapter 3) analysed an archival behavioural data set to validate previous research and test a multivariate method of estimating instantaneous cognitive state (hidden Markov model analysis). We then (Chapter 4) analysed data from our own experiment involving the collection of simultaneous high density behavioural and source-imaged magnetoencephalography neuroimaging data.

We discuss the major findings of each experimental chapter below, and conclude with a discussion of relevant future directions.

5.1 Chapter 2

In Chapter 2, we analysed a large observational data set describing *League of Legends* performance globally and longitudinally. This dataset was first used to examine the relationship between learning rate and final performance (Aung et al., 2018) but no fine-grained analysis of the relationship between practise spacing and overall performance had been performed previously. By comparing the practice trajectories of players that clustered versus distributed their first 100 ranked games, we found that spacing (as opposed to clustering) practice sessions is positively correlated with final achieved performance, extending an established literature on practice scheduling (e.g. Donovan & Radosovich, 1999; Lee & Genovese, 1988).

Previous studies have demonstrated similar effects across various commercial games, also using a combination of large telemetry data sets with data slicing techniques (e.g. Stafford & Dewar, 2014; Stafford et al., 2017; Huang et al., 2017; Stafford & Haasnoot, 2017). However, these studies, as well as our own, operationalised practice spacing as the amount of time elapsed between the first and last game. This conceivably allows for a range of different practice schedules to exist within the same group of players. To iterate on this approach, we used time series clustering techniques to identify prominent clusters of practice schedules that existed within the data, and compared these to our rule-based grouping of practice schedules. This data driven approach revealed that players typically adopt a consistent rhythm of game playing, taking one large break from playing at some point within their 100 games. Interestingly, although players could be clustered reliably depending on the timing of this break, final achieved performance did not depend on the timing of this break, but only the total amount of time taken away from the game.

Although the ability to perform analysis on hundreds of thousands or even millions of subjects is attractive, the downside of using anonymous telemetry data is that we do not have a deep understanding of each subject. In particular, we cannot measure or control even simple factors like age, sex or environment. We also cannot measure physiological correlates of behaviour which are critical for understanding the neuronal basis of cognitive function. One possibility to resolve these issues is to recruit participants online, controlling for demographic and cognitive variables, and then to trace longitudinal performance trajectories through publicly accessible game APIs using participants' game IDs. The availability of wearable, commercial physiological measurement system (for example, smart watches or low-channel count EEG systems) raises the possibility of including simultaneous measurements of heart rate, breathing and even neuronal activity as well as sleep statistics and estimates of circadian rhythm phase and reliability.

Combining cognitive and longitudinal behavioural data in this way, at scale, could

allow us to test the models of talent identification and skill acquisition that have been proposed in previous studies (e.g. Den Hartigh et al., 2016; Gagné, 2004; Simonton, 2014), and improve our understanding of how cognitive ability interacts with sustained behaviour to produce ultimate performance. While the data that we analysed were curated by the developer of *League of Legends* (and therefore contained important data such as MMR that are typically hidden from public view), we can imagine researchers scraping game data through available public APIs, generating even larger data sets that are open sourced.

Additionally, while the analysis in Chapter 2 only tested for the *presence* of the distributed practice effect, larger and more controlled data sets may allow us to identify the optimal amount and timing of breaks for individuals looking to maximise their skill acquisition rates. This would have a direct impact on the huge population of stakeholders in the video game community (e.g., professional esports bodies, coaches, players). But in addition, it may have translational value for adjacent domains such as sport psychology and even for the more general learning literature. As has been noted previously, the link between in-game performance and cognitive function suggests that the analysis of large populations of gamers may also have clinical value in the form of cognitive epidemiology (e.g., Kokkinakis et al., 2017) at a national or international level. Finally, the analysis of remote game telemetry data present great opportunities for citizen science. It already appears that publicly-available data have led to a small "cottage industry" in data analytics in some game communities with relatively sophisticated analytic and visualisation techniques becoming available to individual players wishing to study and improve their own play (e.g., Cavadenti et al., 2016; Phy, 2023; Dotabuff, 2023).

Although the analysis of large observational data sets presents numerous advantages (e.g., Goldstone & Lupyan, 2016), the ability to acquire these data sets is currently limited to the availability of (commercial) games with large populations of players. Previous research has suggested that games must be "manageably complex" (Gray, 2017) to enable studies to reveal deeper insights about the mechanisms of skill acquisition. As we have seen, games such as SF are successful in this regard as they are complex, but also produce behavioural metrics describing moment-to-moment progress across elements that contribute to this complexity (e.g., navigation, damaging mines, damaging the fortress). In contrast, our *League of Legends* data set (which resembles publicly available data through the API) was limited to post-match summary statistics describing various aspects of performance (e.g., damage dealt, gold earned). Without the ability to trace moment-to-moment performance, viable research questions are restricted to particular time scales, which may in turn restrict our potential to reveal deeper insights about skill acquisition. Moreover, *League of Legends* is a team game requiring coordination with 4 team members competing against 5 other players that are affected by their own intra-team dynamics. This likely introduces a large amount of noise, further obscuring the results of studies in this area.

5.2 Chapter 3

To understand the link between performance and cognition in more detail, in the final experimental chapters (Chapters 2, 3) we switched to a lab-based paradigm in which we could study subjects in person. The timescale of study in Chapter 2 was days to weeks but here we asked whether we could identify behavioural correlates of changes at the scale of seconds to minutes. These changes are typically ascribed to spontaneous variations in cognitive state - In chapter 3 we applied a variant of this technique to multivariate behavioural data acquired while subjects played Tetris.

In Chapter 3, we assessed the extent to which Tetris can be used to study cognition, and to address some of the limitations of the previous chapter. Tetris is less complex than many commercial online games, but it is demonstrably engaging and has in recent years seen a resurgence as a popular online game in its own right (O’Callaghan, 2018; Tarantola, 2020; Sweet, 2021). While the commercial version of Tetris has previously been used to assay cognitive ability, recent studies have developed and deployed a laboratory adaptation of the game (i.e., Meta-T; Lindstedt & Gray, 2015) for use as an experimental platform to study cognition. In contrast to our League of Legends data set, this version of Tetris produces describing behaviour (e.g., key-presses) and game-state (e.g., pile structure) at the level of milliseconds. These studies have produced important insights into aspects of Tetris performance (Lindstedt & Gray, 2019), as well as models of decision-making (Sibert et al., 2017; Sibert & Gray, 2018) and reaction time variability in expert players (Denga, 2021).

We analysed a secondary data set describing the performance of 240 participants that played 50 minutes of Tetris in laboratory conditions. Using PCA, we first decomposed the data set to produce four orthogonal components describing separate aspects of Tetris performance. We then confirmed that these components meaningfully describe performance in the sample, first by comparing the trajectories of top and bottom scoring players on these components, and then by regressing ultimate performance on performance components measured in the first game. Thus, given the ability of Tetris to reliably trace multiple orthogonal behavioural variables through time, we find that Tetris is comparable to SF in its ability to enable detailed investigations of cognition. Tetris further improves on this ability as it is a commercially successful game that is inherently engaging, potentially allowing researchers to scale up its use via online deployment and mass recruitment of participants.

Across all of our visualisations of performance trajectories in both chapters 1 and 2, we consistently observed fluctuations in performance both at the individual and group level. While these types of fluctuations in time series are often dismissed as statistical noise, a significant body of work has ascribed these variations to spon-

taneous variations in cognitive state and in particular the tendency of individual subjects to move in and out of states of 'engagement' and 'flow'. Recent work in this area has applied a combination of hidden Markov models (HMMs) and general linear models to study these spontaneous switches in animals performing simple psychophysical tasks or stereotypical mating behaviours (e.g. Ashwood et al., 2022).

We applied a variant of this technique to the multivariate behavioural data set analysed in the present chapter, which resulted in a model of Tetris performance characterised by three states with unique behavioural signatures. Our analysis allows for a deep dive into the behavioural and temporal characteristics of each state. For instance, one state was characterised by high well preparation and points scoring, but short state durations. Another was characterised by high action latencies and sub-optimal decision-making, but also relatively long state durations. Additionally, our modelling approach allows us to estimate how often different participants enter different states and when they do so.

Taken together, we find that Tetris is a useful platform for investigations of cognition. Using Tetris in the laboratory allows recording of multivariate behavioural data and detailed analysis of multiple aspects of performance. While commercial versions of Tetris have limited game modes that produce random sequences of tetrominoes, the adaptation of Tetris we used presently allows the researcher to precisely specify what stimuli are produced. For instance, the experimenter can configure Tetris to produce predictable sequences of tetrominoes, fix the game difficulty, provide visual cues that show players where tetrominoes will land, and modify many other variables to curate experimental conditions of interest.

