



# An empirical investigation into the effects of sleep loss on esports performance

Tim David Smithies

## Publication date

31-12-2023

## Licence

This work is made available under the [CC BY-NC-SA 4.0](#) licence and should only be used in accordance with that licence. For more information on the specific terms, consult the repository record for this item.

## Document Version

1

## Citation for this work (HarvardUL)

Smithies, T.D. (2023) 'An empirical investigation into the effects of sleep loss on esports performance', available: <https://doi.org/10.34961/researchrepository-ul.25102469.v1>.

This work was downloaded from the University of Limerick research repository.

For more information on this work, the University of Limerick research repository or to report an issue, you can contact the repository administrators at [ir@ul.ie](mailto:ir@ul.ie). If you feel that this work breaches copyright, please provide details and we will remove access to the work immediately while we investigate your claim.



# **AN EMPIRICAL INVESTIGATION INTO THE EFFECTS OF SLEEP LOSS ON ESPORTS PERFORMANCE**

**Tim David Smithies, BSc (Hons)**

A thesis submitted to the University of Limerick in fulfilment of the requirement for the  
degree of **Doctor of Philosophy**

**Supervisors:** Professor Mark J. Campbell, Dr Adam J. Toth

Submitted to the University of Limerick, September 2023



# AN EMPIRICAL INVESTIGATION INTO THE EFFECTS OF SLEEP LOSS ON ESPORTS PERFORMANCE

## i. Abstract

Esports (competitive, organised video game play) has risen from obscurity to rival and/or surpass many traditional sports in terms of popularity, viewership, and earnings. As a result, human factors are beginning to be explored in the context of esports, ultimately with the same goals that are pertinent in much traditional sport research; to augment performance, or minimise performance loss. One human factor which has drawn attention within esports literature and practice is sleep. This is largely due to the substantial cognitive demands of esports, combined with the wealth of research linking sleep loss to impeded cognitive performance. Nonetheless, the relationship between sleep loss and esports performance has not been formally investigated to date; this current thesis aims to address this gap in scientific knowledge. Chapter two systematically explores the current scientific literature on how acute sleep restriction impacts the cognitive performance specifically for individuals who engage in cognitively demanding tasks with critical or safety-critical outcomes in their occupation or area of expertise (*Elite Cognitive Performers*). This chapter finds simple cognitive tasks to be most susceptible to sleep loss induced performance hindrance, however performance on complex tasks demanding cognitive flexibility (e.g. task-switching, a cognitive ability deemed highly relevant to esports) also appears potentially sensitive to sleep loss. Chapter three examines the test-retest reliability and presence of practice effects for a shortened version of the Category Switch Task, a task-switching paradigm with unpredictable switches, which allows for the assessment of cognitive performance on a complex task with and without cognitive flexibility demands. Chapter four provides an introduction to the *esport* Rocket League, which is the target esport within the current thesis. Chapter five outlines the identification of performance and rank indicators in the esport Rocket League through use of machine learning methods on a large dataset of in-game data. Performance indicators outlined are metrics targeted within later exploratory analysis on sleep loss and its impact on in-game Rocket League performance. Chapter six outlines key methodological details about the sleep measurement methods and analytical approach used in the subsequent chapter. It includes a bespoke simple imputation approach to deal with missing actigraphy-derived sleep data, which I show to outperform other simple imputation approaches. Chapter seven outlines a study exploring how experimentally induced total sleep deprivation impacts the cognitive and in-game performance of esport players. Cognitive tasks include the Psychomotor Vigilance Task and Category Switch Task, and the esport targeted was Rocket League; chosen due to various properties lending itself strongly to experimental research, as well as access to performance indicators (from chapter five) allowing for analytical depth. I find the overall in-game performance of Rocket League players to not change following ~29 hours of total sleep deprivation, despite increases in sleepiness, and decreases in alertness, motivation, and cognitive performance, immediately prior to esport play. Further exploratory analysis suggests that sleep deprived players may have adopted a simpler or safer (or both) playstyle. Chapter eight combines the findings of chapter seven with expert opinion from professional players, coaches, and analysts, to explore this playstyle change. In this chapter, I find that *simpler* and *safer* playstyles are very much analogous within Rocket League, helping to contextualise my findings with previous sleep loss and decision making literature. Collectively, the chapters within the current thesis provide novel insights into how sleep loss impacts in-game performance within esports, providing further evidence and discourse toward the topic of performance optimisation in esports.

## **ii. Authors Declaration**

I hereby declare that the work in this thesis is my own work, and was completed with the counsel of my supervisors, Dr Adam Toth (Department of Physical Education and Sport Sciences & Lero the Science Foundation Ireland Centre for Software Research), and Professor Mark Campbell (Department of Physical Education and Sport Sciences & Lero the Science Foundation Ireland Centre for Software Research). This work has not been submitted for any academic award at this, or any other third level institution.



Tim D. Smithies



Dr Adam J. Toth



Dr Mark J. Campbell

### **iii. Acknowledgements**

My first gratitude's go to my supervisors, Prof Mark Campbell and Dr Adam Toth. My PhD has been (as it's meant to be) an incredibly tough journey. Admittedly, I definitely made it tougher than it needed be by going down rabbit holes and being far too invested in 'side quests'. Upon reflecting on my journey, I realise just how well you both struck the balance of letting me off my leash to explore, and reeling me in when it was needed. I have learnt countless invaluable lessons from you both on how to conduct research. I value these lessons, the academic guidance you have both provided, and the friendships we have developed equally, and will cherish the time I have spent as your student.

Being involved with Logitech through my employment-based scholarship has been invaluable. It has helped me understand the relevance of research to industry, and how it can be tailored for maximum benefit. I am thankful for the opportunities I have had to provide feedback on programs outside of my own, and I hope that my contributions have been of value. I would like to particularly thank my employment mentor Niall White for his valuable feedback and insight throughout. I must also thank the Irish Research Council for the financial support they have provided to be throughout my PhD journey. The employment-based program has been of tremendous value to me, and I am extremely grateful for the opportunity that it provided.

Aidan, Pat and Lorna were the three I started my PhD with in PESS, and it brings me so much happiness every time I see any of you get a win no matter how big or small, because I know the quality of character behind each of you. To everyone else in PG034 that made it a place to look forward to going to, thank you.

I've been very lucky to have the peers at Lero that I've had; my gratitude extends to Sophie, Yemaya, Sarah, John and Antoine. To Magda and Faz (and at times, Giano!), sharing an office with people as incredible as you was a privilege. Lastly to Niall and Eoin, you probably both know how much you helped me but it's still worth mentioning. Sharing a house with you both (and Orla & Leena of course!) during the bulk of COVID and post-COVID made a very tough situation enjoyable more often than not. Now that I've mentioned the C word, I would like to actually thank COVID for helping me discover a passion for data analysis and computational statistics, which I likely would not found without the delays, hassles, and changes of plans that it caused!

It would be remiss not to mention RLI (particularly Enwrong and Cans), not only for their wholehearted support of my research but for being a community that has provided me with many friendships and cherished moments.

My final 'thank you's go to family. My extended period away from home has made me recognise just how lucky in life I have been to grow up with the people I have. To Sean and Hannah, you are both incredible people, and your trip to visit me this year did more for me mentally than you could know. To Jane, you are a special human being, and I love you very dearly – thank you for your support and patience with me. Lastly, to Mum and Dad, as cliché as this often sounds, you both have been the ultimate source of motivation and inspiration for me throughout my time as a PhD scholar. This is not just because of your tremendous work ethics and unwavering support. How you conduct yourself as human beings is inspirational and is something I genuinely aspire to emulate – thank you.

## iv. Table of Contents

### Contents

i. Abstract .....	iii
ii. Authors Declaration .....	iv
iii. Acknowledgements .....	v
iv. Table of Contents .....	vi
v. List of Figures .....	xii
vi. List of Tables .....	xvii
vii. Abbreviations .....	xviii
viii. Glossary of terms .....	xxii
ix. List of Publications .....	xxvi
x. Other Related Outputs.....	xxviii
Chapter 1. Introduction .....	1
1.1. Thesis scope .....	2
1.2. Introduction outline .....	2
1.3. Esports and human factors .....	3
1.4. An introduction to sleep .....	4
1.5. Sleep in the esports context .....	6
1.6. Sleep loss and cognitive performance – a speedrun.....	8
1.7. Forms of sleep loss .....	13
1.8. A word on circadian rhythms of performance.....	14
1.9. Is acute sleep loss a performance concern in esports? .....	16
1.10. Purpose, research questions, hypotheses .....	18
Chapter 2. The Effect of Sleep Restriction on Cognitive Performance in Elite Cognitive Performers: A Systematic Review .....	20
2.1. Abstract .....	21
2.2. Introduction .....	23

2.3.	Methods .....	27
2.3.1.	Database search strategy .....	27
2.3.2.	Grey Literature & Backward Snowballing .....	28
2.3.3.	Eligibility Criteria .....	29
2.3.4.	Quality Assessment.....	32
2.3.5.	Test/Task Categorisation.....	33
2.3.6.	Categorisation of Training Effect Bias .....	34
2.4.	Results .....	35
2.4.1.	Study Characteristics.....	35
2.4.2.	Low-Salience Test/Task Performance .....	42
2.4.3.	High-Salience Stable Test/Task Performance.....	43
2.4.4.	High-Salience Flexible Test/Task Performance.....	44
2.5.	Discussion .....	46
2.5.1.	Differences of the Effects of Sleep Restriction as a Function of Task Demands.....	46
2.5.2.	Strengths and Limitations .....	48
2.5.3.	Future Directions.....	51
2.5.4.	Conclusion .....	52
2.6.	Linking Section .....	53
2.6.1.	Esport Athletes are ECPs .....	53
2.6.2.	Implications for Experimental Research.....	55
Chapter 3.	Test-Retest Reliability and Practice Effects on a Shortened Version of the Category Switch Task – A Pilot Study .....	57
3.1.	Abstract .....	58
3.2.	Introduction .....	59
3.3.	Methods .....	62
3.3.1.	Participants.....	62
3.3.2.	Procedure.....	63
3.3.3.	Data Processing.....	65

3.3.4.	Statistical Analysis - Test-Retest Reliability .....	65
3.3.5.	Statistical Analysis - Practice Effects .....	65
3.4.	Results .....	67
3.4.1.	Test-Retest Reliability.....	67
3.4.2.	Test-Retest PEs .....	68
3.4.3.	Single Task.....	68
3.4.4.	Switch Costs.....	68
3.4.5.	Mixing Costs.....	69
3.5.	Discussion .....	71
3.5.1.	Test-Retest Reliability.....	71
3.5.2.	Test-Retest Practice Effects .....	72
3.5.3.	Effect of ‘Interval’ .....	73
3.5.4.	Limitations .....	73
3.5.5.	Conclusions.....	74
3.5.6.	Linking Section.....	75
Chapter 4.	An introduction to Rocket League .....	77
4.1.	Rocket League.....	79
4.2.	In-Game Rocket League Metric Categories .....	81
4.2.1.	Offense/ Defence Metrics: .....	82
4.2.2.	Boost Related Metrics:.....	82
4.2.3.	Movement Related Metrics:.....	82
4.2.4.	Positioning Related Metrics:.....	83
Chapter 5.	A Random Forest approach to identify metrics that best predict match outcome and player ranking in the esports Rocket League .....	84
5.1.	Abstract .....	85
5.2.	Introduction .....	86
5.2.1.	Methods.....	90
5.2.2.	Data Processing.....	91
5.2.3.	Analysis 1: Identifying a continuous outcome measure .....	94

5.2.4.	Analysis 2: Obtaining Performance Indicators (PIs) .....	94
5.2.5.	Analysis 3: Obtaining indicators of in-game rank (RIs) .....	95
5.3.	Results .....	98
5.3.1.	Obtaining a continuous performance outcome measure .....	98
5.3.2.	Obtaining indicators of performance (PIs).....	99
5.3.3.	Obtaining indicators of in-game rank (RIs) .....	102
5.4.	Discussion .....	105
5.4.1.	Significance.....	107
5.4.2.	Limitations and Future Research .....	108
5.4.3.	Conclusions .....	109
5.5.	Linking chapter.....	111
Chapter 6.	Key methodologies for the experimental sleep loss study .....	112
6.1.	Using Readibands for Actigraphy-Derived Sleep Outcomes.....	114
6.1.1.	Polysomnography.....	114
6.1.2.	Sleep Diaries .....	114
6.1.3.	Actigraphy.....	115
6.2.	Managing Missing Actigraphy-Derived Sleep Data .....	117
6.2.1.	Brief Background .....	117
6.2.2.	Testing Approach .....	118
6.2.3.	Imputation and Testing Procedure .....	119
6.2.4.	Analysis.....	120
6.2.5.	Outcomes and Discussion .....	121
6.3.	A Description and Discussion of Mixed Effect Models.....	127
Chapter 7.	Don't lose sleep over esports: exploring how total sleep deprivation affects the cognitive and in-game performance of rocket league players.....	131
7.1.	Abstract .....	132
7.2.	Introduction: .....	134
7.3.	Methods:.....	138
7.3.1.	Participants.....	138

7.3.2.	Materials.....	141
7.3.3.	Procedure .....	146
7.3.4.	Statistical Analysis.....	149
7.4.	Results: .....	153
7.4.1.	Participant Pairs & Rank.....	153
7.4.2.	Protocol Adherence.....	154
7.4.3.	Cognitive Performance .....	156
7.4.4.	Subjective Sleepiness, Alertness and Motivation .....	158
7.4.5.	In-Game Performance .....	159
7.5.	Discussion: .....	162
7.5.1.	Limitations .....	166
7.5.2.	Conclusion .....	168
Chapter 8.	The BART effect: Rocket League players appear to play both simpler and safer when sleep deprived.....	169
8.1.	Introduction: .....	170
8.2.	Methods.....	172
8.2.1.	Participants.....	172
8.2.2.	Instruments & Procedure .....	173
8.2.3.	Data Processing and Statistical Analysis .....	173
8.3.	Results .....	175
8.3.1.	Metrics Relevant to Playstyles.....	175
8.3.2.	MANOVA Analysis.....	175
8.4.	Discussion .....	178
8.4.1.	Limitations & Future Research.....	179
Chapter 9.	Discussion .....	181
9.1.	Assessing Performance In-Situ .....	183
9.2.	Why <i>didn't</i> sleep deprivation lead to worsened performance? .....	184
9.3.	Strengths of the overall approach.....	192
9.4.	Limitations and Future Research.....	197

9.4.1.	Generalisability .....	197
9.4.2.	Expertise.....	198
9.4.3.	Forms of Sleep Loss.....	199
9.5.	Practical implications .....	200
9.6.	Conclusions .....	201
Chapter 10.	References .....	203
Chapter 11.	Appendices (separate digital file).....	251

## v. List of Figures

<b>Figure 1-1 A</b> Instances of Lapses on a 10 minute psychomotor vigilance task (PVT), and <b>B</b> subjective sleepiness scores on the Stanford Sleepiness Scale (SSS), across multiple days and for participant groups with varying amounts of sleep opportunity (SO) afforded. In both graphs, higher values denote greater impairment. The grey horizontal line depicts the expected value for each graph following ~36hrs of total sleep deprivation (TSD), denoted by the leftmost vertical dashed line in each graph. In <b>A</b> , the rightmost vertical dashed line denotes the equivalent amount of days (~9) with 4hrs SO required for predicted PVT lapses to be equivalent to that at ~36hrs TSD. This figure is adapted from Van Dongen et al. (2003).	14
<b>Figure 1-2</b> The two-process model, originally described by (Borbély, 1982). This model is used today to describe how sleep history and circadian rhythm interact to influence ones sleep drive, alertness, and performance. Regarding sleep drive, the larger the distance between Process S and Process C, the greater the sleep drive, with this distance being reduced through nighttime sleep.	16
<b>Figure 2-1</b> PRISMA flowchart outlining the eligibility and inclusion process for the current review.	29
<b>Figure 2-2</b> Amount of mean sleep obtained ( $\pm$ SD) within each condition. Note that for studies reporting only “sleep opportunity,” sleep obtained was considered to be the entirety of the reported period. SR = sleep restriction, EM = experimentally manipulated ( $n = 11$ ), Obs = observed ( $n = 3$ ).	36
<b>Figure 2-3</b> Proposed framework explaining the likelihood of sleep restriction affecting cognitive performance for Elite Cognitive Performers.	49
<b>Figure 2-4</b> Skills and experience of which esport athletes possess (left) and the three occupations which best match these skills and experience (right). Figure is from Smithies et al. (2020).	54
<b>Figure 3-1</b> A visual representation of the third or ‘combined’ iteration of the category switch test. For each trial, a <i>living cue</i> or <i>size cue</i> is displayed alone for 150ms. After the 150ms, a word appears on the screen. The participant uses the ‘E’ or ‘I’ key to correctly categorise the word according to the rule associated with the cue. The inter-stimulus interval is 350ms for correct responses and 1500ms for incorrect responses. If the next trial has a different cue, it is a <i>switch</i> trial, whereas if the next trial has the same cue, it is a <i>repeat</i> trial.	64
<b>Figure 3-2</b> Stacked bar charts showing variance decomposition for the six outcome measures. As between-participant variance (black) is analogous to the reported ICC	

values, the thresholds used for defining reliability (poor, moderate, good, and excellent) are also provided. ....67

**Figure 3-3** Bar charts depicting the ANCOVA adjusted mean ( $\pm$ SEM) retest values of the interval groups same day (SD), next day (ND), and next week (NW), adjusted for the ‘baseline’ test value (dotted line), for the six outcome variables of interest: **A** single task RT, **B** single task accuracy, **C** SC RT, **D** SC accuracy, **E** MC RT, and **F** MC accuracy. \*\* denotes  $p < 0.01$ , and \* denotes  $p < 0.05$ ; denotations next to the dotted line indicate a significant test-retest practice effect for the pooled data, and denotations above bars indicate a significant difference between different interval groups at retest. ....70

**Figure 4-1. A** A simplified birds-eye view of a standardised Rocket League map. Blue and orange rectangles denote goal locations. The white circle denotes the location of the ball at the start of the match or immediately following a goal. Yellow circles (black outline) denote the location of *small pads*, and black circles (yellow outline) denote the location of *big pads*. **B** A top-down view of a Rocket League map (“Beckwith Park”), allowing view of the map walls and ceiling. **C** Players view of a kickoff about to commence. Kickoffs place a member of each team at equal distance to the ball, and allow players to move at an identical time. **D** A players view mid-game; in this instance, within an aerial-based attacking play. In **C** and **D**, the current score and time left can be seen in the top of the screen, and the player’s current boost total can be seen in the bottom-right of the screen; this information is present to players throughout the entirety of a Rocket League match. ....80

**Figure 5-1** A density plot showing the distribution of accounts within the Rocket League rank system. Colour shaded areas correspond to the skill brackets, and associated MMRs, considered for the current study. This distribution is as per season 14 of Rocket League, which was the season at the time of the most recent match used in the analysis. ....90

**Figure 5-2** Flowchart depicting the methods of the current study. The three outlined analyses are labelled in blue. The † highlights where average speed was removed due to multicollinearity. ....97

**Figure 5-3** Density plots showcasing **a**: the distributions of goal difference and **b**: in-game score difference as a function of win vs. loss. ....99

**Figure 5-4** Heat map displaying the percentage of the total increase in MSE that can be found when a metric is permuted individually compared to the sum of increase in MSE for all metrics when permuted individually. Only metrics that were significantly important for predicting GD within each raw-score and difference-score model are presented. White squares represent metrics that were not significant for the rank they are

assigned to. **a:** results from raw-score regression models, and **b:** results from difference-score regression models. .... 101

**Figure 5-5** Metrics found to be of significant importance to the classification model created to predict the ranks of individuals playing 1v1 Rocket League, ordered by the increase in mean decrease in accuracy experienced within the model when each metric was permuted. .... 103

**Figure 5-6** Violin plots displaying the means and distributions, within each rank, of the four most important features for predicting rank, **a:** *Time spent at supersonic speed*, **b:** *Time spent on the ground*, **c:** *Shots conceded*, **d:** *Time spent goalside of the ball*..... 104

**Figure 6-1** Line graphs depicting the mean absolute percentage error (APE) of imputed values, obtained using the five outlined imputation approaches, for **A** TST, **B** TASO, and **C** TAW..... 122

**Figure 6-2.** Bland-Altman Plots displaying the agreement between actigraphy-derived Total Sleep Time (TST) values, and values derived from the five considered imputation approaches (with 13 days available per participant): **A** *Diary  $\pm$  Individual Bias*, **B** *Diary Only*, **C** *Proximity Imputation*, **D** *WP Hot Deck*, and **E** *BP Hot Deck*. Blue dashed line represent the mean for TST (imputed) minus TST (actigraphy-derived). Red dashed lines represent 95% Confidence Limits. Black dotted lines represent the 95% bootstrap confidence intervals for these values. .... 123

**Figure 6-3.** Boxplots showing individual tendencies to over/ underestimate total sleep time (TST), in comparison to actigraphy-derived values. Each box represents an individual. Values below the dotted horizontal line resemble a tendency to overestimate TST. \* indicate outliers as determined by Tukey's method (Tukey, 1977)..... 126

**Figure 7-1.** A timeline of the 15-day protocol for all participants within the protocol. The icons within this figure depict the following. Wristband = Readiband; book = Consensus Sleep Diary; coffee mug and beer = caffeine and alcohol; clipboard = subjective measures. Computer with keyboard = computerised cognitive tests. Computer with controller = Rocket League gameplay; bed = sleeping; man running = strenuous activity; light and thermometer = light and temperature-controlled environment. TSD = participants within the sleep deprivation group; CON = participants within the control group. .... 146

**Figure 7-2 A** Rank distribution and pairing of included players. Clear diamonds resemble CON participants, and red diamonds resemble TSD participants. Pairs are denoted by lines joining participants. The x-axis denotes the participants in-game MMR (a proxy for expertise) relative to the esports overall player base at the time of recruitment, such that

lower values resemble a lower-ranked player and vice versa (i.e., 99% denotes a player in the top 1% of players). To better visualize pairs in the top 3% of ranked 1v1 Rocket League players, a magnified display is depicted above the main graph. **B** Relationship between estimated total hours of Rocket League played (x-axis) and player expertise (y-axis). Dashed lines represent 95% CI for line fitted using the linear regression equation. For both A and B, colours represent the in-game rank of the participants (in order from bottom to top; bronze, silver, gold, platinum, diamond, champion, grand champion, supersonic legend)..... 153

**Figure 7-3** Mean ( $\pm$ SD) discrepancy between **A** target bedtime and TASO, and **B** target wake-time and TAW, for each participant within each pair. Participants in CON are denoted by clear diamonds, while TSD participants are denoted by the red diamonds. The green band denotes TASO or TAW within 1hr of the target bed/ wake time, while the red area denotes TASO or TAW outside of that range. .... 155

**Figure 7-4** Box and whisker (min  $\rightarrow$  max) plots showing the group mean TSTs for CON and TSD **A** the night before test sessions, and **B** the mean of the two nights prior to that shown in **A**. .... 155

**Figure 7-5** Frequency distribution of RS ( $1000/RT_{(msec)}$ ) for participants in the **A** CON group and **B** TSD group. I direct the reader to **B**, and note both a leftward shift and leftward skew of the response distribution in the experimental test session for the TSD participants, consistent with previous literature (i.e., Figure 4, Grant et al. (2017)) and demonstrative of both the broadband decrease in RS and increase in lapses observed (trials to the left of the dotted line)..... 157

**Figure 7-6** Box and whisker (min  $\rightarrow$  max) plots showing **A** KSS scores, **B** Alertness VAS scores, and **C** Motivation VAS scores (reverse scored) for CON and TSD participants in the baseline and experimental sessions. .... 159

**Figure 7-7** Violin plots displaying the distribution of **A** GD and **B-F** exploratory PIs for *baseline* and *experimental* sessions, across all 279 matches. Box and whisker (min  $\rightarrow$  max) plots inside the violin plots resemble the distribution of *mean* outcome variables across a test session for a given pair ( $N = 20$  for each box and whisker plot). Diamonds represent pair means for each session, with pair means connected via the dotted lines. The solid red line represents the estimated mean $\pm$ SE from each model used for analysis.. 160

**Figure 8-1.** The five in-game metrics that received the most points, corresponding to *field expert* determined ability to distinguish *playstyle risk* and *playstyle complexity* in 1v1 Rocket League. Metrics labels highlighted in blue or orange are metrics that are also a

top five distinguisher of the other playstyle, which metrics labels highlighted in grey are metrics that are also a top ten distinguisher of the other playstyle ..... 175

**Figure 9-1 A** Bar chart showing the difference in mean reaction time (black), mean within-participant standard deviation or reaction time (i.e., green), and lapse probability (red), for TSD participants following sleep deprivation protocol and when well rested, within each one-minute epoch of the 10-min PVT (inclusive of the first minute of the PVT). Values above 0 resemble greater mean values following sleep deprivation, compared to when rested. Black and green bars are measured on the left y-axis, and red bars are measured on the right y-axis. **B** Bar chart showing the difference in mean reaction time for each trial of the PVT for a randomly-selected TSD participant. Values above zero resemble a trial in which the reaction time was poorer following sleep deprivation, compared to when well rested. Red bars resemble trials which were a lapse within the session while sleep deprived, and blue bars resemble trials which were a lapse within the session when rested. .... 186

## vi. List of Tables

<b>Table 2-1</b> Study Characteristics for included articles, organised by their test/task categorisation and type.....	37
<b>Table 3-1</b> Test-retest reliability results for outcome variables considered. Upper and lower 95% confidence intervals are given in squared brackets for ICC $_{(2,1)}$ .....	67
<b>Table 5-1</b> Predictor metrics obtained through <i>ballchasing.com</i> and subsequent processing. Metrics are time normalised to an average match length (360 seconds) unless provided as a percentage of total time in the original dataset, Metrics are expressed both as <i>raw-score</i> and <i>difference-score</i> except those denoted by a †, which are <i>raw-score</i> only. ....	93
<b>Table 5-2</b> $R^2$ (mean of squared residuals) of the Random Forest models created using the raw and difference-score metrics for each ranks and for all ranks combined.....	100
<b>Table 6-1</b> Number of participants with N amount of days available .....	118
<b>Table 6-2</b> Mean values for actigraphy-derived and imputed sleep variables, assessments of bias, and limits of agreement, for analysis with 13 days of data available per participant. ....	125
<b>Table 7-1.</b> Summary of included participants with reference to the Revised Research Diagnostic Criteria for Defining Normal Sleeping Controls (Beattie et al., 2015).....	139
<b>Table 8-1</b> Eigenvalues, canonical correlation, standard discrimination function coefficients (SDFC) and structure coefficients, for metrics (note that all are considered as difference scores) that distinguish <i>playstyle risk</i> and <i>playstyle complexity</i> in 1v1 Rocket League. The linear equation for the <i>playstyle variables</i> is given at the bottom of the table using raw discrimination function coefficients. ....	177

## **vii. Abbreviations**

<b>Term</b>	<b>Explanation</b>
1v1	One versus One
2v2	Two versus two
3v3	Three versus three
AASM	American Academy of Sleep Medicine
AJT	Adam J. Toth (Author)
ANCOVA	Analysis of Covariates
ANOVA	Analysis of Variances
APE	Absolute Percentage Error
API	Application Processing Interface
AUDIT	Alcohol Use Disorders Identification Test
BART	Balloon Analog Risk Task
BCa	Bias-corrected and Accelerated
BP	Between Participant
CAD	Canadian Dollars
CART	Classification and Regression Trees
CI	Confidence Interval
CON	Control Group
COVID-19	Coronavirus Disease 2019 (SARS-CoV-2)
CRSD	Circadian Rhythm Sleep Disorder
CRT	Choice Reaction Test
CS:GO	Counter Strike: Global Offensive
CSD	Consensus Sleep Diary
CST	Category Switch Task
CTI	Cue-Target Interval
DDA	Descriptive Discriminant Analysis
DOTA2	Defense of the Ancients 2
DSST	Digit Symbol Substitution Test
DTIC	Defence Technical Information Centre
ECP	Elite Cognitive Performer
EEG	Electroencephalography
EMA	Early Morning Awakenings
EM	Experimentally Manipulated

EMG	Electromyography
EOG	Electrooculography
FAST	Fast Alcohol Screening Test
FPS	First Person Shooter
FPT	Fixed Priority Training
GC	Grand Champion
GD	Goal Difference
GDPR	General Data Protection Regulations
GPS	Global Positioning System
HCL	High Cognitive Load
HSD	Honestly Significant Difference
HSDQ	Holland Sleep Disorder Questionnaire
HSF	High-Salience Flexible
HSS	High-Salience Stable
ICC	Intraclass Correlation Coefficient
ICSD	International Classification of Sleep Disorders
IDRT	Individual Differences in Reaction Time
IGSD	In-Game Score Difference
IGT	Iowa Gambling Task
ISI	Inter-Stimulus Interval
KSS	Karolinska Sleepiness Scale
LAN	Local Area Network
LCL	Low Cognitive Load
LLC	Limited Liability Company
LoA	Limits of Agreement
LoL	League of Legends
LRT	Likelihood Ratio Test
LS	Low-Salience
LSD	Least Significant Difference
MANOVA	Multivariate Analysis of Variance
MAR	Missing-at-random
MC	Mixing Cost
MCAR	Missing-completely-at-random
MDA	Mean Decrease in Accuracy
MEM	Mixed Effect Models
MEQ	Horne-Ostberg Morningness Eveningness Questionnaire

MeSH	Medical Subject Headings
MI	Multiple Imputation
MJC	Mark J. Campbell (Author)
MK	Magdalena Kowal (Author)
MMR	Matchmaking Rating
MOBA	Multiplayer Online Battle Arena
MSE	Mean Square Error
NMAR	Not-missing-at-random
N.Y.	New York
NHLBI	National Heart, Lung, and Blood Institute
NR	Niall Ramsbottom (Author)
NS	No Significant Effect
Obs	Observed
OSF	Open Science Framework
OOB	Out-of-bag
PASAT	Paced Auditory Serial Addition Test
PE	Practice Effect
PFC	Prefrontal Cortex
PI	Performance Indicator
PLMS	Periodic Limb Movement Disorder
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSG	Polysomnography
PSQI	Pittsburgh Sleep Quality Index
PVT	Psychomotor Vigilance Task
RF	Random Forest
RI	Rank Indicator
RLS	Restless Leg Syndrome
RRDC	Revised Research Diagnostic Criteria for Defining Normal Sleeping Controls
RS	Response Speed
RT	Reaction Time
SC	Switch Cost
SCN	Suprachiasmatic Nucleus
SD	Standard Deviation (unless otherwise specified, e.g. in quotes)
SDFC	Standard Discrimination Function Coefficients
SE	Standard Error

SE%	Sleep Efficiency (Always with a percentage sign to differentiate with Standard Error)
SNS	Swiss Narcolepsy Scale
SO	Sleep Opportunity
SPSS	Statistical Package for the Social Sciences
SR	Sleep Restriction
SRI	Sleep Regularity Index
SRM	Stimulus Response Mapping
SRT	Serial Reaction Time
SSL	Supersonic Legend
SSS	Stanford Sleepiness Scale
SWS	Slow-Wave Sleep
TASO	Time at Sleep Onset
TAW	Time at Wake
TDS	Tim D. Smithies (Author)
TIB	Time in bed
TSD	Total Sleep Deprivation (can refer to both the TSD group or to TSD as an experience)
TST	Total Sleep Time
U.S.	United States
USD	United States Dollars
VAS	Visual Analog Scale
VPT	Variable Priority Training
VR	Virtual Reality
WASO	Wake After Sleep Onset
WL	Win vs. Loss
WP	Within Participant

## viii. Glossary of terms

Term	Explanation
Agreement	Assessment of <i>closeness</i> between repeated measurements (de Vet et al., 2006).
Bootstrapping	A method of resampling with replacement, normally used to create many simulated samples from a single dataset.
Circadian rhythm	A biological process occurring with a period of approximately 24-hours.
Classification and regression trees	Decision tree algorithms, used to predict categorical (classification) or numerical (regression) outcomes.
Cognitive flexibility	Adjustment of behaviour in response to environmental or task demands (Uddin, 2021).
Cognitive performance	Response/ action on the basis of knowing, learning, or understanding.
Compensatory mechanisms	Within the context of cognitive performance: changes in the brain's response to a task, in an effort to maintain performance in situations where the brains normal response is in some way impaired.
Ecological validity	"The generalisability of experimentally obtained findings to a real-world context, or to the context for which the results directly apply to" (Kihlstrom, 2021; Orne's definition).
Elite Cognitive Performers	Individuals who engage in cognitively demanding tasks with critical or safety-critical outcomes in their occupation or area of expertise.
Esports	"Video/computer games played within the medium of cyberspace competitively" (Campbell et al., 2018; p. 161).
High-Salience Flexible	Tasks involving cognitive flexibility to a significant degree for performance.
High-Salience Stable	Tasks involving complex cognitive functioning (i.e. beyond that of low-salience) but in which performance does not depend on the ability to flexibly shift attention or adapt to changing task dynamics.

Imputation	The process of substituting missing data with an alternate value.
In-situ	For the purposes of this thesis, <i>in-situ</i> refers to performance measured in the environment directly relevant to environment in which it is most applicable to.
Lapse (for PVT)	Responses $\geq 500$ msec.
Local sleep	A state akin to slow-wave sleep (global or 'whole brain') but localised to specific neural assemblies.
Low-Salience	Tasks no distractors and very limited decision-making, and typically require simple, timely responses to a stimulus.
Machine learning	The development of algorithms which can predict outcomes when provided with data.
Metric importance	How important a specific metric is to the performance of the machine learning model built using multiple metrics to predict an outcome.
Mixed Effect Model	Statistical models containing both fixed and random effects.
Mixing Costs	The cost associated with knowing a cue could potentially change in a block of test block, compared to performance when knowing the cue will not change.
Mtry	The number of features randomly selected for each CART entrainment within a random forest model.
Notational analysis	The study of patterns within a match/contest/competition/performance that lead to a successful overall outcome (Hughes & Bartlett, 2002).
Ntree	The number of CARTs created within a random forest model.
Performance indicator	In-match metrics that predict in-game match outcome.
Permutation	The random shuffling of a given feature such that the relationship between the feature and the outcome is broken.
Playstyle	An individual's technique/ strategy employed while playing a game.
Practice effects	Improved performance specifically attributable to repeated engagement with the test (McCaffrey et al., 2000).