While further experimental work is required to validate latent state model of Tetris performance, as a proof of concept we demonstrate how complex behavioural time series can be segmented into discrete states that may tap transient aspects of cognition, such as states of flow (Csikszentmihalyi, 1991) or anxiety. In the current chapter, we investigated states that individuals may shift between during behaviour, but equally interesting may be states that arise before and after periods of performance. For example, combining neuroimaging techniques with latent state models of this variety, it could be possible to identify states that are associated with memory consolidation and skill acquisition. Additionally, state models can be applied to time scales comprising weeks rather than individual games (e.g., Tenison & Anderson, 2017), allowing researchers to test stagelike models of skill acquisition, or may assist in the periodising of training schedules by identifying periods of poor performance or stress.

5.3 Chapter 4

Changes in cognitive state must arise in the brain. The ultimate goal of this thesis was to link analysis of video game data to direct measurements of neuronal activity recorded during the execution of complex psychomotor skill. In Chapter 4 we achieved this by combining the HMM analysis described in Chapter 3 with state-of-the-art high-density, source imaged MEG measurements of neuronal activity in video game players at millisecond resolution. HMM techniques have been used previously to analyze human resting state MEG data (Karapanagiotidis et al., 2020; Vidaurre et al., 2018, e.g.,) but to our knowledge this is the first time that they have been applied to subjects performing an active, non-time-locked cognitive task, and in particular one with high ecological validity.

This was an ambitious project that pioneered an analysis pipeline that has analogues in ongoing work using high-density electrophysiological and behavioural measurements in rodents and humans (Ashwood et al., 2022; Roy et al., 2021; Vidaurre et al., 2019). For simplicity, we chose to analyze a single frequency-domain measure (alpha-band power) measured over multiple MEG sensors because alpha power is well-known to correlate broadly with cognitive engagement (Yamagishi et al., 2003; Sauseng et al., 2005; Foster & Awh, 2019; Peylo et al., 2021, e.g.,). We asked, simply, whether the behavioural states identified using the multivariate HMM methods from Chapter 3 had a neurophysiological correlate in alpha band power.

Our data indicated that they do. This was tested by synchronising our behavioural and neural time series, and then comparing occipital alpha band power between estimates of latent states conditioned on the time series of Tetris performance components that were engineered in Chapter 3. This analysis therefore linked HMM estimates of cognitive state derived from multivariate behavioural data with a direct measure of neuronal activity. Higher alpha band power was identified in the state characterised by difficult structure and suboptimal motor execution. As well as raising important scientific questions about the significance of endogenous alpha rhythms and their potential relation to flow (Csikszentmihalyi, 1991) and disengagement state, this work provides an important link between two different measurement methods.

There are a wide range of additional analyses that we could apply to our existing data and an even wider range of experimental questions that we would like to address using this approach in the future. Using the existing dataset, we could examine the link between other frequency bands (or combinations of frequency bands and cortical locations) and behaviour. As well as studying the neuronal correlates of ongoing cognitive states, we are also interested in the neuronal correlates of state switches: can we predict changes in state switch before they happen from some feature of the neuronal data? In addition, it is well-understood that

a key feature of brain function is communication between different areas (e.g., Sporns et al., 2005; van den Heuvel et al., 2009). This is sometimes measured using a 'connectivity' metric (most simply, as the correlation of response amplitudes between two areas measured over some time period). The high temporal resolution of MEG means that such connectivity metrics are particularly appealing and we plan to examine changes in ongoing neuronal connectivity as a function of behaviour. It would also be possible to extract HMM states directly from the MEG data as outlined in Vidaurre et al. (2018). If cognition and behaviour are linked, a natural prediction would be that the statistics of these latent neuronal states match those derived from behaviour.

Future work could also benefit from advancements in MEG technology such as Optically Pumped Magnetometers (OPM; Tierney et al., 2019). These are wearable alternatives to static, cryogenic MEG systems, and can be placed within millimeters of the subject's scalp, resulting in a three- to five-fold improvement in measurement sensitivity (Seymour et al., 2021). The ability for the subject to move in tandem with the measurement device also enables the study of other complex (non-) video game tasks that might require more movement, such as games played with the mouse and keyboard, without producing the motion artifacts that would arise in traditional MEG systems. These opportunities push the cutting edge of what is currently available in the context of neuroimaging experiments.

5.4 Conclusion

In this thesis, we make significant strides in understanding the relationship between video game performance and cognition by leveraging advances in data science, cognitive neuroscience, and psychology. The primary contributions of this work lie in the development and application of novel methods to analyze complex video game data and elucidate underlying cognitive and neural activity. These methods include clustering techniques for determining player practice schedules, PCA for decomposing performance components, and HMMs for studying spontaneous changes in cognitive states. The use of these analytical techniques enabled the examination of behavioural and neural correlates of performance in both commercial and lab-based games, with a particular focus on *Tetris* as an experimental platform.

The results of this thesis have important implications for current and future research. The identification of latent cognitive states in *Tetris* players and their relation to ongoing neural activity measured by MEG provides a valuable link between cognitive and neurophysiological measures. This line of inquiry paves the way for investigations into other frequency bands, neuronal connectivity, and the prediction of state switches based on neuronal data. Future research could

5 Discussion

also benefit from technological advancements such as Optically Pumped Magnetometers (OPMs) for improved sensitivity and wearable alternatives to static MEG systems, as well as the incorporation of additional measurement modalities such as eye tracking and heart rate monitoring.

Moreover, the study of complex tasks like video games offers the potential to explore a wide range of cognitive and behavioral phenomena, including memory consolidation, skill acquisition, and stage-like models of learning. Researchers could consider applying these methods to other games, requiring more movement or diverse cognitive demands, which would enable a broader understanding of cognitive processes and their neural correlates. In addition to the scientific contributions, the findings of this thesis have practical implications for the development of personalized learning experiences and training schedules based on individual performance and cognitive state fluctuations. By identifying periods of poor performance or stress, tailored interventions could be designed to maximize learning outcomes and promote well-being.

Overall, this thesis represents a step forward in our understanding of the complex interplay between video game performance, cognition, and neural activity. Furthermore, the integration of synchronised behavioural and neuroimaging data open up exciting new avenues for research, with the potential to advance our knowledge of human cognition and inform the design of effective training interventions. We show that fine-grained video game telemetry can be obtained from a variety of sources as well as from a large fraction of the entire world's population. By showing that these telemetry data are unambiguously linked to well-established effects from the skill acquisition literature as well as ongoing changes in neuronal activity, we provide strong support to the idea that video game data is a way of investigating cognitive function at a global scale.

6 Appendix B

6.1 Table of variables used in Principal Component Analysis

| Variable label | Description |
|----------------|---|
| rots | Rotations: The number of zoid-rotations performed in an episode. |
| trans | Translations: The number of times the zoid was moved left or right– translated– in an episode. |
| prop_u_drops | Proportion of user drops: The proportion of top-to-bottom movement that was intentionally dropped by the player in an episode. |
| min_rots_diff | Minimum rotations difference: The difference between the number of rotations used and the number needed to achieve the zoid’s final position. |
| min_trans_diff | Minimum translations difference: The difference between the number of translations used and the number needed to achieve the zoid’s final position. |
| initial_lat | Initial latency: Time elapsed in milliseconds from the start of the episode until the first key-press. |
| drop_lat | Drop latency: Time elapsed in milliseconds from the start of the episode until the player first drops the zoid. |
| avg_lat | Average latency: The mean time between all key-presses in an episode. |

6 Appendix B

| | |
|-------------|--|
| resp_lat | Response latency: The time from the start of the episode in milliseconds until either the zoid is first dropped or the zoid is locked into its final position. |
| mean_ht | Mean height: The mean height among all 10 columns in the pile. |
| max_ht | Maximum height: The maximum height among all 10 columns in the pile. |
| min_ht | Minimum height: The minimum height among all 10 columns in the pile. |
| cd_1 - cd_9 | Column difference (0, 1)–(8, 9): 9 features representing the difference in height between each successive pair of columns. |
| max_diffs | Maximum difference: Maximum difference in heights among cd_1 through cd_9. |
| pits | Pits: The number of empty cells in the pile that are covered from above. |
| pit_depth | Pit depth: The sum of all pits weighted by the number of filled cells above them in a column. |
| pit_rows | Pit rows: The number of rows containing pits. |
| lumped_pits | Lumped pits: A measure of pits considering all adjacent groups of pits to be identical. |
| wells | Wells: The number of empty, uncovered cells with a filled cell on either side. |
| deep_wells | Deep wells: The number of consecutive well segments of depth 3 or more. |
| cuml_wells | Cumulative wells: Weighing each segment of the well heavier as it goes deeper. |
| max_well | Maximum well: The depth of the deepest well. |
| jaggedness | Jaggedness: The perimeter of the top of the pile. |
| col_trans | Column transitions: The number of times a cell changes from open to closed along columns. |
| row_trans | Row transitions: The number of times a cell changes from open to closed along rows. |