Random forest	Machine learning algorithm; specifically, an ensemble of CARTs each trained using a unique bootstrapped data set and random selection of splitting predictor features.
Rank indicator	In-match metrics that predict in-game player rank.
Response Speed	1000/RT(msec).
Rocket League	Rocket League is a <i>vehicular soccer video game</i> (Smithies et al., 2021), commonly played as an esports. For a further description, see <b>Chapter 4</b> .
Scrimmaging (scrims)	Practice through organised esports play against other individuals or teams, performed in a way to directly mimic competition.
Sleep	“A reversible behavioural state of perceptual disengagement from an unresponsiveness to the environment” (Sullivan et al., 2021, p. 16).
Sleep Disruption	Frequent arousals and wake periods during nighttime periods, which may reduce sleep quantity but also reduce sleep quality.
Sleep loss	Obtaining less sleep than is optimal. Acute sleep loss can be categorised into three categories: Total Sleep Deprivation, Sleep Restriction, and Sleep Disruption (Reynolds and Banks, 2010).
Sleep Onset Latency (SOL)	Length of time from time at lights out to time of sleep onset.
Sleep Opportunity	The amount of time allowed for participant sleep within a study.
Sleep Restriction	The reduction of sleep quantity below that normally experienced for one or more nights.
Switch costs	The cost of responding to a changing cue, compared to performance when a cue within a task remains the same.
Task-Switching	The ability to rapidly/ efficiently shift one’s attention or cognitive resources between two or more tasks.
Test-retest interval	The length of time between test administrations in a test-retest design study.
Test-retest reliability	The degree to which individual's performance can be distinguished from each other across two administrations of a test.

Total Sleep Deprivation (TSD)	The total elimination of sleep, normally for 24 hours or more.
Total Sleep Time (TST)	Length of from time of sleep onset to time at wake, minus the wake after sleep onset.
Vigilance decrement (time-on-task effect)	Task performance becomes poorer and more variable over prolonged/ sustained bouts.
Wake After Sleep Onset (WASO)	Length of time awake after sleep onset and before wake time.

## ix. List of Publications

### Publications forming chapters within this thesis:

- **Smithies, T. D.**, Toth, A. J., Dunican, I. C., Caldwell, J. A., Kowal, M., & Campbell, M. J. (2021). The Effect of Sleep Restriction on Cognitive Performance in Elite Cognitive Performers: A Systematic Review. *Sleep*, 44(7), zsab008. DOI: <https://doi.org/10.1093/sleep/zsab008>
- **Smithies, T. D.**, Campbell, M. J., Ramsbottom, N., & Toth, A. J. (2021). A Random Forest approach to identify metrics that best predict match outcome and player ranking in the esports Rocket League. *Scientific reports*, 11(1), 1-12. DOI: <https://doi.org/10.1038/s41598-021-98879-9>
- **UNDER REVIEW: Smithies, T. D.**, Toth, A. J., Campbell, M. J. (2023). Test-Retest Reliability and Practice Effects on a Shortened Version of the Category Switch Task.
- **UNDER REVIEW: Smithies, T. D.**, Toth, A. J., Campbell, M. J. (2023). Don't lose sleep over esports: exploring how total sleep deprivation effects the cognitive and in-game performance of rocket league players.

### Publications completed during PhD however not forming chapters:

- **Smithies, T. D.**, Toth, A. J., Conroy, E., Ramsbottom, N., Kowal, M., & Campbell, M. J. (2020). Life after esports: a grand field challenge. *Frontiers in Psychology*, 11, 883. DOI: <https://doi.org/10.3389/fpsyg.2020.00883>
- Kowal, M., Conroy, E., Ramsbottom, N., **Smithies, T. D.**, Toth, A. J., & Campbell, M. J. (2021). Gaming your mental health: A narrative review on mitigating symptoms of depression and anxiety using commercial video games. *JMIR serious games*, 9(2), e26575. <https://doi.org/10.2196/26575>
- **UNDER REVIEW: Campbell, M. J., Jenny, S. E., Cregan, S., Smithies, T. D.** (2023). Chapter 2.4: General Recommendations for Esports Research. In Jenny, S. E., Besombes, N., Brock, T., Cote, A., Scholz, T. M., *Routledge Handbook of Esports*.

### Conference Publications:

- **Smithies, T. D.**, Campbell, M. J., Ramsbottom, N., & Toth, A. J. (2021). A Random Forest approach for uncovering performance and in-game rank indicators

for Rocket League [Online Oral Presentation]. *9th International Performance Analysis Workshop and Conference & 5th International Conference of Computer Science in Sports Conference*, 30<sup>th</sup> – 31<sup>st</sup> August.

- **Smithies, T. D.**, Toth, A. J., Campbell, M. J. (2022). Don't lose sleep over esports: exploring how total sleep deprivation affects cognitive and in-game performance of rocket league players [Poster Presentation]. *The 26th Conference of the European Sleep Research Society*. 27<sup>th</sup> – 30<sup>th</sup> September.
- **Smithies, T. D.**, Toth, A. J., Campbell, M. J. (2023). Uncovering the optimal simple imputation approach for missing actigraphy-derived sleep outcomes [Oral Presentation]. *The 7th Annual All-Ireland Postgraduate (AIPG) Conference in Sport Sciences, Physical Activity & Physical Education*. 26<sup>th</sup> May.
- **UPCOMING: Smithies, T. D.**, Toth, A. J., Campbell, M. J. (2023). The Impact of Total Sleep Deprivation on Performance in the Esport 'Rocket League [Symposium Speaker & Poster Presentation]. *Sleep DownUnder 2023*. 8<sup>th</sup>- 11<sup>th</sup> November.

## x. Other Related Outputs

### Other Conference/ Seminar Presentations - International:

- **Smithies, T. D.** (2020) The Effect of Sleep Restriction on Cognitive Performance in Elite Cognitive Performers: A Systematic Review [Keynote Speaker, Online]. *University of California, Irvine (UCI) Esports Conference 2020*. 8<sup>th</sup> October.
- **Smithies, T. D.** (2021) The Effect of Sleep Restriction on Cognitive Performance Among Elite Cognitive Performers: A Systematic Review [Online Oral Presentation]. *Sleep4Performance Seminar 2021*. 12<sup>th</sup> August.
- **Smithies, T. D.** (2022) Fatigue in Operational Environments – Insights from Esports & Other Contexts [Oral Presentation]. *SIESTA Research Group Seminar*. 8<sup>th</sup> December.
- **Smithies, T. D.** (2023) The Effects of Sleep Deprivation on Esports Performance [Online Oral Presentation]. *Sleep4Performance Seminar 2023*. 21<sup>st</sup> June.
- **UPCOMING: Smithies, T. D., Toth, A. J., Campbell, M. J.** (2023). Circadian advantage in esports; when is the best time to play? [Poster Presentation]. *National Sport and Human Performance Conference 2023*. 29<sup>th</sup> September.

### Other Conference/ Seminar Presentations - Domestic:

- **Smithies, T. D.** (2021) Does losing sleep affect the performance of elite performers? [Online Oral Presentation]. *Pint of Science IE*. May 18<sup>th</sup>.
- **Smithies, T. D.** (2022) Don't Lose Sleep Over Esports. *Thesis-in-Three (University of Limerick EHS Faculty Heat)*. March 31<sup>st</sup>.

### Podcasts, Popular Science & Blog Articles:

- University of Limerick Physical Education and Sport Sciences Blog (2021/ 22)
  - o [It's Time to Play: Discussing Circadian Rhythm and Esport Performance](#)
  - o [Playing Action Video-Games for Benefit on the Pitch](#)
  - o [Playing Sport-Based Games for IRL Benefit](#)
- [Esports Research Network 'Esports Research Report'](#) (2021) [Podcast]
- [Adamas Esports 'Path to Pro'](#) (2021) [Podcast]
- [Synapse Performance Podcast: The Role of Sports Science in E-Sports](#) (2021) [Podcast]
- [Irish Research Council 'Spotlight on Research'](#) [Research Showcase Article] (2021)

## **Chapter 1.      Introduction**

## **1.1. Thesis scope**

The main objective of the work outlined within this thesis is to explore if and how acute sleep loss may impact performance in the world of esports. Throughout the thesis, the definition of esports provided by Campbell et al. (2018), “video/computer games played within the medium of cyberspace competitively” (p. 161), is used as the working definition for esports.

The thesis objective is addressed in a multidisciplinary fashion and with use of a variety of approaches. Firstly, I attempt to gain insight into how sleep loss may impact esports performance by first exploring how sleep loss affects both the cognitive and occupation specific performance of individuals with work demands comparable to those within esports. This is undertaken by way of a systematic review. From this review, I outline task switching as a cognitive function of interest, and subsequently assess the presence of practice effects and reliability of a task switching test, the Category Switch Task (CST), in a test-retest design administration. These pieces of work were performed within the context of a perceived importance of cognitive factors for in-game esports performance (Campbell et al., 2018), an idea supported by large bodies of work demonstrating experienced action video gamers to have improved performance in many cognitive domains relative to their peers, and practice in action video game play leading to improved cognitive performance (Bediou et al., 2018; Bediou et al., 2023; Toth et al., 2020).

Beyond looking at cognitive performance, a direct measure of in-game esports performance was sought. Firstly, *Rocket League* was identified as a suitable esports to assess performance in, within an experimental design. A machine learning notational analysis was performed within this esports to uncover in-game metrics directly relevant to performance. Lastly, an experimental protocol was undertaken, in which the impact of a total sleep deprivation protocol on cognitive and in-game *Rocket League* performance (using measures explored in the abovementioned work) was assessed.

## **1.2. Introduction outline**

The following sections within this chapter will describe many of the themes directly relevant to the overarching topic of acute sleep loss and its influence on esports performance. Firstly, the world of esports will be briefly introduced. Following this, the introduction will hone in on the topic of sleep; it will discuss the perceived adverse impact

of sleep loss on esports performance, theories that have been developed to explain how sleep loss impacts cognitive performance, different types of sleep loss, and briefly discuss the role of circadian rhythms. Three key considerations when defining the relationship between sleep loss and esports performance will be introduced and outlined in preparation for later chapters. The chapter will conclude by listing the thesis purpose, research questions, aims and hypotheses explored.

### **1.3. Esports and human factors**

Esports, already massively popular, are comfortably the fastest growing competitive activity worldwide. Here and throughout this thesis, I adopt Campbell et al. (2018)'s definition of esports, being "video/computer games played within the medium of cyberspace competitively" (p. 161), though I note the existence of countless definitions for esports, each with their similarities and nuances (see Cranmer et al., 2021 for a collation of definitions). There is considerable and ongoing debate as to esports' status as a sport, ignited by inclusions/ exclusions of esports within sport specific arenas such as the Olympics (Olympics.com, 2023; Ribeiro et al., 2023; Todt et al., 2020), and Commonwealth Games (Olympics.com, 2022; Tidy, 2022); I will avoid this debate within this thesis, but will point to numerous published articles (Cranmer et al., 2021; Franks & King, 2023; García & Murillo, 2020; Hallmann & Giel, 2018; Hamari & Sjöblom, 2017; Holden et al., 2017; Jenny et al., 2017; Reitman et al., 2019) for nuanced discussion of this topic. Importantly, the term *esports* is analogous to *sport* in that it encompasses a multitude of games, with different dynamics, strategies, cognitive and physical demands. Continuing this analogy, esports can be categorised into *genres* according to these characteristics; and again like traditional sport, differences and debates over classification methods exist (Apperley, 2006; Jang & Byon, 2020; Jonasson & Thiborg, 2010; Toth, Conroy, et al., 2021).

Esports are an integral part of the juggernaut gaming industry, which has a projected market value of €375billion in 2023 (Statista, 2023). The staggering value of esports is attributable primarily to the large and dedicated fanbases they foster, with viewership estimates exceeding one billion individuals in 2020 (Ahn et al., 2020) (and growing yearly). Hence, esports present as an enticing medium for publicity, attracting large investment from major companies such as Microsoft, Coca-Cola, Amazon, and Tencent (Marques, 2019). For successful esports organisations, these financial outlays are not trivial, with the ten most valuable esports organisations combined being valued at

approximately ~€3.21billion in 2022 (Knight, 2022). For top players within popular esports, financial rewards can be lucrative, with over 135 players (as of June 20, 2023) earning more than €1million in competition earnings alone (Esports Earnings, 2023a), and with average yearly contracts exceeding \$400,000USD within the top-tier of League of Legends (Studholme, 2023). However, it should be noted that there is an extreme skewing in esports prize money distribution, where winners receive a disproportionate amount of earnings compared to other competitors (Coates & Parshakov, 2016), alongside extremely poor job security and career lifespan (discussed by Smithies et al., 2020), altogether placing an extreme emphasis on performance maximisation for esports players and organisations alike.

In light of this, there is ever increasing interest towards understanding the human factors that influence esports performance in order to maximise chances for success; both in field-based and laboratory settings. To demonstrate the latter, I turn to a database of 566 journal articles, all of which include the word *esport\** (with the asterix meaning truncation here and throughout; so this search would include perform, performs, performance etc.) in their title or abstract and are indexed by the large multidisciplinary databases Embase, Ovid Medline, and Web of Science as of April 23, 2023 (this database was created for a book chapter on research methods in esports; see **x. List of Publications**). This database includes articles published between 2013 and 2023, and covers all topics, including arts, economics, engineering, law, philosophy, psychology, social sciences, and sport sciences. Of the 566 articles, 134 mention performance (*perform\**) in their title and abstract, emphasising that (unsurprisingly) performance is a key theme within much of esports research. One human factor which has received interest within the short lifespan of esports research is sleep. As many as 25 peer-reviewed scientific articles (from the abovementioned database) discuss sleep (*sleep\**) in the title or abstract as of June 20, 2023.

#### **1.4. An introduction to sleep**

For most purposes, sleep can be simply defined as “a reversible behavioural state of perceptual disengagement from an unresponsiveness to the environment” (Sullivan et al., 2021, p. 16)<sup>†</sup>. In humans, sleep is mostly and normally (but not exclusively) experienced

<sup>†</sup>A universally agreed upon definition of sleep does not exist, which is a persistent issue in sleep research and medicine, particularly when expressing or lobbying for more emphasis of sleep in population health conversations and within major health organisations. For a brief introduction to the difficulty of defining sleep, see Siegel (2021).

nocturnally and monophasically (i.e. one ‘block’ of sleep per 24-hour cycle, as opposed to multiple dispersed blocks or ‘polyphasic’ sleep) in a recumbent posture, with closed eyes and a marked reduction of behavioural activity.

While initial forays into human sleep research largely concluded that sleep was a passive and idling state (see Pelayo & Dement, 2021 for a detailed history of sleep research), a paper published in 1953 (Aserinsky & Kleitman, 1953) and a cascade of subsequent research has ultimately led to the key understanding that sleep has two distinct sleep states; REM and non-Rapid Eye Movement (NREM) Sleep. NREM sleep can be further broken down into stage N1 (colloquially referred to as ‘light sleep’, and normally accounting for 2-5% of sleep in healthy young adults), N2 (normally 45-55% of sleep), and N3 (often call slow wave sleep or SWS in human sleep literature, and colloquially referred to as ‘deep sleep’; normally 10-20% of sleep)<sup>†</sup>. Together with REM (normally 20-25% of sleep), these are the four sleep ‘stages’. Broadly speaking, in humans, stage N2 and N3 sleep are most commonly associated with memory and neuroplasticity, with stage N3 additionally being associated with physical repair, growth, and immune system functioning, while REM sleep is associated with emotional memory consolidation and emotional regulation. A high proportion of stage N1 sleep is often associated with disordered sleep (i.e. obstructive sleep apnoea or periodic leg movement disorder; (Sullivan et al., 2021)).

In normal human sleep, these sleep stages are progressed through in a somewhat predictable pattern, which last for ~90 minute sleep ‘cycles’ (Dement & Kleitman, 1957). However, the proportion of certain sleep stages within each sleep cycle changes (predictably) across a night of sleep, such that most stage N3 sleep (linked to Process S, see **section 1.8**) is experienced within the first third of a nighttime sleep bout, and REM sleep (linked to Process C, see **section 1.8**) bouts are generally longest within the last third of a nighttime sleep bout (Sullivan et al., 2021). The pattern of one’s transition through sleep stages is referred to as sleep architecture.

The amount of nighttime sleep needed for optimal functioning contains considerable individual variability due to genetic factors (Franken et al., 2001), and is heavily influenced by extraneous factors such as prior daytime activity (Horne & Minard, 1985)

<sup>†</sup>The American Academy of Sleep Medicine (AASM) guidelines for defining and scoring NREM sleep stages changed in 2007; previously, NREM sleep was separated into four stages, with ‘slow wave sleep’ having two distinct stages (Stage 3 and Stage 4) instead of just one (N3). Thus, the sleep stages provided within the description about are not necessarily reflective of those used in past or contemporary literature.

or prior sleep history. However, the National Sleep Foundation provide a general recommendation (derived through a systematic review and Delphi type consensus voting) for young adults of seven to nine hours of nighttime sleep (Hirshkowitz et al., 2015). The National Health Foundation have also provided recommendations for various sleep quality related measures, including sleep onset latency (SOL), number of nighttime awakenings, wake after sleep onset (WASO), sleep efficiency (SE%), and proportions of nighttime sleep spent in different sleep stages (Ohayon et al., 2017). It is noted that the recommendations are provided across a spectrum of ages, owing to the well understood changes to normal human sleep across the lifespan (Miner & Lucey, 2021).

While for the majority of the thesis sleep will be discussed with respect to its relationship to task performance (i.e. esports performance), sleep (and the myriad of relevant themes around it; sleep disorders, sleep quantity and quality, sleep timing and variability, circadian factors etc.) is a component of the human experience which's importance spans well beyond its influence on performance. Indeed, reduced sleep quality/ quantity, disturbed and/ or disordered sleep have been linked to reduced academic performance (Dewald et al., 2010) and quality of life (Baldassari et al., 2008), and an increased risk of obesity and diabetes (Anothaisintawee et al., 2016; Cappuccio et al., 2008; Fatima et al., 2016; Itani et al., 2017), hypertension (Itani et al., 2017; Wang et al., 2015), stroke (Johnson & Johnson, 2010), markers of systemic inflammation (Irwin et al., 2016; Nadeem et al., 2013), memory impairment (Wallace & Bucks, 2013), dementia (Shi et al., 2018), work injuries (Uehli et al., 2014), motor vehicle crash (Tregear et al., 2009), risky behaviours (Short & Weber, 2018), suicidal ideation and behaviours (Chiu et al., 2018; Harris et al., 2020; Liu et al., 2020; Pigeon et al., 2012), and all-cause mortality (i.e. mortality, irrespective of reason) (Cappuccio et al., 2010; da Silva et al., 2016; Gallicchio & Kalesan, 2009; Itani et al., 2017). Lastly, there is an increasing understanding of the importance of sleep regularity, with a recent analysis of ~61,000 individuals finding it to be a better predictor of all-cause mortality than sleep duration (Windred et al., 2023).

### **1.5. Sleep in the esports context**

Despite the well understood importance of sleep for human health, dialogue around sleep within an esports context remains primarily focused on esports performance. Indeed, an aptly titled seminal article regarding sleep and esports “Sleep and performance in Eathletes: for the win!” explored potential risk factors of sleep disturbances in esports,

however did so from a perspective that such disturbances may lead to esports performance deficits (Bonnar, Castine, et al., 2019). The authors argued that while sleep is being increasingly recognised as a factor which impacts traditional sport performance, esports performance is more greatly founded in cognitive abilities than almost all traditional sports (which rely on a higher degree of physical performance). As cognitive performance has been demonstrated to degrade more greatly than physical performance with sleep loss (i.e., Fullagar et al., 2015; How et al., 1994; Pilcher & Huffcutt, 1996), it is logical to hypothesize that sleep loss may have a greater detrimental impact on esports performance relative to performance in traditional sports. This logic is extended within many reviews and original research articles around sleep and esports (Bonnar, Lee, et al., 2019; Bonnar et al., 2022; Goulart et al., 2023; S. Lee et al., 2021; Moen et al., 2022; Sanz-Milone et al., 2021). Clearly, researchers, practitioners, and professionals have an interest in optimising esports athletes sleep that is motivated (at least to a large degree) by performance optimisation reasons.

Additional reasons for sleep being highlighted as a performance are (a) a series of studies reporting concerning sleep behaviours at a group level within esports populations, and (b) the existence of risk factors for reduced/ disturbed sleep which includes those already highlighted for traditional sport athletes (Walsh et al., 2021) but also which spans beyond this. I will briefly summarise the patterns in esports athlete sleep behaviours below, but I implore interested readers to read the original articles cited for further context and information.

Generally, esports athletes experience habitual nighttime total sleep times (TSTs; 6.5-8 hours) comparable to those described for other young adult populations (Bonnar et al., 2022; Gomes et al., 2021; S. Lee et al., 2021; Moen et al., 2022; Thomas et al., 2019). However, most notable is severely delayed sleep onset and wake times; while some studies have reported already delayed group means/ medians of ~2am and ~10am respectively (Gomes et al., 2021; Moen et al., 2022; Sanz-Milone et al., 2021), others have reported even later mean habitual sleep onset ~5am and wake ~12pm times, with evidence of regional differences regarding these values (Bonnar et al., 2022; S. Lee et al., 2021). Unsurprisingly, a high proportion (~60%) of esports athletes present as evening chronotype (Gomes et al., 2021; Sanz-Milone et al., 2021). I note that while TST appears largely normal, values on sleep quality measures such as SOL, WASO (both >30min; Bonnar et al. (2022); S. Lee et al. (2021)) and SE% (as low as 68%; Moen et al. (2022))

are below those recommended by the National Sleep Foundation for young adults (Ohayon et al., 2017).

Regarding risk factors for reduced/ disturbed sleep, I firstly note that with the exception of early morning training, all traditional sport athlete specific risk factors outlined in a 2021 British Journal of Sports Medicine sleep and athlete consensus statement (pre-competition cognitive arousal, long-haul travel and unfamiliar sleeping environment following such travel, night competition, and high training loads; Walsh et al. (2021; see Figure 1), as well as caffeine use during competition, appear to be relevant to some degree in elite esports. However, the esports environment seems to result in further unique risk factors. These are discussed in depth by Bonnar, Castine, et al. (2019) and Bonnar, Lee, et al. (2019), however to highlight some highly pertinent factors; esports are (a) experienced through blue-light emitting computer monitors which, when used during the evening/ nighttime, may suppress endogenous melatonin secretion, potentially delaying circadian phase and reducing sleep quality/ quantity (Green et al., 2017; Schöhlhorn et al., 2023), and (b) cognitively/ physiologically arousing by design. The combination of these two factors have been previously highlighted as a mechanism potentially explaining gaming frequency/ duration and poor sleep outcomes (Kemp et al., 2021). Behavioural factors likely further compound this risk; there appears to be a ‘culture’ within elite esports which promotes (and may necessitate at times†) play into the late night/ early morning (Bonnar, Lee, et al., 2019; Lee et al., 2020), and simultaneously, esports athletes are seemingly quite unwilling to participate in sleep monitoring/ hygiene practices (Bonnar et al., 2023).

## **1.6. Sleep loss and cognitive performance – a speedrun**

While the notion that sleep loss impacts cognitive performance to a greater degree than physical performance is true (or, at the least, there is far more robust evidence for effects on cognitive vs. physical performance), this statement is a large generalisation made towards a very rich and diverse scientific field. *Cognitive performance* refers not to one domain but a plethora of abilities, tested using a plethora of means. Despite the fact that with every year of scientific enquiry we develop a stronger understanding of the

†Much of esports practice is undertaken through scrimmaging (scrims) against other teams. The timing of this practice is limited by the availability of other teams. Hence, avoiding late night scrims (which are often the norm; Bonnar, Lee, et al. (2019)) may limit quality training opportunities. I conclude this point by noting that ~50% of coaches/ support staff consider night training schedules to be a condition that impacts the sleep of their esports athletes (Bonnar et al., 2023).

mechanisms underlying the effect of sleep loss on performance for various cognitive domains, there is still a great deal which is not yet understood. As a result, many (competing and complementary) theories persist, which attempt to shed light on such mechanisms. The purpose of this section is to provide a brief overview of current understanding of sleep loss and cognitive performance, some discussion of which is expanded upon in future chapters.

Regarding cognitive domains which are more/ less sensitive to sleep loss, multiple meta-analyses (alongside earlier seminal work) have shown a general trend that as *task complexity* increases, the adverse effect of sleep loss decreases (Glenville et al., 1978; Harrison & Horne, 2000; Lim & Dinges, 2010; Lowe et al., 2017; Pilcher & Huffcutt, 1996; Wickens et al., 2015). This is exemplified by the fact that the test most universally used to assess cognitive performance changes in sleep research (owing largely to its sensitivity to sleep loss) is the psychomotor vigilance task (PVT); a task which only has one possible stimulus, and one possible response mechanism. This is not to state that performance on highly complex tasks is immune to sleep loss induced performance deficits; rather, effects tend to be smaller, can be more easily compensated for, and appear meaningful following only more severe bouts of sleep loss. As stated by Lim and Dinges (2010) (following a meta-analysis on sleep deprivation and cognitive performance), “although total SD [total sleep deprivation; discussed in **Section 1.7**] does produce statistically significant differences in most cognitive domains, the largest effects are seen in tests of simple, sustained attention” (p. 13). Among more complex tasks, one aspect of cognition which has received increased attention in the past decade for its supposed sensitivity to sleep loss is *cognitive flexibility* (i.e., Harrison and Horne (2000); Honn et al., 2019); Lawrence-Sidebottom et al. (2020); Stenson et al. (2023); Whitney et al. (2015); Whitney et al. (2019); Whitney et al. (2017)). Cognitive domains and task characteristics which are sensitive/ robust to sleep loss (in particular, acute sleep restriction) will be discussed in greater detail in **Chapter 2**.

As previously mentioned, the mechanisms for how sleep loss impact aspects of cognition are not completely understood, and hence are constantly a subject of research. Regarding simple attention task performance, three predominant (non-exhaustive and non-exclusive) theories have been proposed and expanded upon throughout the history of sleep research. All three of these theories provide explanation as to why sleep loss leads to worsened performance on simple attention taxing tasks. The first of these theories is the *lapse hypothesis* (Williams et al., 1959). This hypothesis expands upon work as far

back as 100 years ago (Bills, 1931; Bjerner, 1949; Kleitman, 1923; Lee & Kleitman, 1923; Warren & Clark, 1937; Williams et al., 1959) and suggests that sleep loss results in brief *blips* in physiological arousal (called *blocks* in seminal research), and hence periods of reduced responsiveness to stimuli. The second theory, called the *state instability hypothesis*, extends this logic by positing that global (i.e., entire brain) sleep and wake states (governed by interactions between sleep history and circadian processes, which are elaborated on in **Section 1.8**) and top-down compensatory effort to maintain wakefulness, are highly unstable and can rapidly fluctuate when sleep drive is high, resulting in variable performance (and increased frequency of these *blips* or *lapses*) (Doran et al., 2001). This theory helps to explain the effects of sleep loss on task performance at a *global* (i.e., whole brain) level. More recently, a third theory has been presented, the *local* sleep hypothesis, which suggests that network circuitry which is constantly used due to repetitive task demands (i.e., use-dependent) may be faster to enter a local sleep state (a state akin to slow-wave sleep but localised to specific neural assemblies; see Krueger et al. (2008) and Vyazovskiy et al. (2011)) when prior sleep loss is experienced (Hudson et al., 2020; Van Dongen, Belenky, et al., 2011). The latter two theories help explain how sleep loss expedites (makes occur earlier) and exacerbates (makes the effect stronger) the vigilance decrement (or *time-on-task* effect; stating that task performance becomes poorer and more variable over prolonged/ sustained bouts). The theories also help explain an increase in errors on time-sensitive tasks and the dramatic increase in responses >2 times slower than average on the PVT for example, but do not explain the observed general slowing of responses across the entire administration of the test, as demonstrated by the sensitivity of the fastest 10% response time measure of the PVT to sleep loss (Basner & Dinges, 2011; Belenky et al., 2003; Dinges & Kribbs, 1991; Dinges & Powell, 1988, 1989; Loh et al., 2004).

These theories also do not explain the observed decreases in cognitive performance from sleep loss which are dissociable from decreases in vigilance/ general alertness. For explanations of these performance decrements, neuroimaging techniques have been utilised. From such research, some notable findings have included reductions in brain activation within the central executive (dorsolateral prefrontal (PFC) cortex, intraparietal sulcus, posterior parietal areas) and salience (insula, medial frontal cortex) networks (Krause et al., 2017; Ma et al., 2015).

Additionally, the functional connectivity of various brain areas (especially those with connections to the hippocampus) have been reported to be altered under sleep loss, as has

the functional segregation of networks; referring to the ability to simultaneously engage some networks and disengage others when desirable for a specific function. In the case of sleep loss, diminished ability to disengage the default mode network is observed during attention-based tasks. Changes in activation levels to reward and losses have also been noted following sleep loss. Given the breadth and complexity of this topic, I will not expand further but instead guide readers to two key reviews within this area (Krause et al., 2017; Ma et al., 2015). Such changes in the activation of brain areas and networks could be expected to result in worsened performance within cognitive domains actively requiring them, such as attentional capacities, working memory, and executive functioning. However, research has demonstrated that performance on many complex cognitive tasks appears robust to mild bouts of sleep loss<sup>†</sup>.

By again turning to neuroimaging, results from studies have suggested that increased activation of certain brain areas (in particular, parietal areas, thalamus and frontostriatal circuitry) during cognitive performance tasks following sleep loss may contribute to the observed performance preservation on more complex tasks under mild bouts of sleep loss (i.e., Chuah et al., 2006; Drummond et al., 2000; Drummond et al., 2004; Drummond et al., 2005; Nakashima et al., 2018). Overall, it is hypothesised that additional resources, both within and outside of the task-specific areas/ networks, are recruited to overcome the diminished efficiency of those normally used for task performance. This understanding is consistent with explanations of task-specificity and individual differences within the aforementioned local sleep hypothesis. Networks required for the performance of different tasks may have varying levels of redundancy/ spare capacity that can be called upon within a bout of sleep loss, and individuals may have varying amounts of redundancy available within the same given network (Hudson et al., 2020).

However, some researchers argue that compensatory mechanisms may function to preserve some cognitive functions at the expense of others. Within their *dynamic attentional control framework*, Whitney et al. (2019) propose that the compensatory mechanisms specific to frontostriatal circuitry prioritise the maintenance of task-relevant information over the changing or updating of such information, such that performance on tasks *not* demanding attentional shifts or adaptation to changing task requirements (i.e., stable tasks) will be maintained, while performance on tasks with these requirements (i.e., flexible tasks) will be degraded. This is purportedly through an increase in tonic striatal

<sup>†</sup>I use  $\leq 36$ hrs as the cutoff for *mild* bouts of total sleep deprivation here and throughout, due its frequent use (i.e., Harrison & Horne, 1999; Horne & Pettitt, 1985; Whitmore & Fisher, 1996) and due to bouts beyond this being considered only in highly specialised circumstances (i.e., military operations).

dopamine levels leading to poorer phasic dopamine signalling, which is considered critical for cognitive flexibility (Grace, 2000; Whitney et al., 2019).

It is critical to outline some factors which may influence the extent to which performance compensation occurs under conditions of sleep loss. As performance maintenance following sleep loss is associated with compensatory recruitment of cognitive resources, it can be thought of as an increase in effort, and therefore *cost*, within a *cost-benefit* neuroeconomic framework (this idea is described in detail by Massar, Lim and Huettel (2019)). Simply, a cost-benefit neuroeconomic framework posits that actions taken are proportional to a subjective value, which is determined by an individual's perception of perceived benefit (task completion, victory, monetary rewards) and perceived cost (time, risk, effort). This increase in cost associated with sleep loss is neatly demonstrated by studies reporting larger monetary rewards to be required to convince participants to engage in effortful cognitive tests following sleep loss, when compared when participants are well rested (Libedinsky et al., 2013; Massar, Lim, Sasmita, et al., 2019). Through this lens, factors that increase one's intrinsic or extrinsic motivation may help to promote the maintenance of performance under sleep loss by increasing the perceived *benefit*. I note that these factors can be task-specific (i.e., how boring/ monotonous the task is, or whether feedback is provided) or context specific (whether a competition aspect is introduced, whether a monetary reward is available, or whether outcomes are critical for the safety of self or others). Regarding task specific drivers of motivation, I highlight the following quote from Harrison and Horne (2000) (p. 236); "the prevailing view in SD [sleep deprivation] research is that high-level complex skills are relatively unaffected by SD because of the interest they generate and the implicit encouragement for participants to apply compensatory effort to overcome their sleepiness".

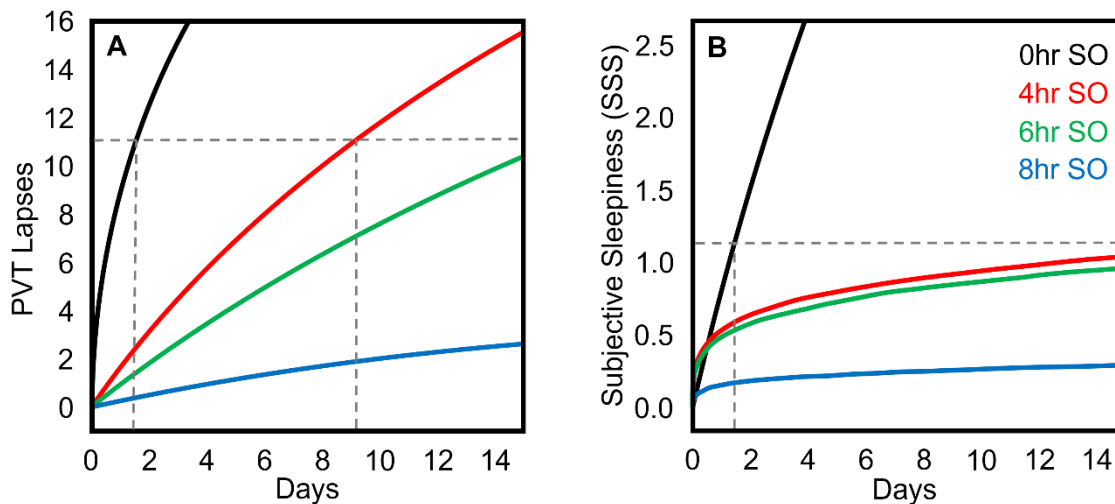
However, sleep loss can in turn reduce task-specific motivation (Mathew et al., 2021; Mikulincer et al., 1989; Odle-Dusseau et al., 2010), perhaps as a shift of prioritisation from task completion toward sleep-preparatory behaviours (Axelsson et al., 2020). Furthermore, while compensatory effort can maintain performance under sleep loss, there is an overall tendency to engage in less effortful tasks, actions, or strategies, when the choice is provided (Engle-Friedman et al., 2010; Engle-Friedman et al., 2003; Sullan et al., 2021). Lastly, it is important to stress that once a certain amount of sleep loss is achieved, performance will diminish in spite of any compensatory mechanisms (though most experimentally manipulated sleep loss studies do not reach this threshold; Dinges and Kribbs (1991); Horne and Pettitt (1985)).