6.1 Table of variables used in Principal Component Analysis

| | |
|----------------|---|
| pa row_trans | Row Transitions: Number of times a cell changes from open to closed along rows. |
| pattern_div | Pattern Diversity: Measure comparing the pattern of empty and filled cells in each column, and the same for each row. |
| weighted_cells | Weighted Cells: Count of the total number of filled cells in all columns, each weighted by its own height. |
| landing_height | Landing Height: Height of the bottom of the zoid's final position. |
| matches | Matches: Number of edges of the zoid in its final position that border a filled cell. |
| d_max_ht | Delta Maximum Height: Change in the max_height score after placing the zoid and clearing any filled lines. |
| d_pits | Delta Pits: Change in the pits score after placing the zoid and clearing any filled line |
| | |

References

- (2021), Mmr, rank, and lp. URL <https://support-leagueoflegends.riotgames.com/hc/en-us/articles/4405781372051-MMR-Rank-and-LP>.
- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2015), TensorFlow: Large-scale machine learning on heterogeneous systems. URL <https://www.tensorflow.org/>.
- Ackerman, P. L. (2014), Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence*, 45(1):6–17.
- Adams, J. A. (1987), Historical Review and Appraisal of Research on the Learning, Retention, and Transfer of Human Motor Skills. *Psychological Bulletin*, 101(1):41–74, URL <http://dx.doi.org/10.1037/0033-2909.101.1.41>.
- Agarwal, T., Burghardt, K., & Lerman, K. (2017), On quitting: Performance and practice in online game play. In *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, Icwsm, (pp. 452–455).
- Allen, K. R., Brändle, F., Botvinick, M. M., Fan, J., Gershman, S. J., gopnik, a., Griffiths, T. L., Hartshorne, J. K., Hauser, T. U., Ho, M. K., & et al. (2023), Using games to understand the mind. *PsyArXiv*, URL <http://dx.doi.org/10.31234/osf.io/hbsvj>.
- Amunts, K. (2018), *Brodmann Areas*, (pp. 1–4). Cham: Springer International Publishing, URL http://dx.doi.org/10.1007/978-3-319-16999-6_3341-1.
- Anderson, J. R. (2002), Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, 26(1):85–112.

References

- Anderson, J. R. (2007), *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.
- Anderson, J. R., Betts, S., Bothell, D., Hope, R., & Lebiere, C. (2019), Learning rapid and precise skills. *Psychological Review*, 126(5):727–760.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004), An integrated theory of mind. *Psychological review*, 111(4):1036–1060.
- Anderson, J. R., Bothell, D., Fincham, J. M., Anderson, A. R., Poole, B., & Qin, Y. (2011), Brain regions engaged by part- and whole-task performance in a video game: A model-based test of the decomposition hypothesis. *Journal of Cognitive Neuroscience*, 23(12):3983–3997.
- Anderson, J. R., Bothell, D., Fincham, J. M., & Moon, J. (2016), The sequential structure of brain activation predicts skill. *Neuropsychologia*, (pp. 94–106).
- Ashwood, Z. C., Roy, N. A., Stone, I. R., International Brain Laboratory, Urai, A. E., Churchland, A. K., Pouget, A., & Pillow, J. W. (2022), Mice alternate between discrete strategies during perceptual decision-making. *Nat Neurosci*, 25(2):201–212.
- Aung, M., Bonometti, V., Drachen, A., Cowling, P., Kokkinakis, A. V., & Wade, A. R. (2018), Predicting skill learning in a large, longitudinal MOBA dataset. In *Proceedings of the IEEE Computational Intelligence in Games*, (pp. 1–7).
- Baddeley, D., & Longman, D. J. (1978), The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, 21(8):627–635, URL <http://dx.doi.org/10.1080/00140137808931764>.
- Bagherzadeh, Y., Baldauf, D., Pantazis, D., & Desimone, R. (2020), Alpha synchrony and the neurofeedback control of spatial attention. *Neuron*, 105(3):577–587.e5.
- Baker, J., Côté, J., & Abernethy, B. (2003), Sport-specific practice and the development of expert decision-making in team ball sports. *Journal of Applied Sport Psychology*, 15(1):12–25, URL <http://dx.doi.org/10.1080/10413200305400>.
- Baker, J., & Young, B. (2014), 20 Years Later: Deliberate Practice and the Development of Expertise in Sport. *International Review of Sport and Exercise Psychology*, 7(1):135–157, URL <http://dx.doi.org/10.1080/1750984X.2014.896024>.
- Baldassarre, G., Stafford, T., Mirolli, M., Redgrave, P., Ryan, R., & Barto, A.

- (2014), Intrinsic motivations and open-ended development in animals, humans, and robots: An overview. *Frontiers in psychology*, 5:985, URL <http://dx.doi.org/10.3389/fpsyg.2014.00985>.
- Bányai, F., Griffiths, M. D., Király, O., & Demetrovics, Z. (2019), The Psychology of Esports: A Systematic Literature Review. *Journal of Gambling Studies*.
- Basak, C., Voss, M. W., Erickson, K. I., Boot, W. R., & Kramer, A. F. (2011), Regional differences in brain volume predict the acquisition of skill in a complex real-time strategy videogame. *Brain and Cognition*, 76(3):407–414.
- Bauer, K. N., Brusso, R. C., & Orvis, K. A. (2012), Using adaptive difficulty to optimize videogame-based training performance: The moderating role of personality. *Military Psychology*, 24(2):148–165.
- Bavelier, D., Green, C., Pouget, A., & Schrater, P. (2012), Brain plasticity through the life span: Learning to learn and action video games. *Annual review of neuroscience*, 35:391–416, URL <http://dx.doi.org/10.1146/annurev-neuro-060909-152832>.
- Bell, A. J., & Sejnowski, T. J. (1995), An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6):1129–1159.
- Benjamin, A. S., & Tullis, J. G. (2010), What makes distributed practice effective? *Cognitive Psychology*, 61:228–247.
- Berg, W. K., & Byrd, D. L. (2002), The tower of london spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology*, 24(5):586–604, URL <http://dx.doi.org/10.1076/jcen.24.5.586.1006>.
- Berger, H. (1929), Über das elektrenkephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87:527–570.
- Blumen, H. M., Gopher, D., Steinerman, J. R., & Stern, Y. (2010), Training cognitive control in older adults with the Space Fortress game: The role of training instructions and basic motor ability. *Frontiers in Aging Neuroscience*.
- Bonny, J., Scanlon, M., & Castaneda, L. (2020), Variations in psychological factors and experience-dependent changes in team-based video game performance. *Intelligence*, 80, URL <http://dx.doi.org/10.1016/j.intell.2020.101450>.
- Boot, W. R. (2015), Video games as tools to achieve insight into cognitive pro-

References

- cesses. *Topics in Cognitive Science*, 6(3):1–3, URL <http://dx.doi.org/10.3389/fpsyg.2015.00003>.
- Boot, W. R., Basak, C., Erickson, K. I., Neider, M., Simons, D. J., Fabiani, M., Gratton, G., Voss, M. W., Prakash, R., Lee, H., Low, K. A., & Kramer, A. F. (2010), Transfer of skill engendered by complex task training under conditions of variable priority. *Acta Psychologica*, 135(3):349–357.
- Boot, W. R., Sumner, A., Towne, T. J., Rodriguez, P., & Anders Ericsson, K. (2017), Applying aspects of the expert performance approach to better understand the structure of skill and mechanisms of skill acquisition in video games. *Topics in Cognitive Science*, 9(2):413–436.
- Borst, J. P., & Anderson, J. R. (2013), Using model-based functional mri to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Sciences*, 110(5):1628–1633, URL <http://dx.doi.org/10.1073/pnas.1221572110>.
- Brutsaert, T. D., & Parra, E. J. (2016), Nature versus Nurture in Determining Athletic Ability. *Medicine and Sport Science*, 61:15–28.
- Bryan, L. W., & Harter, N. (1897), Studies in the Physiology and Psychology of the Telegraphic Language. *The Psychological Review*, 4(1):27–53.
- Bryan, L. W., & Harter, N. (1899), Studies on the telegraphic language: The acquisition of a hierarchy of habits. *Psychological Review*, 6(4):345–375.
- Brändle, F., Binz, M., & Schulz, E. (2021), Exploration beyond bandits. *PsyArXiv*, URL <http://dx.doi.org/10.31234/osf.io/9fnmj>.
- Brändle, F., Stocks, L. J., Tenenbaum, J., Gershman, S. J., & Schulz, E. (2022), Intrinsically motivated exploration as empowerment. *PsyArXiv*, URL <http://dx.doi.org/10.31234/osf.io/ybs7g>.
- Bull, R., Espy, K. A., & Senn, T. E. (2004), A comparison of performance on the towers of london and hanoi in young children. *Journal of Child Psychology and Psychiatry*, 45(4):743–754, URL <http://dx.doi.org/https://doi.org/10.1111/j.1469-7610.2004.00268.x>.
- Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016), The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence*, 59:72–83.
- Calhoun, A., Pillow, J., & Murthy, M. (2019), Unsupervised identification of