## 1.7. Forms of sleep loss

Until now, I have intentionally used the vague term *sleep loss* throughout. However in reality, just like cognitive performance, sleep loss is not a uniform concept but can be experienced in a multitude of forms. These forms are neatly divided into three categories by Reynolds and Banks (2010); total sleep deprivation (TSD; also sometimes called extended wakefulness, or simply sleep deprivation), sleep restriction (SR; sometimes called partial sleep deprivation), and sleep disruption (or sleep fragmentation). TSD refers to the total elimination of sleep, normally for 24 hours or more. It is by far the most commonly studied form of sleep loss in experimental research. SR refers to the reduction of sleep quantity below that normally experienced for one or more nights. SR is much more commonly experienced than TSD for both the general population (Banks & Dinges, 2007) and for individuals in many specialised occupations and environments (Caldwell et al., 2012; Capaldi et al., 2019), however is far *less* commonly studied than TSD, predominately due to the extreme logistical boundaries of studies imploring multiple days of SR (Banks & Dinges, 2007). To avoid confusion, throughout the manuscript I will refer to one day to two weeks of SR as *acute SR*, and SR spanning beyond two weeks as *chronic SR*. Sleep disruption specifically refers to frequent arousals and wake periods during nighttime periods, which may reduce sleep quantity but also reduce sleep quality. While sleep disruption can certainly hinder performance (Bonnet & Arand, 2003; Kahn et al., 2014), it is generally associated with disordered sleep (Reynolds & Banks, 2010), and hence is largely outside the scope of the current thesis.

Though both TSD and acute SR are forms of sleep loss, there are two key differences in the effects they tend to produce. Firstly, while subjectively reported alertness and sleepiness tend to closely follow objectively measured vigilance under conditions of TSD, these subjective outcomes underestimate performance degradation under SR (Banks et al., 2010; Belenky et al., 2003; Van Dongen et al., 2003; also see Figure 1-1). Secondly, performance recovery following SR takes much longer than a bout of TSD inducing an identical performance deficit (Banks et al., 2010; Belenky et al., 2003); a finding with fundamental implications for (bio)mathematical modelling of performance under conditions of sleep loss (McCauley et al., 2009). These two findings combined have led to extreme concern regarding the performance, safety and wellbeing of individuals operating in critical or safety-critical environments; essentially, SR can result in a situation where people aren't fully aware of their impairment, and take many days to

recover to baseline. Despite these differences between TSD and SR and nuances in the biological mechanisms underpinning impairment from each, the effects of TSD and SR on task performance during the actual bout of sleep loss appear to be similar and equatable (Banks & Dinges, 2007; Van Dongen et al., 2003; Figure 1-1). An example of such can be found as Figure 1-1, which shows data equating performance within PVT and subjective sleepiness as measured by the Stanford Sleepiness Scale (SSS) following TSD and bouts of SR of various severities.



**Figure 1-1** **A** Instances of Lapses on a 10 minute psychomotor vigilance task (PVT), and **B** subjective sleepiness scores on the Stanford Sleepiness Scale (SSS), across multiple days and for participant groups with varying amounts of sleep opportunity (SO) afforded. In both graphs, higher values denote greater impairment. The grey horizontal line depicts the expected value for each graph following ~36hrs of total sleep deprivation (TSD), denoted by the leftmost vertical dashed line in each graph. In **A**, the rightmost vertical dashed line denotes the equivalent amount of days (~9) with 4hrs SO required for predicted PVT lapses to be equivalent to that at ~36hrs TSD. This figure is adapted from Van Dongen et al. (2003).

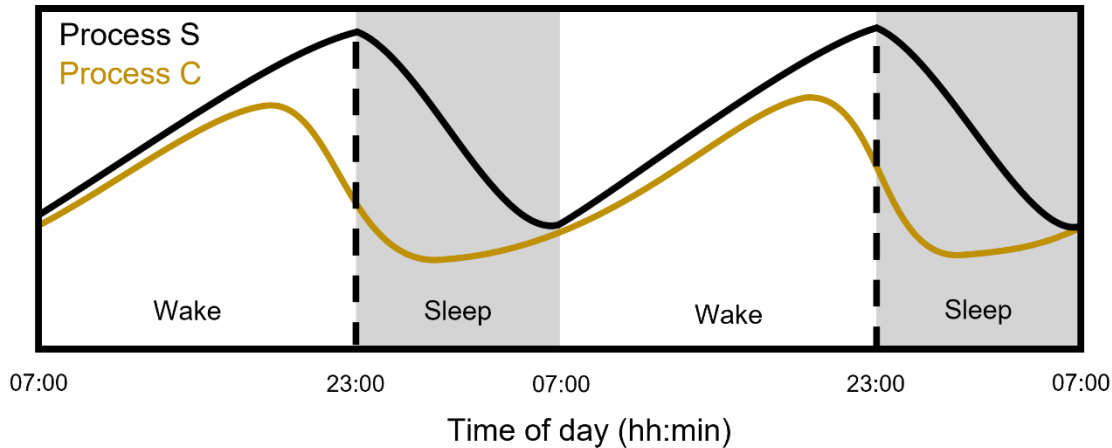
## 1.8. A word on circadian rhythms of performance

While the current thesis concerns itself primarily with sleep loss (in relation to effects on cognitive and in-game performance among esports players), it would be remiss to not briefly introduce and discuss the role of circadian rhythms on cognitive performance. Circadian (a word derived from the Latin phrase “Circa diem”, meaning “around a day”)

rhythm refers to biological processes occurring with a period of approximately 24-hours. Within the human body, circadian rhythm's can be thought of as an orchestra – where individual cells throughout the body are the musicians, each with nuanced differences in natural rhythms if left without a conductor, who is present to synchronise the pace and tempo of the players (Morrow & Harrington, 2020). Within this analogy, the conductor is the suprachiasmatic nucleus (SCN) located within the anterior hypothalamus, with eventual influence over almost all cells within the human body (Foster & Kreitzman, 2014). The primary role of the circadian rhythm is to augment the timing of biological processes, such that they occur at the most ideal *time of day*. Circadian rhythms are entrained by *zeitgebers* (German for *time giver*, and referring to external influences), however are ultimately endogenous and are still present with a ~24hr period without zeitgeber influence. By far the most influential zeitgeber is light (received through the eyes), with the intensity and spectral properties of light further influencing its impact on human biology (Roenneberg et al., 2013).

Sleep and circadian rhythms affect human neurobiology, subjective experience, and objective performance, both independently and synergistically. This relationship is described by the *two-process model*, which was proposed by Borbély (1982) and is still widely used today in describing sleep-wake regulation, sleep architecture, and alertness/human performance. The two-process model describes an interaction between the homeostatic process (Process S), a process which builds (normally modelled linearly) with every moment awake and diminishes with sleep, and the circadian process (Process C), referring to circadian rhythm (Figure 1-2). Subjective alertness and performance on cognitive tasks tends to be greatest when Process S is low, and poorest when Process S is high, and peaks (*acrophases*) and valleys (*bathyphases*) over the period of Process C (Dijk et al., 1992; Van Dongen & Dinges, 2003). The distribution of task-performance acrophases and bathyphases across a day is related to one's chronotype (i.e., Rae et al., 2015), or highly individualised preferences in rest/ activity timing that are influenced by both biological and environmental factors (Kunorozva et al., 2017; Montaruli et al., 2021; Roenneberg et al., 2003; Shawa et al., 2018). While both Process S and Process C affect humans independently, they also interact, such that when sleep pressure is high, the influence of Process C on subjective and objective markers of performance appears to be greatly amplified (Dijk et al., 1992; Van Dongen & Dinges, 2003). Task specificity of performance acrophases and bathyphases can also be explained within the framework of the two-process model. The current best evidence regarding task-specificity of peak

performance acrophases and bathyphases (for cognitive predominant tasks) indicates task-independence across one's circadian rhythm, with temporal changes between task acrophases/ bathyphases resulting from differential influences of Process S (Johnson et al., 1992; Monk et al., 1997; Muck et al., 2022).



**Figure 1-2** The two-process model, originally described by (Borbély, 1982). This model is used today to describe how sleep history and circadian rhythm interact to influence ones sleep drive, alertness, and performance. Regarding sleep drive, the larger the distance between Process S and Process C, the greater the sleep drive, with this distance being reduced through nighttime sleep.

### 1.9. Is acute sleep loss a performance concern in esports?

Circling back to the presumptions of researchers and esports athletes outlined alike in **Section 1.5**, that sleep loss negatively impacts esports performance; it is clear that relationships between sleep loss and performance (in any element) are highly nuanced and multifactorial. The literature does not suggest that any amount of sleep loss will *certainly* negatively impact esports performance. Such complexities warrant further investigation into pertinent themes around sleep loss and esports performance, as well as an empirical investigation into the relationship itself. As such, the **overall purpose** of this thesis is to shed light on if and how sleep loss impacts the ability of esports athletes to play at their best. To address this purpose, three key investigations are outlined in the current thesis:

The first key investigation was spawned from a desire to understand how the current literature would suggest sleep loss to impact esports performance, with prior knowledge that such research had not yet taken place to date. To execute this investigation, I

performed a systematic search and review of the scientific literature (alongside grey literature sources, as per systematic search protocol) of studies exploring the impact of sleep restriction on cognitive performance; however, key criteria were placed on population and outcomes explored, in attempt to maximise relevancy to esports. Specifically, I included studies only using *Elite Cognitive Performer* populations; that is, populations within occupations or positions in which they are required to perform cognitively demanding tasks with critical or safety-critical outcomes. Furthermore, I included performance on occupation-specific cognitively demanding tasks as outcomes of consideration. SR was chosen as the sleep loss mode of choice (compared to TSD) due to its increased likelihood of being experienced by esports athletes. This review is provided as **Chapter 2**, and published in the peer-review journal *Sleep*.

The second key investigation spawned from the understanding that in order to explore how sleep loss impacts esports performance, an ability to rigorously measure esports performance is first required. The ability to obtain relevant measures of overall game outcome were important, but obtaining an understanding of in-game measures which actually influence performance may help shed light on any whether in-game strategy changes occur under conditions of sleep loss. While game outcome and performance indicator metrics were already present for major multiplayer online battle arena (MOBA) esports (Novak et al., 2020; Xia et al., 2017), these esports are not particularly conducive for use in experimental research, due to long and unpredictable match lengths and team-based competition. Rocket League, however, is an esports particularly suitable for experimental research, owing to the fact that match lengths are short and predictable, it can involve solo competitive game play (i.e., 1v1), and boasts exceptional data availability. Since Rocket League game outcome and performance metrics have not been previously examined, **Chapter 5** will outline my use of a contemporary machine learning approach to identify these metrics; this work is also published in the peer-review journal *Scientific Reports*.

The third key investigation is into how experimentally induced sleep loss (in the form of TSD) impacts both the cognitive and in-game performance of *Rocket League* players. This study makes direct use of the investigations preceding it (through its choice of cognitive tasks & in-game performance measures used). It also directly addresses the assertions made within previous sleep and esports literature (Bonnar, Castine, et al., 2019; Bonnar, Lee, et al., 2019; Bonnar et al., 2022; Goulart et al., 2023; Kemp et al., 2021; S. Lee et al., 2021; Moen et al., 2022; Sanz-Milone et al., 2021), and answers the call for

research into sleep loss and its direct effects on in-game performance (S. Lee et al., 2021). It is, to date, the only formal investigation into experimentally induced sleep loss and esports performance, and is described in **Chapter 7**.

### **1.10.Purpose, research questions, hypotheses**

The overarching aim of the thesis is to understand how acute sleep loss may impact esports performance. Specific research questions covered, and the hypotheses associated, are listed below:

#### **Chapter 2:**

**Research Question:** How does the current scientific literature suggest that acute sleep restriction impacts both the cognitive and occupation specific cognitively demanding tasks, for individuals in populations who must perform such tasks with critical outcomes as part of their occupation (i.e. Elite Cognitive Performers)?

**Hypothesis:** No formal hypothesis is provided.

#### **Chapter 3:**

**Research Question/s:** What is the test-retest reliability of the Category Switch Task (CST) on various reaction time and error-based outcome measures? Which outcome measures on the Category Switch Task (CST) experience test-retest practice effects? Do such practice effects vary as a function of test-retest interval (*same day, next day, next week*)?

**Hypotheses:** Performance in all outcome measures would improve from test to retest, and shorter test-retest intervals would be conducive to larger practice effects than longer test-retest intervals.

#### **Chapter 5:**

**Research Question/s:** What is a suitable match outcome measure in 1v1 Rocket League? Which in-game metrics best predict this match outcome measure, and how do these vary as a function of player ability level? Which in-game metrics best predict player ability level?

**Hypotheses:** No formal hypotheses are provided.

## **Chapter 6:**

**Research Question/s:** What is the optimal simple imputation strategy for missing actigraphy-derived sleep data?

**Hypotheses:** No formal hypotheses are provided.

## **Chapter 7:**

**Research Question/s:** Does a ~29 hours of total sleep deprivation (TSD) reduce worsen the cognitive performance (vigilance and task-switching performance) of Rocket League players? Does such a bout of TSD reduce in-game Rocket League performance? Are Rocket League performance indicators impacted by this bout of TSD?

**Hypotheses:** Cognitive and in-game Rocket League performance will be negatively impacted by ~29 hours TSD.

## **Chapter 8:**

**Research Question/s:** Which Rocket League in-game metrics are perceived to best differentiate both *safe vs. risky* and *simple vs. complex* playstyles? Do a *safer* or *simpler* (or both) strategy changes best explain variance between Rocket League matches played with both participants rested vs. one participant sleep deprived?

**Hypotheses:** There will be substantial overlap between in-game metrics that distinguish *playstyle risk* and *playstyle complexity*, and both perceived *simpler* and *safer* strategy derived metrics will explain changes in Rocket League matches played with both participants rested vs. one participant sleep deprived.

## **Chapter 2. The Effect of Sleep Restriction on Cognitive Performance in Elite Cognitive Performers: A Systematic Review**

This chapter has been published in a modified format in *Sleep*:

Smithies, T. D., Toth, A. J., Dunican, I. C., Caldwell, J. A., Kowal, M., & Campbell, M. J. (2021). The effect of sleep restriction on cognitive performance in elite cognitive performers: a systematic review. *Sleep*, 44(7), zsab008. DOI: <https://doi.org/10.1093/sleep/zsab008>

Changes to the abovementioned publication for the purposes of this thesis are outlined below:

- Change in referencing style (article version is in numbered format).
- References to supplementary files are changed to the appropriate location within the appendix.
- Words emphasised using quotation marks were changed to be emphasised using italics, in line with the thesis format.
- The words *Figure* and *Table* in in-text references to figures was capitalised. Furthermore, figure/ table numbering convention was changed in line with the thesis format.
- Abbreviations representing authors were changed where needed to avoid conflict with other abbreviations used throughout the thesis.
- Addition of a *linking section* for the purpose of thesis flow.
- Minor amendments have been made based on examiner correction suggestions.

## 2.1. Abstract

**Study Objectives:** To synthesise original articles exploring the effects of sleep restriction on cognitive performance specifically for *Elite Cognitive Performers*; i.e. those who engage in cognitively demanding tasks with critical or safety-critical outcomes in their occupation or area of expertise.

**Methods:** Backward snowballing techniques, grey literature searches, and traditional database searches (Embase, MEDLINE, Web of Science, Google Scholar, PSYCinfo, and SportDiscus) were used to obtain relevant articles. A quality assessment was performed, and risk of training effects was considered. Results were narratively synthesised. Fourteen articles fit the criteria. Cognitive outcomes were divided into three categories defined by whether cognitive demands were ‘low-salience’, ‘high-salience stable’, or ‘high-salience flexible’.

**Results:** Low-salience tests (i.e., psychomotor vigilance tasks & serial reaction tests), mainly requiring vigilance and rudimentary attentional capacities, were sensitive to sleep restriction, however this did not necessarily translate to significant performance deficits on low-salience occupation-specific task performance. High-salience cognitive outcomes were typically unaffected unless when cognitive flexibility was required.

**Conclusions:** Sleep Restriction is of particular concern to occupations whereby individuals perform (a) simple, low-salience tasks or (b) high-salience tasks with demands on flexible allocation of attention and working memory, with critical or safety-critical outcomes.

**Keywords:** *vigilance, cognitive flexibility, occupation, safety-critical, attention, sleep restriction.*

**Statement of Significance:** Sleep restriction is considered a significant concern to performance on cognitively demanding tasks within occupations that involve such tasks (i.e. pilots, air traffic controllers, surgeons, medical residents, emergency responders, process operators, athletes). However, no review to date has focused specifically on these populations, outlining the results of research exploring how the performance of these individuals is impacted by sleep restriction. Our review systematically searches for and narratively synthesizes the current literature to date within these populations, and outlines how cognitive tests and occupational tasks of different demands are differentially

impacted by sleep restriction. Lastly, the review shows that more work is needed that examines the impact of sleep restriction on cognitive flexibility within these populations.