- the internal states that shape natural behavior. *Nature Neuroscience*, 22:1–10, URL <http://dx.doi.org/10.1038/s41593-019-0533-x>.
- Campitelli, G., Connors, M. H., Bilalic, M., & Hambrick, D. Z. (2015), Psychological perspectives on expertise. *Frontiers in Psychology*, 6(MAR):1–4.
- Campitelli, G., & Gobet, F. (2011), Deliberate practice: Necessary but not sufficient. *Current Directions in Psychological Science*, 20(5):280–285.
- Castaneda, L., Sidhu, M. K., Azose, J. J., & Swanson, T. (2016), Game play differences by expertise level in Dota 2, a complex multiplayer video game. *Journal of Gaming and Computer-Mediated Simulations*, 8(4):1–24.
- Cavadenti, O., Codocedo, V., Boulicaut, J.-F., & Kaytoue, M. (2016), What did i do wrong in my moba game? mining patterns discriminating deviant behaviours. In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, (pp. 662–671).
- Chabris, C. F. (2017), Six Suggestions for Research on Games in Cognitive Science. *Topics in Cognitive Science*, 9(2):497–509.
- Chang, C.-C., & Lin, C.-J. (2011), Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27.
- Charness, N. (2017), What Has the Study of Digital Games Contributed to the Science of Expert Behavior? *Topics in Cognitive Science*, 9(2):510–521, URL <http://dx.doi.org/10.1111/tops.12259>.
- Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005), The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, 19(2):151–165.
- Chase, W. G., & Simon, H. A. (1973), Perception in chess. *Cognitive Psychology*, 4(1):55–81.
- Chen, Z. (2015), Estimating latent attentional states based on simultaneous binary and continuous behavioral measures. *Comput Intell Neurosci*, 2015:493769.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981), Categorization and representation of physics problems by experts and novices*. *Cognitive Science*, 5(2):121–152, URL http://dx.doi.org/https://doi.org/10.1207/s15516709cog0502_2.
- Chollet, F., et al. (2015), Keras. URL <https://github.com/fchollet/keras>.

References

- Cohen, M., & Maunsell, J. (2009), Attention improves performance primarily by reducing interneuronal correlations. *Nature neuroscience*, 12:1594–600, URL <http://dx.doi.org/10.1038/nn.2439>.
- Coughlan, G., Coutrot, A., Khondoker, M., Minihane, A.-M., Spiers, H., & Hornberger, M. (2019), Toward personalized cognitive diagnostics of at-genetic-risk alzheimer's disease. *Proceedings of the National Academy of Sciences*, 116(19):9285–9292, URL <http://dx.doi.org/10.1073/pnas.1901600116>.
- Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R., Wiener, J. M., Hölscher, C., Hornberger, M., & Spiers, H. J. (2022), Entropy of city street networks linked to future spatial navigation ability. *bioRxiv*, URL <http://dx.doi.org/10.1101/2020.01.23.917211>.
- Coutrot, A., Silva, R., Manley, E., Cothi, W., Sami, S., Bohbot, V., Wiener, J., Hölscher, C., Dalton, R., Hornberger, M., & Spiers, H. (2018), Global determinants of navigation ability. *Current Biology*, 28, URL <http://dx.doi.org/10.1016/j.cub.2018.06.009>.
- Csikszentmihalyi, M. (1991), *Flow: The Psychology of Optimal Experience*. New York, NY: Harper Perennial.
- Dale, A., Fischl, B., & Sereno, M. I. (1999), Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194.
- Deakin, J., & Cobley, S. (2003), A search for deliberate practice: An examination of the practice environments in figure skating and volleyball. In J. Starkes, & K. A. Ericsson (Eds.), *Expert performance in sport: Recent advances in research on sport expertise*, (pp. 115–135), Champaign, IL: Human Kinetics.
- Deary, I., Liewald, D. D., & Nissan, J. (2010), A free, easy-to-use, computer-based simple and four-choice reaction time programme: The deary-liewald reaction time task. *Behavior research methods*, 43:258–68, URL <http://dx.doi.org/10.3758/s13428-010-0024-1>.
- Demediuk, S., York, P., Drachen, A., Walker, J. A., & Block, F. (2019), Role identification for accurate analysis in dota 2. In *AIIDE*.
- Den Hartigh, R., van der Steen, S., Meij, M., Van Yperen, N., Gernigon, C., & van geert, P. (2014), Characterising expert representations during real-time action: A skill theory application to soccer. *Journal of Cognitive Psychology*, 26, URL <http://dx.doi.org/10.1080/20445911.2014.955028>.
- Den Hartigh, R. J., Van Dijk, M. W., Steenbeek, H. W., & Van Geert, P. L.

- (2016), A dynamic network model to explain the development of excellent human performance. *Frontiers in Psychology*, 7(532).
- Denga, R. (2021), Need for speed: Applying ex-gaussian modeling techniques to examine intra-individual reaction time variability in expert tetris players. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43.
- Destefano, M., & Gray, W. (2016), Where should researchers look for strategy discoveries during the acquisition of complex task performance? the case of space fortress. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.
- Destefano, M., & Gray, W. D. (2008), *Progress report on Pygame Space Fortress*. Troy, NY: Rensselaer Polytechnic Institute.
- Donchin, E. (1995), Video games as research tools: The space fortress game. *Behavior Research Methods, Instruments, & Computers*, 27:217–223, URL <http://dx.doi.org/10.3758/BF03204735>.
- Donovan, J. J., & Radosevich, D. J. (1999), A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, 84(5):795–805, URL <http://dx.doi.org/10.1037/0021-9010.84.5.795>.
- Dotabuff (2023), About Dotabuff. URL <https://www.dotabuff.com/pages/about>.
- Drachen, A., Yancey, M., Maguire, J., Chu, D., Wang, I. Y., Mahlmann, T., Schubert, M., & Klabajan, D. (2014), Skill-Based Differences in Spatio-Temporal Team Behaviour in Defence of The Ancients 2 (DotA 2). In *2014 IEEE Games, Media, Entertainment (gem)*, volume 2, (pp. 1–8), IEEE Consumer Elect Soc.
- Ebbinghaus, H. (1885/1964), *Memory: A contribution to experimental psychology*. Oxford, England: Dover.
- Eccles, D. W., & Arsal, G. (2017), The think aloud method: what is it and how do i use it? *Qualitative Research in Sport, Exercise and Health*, 9(4):514–531, URL <http://dx.doi.org/10.1080/2159676X.2017.1331501>.
- Erickson, K. I., Boot, W. R., Basak, C., Neider, M. B., Prakash, R. S., Voss, M. W., Graybiel, A. M., Simons, D. J., Fabiani, M., Gratton, G., & Kramer, A. F. (2010), Striatal volume predicts level of video game skill acquisition. *Cerebral Cortex*, 20(11):2522–2530.
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993), The Role of Deliber-

References

- ate Practice in the Acquisition of Expert Performance. *Psychological Review*, 100(3):363–406, URL <http://dx.doi.org/10.1037//0033-295x.100.3.363>.
- Ericsson, K. A., & Lehmann, A. C. (1996), Expert and exceptional performance: evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, 47:273–305, URL <http://dx.doi.org/10.1146/annurev.psych.47.1.273>.
- Ericsson, K. A., & Smith, J. (Eds.) (1991), *Toward a general theory of expertise: Prospects and limits*. New York, NY, US: Cambridge University Press.
- Esguerra, T. (2021), Worlds 2021 hits record-breaking 4 million peak viewership during finals between edg and dwg kia. URL <https://dotesports.com/league-of-legends/news/worlds-2021-hits-record-breaking-4-million-peak-viewership-during-finals-between-edg-and-dwg-kia>.
- Eyolfjöldottir, E., Branson, K., Yue, Y., & Perona, P. (2017), Learning recurrent representations for hierarchical behavior modeling. In *International Conference on Learning Representations*.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., & Dale, A. M. (2004), Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, 14(1):11–22, URL <http://dx.doi.org/10.1093/cercor/bhg087>.
- Fitts, P. M., & Posner, M. I. (1967), *Human performance*. Brooks/Cole Pub. Co.
- Foster, J. J., & Awh, E. (2019), The role of alpha oscillations in spatial attention: limited evidence for a suppression account. *Current Opinion in Psychology*, 29:34–40, URL <http://dx.doi.org/https://doi.org/10.1016/j.copsyc.2018.11.001>.
- Foster, J. J., Sutterer, D. W., Serences, J. T., Vogel, E. K., & Awh, E. (2017), Alpha-band oscillations enable spatially and temporally resolved tracking of covert spatial attention. *Psychological Science*, 28(7):929–941, URL <http://dx.doi.org/10.1177/0956797617699167>.
- Foxe, J., & Snyder, A. (2011), The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, 2, URL <http://dx.doi.org/10.3389/fpsyg.2011.00154>.
- Frerejean, J., Velthorst, G. J., van Strien, J. L., Kirschner, P. A., & Brand-Gruwel, S. (2019), Embedded instruction to learn information problem solving:

- Effects of a whole task approach. *Computers in Human Behavior*, 90:117–130, URL <http://dx.doi.org/https://doi.org/10.1016/j.chb.2018.08.043>.
- Furlough, C. S., & Gillan, D. J. (2018), Mental models: Structural differences and the role of experience. *Journal of Cognitive Engineering and*, 12(4):269–287.
- Fuster, J. M. (2004), Upper processing stages of the perception–action cycle. *Trends in Cognitive Sciences*, 8(4):143–145, URL <http://dx.doi.org/https://doi.org/10.1016/j.tics.2004.02.004>.
- Gagné, F. (2004), Transforming gifts into talents: The DMGT as a developmental theory. *High Ability Studies*, 15(2):119–147.
- Galton, F. (1869), *Hereditary genius: An inquiry into its laws and consequences*. Macmillan and Co.
- Gaschler, R., Progscha, J., Smallbone, K., Ram, N., & Bilalic, M. (2014), Playing off the curve - testing quantitative predictions of skill acquisition theories in development of chess performance. *Frontiers in Psychology*, 5(AUG):1–11, URL <http://dx.doi.org/10.3389/fpsyg.2014.00923>.
- Gegenfurtner, A., & Seppänen, M. (2013), Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, 63:393–403, URL <http://dx.doi.org/https://doi.org/10.1016/j.compedu.2012.12.021>.
- Georgiades, E., Klissouras, V., Baulch, J., Wang, G., & Pitsiladis, Y. (2017), Why nature prevails over nurture in the making of the elite athlete. *BMC Genomics*, 18(Suppl 8):59–66.
- Gladwell, M. (2008), *Outliers: the story of success*. New York: Little, Brown and Company.
- Gobet, F. (2017), Allen newell’s program of research: The video-game test. *Topics in Cognitive Science*, 9(2):522–532, URL <http://dx.doi.org/https://doi.org/10.1111/tops.12265>.
- Gobet, F., & Campitelli, G. (2007), The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology*, 43(1):159–172.
- Goldstone, R. L., & Lupyan, G. (2016), Discovering Psychological Principles by Mining Naturally Occurring Data Sets. *Topics in Cognitive Science*, 8(3):548–568, URL <http://dx.doi.org/10.1111/tops.12212>.

References

- Goodfellow, I. J., Bengio, Y., & Courville, A. (2016), *Deep Learning*. Cambridge, MA, USA: MIT Press.
- Gopher, D., Weil, M., & Siegel, D. (1989), Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica*, 71:147–177, URL [http://dx.doi.org/10.1016/0001-6918\(89\)90007-3](http://dx.doi.org/10.1016/0001-6918(89)90007-3).
- Grabner, R. H. (2014), The role of intelligence for performance in the prototypical expertise domain of chess. *Intelligence*, 45(1):26–33.
- Gray, W. (2017), Game-xp: Action games as experimental paradigms for cognitive science. *Topics in Cognitive Science*, 9, URL <http://dx.doi.org/10.1111/tops.12260>.
- Gray, W. D., & Lindstedt, J. K. (2017), Plateaus, dips, and leaps: Where to look for inventions and discoveries during skilled performance. *Cognitive Science*, 41(7):1838–1870.
- Green, C., & Bavelier, D. (2015), Action video game training for cognitive enhancement. *Current Opinion in Behavioral Sciences*, 8, URL <http://dx.doi.org/10.1016/j.cobeha.2015.04.012>.
- de Groot, A. D. (1978), *Thought and choice in chess*, 2nd ed. Oxford, England: Mouton.
- Haider, H., & Frensch, P. A. (2002), Why Aggregated Learning Follows the Power Law of Practice When Individual Learning Does Not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning Memory and Cognition*, 28(2):392–406, URL <http://dx.doi.org/10.1037/0278-7393.28.2.392>.
- Hämäläinen, M. S., & Ilmoniemi, R. J. (1994), Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & Biological Engineering & Computing*, 32(1):35–42.
- Hambrick, D. Z., Altmann, E. M., Oswald, F. L., Meinz, E. J., & Gobet, F. (2014a), Facing facts about deliberate practice. *Frontiers in Psychology*, 5(JUL):10–11.
- Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014b), Deliberate practice: Is that all it takes to become an expert? *Intelligence*, 45(1):34–45.
- Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014c), Deliberate practice: Is that all it takes to become

- an expert? *Intelligence*, 45(1):34–45, URL <http://dx.doi.org/10.1016/j.intell.2013.04.001>.
- Hamilton, W. G. (1986), Physical prerequisites for ballet dancers. *The Journal of Musculoskeletal Medicine*, 3:61–66.
- Harwell, K. W., Boot, W. R., & Ericsson, K. A. (2018), Looking behind the score: Skill structure explains sex differences in skilled video game performance. *PLoS ONE*, 13(5):1–18, URL <http://dx.doi.org/10.1371/journal.pone.0197311>.
- Heathcote, A., Brown, S., & Mewhort, D. J. (2000), The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review*, 7(2):185–207, URL <http://dx.doi.org/10.3758/BF03212979>.
- Heinemann, L., Kleinschmidt, A., & Müller, N. G. (2009), Exploring bold changes during spatial attention in non-stimulated visual cortex. *PLOS ONE*, 4(5):1–8, URL <http://dx.doi.org/10.1371/journal.pone.0005560>.
- Heitz, R. P. (2014), The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, 8, URL <http://dx.doi.org/10.3389/fnins.2014.00150>.
- Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998), Team Sports and the Theory of Deliberate Practice. *Journal of Sport and Exercise Psychology*, 20:12–34.
- Herbrich, R., Minka, T., & Graepel, T. (2007), TrueSkill: A Bayesian Skill Rating System. *Advances in Neural Information Processing Systems*, URL <http://papers.nips.cc/paper/3079-trueskilltm-a-bayesian-skill-rating-system.pdf>.
- van den Heuvel, M. P., Mandl, R. C., Kahn, R. S., & Hulshoff Pol, H. E. (2009), Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping*, 30(10):3127–3141.
- HMMLEARN. (2022), hmmlearn. URL <https://hmmlearn.readthedocs.io/en/latest/>.
- Hochreiter, S., & Schmidhuber, J. (1997), Long short-term memory. *Neural Computation*, 9(8):1735–1780.
- Hodges, N. J., Starkes, J. L., Nananidou, A., Kerr, T., & Weir, P. L. (2004), Predicting performance times from deliberate practice hours for triathletes and

References

- swimmers: What, when, and where is practice important? *Journal of Experimental Psychology: Applied*, 10(4):219–237.
- Howe, M. J. A., Davidson, J. W., & Sloboda, J. A. (1998), Natural born talents undiscovered. *Behavioral and Brain Sciences*, 21(3):432–437.
- Huang, J., Yan, E., Cheung, G., Nagappan, N., & Zimmermann, T. (2017), Master Maker: Understanding Gaming Skill Through Practice and Habit From Gameplay Behavior. *Topics in Cognitive Science*, 9(2):437–466, URL <http://dx.doi.org/10.1111/tops.12251>.
- Huang, J., Zimmermann, T., Nagappan, N., Harrison, C., & Phillips, B. (2013), Mastering the art of war: How patterns of gameplay influence skill in halo. *Conference on Human Factors in Computing Systems - Proceedings*, (pp. 695–704), URL <http://dx.doi.org/10.1145/2470654.2470753>.
- Hughes, M., Day, E., Wang, X., Schuelke, M., Arsenault, M., Harkrider, L., & Cooper, O. (2013), Learner-controlled practice difficulty in the training of a complex task: Cognitive and motivational mechanisms. *Journal of Applied Psychology*, 98(1):80–98, URL <http://dx.doi.org/10.1037/a0029821>.
- Johanson, C., Gutwin, C., Bowey, J. T., & Mandryk, R. L. (2019), Press Pause when you Play: Comparing Spaced Practice Intervals for Skill Development in Games. In *CHI Play'19: Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, (pp. 169–184), ACM SIGCHI; Assoc Comp Machinery, URL <http://dx.doi.org/10.1145/3311350.3347195>.
- Johanson, C., & Mandryk, R. (2016), Scaffolding player location awareness through audio cues in first-person shooters. In *Conference on Human Factors in Computing Systems - Proceedings*, (pp. 3450–3461), URL <http://dx.doi.org/10.1145/2858036.2858172>.
- Johnston, K., Wattie, N., Schorer, J., & Baker, J. (2018), Talent Identification in Sport: A Systematic Review. *Sports Medicine*, 48(1):97–109.
- Karapanagiotidis, T., Vidaurre, D., Quinn, A., Vatansever, D., Poerio, G., Turnbull, A., Ho, N., Leech, R., Bernhardt, B., Jefferies, E., Margulies, D., Nichols, T., Woolrich, M., & Smallwood, J. (2020), The psychological correlates of distinct neural states occurring during wakeful rest. *Scientific Reports*, 10(1), URL <http://dx.doi.org/10.1038/s41598-020-77336-z>.
- Kenreck, T. (2012), “League of Legends” players log 1 billion hours a month. URL <https://www.nbcnews.com/tech/tech-news/league-legends-players-log-1-billion-hours-month-flna1c6423906>.