## 2.2. Introduction

Optimal cognitive functioning is fundamental for performance within many work environments. In select safety-critical occupations, the ability to perform complex, cognitively demanding tasks within unpredictable circumstances is integral to operational success. Active military personnel (Serfaty et al., 1997), aviation pilots (Adams & Ericsson, 2000), air traffic controllers (Hilburn, 2004), emergency responders (Paton & Flin, 1999), surgeons and medical practitioners (Patel et al., 1996; Schmidt et al., 1990), and process operators in potentially dangerous environments (i.e., mines, power plants, oil refineries) (Mumaw et al., 1994) are all examples of individuals involved in such safety-critical professions. Additionally, while elite athletes do not engage in safety-critical work, optimal cognitive functioning (i.e. attention, executive functioning, decision making) within time-constrained and unpredictable environments is often integral for elite performance (Janelle & Hillman, 2003; Williams et al., 2011). Individuals within these professions must exhibit cognitive expertise not normally present within the general population for operational success, given the complexities and cognitive demands embedded within the tasks involved. Individuals in some of the professions mentioned (i.e., athletes, pilots, air traffic controllers) have been shown to demonstrate enhanced cognitive performance compared to the general population not only within the context of their area of expertise, but also through laboratory testing (Arbula et al., 2016; O'Hare, 1997; Voss et al., 2010; Yildiz et al., 2014), though see an article by Taylor and colleagues (Taylor et al., 2005) for a contrary finding), particularly during task-switching, multitasking and attentionally demanding task paradigms. As a result of the aforementioned cognitive demands and the observed performance benefits these individuals may possess, we refer to them here collectively as *Elite Cognitive Performers (ECPs)*.

Sleep quantity has been identified as a key moderator of cognitive performance (Durmer & Dinges, 2005; Lim & Dinges, 2008, 2010; Lowe et al., 2017). To date, most sleep quantity research has concerned itself with total sleep deprivation (TSD; a total elimination of sleep obtained during a specified time period), primarily due to the time and cost efficiency of their designs (Banks & Dinges, 2007). However, TSD is uncommon ecologically, whereas sleep restriction (SR), referring to a moderate reduction in the amount of sleep across one or more nights (~2-6hr sleep obtained per night), is far more commonly experienced both by the general population (Banks & Dinges, 2007) and by

ECPs (Caldwell et al., 2012; Capaldi et al., 2019). The fact that SR is more frequently experienced than TSD, and that each affects human neurobiology differently (Banks & Dinges, 2007), has led more recent work to specifically focus on understanding the effects of SR on cognitive performance. In addition to the reviews assessing the effects of SR on cognitive performance among youth (Lundahl et al., 2015) and adolescent (de Bruin et al., 2017) populations, experimental sleep dose-response studies, such as those conducted by Belenky and colleagues (Belenky et al., 2003), Jewett, Dijk, Kronauer and Dinges (Jewett et al., 1999), Van Dongen, Maislin, Mullington and Dinges (Van Dongen et al., 2003), and Banks, Van Dongen, Maislin and Dinges (Banks et al., 2010), have provided comprehensive insight into the effects of SR on cognition. The results of studies such as by Belenky and colleagues (Belenky et al., 2003) and by Van Dongen and colleagues (Van Dongen et al., 2003), as well as other experimental research, have informed the creation of biomathematical fatigue models, used in safety-critical environments to identify periods of risk and, guide mitigation, and maximise performance (Hursh et al., 2004). Recently, Lowe, Safati, and Hall (Lowe et al., 2017), in a meta-analysis investigating the effects of SR on cognitive performance, found SR to impact *sustained attention* tasks more than increasingly complex tasks assessing performance in other cognitive domains across numerous populations and age groups. This finding corroborates those of Wickens, Hutchins, Laux and Sebok (Wickens et al., 2015), who noted that simple cognitive task performance is more greatly impacted by sleep loss, as well as earlier seminal research outlining the comparatively greater effects of sleep loss on simple tasks (Glenville et al., 1978).

That performance on simple tasks appears selectively hindered by SR initially seems counter-intuitive, as prefrontal cortex (PFC; integral to executive functioning) activation is decreased by sleep loss (Krause et al., 2017; Ma et al., 2015). However, imaging studies (using functional magnetic resonance imaging) have found strong evidence for increased recruitment of frontostriatal circuits and additional brain areas coinciding with the maintenance of performance during increasingly complex and engaging cognitive tests despite decreased PFC activation (Beebe et al., 2009; Chuah et al., 2006; Drummond et al., 2004; Drummond et al., 2005; Krause et al., 2017). Through this lens, simple attentional test performance tends not to receive similar compensation due to a lack of arousal, stemming from the low stimulus/salience nature of such tests (Harrison & Horne, 2000; Whitney et al., 2019). Recent work has suggested that these compensatory mechanisms function in a way so as to give preference to task information already present

within working memory, helping to maintain focus and attentional strategy throughout the task (i.e., cognitive stability). However, the trade-off appears to be that the ability to alter this information within working memory (i.e., cognitive flexibility), necessary for when attention needs to be shifted when a task dynamic changes (as is common within real-world tasks), is impeded (Whitney et al., 2019).

Despite the abovementioned literature outlining the effects of SR among general populations, it is less clear how SR affects the cognitive performance of ECPs or whether this group is differentially affected by SR. The importance of studying this group independently from the general population is three-fold. Firstly, optimal cognitive performance is arguably more important for ECPs than for the general population, as errors or inadequate performance can have critical outcomes, ranging from loss of competition for high-level athletes, to loss of life in safety-critical occupations. Numerous high-profile catastrophes have involved human errors linked to sleep loss, such as the fatal decision to launch the Space Shuttle Challenger in 1986. In the report on the *Presidential Commission on the Space Shuttle Challenger Accident* (1986), it was stated that prior to an important teleconference regarding the decision to launch (a decision proving to result in seven casualties), “key managers obtained only minimal sleep the night before the teleconference” (p. G5), which may have led to poor judgement contributing to the fatal decision to launch. Another example is the pervasiveness of fatigue in aviation, where it is estimated that fatigue contributes to 4-8% of aviation catastrophes (Caldwell, 2005).

Secondly, ECPs are at an increased risk of experiencing SR due to their occupational requirements. For example, sleep opportunity can be sparse and unpredictable throughout military combat operations, while other military-specific stressors, such as watch duty and field-based exercises, result in the frequent occurrence of SR (Capaldi et al., 2019). Commercial pilots often have demanding schedules, are constantly exposed to rapid time-zone changes, and often must obtain night-time sleep in uncomfortable cockpit environments, resulting in regularly experienced SR. Rapidly changing work schedules are common for air traffic controllers, causing drastic reductions in sleep quantity, with some operating with as little as 2 hours of sleep at times (Signal & Gander, 2007). Irregular and demanding shift work schedules can lead to SR for emergency medical practitioners (Cheng et al., 2014). Finally, elite athletes can experience SR due to the timing and intensity of training and competition schedules, as well as air-travel requirements, especially when travelling over multiple time-zones (L. Gupta et al., 2017).

Thirdly, contemporary literature has suggested that ECPs at a group level may demonstrate an increased resistance to the effects of sleep loss on cognitive performance. For example, one study found a group of seven active-duty F117 fighter pilots to have greater baseline global cortical activation compared to non-pilots during a working memory task, which then positively correlated with performance on a flight simulator task after 37 hours of TSD (Caldwell et al., 2005); however the authors advocated for further research on larger samples to validate such a finding. In reference to this, some authors have discussed the idea that naturally tolerant individuals to sleep loss may either *self-select* into, or that vulnerable individuals may *self-select* out of, active military professions due to the necessity of maintaining performance following sleep loss (Caldwell et al., 2012; Van Dongen & Belenky, 2009; Van Dongen, Caldwell, et al., 2011). Similar theories have been posited to explain a lack of performance degradation following sleep loss among medical residents (Schlosser et al., 2012; Veasey et al., 2002). It is noted that individual differences in tolerance to sleep loss within elite groups such as the U.S. Air Force are still present (Van Dongen et al., 2006).

Together, the importance of optimal cognitive performance for ECPs, their increased risk towards experiencing SR, and their potential increased tolerance to the performance effects of SR at a group level, all make the study of the effects of SR on cognitive performance in ECPs worthwhile. To date, no attempt has been made to review the existing literature examining the effects of SR on the cognitive performance of ECPs. As a result, the purpose of this review is to synthesize and summarize the existing literature explicitly examining the effect of SR on cognitive and occupation-specific performance among ECPs.

## 2.3. Methods

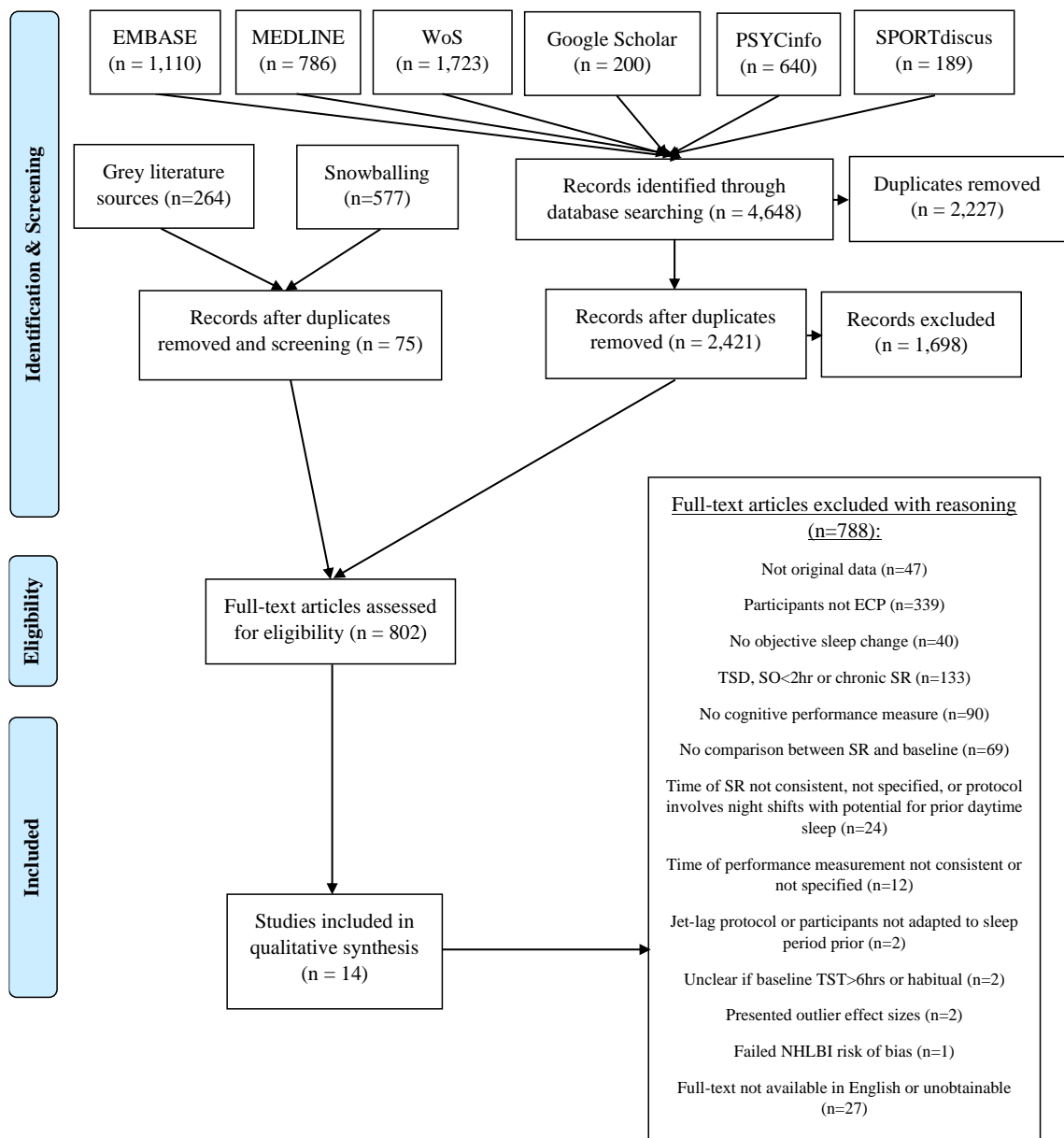
### 2.3.1. Database search strategy

This review was not registered prior to its undertaking. Included articles did not have to be published in peer-review scientific journals to be considered. Articles included for the current review were obtained through an exhaustive systematic search, in accordance with the updated PRISMA guidelines (Page et al., 2020). Embase, MEDLINE (Ovid MEDLINE(R) and ePub ahead of print, in-process & other non-indexed citations, daily and versions(R)), Web of Science (Core Collection), and Google Scholar databases were searched, as the combination of these four databases presents superior sensitivity/specificity trade-off for systematic searches (Bramer et al., 2017). Subject-specific databases APA, PSYCinfo, and SportDiscus (both EBSCO host) were also queried to add further sensitivity to the search. Searches using these databases took place on 27/01/2020, except for Google Scholar, which took place the next day. The exact syntax used for each primary database can be found as appendix 2.1. The search strategy for each database involved identifying *key-words* (22 total) within titles and abstracts pertaining to motor or cognitive abilities, or performance, and combining them with words pertaining to SR (5 total), with the exclusion of words related to animal studies, clinical conditions, or reviews. Controlled vocabulary terms (MeSH/EMTREE) were explored and used as exploded terms (searching for the particular word as well as the more specific words that stem from it within the given organisation system) where relevant in databases that allowed for them. Inbuilt database filters were used where available to remove studies specifically investigating nonhuman subjects, children, or the elderly; no date or language restrictions were enforced. TDS performed the search and screening described.

All identified article references were extracted and exported into Endnote version 9.2 (Clarivate Analytics), except for those found via Google Scholar, where only the first 200 references (when searched by relevance; as per Bramer et al. (2017)) were extracted. Overall, 4,648 articles were identified through this search process, with 2,421 remaining once duplicates had been removed (see Figure 2-1).

### 2.3.2. Grey Literature & Backward Snowballing

As some research concerning the effects of SR on performance among ECPs may not have been detected by the above database searches, an additional grey literature search was performed in addition to the use of *backward snowballing* techniques. Five sources of grey literature were queried; two conventional search engines (Google, duckduckgo), two grey literature specific databases (OpenGrey and Science.gov), and the Defence Technical Information Centre (DTIC). These searches took place between 31/01/2020 and 04/02/2020. For these searches, similar terms to those used in the primary database searches were used (see appendix 2.2 for the exact syntax used for each grey literature database search). For the DTIC search, the first 100 results were investigated, while for the other grey literature sources, the first 50 were investigated (or less, if less than 50 results appeared), in a similar fashion to that discussed for Google Scholar by Bramer, Rethlefsen, Kliejnen and Franco (Bramer et al., 2017). *Backward snowballing* refers to a technique where the reference lists of previously identified reviews or journal articles within a relevant topic are searched to obtain further relevant articles (Wohlin, 2014). Due to prior knowledge that many studies conducted in defence institutes are not published in peer-reviewed journals and are therefore not identified by primary database searches, reviews focussing on such studies were targeted for backward snowballing. Additionally, the references of two reviews on the effects of SR on cognition in the general population were also searched, as they were considered to be the closest in content to the current review. Overall, the reference sections of five reviews and one annotated bibliography were searched for relevant studies (Belenky et al., 1987; Grandou et al., 2019; Lowe et al., 2017; Miller et al., 2007; Vrijotte et al., 2016; Wickens et al., 2015). Backward snowballing was manually performed by TS. In total, 264 articles identified based on their title and abstract through the grey literature searches, and 577 articles identified through backward snowballing were screened (Figure 2-1).



**Figure 2-1** PRISMA flowchart outlining the eligibility and inclusion process for the current review.

### 2.3.3. Eligibility Criteria

The titles and abstracts obtained from the primary database search, the grey literature search, and through backward snowballing were screened and excluded only if they unambiguously did not fit the following eligibility criteria:

1. *Data.* Articles must present original data; reviews or articles re-presenting previously available data were excluded.
2. *Population.* Participants must have been ECPs: that is, they must be members of the military (e.g. army, navy, air-force, special forces), in aviation (specifically pilot and air traffic controllers), medical personnel (physicians, surgeons, anaesthesiologists, residents etc.), alternate emergency responders (police, firefighters etc.), process operators in a high-risk environment (i.e., mines, oil-rigs, power plants etc.), or elite athletes (highly trained and competing regularly in their given sport). Data for participants that was influenced by the use of alcohol or psychostimulants was excluded with the exception of habitual caffeine or tobacco use.
3. *Baseline Sleep Condition.* Each study must have compared a SR condition to a baseline condition. This could have been conducted through a repeated measures design (each participant is exposed to both baseline and SR conditions) or an independent group design (participants exposed to a SR condition are compared to participants exposed to a baseline condition). For repeated measures designs, baseline conditions must have been conducted either before the SR condition or multiple days after SR was experienced (two-days recovery for every one-day SR), to account for delayed recovery of cognitive performance noted following SR (Banks et al., 2010; Belenky et al., 2003); where both are provided, only the baseline prior to SR condition was considered. Where a mean TST or sleep opportunity value is provided, it must be at least 6 hours (TST > 6 hrs) and at least 2 hours longer than one or more nights in the SR condition; otherwise, it must be clearly stated that baseline sleep was habitual or unhindered.
4. *Intervention.* Articles must have included SR conditions within their protocols that involved 1-7 nights whereby sleep was restricted to between 2 and 6 hours of sleep opportunity or mean sleep obtained. Sleep restriction must have been either experimentally induced or resulting from an abrupt event directly causing SR to occur (e.g., 24-hour overnight shift). Sleep undertaken during an overnight shift was only considered if sleep was not likely to have occurred earlier during the same day, hence 24-hour shifts were considered if participants obtained some sleep throughout the night, however night-only shifts were not. An example of a *near-miss* article that fulfilled all other criteria but was not included due to this point was by Szelenberger, Piotrowski

and Dąbrowska (Szelenberger et al., 2005); this article was not included as it was unclear whether the sleep loss condition was due to sleep during a night-only shift allowing for prior daytime sleep or from a 24-hour overnight shift. For protocols involving multiple nights of SR, all periods of sleep (sleep onset or wake time) must have commenced within the same three-hour time window within the recurring 24-hour cycle. This was implemented to minimise any confounding circadian phase-shifting effects on cognitive performance (Burgess et al., 2013; Santhi et al., 2007). Similarly, any studies where protocols involved participants travelling across three or more time-zones were excluded to eliminate any confounding effects of jet-lag (Waterhouse et al., 2007). If multiple SR conditions were presented within the same article, only the SR condition involving night-time sleep was included. Further, daytime sleep periods were only considered if it was explicitly clear that participants were adapted to diurnal sleep prior to measurement. Sleep restriction interventions must have been monitored using sleep diaries or subjective recollections provided the day immediately following sleep, objective sleep measurement techniques (actigraphy, polysomnography (PSG) etc.), or enforced in an experimental setting through limiting sleep opportunity. If multiple sleep measurement techniques were implemented, preference for reporting sleep obtained was given to the gold standard PSG, followed by actigraphy, and finally subjective recollections.

5. Outcome. Articles must have evaluated cognitive performance using a validated neuropsychological test or an occupation-specific cognitively demanding task. Testing following SR interventions had to occur within the same three hour window as testing following the baseline condition, to minimise the influence of circadian factors on performance (Mollicone et al., 2010). For reviews on the effects of circadian factors on cognitive performance, see articles by Carrier and Monk (Carrier & Monk, 2000); Valdez, Reilly and Waterhouse (Valdez et al., 2008); and Van Dongen and Dinges (Van Dongen & Dinges, 2000). Additionally, sufficient information must have been provided within the manuscript for each test or task (or be freely available if commonly used) to allow for classification of test (classification described further below).

Although effect sizes are not presented within this review, they were calculated for each relevant measure. In doing so, we observed two studies presenting effect sizes on performance effects of SR that were highly improbable (hedges'  $g > 3$  and greater than double the next largest effect size observed by a separate article within the test category). Due to their improbability, these two studies (Daaloul et al., 2019; Taheri & Irandoust, 2019) were removed from consideration in this review.

Following exclusions based on titles and abstracts, the full texts of the 802 remaining articles were screened and excluded if they did not satisfy any of the criteria described above, if articles were written in a language other than English with no translation available, or if full-texts were not present (i.e. conference abstracts). Nine corresponding authors of articles were contacted, as the results of these articles could not be included in the current state, however with clarification of population, methodologies, or results, they may have fit the criteria for the review. Unfortunately however, only one author responded, confirming that the relevant article was not suitable for consideration here. Following full-text exclusions, fifteen articles were assessed for quality.

#### **2.3.4. Quality Assessment**

Study quality was assessed using the specific study design tools from the National Heart, Lung, and Blood Institute (NHLBI, 2014). These criteria were chosen as the NHLBI provides multiple checklists which differ depending on study design and because they include the only standard assessment tool specifically catered for assessing repeated-measures designs within systematic reviews. These tools have been developed by expert panels, are intuitive and easy-to-use for researchers, and have been used within systematic reviews previously (Frestad & Prescott, 2017; A. Gupta et al., 2017; Saltzman & Liechty, 2016). For the thirteen studies with a repeated measures design, the *Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group* checklist was used, and for the two independent-group designs the *Quality Assessment of Controlled Intervention Studies* was used (see appendix 2.3 for the checklists in tabular form). In the latter, criteria regarding the blinding of participants to the intervention were excluded due to the practical difficulties of doing so within SR protocols. Included articles were assessed independently by TDS and AJT, with agreement being reached through consensus. Using the checklists and their accompanying guidelines, articles were given a rating of *good*, *fair*, or *poor*, with *poor* articles being removed from further consideration. Overall, seven studies were assessed as *good*, eight studies were assessed as *fair*, and one study was

assessed as *poor*. For the study assessed as *poor*, this was due to a >15% difference in drop-out rate between groups, constituting a *fatal flaw* and mandating a *poor* rating according to the tool (Naitoh et al., 1987). This study was thus not included further in the review.

Overall, fourteen articles were included in the review (Figure 2-1) and were categorised according to the task used to evaluate cognitive performance (cognitive tests or performance in cognitively demanding occupation-specific tasks), as well as the risk of performance bias due to training effects.

### **2.3.5. Test/Task Categorisation**

The tests used within each of the fourteen articles to evaluate cognitive or occupation-specific performance were categorised as *low-salience* (LS), *high-salience stable* (HSS), or *high-salience flexible* (HSF). The *low-salience* (LS) category included simple attention-based tests that involved no distractors, very limited decision-making, and typically required simple, timely responses to a stimulus. Performance was dependent on vigilance and simple attentional capacity. Low-salience tests included the psychomotor vigilance task (PVT) and serial reaction time (SRT) tests. Occupation-specific tasks were coded as *low-salience* if performance on the task primarily depended on vigilance and maintenance of simple attentional capacity. An example of such would be a vigilance rifle task, where stimuli is interspersed over very long periods of waiting, the response (shoot) is always the same, and the main determinant of performance is clearly how long the individual can maintain vigilant attention.

The *high-salience stable* (HSS) category included tests which are typically used to evaluate more complex cognitive functioning. However, performance on these tests did not depend on the ability to flexibly shift attention or adapt to changing task dynamics (i.e. cognitive flexibility). High-salience stable tests included working memory tasks, grammatical reasoning tests, and the digit symbol substitution test (DSST). Occupation-specific tasks were similarly coded as HSS if performance primarily depended on more complex cognitive functions without requiring task switching or adapting to changing dynamics. This could include psychomotor dominant tasks (skilled sport performance, surgery skill performance) as well as tasks such as friend foe discrimination tasks where the features discriminating friends and foes remain constant throughout.

The *high-salience flexible* (HSF) category consisted of complex/higher-salience cognitive tests & occupation-specific tasks that required cognitive flexibility and/or task switching ability for optimal performance. Examples of high-salience flexible tests included task-switching tests, multitask tests, and tests where the nature of targets could change unpredictably throughout the test. The categorization of both neuropsychological tests and ECP tasks was performed independently by two researchers (TDS and MK); where there was disagreement, consensus was reached upon consultation with AJT and MJC.

### **2.3.6. Categorisation of Training Effect Bias**

In order to assess the degree to which repeated-measure study designs risked confounding the effect of SR on cognitive ability by showing a training effect on cognitive performance, TDS and NR reviewed the included repeated-measures design articles, rating them as having a *no risk*, *low-to-moderate risk*, or *moderate-to-high risk* of training effects, with consensus being reached through discussion. Repeated-measures studies were considered *no risk* if the order between baseline and SR measurements was counterbalanced or if PVT was the performance outcome measure, due to thorough demonstration of robustness of PVT to training effects (Basner et al., 2017). Studies were considered *low-to-moderate risk* if no more than three testing sessions were administered, and *moderate-to-high risk* if more than three testing sessions were administered where the order of baseline and SR conditions were not counterbalanced between participants. Of the fourteen remaining studies, nine had *no-risk*, one had *low-to-moderate risk*, and four had *moderate-to-high risk* of training effects biasing results.

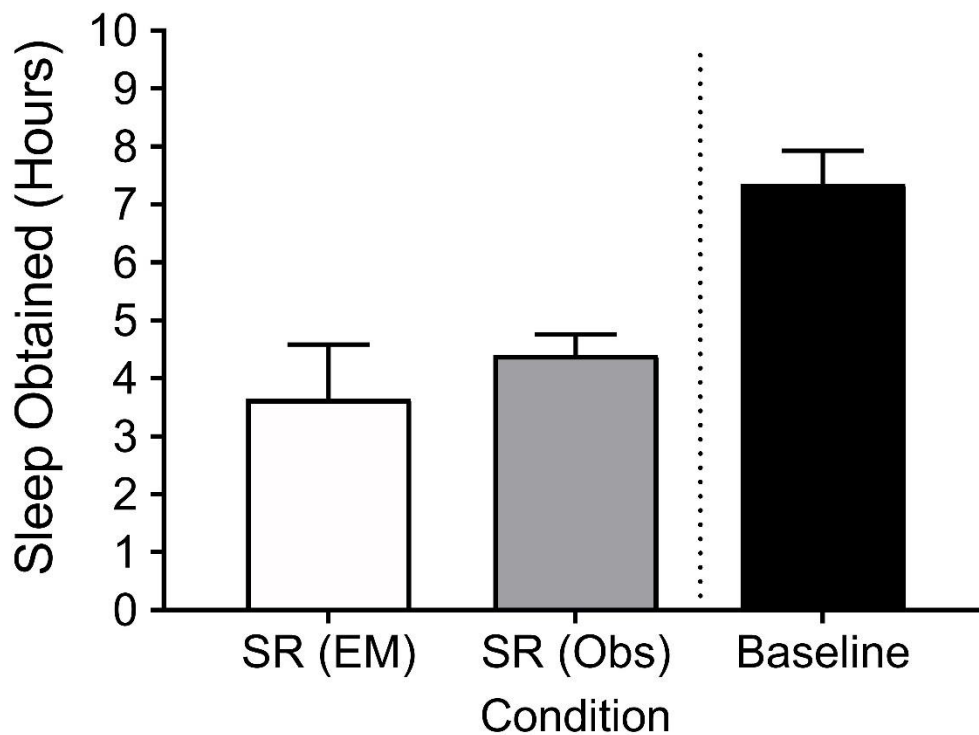
The number and age of participants in the article, occupation, nature and measurement method of SR and baseline conditions, performance test/task used, whether a significant difference was found between performance in conditions, risk of training effect bias, and quality assessment, was coded for each included article and is presented in this review. Results of the review are synthesised narratively below.

## **2.4. Results**

### **2.4.1. Study Characteristics**

The details pertaining to each study included in this review can be found in Table 2-1 (also provided in spreadsheet form for readability as appendix 2.4). The fourteen included articles were published between 1985-2019 with thirteen of the articles published in peer-reviewed journals and one article published as a technical report within the Naval Medical Research Unit (Hartzler et al., 2015). A total of 246 participants (41 females) were tested across all studies, with each individual study sampling 18 participants on average ( $SD=10$ ). The mean age of participants was not provided in each study (with some instead opting to only provide an age range), however for those that did provide this data (11 of 14), the mean age ranged from 19-30 years. The absence of mean age information in some studies was not considered particularly troublesome as similar effect sizes have been noted for the effects of SR on cognitive performance for individuals aged from 18-59 (Lowe et al., 2017).

Six studies included military personnel as participants, three tested medical service workers, four tested elite or highly-trained athletes, and one tested oil refinery process operators. Nine studies assessed performance following only one day of SR, one study following two days of SR, two studies following three days of SR, one study following four days of SR, and one study each following four days and six days of SR respectively. Two studies used polysomnography (PSG) to measure the sleep of participants, five used actigraphy, two used subjective recollection, and five used enforced restriction of sleep opportunity (SO) within a laboratory. Eleven studies experimentally manipulated the SR protocol (mean reported sleep obtained across articles  $\approx 3.6 \pm 0.9$  hr per night SR), while the remaining three all observed sleep obtained during a 24-hour overnight shift (mean reported sleep obtained across articles  $\approx 4.4 \pm 0.4$  hr per night SR). Mean baseline sleep duration was approximately  $7.3 \pm 0.6$  hr (Figure 2-2); note that the report by Hartzler, Chandler, Levin and Turnmire (Hartzler et al., 2015) was not included in this baseline mean calculation, due to reporting “unhindered” baseline sleep rather than a quantity.



**Figure 2-2** Amount of mean sleep obtained ( $\pm$ SD) within each condition. Note that for studies reporting only “sleep opportunity,” sleep obtained was considered to be the entirety of the reported period. SR = sleep restriction, EM = experimentally manipulated ( $n = 11$ ), Obs = observed ( $n = 3$ ).

Assessed using the NHLBI checklists, strengths generally included well-defined research questions, thoroughly described procedures, and minimal participant drop-out. Common weaknesses included a lack of evidence provided on the validity of performance measures and outcomes used, not providing information of whether test administrators were blinded to the condition of the participants, and a lack of consideration of statistical power when determining sample size (although in many cases, the sample size was likely limited by the availability of participants, given the specialised populations).

**Table 2-1** Study Characteristics for included articles, organised by their test/task categorisation and type.

Authors (Year)	Population	Occupation	Test/Task Categorisation	Cognitive Test or Occupation-specific task	Cognitive Test Used	Measure	Baseline Sleep Protocol	Sleep Restriction (SR) Protocol	Result	Risk of Training Effects Biasing Results
Englund, Ryman, Naitoh & Hodgdon (1985)	22 marine corps	Military	LS	Cognitive Test	Alpha-Numeric Visual Vigilance Task	% correct	8hr SO	3hr SO	↓	Moderate-to-high
					4-Choice SRT	% correct			NS	
Gillberg and Akerstedt (1994)	7 military conscripts				6-minute Visual SRT	Response Time (1/RT)	8hr SO	4hr Undisturbed SO	↓	None
Hartzler, Chandler, Levin & Turnmire (2015)	24 naval aviation preflight training program participants				PVT	Lapses (reaction time > 500ms)	"Unhindered sleep"	4hr SWS-suppressed SO	↓	None
								1 night 4hr SO	↓	
								2 nights 4hr SO	↓	
								3 nights 4hr SO	↓	
								4 nights 4hr SO	↓	
								1 night 4hr SO	↓	
								2 nights 4hr SO	↓	
								3 nights 4hr SO	↓	
Romdhani et al. (2019)	14 elite judo athletes	Athlete			SRT	Response Time (1/RT)	8hr SO	4 nights 4hr SO	↓	None
								1 night 4hr SO (early wake)	↓	
								1 night 4hr SO (late sleep onset)	NS	

Roberts, Teo, Aisbett & Warmington (2019)	9 trained cyclists or triathletes			2-choice SRT				1 night 4hr SO (early wake)	NS	None
								1 night 4hr SO (late sleep onset)	↓	
					10-minute PVT	Lapses (reaction time > 500ms)	7.1(0.8)hr sleep	1 night 4.7hr sleep	NS	
								2 nights 4.7-4.8hr sleep	NS	
								3 nights 4.7-4.9hr sleep	↓	
								1 night 4.7hr sleep	NS	
						Response Time (1/RT)	7.1(0.8)hr sleep	2 nights 4.7-4.8hr sleep	↓	
								3 nights 4.7-4.9hr sleep	↓	
								1 night 4.7hr sleep	NS	
								2 nights 4.7-4.8hr sleep	NS	
								3 nights 4.7-4.9hr sleep	↓	
								1 night 4.7hr sleep	↓	
								2 nights 4.7-4.8hr sleep	↓	
								3 nights 4.7-4.9hr sleep	↓	
Mah, Sparks, Samaan, Souza & Luke (2019)	10 elite OR highly trained and actively competing cyclists			10-minute PVT		Reaction Time	7 nights of mean 6.7(0.7) sleep	3 nights of 3.7(0.2)hr sleep	↓	None
									↓	
									↓	
Saxena & George (2005)	13 medical residents	Medical		5-minute PVT		Slowest 10% Reaction Time	7.6(3.0)hr sleep	4.8(2.4)hr sleep	NS	None
									NS	
Sallinen et al. (2004)		Process Operators		10-Choice SRT		Reaction Time	7.1-7.4(0.6-0.9)hr sleep	3.6-3.7(0.1-0.2)hr sleep	NS	None

	12 process operators at an oil refinery					Slowest 10% Reaction Time			NS	
Haslam (1985)	6 trained infantrymen	Military		<i>Occupation-Specific Task</i>	<i>Marksmanship: Vigilance Rifle Shooting</i>	<i>Number of Hits</i>	2 nights 7.25hr SO	6 nights 4hr SO	NS	Moderate-to-high
Hartzler, Chandler, Levin & Turnmire (2015)	24 naval aviation preflight indoctrination program participants				<i>Flight Simulator</i>	<i>Total Lapse Time</i>	Unhindered sleep	1 night 4hr SO 2 nights 4hr SO 3 nights 4hr SO 4 nights 4hr SO	↓ ↓ ↓ ↓	Moderate-to-high
Sallinen et al. (2004)	12 process operators at an oil refinery	Process Operators			<i>Simulated Distillation Process – Monotonous Workday Simulation</i>	<i>Periods of Nil Production</i>	7.4(0.6)hr sleep	3.6(0.2)hr sleep	NS	None
Haslam (1985)	6 trained infantrymen	Military	HSS	Cognitive Test	Adapted Williams Word Memory Test	Number Correct	2 nights 7.25hr SO	6 nights 4hr SO	↓	Moderate-to-high
					15-minute Addition Test	Number Correct Number of Errors			NS NS	
Englund, Ryman, Naitoh & Hodgdon (1985)	22 marine corps (11 exercise, 11 non exercise)				Baddeleys Logical Reasoning Test	% correct	8hr SO	3hr SO	NS	Moderate-to-high
					Williams Auditory Word Memory Test	% correct			NS	
					Gates-Peardon Reading Exercise - "Remembering Details"	Number Correct			NS	
					Gates-Peardon Reading Exercise - "Section About"				NS	
					Gates-Peardon Reading Exercise - "Following Direction"				NS	
					Miller Reading Efficiency Test	Number of Lines Completed			NS	