- Kim, J., Keegan, B. C., Park, S., & Oh, A. (2016), The Proficiency-Congruency Dilemma: Virtual Team Design and Performance in Multiplayer Online Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, (pp. 4351–4365), URL <http://dx.doi.org/10.1145/2858036.2858464>.
- Kim, Y. J., Engel, D., Woolley, A. W., Lin, J. Y.-T., McArthur, N., & Malone, T. W. (2017), What Makes a Strong Team? In *Proc. 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, (pp. 2316–2329), URL <http://dx.doi.org/10.1145/2998181.2998185>.
- Kingma, D. P., & Ba, J. (2014), Adam: A method for stochastic optimization.
- Kokkinakis, A. V., Cowling, P. I., Drachen, A., & Wade, A. R. (2017), Exploring the relationship between video game expertise and fluid intelligence. *PLoS ONE*, 12(11):1–15, URL <http://dx.doi.org/10.1371/journal.pone.0186621>.
- Kunkel, D., Yan, Z., Craigmile, P. F., Peruggia, M., & Zandt, T. V. (2020), Hierarchical hidden markov models for response time data. *Computational Brain & Behavior*, 4:70–86.
- Kurtz, S., & Lee, T. (2003), Part and whole perceptual-motor practice of a polyrhythm. *Neuroscience letters*, 338:205–8, URL [http://dx.doi.org/10.1016/S0304-3940\(02\)01394-0](http://dx.doi.org/10.1016/S0304-3940(02)01394-0).
- Landfried, G., Slezak, D. F., & Mocskos, E. (2019), Faithfulness-boost effect: Loyal teammate selection correlates with skill acquisition improvement in online games. *PLoS ONE*, 14(3).
- Large, A., Bediou, B., Cekic, S., Hart, Y., Bavelier, D., & Green, C. (2019), Cognitive and behavioral correlates of achievement in a complex multi-player video game. *Media and Communication*, 7(4):198–212, URL <http://dx.doi.org/10.17645/mac.v7i4.2314>.
- Laserface (2022), Matchmaking Guide. URL <https://support-leagueoflegends.riotgames.com/hc/en-us/articles/201752954-Matchmaking-Guide>.
- Lee, H., Boot, W. R., Baniqued, P. L., Voss, M. W., Prakash, R. S., Basak, C., & Kramer, A. F. (2015), The relationship between intelligence and training gains is moderated by training strategy. *PLoS ONE*, 10(4).
- Lee, H., Voss, M. W., Prakash, R. S., Boot, W. R., Vo, L. T. K., Basak, C., VanPatter, M., Gratton, G., Fabiani, M., & Kramer, A. F. (2012), Videogame

References

- training strategy-induced change in brain function during a complex visuomotor task. *Behavioural Brain Research*, 232(2):348–357.
- Lee, T. D., & Genovese, E. D. (1988), Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, 59(4):277–287, URL <http://dx.doi.org/10.1080/02701367.1988.10609373>.
- Li, K. T., Liang, J., & Zhou, C. (2021), Gamma oscillations facilitate effective learning in excitatory-inhibitory balanced neural circuits. *Neural Plasticity*, 2021:6668175, URL <http://dx.doi.org/10.1155/2021/6668175>.
- Li, X., Huang, L., Li, B., Wang, H., & Han, C. (2020), Time for a true display of skill: Top players in League of Legends have better executive control. *Acta Psychologica*.
- Liang, P., Jia, X., Taatgen, N., Borst, J. P., & Li, K. (2016), Activity in the frontoparietal network indicates numerical inductive reasoning beyond calculation: An fmri study combined with a cognitive model. *Scientific reports*, 6:25976.
- Lim, H., & Yen, J. (2004), Preliminary study of attention control modeling in complex skill training environments. *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, 3060:120–132, URL http://dx.doi.org/10.1007/978-3-540-24840-8_9.
- Lindstedt, J., & Gray, W. (2015), Meta-t: Tetris as an experimental paradigm for cognitive skills research. *Behavior research methods*, 47.
- Lindstedt, J., & Gray, W. (2019), Distinguishing experts from novices by the mind's hand and mind's eye. *Cognitive Psychology*, 109:1–25, URL <http://dx.doi.org/10.1016/j.cogpsych.2018.11.003>.
- Lindstedt, J. K., & Gray, W. D. (2013), Extreme expertise: Exploring expert behavior in tetris. *Cognitive Science*, 35.
- van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., & Forstmann, B. U. (2011), Neural correlates of trial-to-trial fluctuations in response caution. *Journal of Neuroscience*, 31(48):17488–17495, URL <http://dx.doi.org/10.1523/JNEUROSCI.2924-11.2011>.
- Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014), Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions: A Meta-Analysis. *Psychological Science*, 25(8):1608–1618.
- Macnamara, B. N., & Maitra, M. (2019), The role of deliberate practice in ex-

- pert performance: revisiting Ericsson, Krampe & Tesch-Romer (1993). *Royal Society Open Science*, 6(190327):190327, URL <http://dx.doi.org/10.1098/rsos.190327>.
- Magill, R., & Anderson, D. (2017), *Motor learning and control: concepts and applications*. New York, NY: McGraw-Hill Education.
- Mane, A., & Donchin, E. (1989), The space fortress game. *Acta Psychologica*, 71:17–22.
- Matuszewski, P., Dobrowolski, P., & Zawadzki, B. (2020), The Association Between Personality Traits and eSports Performance. *Frontiers in Psychology*, 11, URL <http://dx.doi.org/10.3389/fpsyg.2020.01490>.
- McAndrew, C., & Gore, J. (2013), Understanding preferences in experience-based choice: A study of cognition in the “wild”. *Journal of Cognitive Engineering and Decision Making*, 7(2):179–197, URL <http://dx.doi.org/10.1177/1555343412463922>.
- McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018), Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, URL <http://dx.doi.org/10.21105/joss.00861>.
- Wes McKinney (2010), Data Structures for Statistical Computing in Python. In Stéfan van der Walt, & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference*, (pp. 56–61), URL <http://dx.doi.org/10.25080/Majora-92bf1922-00a>.
- Mitchell, J., Sundberg, K., & Reynolds, J. (2009), Spatial attention decorrelates intrinsic activity fluctuations in macaque area v4. *Neuron*, 63:879–88, URL <http://dx.doi.org/10.1016/j.neuron.2009.09.013>.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009), Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *BMJ*, 339(7):e1000097, URL <http://dx.doi.org/10.1136/bmj.b2535>.
- Moll, P., Frick, V., Rauscher, N., & Lux, M. (2020), How players play games: Observing the influences of game mechanics. In *MMVE 2020 - Proceedings of the 2020 International Workshop on Immersive Mixed and Virtual Environment Systems - Part of MMSys 2020*, (pp. 7–12), URL <http://dx.doi.org/10.1145/3386293.3397113>.
- Mora-Cantalops, M., & Ángel Sicilia, M. (2018), Moba games: A literature review. *Entertainment Computing*, 26:128–138, URL <http://dx.doi.org/https://doi.org/10.1016/j.entcom.2018.02.005>.

References

- Murphy, G. L., & Wright, J. (1984), Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10:144–155.
- Nagorsky, E., & Wiemeyer, J. (2020), The structure of performance and training in esports. *PLoS ONE*, 15(8 August 2020), URL <http://dx.doi.org/10.1371/journal.pone.0237584>.
- Nyhus, E., & Curran, T. (2010), Functional role of gamma and theta oscillations in episodic memory. *Neurosci Biobehav Rev*, 34(7):1023–1035.
- O’Callaghan, J. (2018), The bizarre history of ’tetris. *Live Science*, URL <https://www.livescience.com/63500-tetris-history.html>.
- van Opheusden, B., Galbiati, G., Kuperwajs, I., Bnaya, Z., li, Y., & Ma, W. J. (2021), Revealing the impact of expertise on human planning with a two-player board game. *PsyArXiv*, URL <http://dx.doi.org/10.31234/osf.io/rhq5j>.
- Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006), Theta and gamma oscillations predict encoding and retrieval of declarative memory. *Journal of Neuroscience*, 26(28):7523–7531, URL <http://dx.doi.org/10.1523/JNEUROSCI.1948-06.2006>.
- Palva, S., & Palva, J. M. (2011), Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Frontiers in psychology*, 2:204, URL <http://dx.doi.org/10.3389/fpsyg.2011.00204>.
- Pedraza-Ramirez, I., Musculus, L., Raab, M., & Laborde, S. (2020), Setting the scientific stage for esports psychology: a systematic review. *International Review of Sport and Exercise Psychology*, 13(1):319–352, URL <http://dx.doi.org/10.1080/1750984X.2020.1723122>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011), Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Peylo, C., Hilla, Y., & Sheineman2009, P., auseng (2021), Cause or consequence? alpha oscillations in visuospatial attention. *Trends in Neurosciences*, 44(9):705–713, URL <http://dx.doi.org/10.1016/j.tins.2021.05.004>.
- Pfurtscheller, G. (2003), Induced oscillations in the alpha band: Functional meaning. *Epilepsia*, 44(s12):2–8, URL <http://dx.doi.org/https://doi.org/10.1111/j.0013-9580.2003.12001.x>.

- Phy, E. (2023), IMP Decoding Your Performance. URL <https://medium.com/stratz/imp-decoding-your-performance-c251dcb42b93>.
- Piller, B., Johanson, C., Phillips, C., Gutwin, C., & Mandryk, R. L. (2020), Is a change as good as a rest? comparing breaktypes for spaced practice in a platformer game. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '20*, (p. 294–305), New York, NY, USA: Association for Computing Machinery, URL <http://dx.doi.org/10.1145/3410404.3414225>.
- Pluss, M. A., Bennett, K. J. M., Novak, A. R., Panchuk, D., Coutts, A. J., & Fransen, J. (2019), Esports: The Chess of the 21st Century. *Frontiers in Psychology*, 10:156.
- Poldrack, R. A., Sabb, F. W., Foerde, K., Tom, S. M., Asarnow, R. F., Bookheimer, S. Y., & Knowlton, B. J. (2005), The neural correlates of motor skill automaticity. *Journal of Neuroscience*, 25(22):5356–5364.
- Popescu, A. T., Popa, D., & Paré, D. (2009), Coherent gamma oscillations couple the amygdala and striatum during learning. *Nature Neuroscience*, 12(6):801–807, URL <http://dx.doi.org/10.1038/nn.2305>.
- Prakash, R. S., De Leon, A. A., Mourany, L., Lee, H., Voss, M., Boot, W. R., Basak, C., Fabiani, M., Gratton, G., & Kramer, A. F. (2012), Examining neural correlates of skill acquisition in a complex videogame training program. *Frontiers in Human Neuroscience*.
- Rahman, R., & Gray, W. D. (2020), Spotlight on dynamics of individual learning. *Topics in Cognitive Science*, 12(3):975–991, URL <http://dx.doi.org/10.1111/tops.12512>.
- Reggente, N., Essoe, J., Aghajan, Z., Tavakoli, A., McGuire, J., Suthana, N., & Rissman, J. (2018), Enhancing the ecological validity of fmri memory research using virtual reality. *Frontiers in Neuroscience*, 12(JUN), URL <http://dx.doi.org/10.3389/fnins.2018.00408>.
- Renart, A., & Machens, C. K. (2014), Variability in neural activity and behavior. *Curr Opin Neurobiol*, 25:211–220.
- Richels, K., Day, E., Jorgensen, A., & Huck, J. (2020), Keeping Calm and Carrying On: Relating Affect Spin and Pulse to Complex Skill Acquisition and Adaptive Performance. *Frontiers in Psychology*, 11, URL <http://dx.doi.org/10.3389/fpsyg.2020.00377>.

References

- Rohlcke, S., Backlund, C., Sorman, D. E., & Jonsson, B. (2018), Time on task matters most in video game expertise. *PLoS ONE*, 13(10):e0206555.
- Roy, N. A., Bak, J. H., Akrami, A., Brody, C. D., & Pillow, J. W. (2021), Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron*, 109(4):597–610.e6, URL <http://dx.doi.org/https://doi.org/10.1016/j.neuron.2020.12.004>.
- Salmon, P. (2003), How do we recognise good research? *The Psychologist*, 16(1):24–27.
- Samson, A., Simpson, D., Kamphoff, C., & Langlier, A. (2017), Think aloud: An examination of distance runners' thought processes. *International Journal of Sport and Exercise Psychology*, 15(2):176–189, URL <http://dx.doi.org/10.1080/1612197X.2015.1069877>.
- Sangster, M.-D., Mendonca, D., & Gray, W. (2016), Big data meets team expertise in a dynamic task environment. In *Proceedings of the Human Factors and Ergonomics Society*, (pp. 158–162), URL <http://dx.doi.org/10.1177/1541931213601036>.
- Sapienza, A., Zeng, Y., Bessi, A., Lerman, K., & Ferrara, E. (2018), Individual performance in team-based online games. *Royal Society Open Science*, 5(6), URL <http://dx.doi.org/10.1098/rsos.180329>.
- Satopaa, V., Albrecht, J. R., Irwin, D. E., & Raghavan, B. (2011), Finding a "kneedle" in a haystack: Detecting knee points in system behavior. *2011 31st International Conference on Distributed Computing Systems Workshops*, (pp. 166–171).
- Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W. R., & Birbaumer, N. (2005), A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur J Neurosci*, 22(11):2917–2926.
- Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007), Utilization of the pico framework to improve searching pubmed for clinical questions. *BMC Medical Informatics and Decision Making*, 7(1):16, URL <http://dx.doi.org/10.1186/1472-6947-7-16>.
- Schmidt, R. A., & Lee, T. D. (2019), *Motor control and learning: A behavioral emphasis, 6th ed.* Champaign, IL, US: Human Kinetics.
- Sculley, D. (2010), Web-scale k-means clustering. In *Proceedings of the 19th International Conference on World Wide Web*, (p. 1177–1178), New York, NY,

- USA: Association for Computing Machinery, URL <http://dx.doi.org/10.1145/1772690.1772862>.
- Seabold, S., & Perktold, J. (2010), statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*.
- Segal, D. (2014), Behind League of Legends, E-Sports's Main Attraction. URL <https://www.nytimes.com/2014/10/12/technology/riot-games-league-of-legends-main-attraction-esports.html>.
- Seo, Y. (2016), Professionalized consumption and identity transformations in the field of eSports. *Journal of Business Research*, 69(1):264–272, URL <http://dx.doi.org/10.1016/j.jbusres.2015.07.039>.
- Seymour, R., Alexander, N., Mellor, S., O'Neill, G., Tierney, T., Barnes, G., & Maguire, E. (2021), Interference suppression techniques for opm-based meg: Opportunities and challenges. *NeuroImage*, 247:118834.
- Shapiro, K., & Raymond, J. (1989), Training of efficient oculomotor strategies enhances skill acquisition. *Acta Psychologica*, 71(1-3):217–242.
- Sharma, A., Johnson, R., Engert, F., & Linderman, S. (2018), Point process latent variable models of larval zebrafish behavior. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 31, Curran Associates, Inc.
- Sibert, C., & Gray, W. (2018), The tortoise and the hare: Understanding the influence of sequence length and variability on decision-making in skilled performance. *Computational Brain & Behavior*, 1, URL <http://dx.doi.org/10.1007/s42113-018-0014-4>.
- Sibert, C., Gray, W., & Lindstedt, J. (2017), Interrogating feature learning models to discover insights into the development of human expertise in a real-time, dynamic decision-making task. *Topics in cognitive science*, 9(2), URL <http://dx.doi.org/10.1111/tops.12225>.
- Simonton, D. K. (1999), Talent and Its Development: An Emergent and Epigenetic Model. *Psychological Review*, 106(3):435–457.
- Simonton, D. K. (2014), Creative performance, expertise acquisition, individual differences, and developmental antecedents: An integrative research agenda. *Intelligence*, 45(1):66–73.
- Smolen, P., Zhang, Y., & Byrne, J. H. (2016), The right time to learn: Mech-

References

- anisms and optimization of spaced learning. *Nature Reviews Neuroscience*, 17(2):77–88, URL <http://dx.doi.org/10.1038/nrn.2015.18>.
- Sonkusare, S., Breakspear, M., & Guo, C. (2019), Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends in Cognitive Sciences*, 23, URL <http://dx.doi.org/10.1016/j.tics.2019.05.004>.
- Sporns, O., Tononi, G., & Kötter, R. (2005), The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1(4):null, URL <http://dx.doi.org/10.1371/journal.pcbi.0010042>.
- Stafford, T., Devlin, S., Sifa, R., & Drachen, A. (2017), Exploration and Skill Acquisition in a Major Online Game. *The 39th Annual Meeting of the Cognitive Science Society*, 9(2):485–496.
- 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, URL <http://dx.doi.org/10.1177/0956797613511466>.
- Stafford, T., & Haasnoot, E. (2017), Testing sleep consolidation in skill learning: a field study using an online game. *Topics in Cognitive Science*, 9(2):485–496.
- Stafford, T., & Vaci, N. (2022), Maximizing the potential of digital games for understanding skill acquisition. *Current Directions in Psychological Science*, 31(1):49–55, URL <http://dx.doi.org/10.1177/09637214211057841>.
- Sundberg, K. A., Mitchell, J. F., Gawne, T. J., & Reynolds, J. H. (2012), Attention influences single unit and local field potential response latencies in visual cortical area V4. *J Neurosci*, 32(45):16040–16050.
- Swanson, R. A., & Law, B. D. (1993), Whole-part-whole learning model. *Performance Improvement Quarterly*, 6(1):43–53, URL <http://dx.doi.org/https://doi.org/10.1111/j.1937-8327.1993.tb00572.x>.
- Sweet, J. (2021), The revolution in classic tetris. *The New Yorker*, URL <https://www.newyorker.com/culture/cultural-comment/the-revolution-in-classic-tetris>.
- Tadel, F., Bock, E., Niso, G., Mosher, J. C., Cousineau, M., Pantazis, D., Leahy, R. M., & Baillet, S. (2019), Meg/eeg group analysis with brainstorm. *Frontiers in Neuroscience*, 13, URL <http://dx.doi.org/10.3389/fnins.2019.00076>.
- Tanaka, S., Ikeda, H., Kasahara, K., Kato, R., Tsubomi, H., Sugawara, S., Mori, M., Hanakawa, T., Sadato, N., Honda, M., & Watanabe, K. (2013), Larger Right Posterior Parietal Volume in Action Video Game Experts: A Behavioral

- and Voxel-Based Morphometry (VBM) Study. *PLoS ONE*, 8(6), URL <http://dx.doi.org/10.1371/journal.pone.0066998>.
- Tarantola, A. (2020), How gen z is pushing nes tetris to its limits. *Engadget*, URL <https://www.engadget.com/teens-tetris-and-rolling-130002013.html>.
- pandas development team, T. (2020), pandas-dev/pandas: Pandas 1.0.3. URL <http://dx.doi.org/10.5281/zenodo.3715232>.
- Tenison, C., & Anderson, J. (2016), Modeling the distinct phases of skill acquisition. *Journal of experimental psychology. Learning, memory, and cognition*, 42:749–767, URL <http://dx.doi.org/10.1037/xlm0000204>.
- Tenison, C., & Anderson, J. R. (2017), The impact of practice frequency on learning and retention. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, (pp. 1175–1180).
- Thompson, J., Blair, M., Chen, L., & Henrey, A. (2013), Video Game Telemetry as a Critical Tool in the Study of Complex Skill Learning. *PLoS ONE*, 8(9):1–12, URL <http://dx.doi.org/10.1371/journal.pone.0075129>.
- Thompson, J., Blair, M., & Henrey, A. (2014), Over the hill at 24: Persistent age-related cognitive-motor decline in reaction times in an ecologically valid video game task begins in early adulthood. *PLoS ONE*, 9(4):e94215, URL <http://dx.doi.org/10.1371/journal.pone.0094215>.
- Thompson, J. J., McColeman, C. M., Blair, M. R., & Henrey, A. J. (2019), Classic motor chunking theory fails to account for behavioural diversity and speed in a complex naturalistic task. *PLOS ONE*, 14(6):1–24, URL <http://dx.doi.org/10.1371/journal.pone.0218251>.
- Thompson, J. J., McColeman, C. M., Stepanova, E. R., & Blair, M. R. (2017), Using video game telemetry data to research motor chunking, action latencies, and complex cognitive-motor skill learning. *Topics in Cognitive Science*, 9(2):467–484, URL <http://dx.doi.org/https://doi.org/10.1111/tops.12254>.
- Thorndike, E. L. (1921), *The psychology of learning*. New York: Teachers College, Columbia Univ.
- Tierney, T. M., Holmes, N., Mellor, S., López, J. D., Roberts, G., Hill, R. M., Boto, E., Leggett, J., Shah, V., Brookes, M. J., Bowtell, R., & Barnes, G. R. (2019), Optically pumped magnetometers: From quantum origins to multi-channel magnetoencephalography. *NeuroImage*, 199:598–608.
- Towne, T., Ericsson, K., & Sumner, A. (2014), Uncovering mechanisms in video

References

- game research: Suggestions from the expert-performance approach. *Frontiers in psychology*, 5:161, URL <http://dx.doi.org/10.3389/fpsyg.2014.00161>.
- Towne, T. J., Boot, W. R., & Ericsson, K. A. (2016), Understanding the structure of skill through a detailed analysis of Individuals' performance on the Space Fortress game. *Acta Psychologica*, 169:27–37, URL <http://dx.doi.org/10.1016/j.actpsy.2016.05.006>.
- Vaci, N., Edelsbrunner, P., Stern, E., Neubauer, A., Bilalic, M., & Grabner, R. (2019), The joint influence of intelligence and practice on skill development throughout the life span. *Proceedings of the National Academy of Sciences*, 116, URL <http://dx.doi.org/10.1073/pnas.1819086116>.
- Vallat, R. (2018), Pingouin: statistics in python. *The Journal of Open Source Software*, 3(31):1026.
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011), The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22.
- Van Rij, J., Van Rijn, H., & Hendriks, P. (2010), Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language*, 37(3):731–766, URL <http://dx.doi.org/10.1017/S0305000909990560>.
- Van Rossum, G., & Drake, F. L. (2009), *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Vidaurre, D., Abeysuriya, R., Becker, R., Quinn, A. J., Alfaro-Almagro, F., Smith, S. M., & Woolrich, M. W. (2018), Discovering dynamic brain networks from big data in rest and task. *NeuroImage*, 180:646–656, URL <http://dx.doi.org/https://doi.org/10.1016/j.neuroimage.2017.06.077>.
- Vidaurre, D., Myers, N. E., Stokes, M., Nobre, A. C., & Woolrich, M. W. (2019), Temporally Unconstrained Decoding Reveals Consistent but Time-Varying Stages of Stimulus Processing. *Cerebral Cortex*, 29(2):863–874, URL <http://dx.doi.org/10.1093/cercor/bhy290>.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., & SciPy 1.0 Contributors (2020), SciPy 1.0: Fundamental

- Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, URL <http://dx.doi.org/https://doi.org/10.1038/s41592-019-0686-2>.
- Vo, L., Walther, D., Kramer, A., Erickson, K., Boot, W., Voss, M., Prakash, R., Lee, H., Fabiani, M., Gratton, G., Simons, D., Sutton, B., & Wang, M. (2011), Predicting individuals' learning success from patterns of pre-learning MRI activity. *PLoS ONE*, 6(1), URL <http://dx.doi.org/10.1371/journal.pone.0016093>.
- Voss, M. W., Prakash, R. S., Erickson, K. I., Boot, W. R., Basak, C., Neider, M. B., Simons, D. J., Fabiani, M., Gratton, G., & Kramer, A. F. (2012), Effects of training strategies implemented in a complex videogame on functional connectivity of attentional networks. *NeuroImage*, 59(1):138–148, URL <http://dx.doi.org/https://doi.org/10.1016/j.neuroimage.2011.03.052>.
- de Vries, I. E. J., Marinato, G., & Baldauf, D. (2021), Decoding object-based auditory attention from source-reconstructed meg alpha oscillations. *Journal of Neuroscience*, 41(41):8603–8617, URL <http://dx.doi.org/10.1523/JNEUROSCI.0583-21.2021>.
- Vygotsky, L. S. (1978), *Mind and society: The Development of Higher Mental Processes*. Cambridge, MA: Harvard University Press.
- Wang, Y., Moon, J. M., Fu, W.-T., Boot, W., Erickson, K., & Kramer, A. (2010), Effects of Varied Priority Training on Complex Perceptual-Motor Learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32, (pp. 2536–2541).
- Ward, G., & Allport, A. (1997), Planning and problem solving using the five disc tower of london task. *The Quarterly Journal of Experimental Psychology Section A*, 50(1):49–78, URL <http://dx.doi.org/10.1080/713755681>.
- Ward, P. (2002), *The development of perceptual-cognitive expertise*. Ph.D. thesis, Liverpool John Moores University.
- Ward, P., Hodges, N. J., Starkes, J. L., & Williams, M. A. (2007), The road to excellence: Deliberate practice and the development of expertise. *High Ability Studies*, 18(2):119–153, URL <http://dx.doi.org/10.1080/13598130701709715>.
- White, H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838.
- Wickens, C. D., Hutchins, S., Carolan, T., & Cumming, J. (2013), Effectiveness of part-task training and increasing-difficulty training strategies: A meta-

References

- analysis approach. *Human Factors*, 55(2):461–470, URL <http://dx.doi.org/10.1177/0018720812451994>.
- Wiggins, M., Stevens, C., Howard, A., Henley, I., & O'hare, D. (2002), Expert, intermediate and novice performance during simulated pre-flight decision-making. *Australian Journal of Psychology*, 54(3):162–167, URL <http://dx.doi.org/10.1080/00049530412331312744>.
- Wiltschko, A., Johnson, M., Iurilli, G., Peterson, R., Katon, J., Pashkovski, S., Abaira, V., Adams, R., & Datta, S. (2015), Mapping sub-second structure in mouse behavior. *Neuron*, 88(6):1121–1135, URL <http://dx.doi.org/https://doi.org/10.1016/j.neuron.2015.11.031>.
- Yamagishi, N., Callan, D. E., Goda, N., Anderson, S. J., Yoshida, Y., & Kawato, M. (2003), Attentional modulation of oscillatory activity in human visual cortex. *NeuroImage*, 20(1):98–113, URL [http://dx.doi.org/https://doi.org/10.1016/S1053-8119\(03\)00341-0](http://dx.doi.org/https://doi.org/10.1016/S1053-8119(03)00341-0).
- 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*, 9:98, URL <http://dx.doi.org/10.3389/fnhum.2015.00098>.