Reimann, Manz, Prieur, Reichmann & Ziemssen (2009)	32 neurology residents			Paced Auditory Serial Addition Test (PASAT)	Proportion of Items Correctly Answered	6.5(6.0-7.0)hr sleep	4.3(2.8-4.6)hr sleep observed on 24hr overnight shift	NS	N/A (independent-group design)	
Schlosser et al. (2012)	38 surgeons			d2-Paper-Pencil test	Main Concentration inTdex	6.7(0.2)hr sleep	4.1(0.3)hr sleep observed on 24hr overnight shift	↑	Low-to-moderate	
Sallinen et al. (2004)	12 process operators at an oil refinery	Process Operators		Subtraction Test	Reaction Time Slowest 10% Reaction Time	7.1-7.4(0.6-0.9)hr sleep	3.6-3.7(0.1-0.2)hr sleep	NS NS	None	
Haslam (1985)	6 trained infantrymen	Military	<i>Occupation-Specific Task</i>	<i>10-minute Map Grip Reference Encoding/Decoding</i>	<i>Number Correct</i>	2 nights 7.25hr SO	6 nights 4hr SO	NS	Moderate-to-high	
					<i>Number of Errors</i>			NS	Moderate-to-high	
					<i>Marksmanship: Grouping Capacity</i>	<i>Shooting Accuracy</i>			NS	Moderate-to-high
					<i>Air Defense Game</i>	<i>Average Range of Intercept</i>	8hr SO	3hr SO	NS	Moderate-to-high
Englund, Ryman, Naitoh & Hodgdon (1985)	22 marine corps (11 exercise, 11 non exercise)									
Smith et al. (2019)	15 active duty soldiers			<i>Marksmanship: Friend vs. Foe Discrimination Task - "Low Cognitive Load (LCL)"</i>	<i>Errors (incorrect response to friend or foe target)</i>  <i>Accuracy on hitting foes (%)</i>	7.7(0.1)hr sleep	1 Night 2hr SO 2 Nights 2hr SO 1 Night 2hr SO 2 Nights 2hr SO	NS NS NS NS	Moderate-to-high	
				<i>Marksmanship: Army Record Fire Task</i>	<i>Accuracy</i>		1 Night 2hr SO 2 Nights 2hr SO	NS NS		
Reyner and Horne (2013)	28 first or second team univeristy tennis players	Athlete		<i>Tennis Serving Accuracy</i>	<i>Hits Within a Designated Area</i>	6.6-7.8(SE=0.1-0.2)hr sleep	4.3-5.4(SE=0.1)hr sleep	↓	None	
Schlosser et al. (2012)	38 surgeons	Medical		<i>LapSim Low-Fidelity Tasks</i> <i>High-Fidelity Intracorporeal Suturing</i> <i>High Fidelity Chole-Cystectomy</i>	<i>Composite Performance Score (%)</i>	6.7(0.2)hr sleep	4.1(0.3)hr sleep observed on 24hr overnight shift	↑ ↑ ↑	Low-to-moderate	

Sallinen et al. (2008)	16 military conscripts	Military	HSF	Cognitive Test	Brain@work Multitask	Score obtained relative to highest possible score obtainable for the individual	8.0(0.4)hr sleep	2.1(0.1)hr sleep	↓	None
Hartzler, Chandler, Levin & Turnmire (2015)	24 naval aviation preflight indoctrination program participants				Dual <i>n</i> -back	Dual <i>n</i> -back Metric	Unhindered sleep	1 night 4hr SO 2 nights 4hr SO 3 nights 4hr SO 4 nights 4hr SO	↑ ↑ ↑ ↑	Moderate-to-high
Smith et al. (2019)	15 active duty soldiers			<b>Occupation-Specific Task</b>	<b>Marksanship: Friend vs. Foe Discrimination Task, "High Cognitive Load (HCL)"</b>	<b>Errors</b>  <b>High Value Target Detections</b>  <b>Accuracy on hitting foes (%)</b>	7.7(0.1)hr sleep	1 Night 2hr SO 2 Nights 2hr SO 1 Night 2hr SO 2 Nights 2hr SO 1 Night 2hr SO 2 Nights 2hr SO	↓ ↓ NS ↓ NS NS	Moderate-to-high
Sallinen et al. (2004)	12 process operators at an oil refinery	Process Operators			<b>Simulated Distillation Process – Busy Workday Simulation</b>	<b>Amount of Time with Nil Production</b>	7.1(0.9)hr sleep	3.7(0.1)hr sleep	NS	None

*Note.* ↓: significant negative effect of SR condition, ↑: significant positive effect of SR condition, NS: no significant effect of SR condition. LS: Low-salience, HSS: High-salience stable, HSF: High-salience flexible, PVT : psychomotor vigilance task, SRT: serial reaction test, CRT: choice reaction test, SO: Sleep opportunity provided, SWS: slow-wave sleep. Variance for sleep measures is standard deviation except when specified using 'SE' for standard error and is given in brackets following the value. Bolded cognitive tasks and outcomes are occupation/expertise specific performance measures.

## **2.4.2. Low-Salience Test/Task Performance**

### **2.4.2.1. Descriptive Information**

Nine studies investigated *low-salience* cognitive or occupation-specific task performance following SR among 117 participants. Four of these studies tested military personnel, three tested elite or highly-trained athletes, one tested medical residents, and one tested oil refinery process operators. Five studies examined the effect of only a single night of SR on performance, while the remaining studies examined the effect of multiple consecutive nights of SR on performance. Three studies incorporated occupation-specific tasks; a marksman vigilance task, a distillation simulation task (monotonous workday condition), and a flight simulator lapse task.

### **2.4.2.2. Findings**

Only performance on the flight simulator task (deviation from a *simple flight profile*) was found to be significantly weakened by SR (Hartzler et al., 2015). In the other two studies, the vigilance of trained infantrymen was found to be unaffected while performing a shooting task following six consecutive nights of SR (4hr SO) (Haslam, 1985), and no significant performance change was found on a simulated distillation task among experienced oil-refinery process operators following one night of ~3.5hr TST (Sallinen et al., 2004). Among the eight studies testing cognitive abilities directly, Englund, Ryman, Naitoh and Hodgdon (Englund et al., 1985) observed a significant performance decrement among a sample of U.S. marines on the alpha-numeric visual vigilance task, but not on a four-choice SRT, following one night of SR (3hr SO). Gillberg and Åkerstedt (Gillberg & Åkerstedt, 1994) found response times on a SRT were not significantly affected by SR when tested at 08:00, but were significantly worsened when tested at 14:00 or 20:30, as well as when the results from all time points were combined; this was regardless of whether the four hours of SO allocated were undisturbed or manipulated so that participants could obtain minimal slow-wave sleep. Among a sample of naval aviation trainees, Hartzler and colleagues (Hartzler et al., 2015) found the number of lapses (reaction time >500ms) increased during each night of SR experienced (4-nights of 4hr SO), with an increased overall response time of the slowest 10% attempts on the PVT compared to baseline following SR. Romdhani and colleagues (Romdhani et al., 2019) found that one night of SR (4hr SO) slowed reaction times of judo athletes on (a)

a SRT when sleep was restricted by initiating an early wake time but not when delaying sleep onset, and (b) a *choice* reaction time task when sleep was restricted by delaying sleep onset but not when initiating an early wake time. Roberts, Teo, Aisbett and Warmington (Roberts, Teo, Aisbett, et al., 2019) found a significant increase in PVT lapses and response times among highly trained cyclists and triathletes following three days of SR (~4.5 to 5hr TST) when compared both to the day before the first night of SR (6.5hr TST) or the equivalent day within a baseline condition (~6.5 to 7hr TST), additionally finding differences in lapses following two days of SR and in response times following both one and two days of SR when compared to the equivalent baseline condition. Mah and colleagues (Mah et al., 2019) similarly found the vigilance of ten elite (or highly trained & actively competing) cyclists (PVT reaction time, inverse reaction time & fastest 10% reaction time) to be adversely affected by three nights of ~4hr SO. Interestingly, no significant differences in LS test performance (PVT, serial reaction time task) were found between SR (~5hr TST) and baseline conditions for medical residents (Saxena & George, 2005), nor among oil refinery process operators following one night of ~3.5hr TST (Sallinen et al., 2004).

### **2.4.3. High-Salience Stable Test/Task Performance**

#### **2.4.3.1. Descriptive Information**

Seven studies examined the effects of SR on cognitive and occupation-specific *high-salience stable* tasks among 153 participants (Table 2-1). Three of these studies tested military personnel, one tested highly-trained athletes, two tested surgeons or medical residents, and one tested oil refinery process operators. Five studies implemented only a single night of SR, while two involved multiple consecutive nights of SR. Five studies incorporated occupation-specific tasks, including marksmanship accuracy tasks, a marksmanship friend vs. foe discrimination task (low-cognitive load condition), an “air defense” game, a map-grid encoding/decoding task, a tennis serving accuracy protocol, and a VR-surgery simulator task.

#### **2.4.3.2. Findings**

Two studies found SR to significantly decrease performance relative to a baseline condition. Haslam (1985) found significant deterioration in the number of correctly recalled items by trained infantrymen on a word memory test throughout six nights of SR (4hr SO). Conversely, Englund and colleagues (Englund et al., 1985) found no effect of

one night of SR (3hr SO) on the immediate recall of marine corps on an almost identical task to that in Haslam (1985), as well as the ability to immediately recall details in a short reading task. Performance on the d2-paper-pencil test (selective attention) was found to significantly improve following SR (one night of ~4hr TST) relative to a previously taken baseline among surgeons (Schlosser et al., 2012). This study additionally found performance on surgery skills of varying complexity (as well as “economy of motion” measures which comprised these overall composite scores) on a VR-surgery simulator to *improve* following SR. It is noted however that in measurements taken 24hr after the SR condition testing, d2-paper-pencil test performance improved again from the SR condition, and performance on two of the three surgery performance measures (Low-Fidelity Task and Chole-Cystectomy performance score) was maintained from the SR condition and also significantly better than in the baseline condition. When examining studies that evaluated occupation-specific task performance, Reyner and Horne (Reyner & Horne, 2013) found serving accuracy of semi-elite tennis players to be hindered by one night of SR (~4.5 to 5.5hr TST), whereas all other studies utilising occupation-specific HSS measures failed to detect significant performance differences between baseline and SR conditions.

#### **2.4.4. High-Salience Flexible Test/Task Performance**

##### **2.4.4.1. Descriptive Information**

Four studies examined performance outcomes on *high-salience flexible* cognitive and occupation-specific tasks among 67 participants. Three of these studies tested military personnel, and one tested oil refinery process operators. Two of these studies investigated only a single night of SR, while the other two implemented multiple consecutive nights of SR. Two studies incorporated occupation-specific tasks including a military marksmanship friend vs. foe discrimination task (high-cognitive load condition), and an oil-refinery distillation simulation task, similar to the task mentioned previously but with increased cognitive demand and functional instability embedded within it.

##### **2.4.4.2. Findings**

Among those studies investigating cognitive test performance, Sallinen and colleagues (Sallinen et al., 2008) tested 16 military conscripts on a multitask test (Brain@Work) involving four cognitive tests performed simultaneously, and reported a significant decrease in performance following SR (one night ~2hr TST) relative to baseline.

Alternatively, Hartzler and colleagues (Hartzler et al., 2015) used an adaptive dual *n*-back measure requiring simultaneous attention towards both visual and auditory stimuli, and found *improvement* compared to baseline on all four days of SR (4hr SO). However, the authors concluded that “practice effects were evident throughout the study” (p. 24) and these likely confounded the results. For studies investigating occupation-specific performance, Smith and colleagues (C. D. Smith et al., 2019) found one and two nights of SR (2hr SO) led to an increase in the number of errors made on a high cognitive load (HCL) challenge of the marksmanship friend vs. foe discrimination task. In this task, colours that represented friends and foes changed frequently, requiring participants to flexibly adapt to the task details for correct completion. Notably, in the low cognitive load (LCL) version of this challenge, whereby colour coding was held constant (HSS task), error rate did not significantly differ between baseline and SR conditions. Additionally for high value target detection, a measure present only for the HCL version of the task, percentage detection rate was impaired after 2-nights of SR. No effect of SR was found in either version of the task for the marksmanship accuracy in shooting foes. Lastly, time at nil production, the main performance outcome referring to a lack of activity occurring within the *busy* condition of a simulated distillation process, was not found to change significantly between SR (one night of ~3.5hr TST) and baseline conditions for oil refinery process operators (Sallinen et al., 2004). Again, this task was identical to the *monotonous* condition in the simulated distillation included in the LS table except for a greatly increased depletion speed and the addition of functional instability.

## 2.5. Discussion

Our systematic review explored the literature examining the effect of SR on cognitive and task-related performance specifically among *Elite Cognitive Performers* (individuals within occupations that (a) have cognitive demands exceeding the norm and (b) have critical outcomes associated with these demands). In doing so, we aimed to provide an indication of how SR may affect the cognitive performance of those for whom the outcomes are arguably more important than is often the case within the general population, with a degree of specificity not previously available within the literature. Overall, this review found that the performance of this select group on monotonous, low-salience tasks is often poorer following SR, and while performance on more complex, yet cognitively stable, tasks is usually maintained, performance may be more prone to decline when the task involves adaptation to changing goal-oriented information and a shifting of attention.

### 2.5.1. Differences of the Effects of Sleep Restriction as a Function of Task Demands

In this review, we found performance on simple tests designed to measure vigilance and rudimentary attentional capacity (whether occupation-specific or not) was most commonly hindered by SR. This corroborates findings from a meta-analysis on the wider population (Lowe et al., 2017) and is consistent with the effects found following TSD (Lim & Dinges, 2010). The ability to maintain attention in low-salience circumstances is integral to many components of safety-critical work (i.e., monitoring human-machine interfaces or environment), however two of the three studies using low-salience occupation-specific task performance found no significant deterioration resulting from SR. These two articles however had factors within their design or within the outcomes themselves that could have confounded results. Specifically, Sallinen and colleagues (Sallinen et al., 2004), who tested performance on a simulated distillation task among oil refinery workers, reported that their failure to find an effect of one night of SR on participants' *monotonous* and *busy* workday (the latter coded as HSF) may be due to the performance task being "too rough to demonstrate a significant sleep debt-related effect" (p. 293). The other task was a vigilance rifle shooting task for trained infantrymen following six nights of SR (Haslam, 1985). This article stated that the infantrymen were to receive long-weekend leave if they "maintained a certain standard" of performance on

“key tasks” (p. 91), likely raising the external motivation to maintain performance in spite of SR (although it is unclear which were the “key tasks” in this study). Such external motivation likely increased task engagement, which purportedly facilitates the maintenance of performance in spite of sleep loss (Harrison & Horne, 2000). Along these same lines, one of the few studies that did not find a difference in PVT performance following SR rewarded the highest-performing participants with \$100CAD (Saxena & George, 2005). Hence, it appears that context and motivation are key moderators of the effect of SR on the performance of simpler tasks; if task engagement is promoted (through task demand, external motivation etc.) performance is likely to be maintained, whereas if there is a lack of exogenous factors promoting task engagement, ECPs, like the general population, will show degraded performance on tasks prioritising simple attention.

In stark comparison to studies investigating the effect of SR on simple cognitive performance, studies using evaluation tools that required more complex cognitive processes that rewarded cognitive stability *almost unanimously* reported no effect of SR. The two studies which reported SR to negatively affect performance tested the immediate recall of infantrymen on a working memory task (Haslam, 1985), and university representative tennis players on a tennis serving accuracy test respectively (Reyner & Horne, 2013). The former suggests that immediate recall is vulnerable to SR, however it is to be noted that the immediate recall of marine corps in a separate study (Englund et al., 1985) with a similar procedure to Haslam (1985), as well as the ability to immediately recall details in a short reading task, was not significantly affected by SR. The latter finding would suggest perhaps that skilled, psychomotor performance outcomes could be vulnerable to SR, however Schlosser and colleagues (Schlosser et al., 2012) contrastingly noted an *increase* in performance on simulated surgery tasks of differing complexity following SR and compared to a previously taken baseline condition. Investigating tasks with similar cognitive demands, Haslam (1985) and Smith and colleagues (C. D. Smith et al., 2019) did not report a significant effect of six or two nights of SR respectively on marksmanship tasks specifically measuring accuracy. Taken altogether, these results suggest that acute SR alone is unlikely to negatively affect the performance of ECPs on complex cognitively stable tasks when they are related to their area of expertise. In a practical sense, this could be interpreted to closed skills for elite athletes (i.e., a free-throw or a golf putt), performance of routine fine-motor, yet demanding surgery task, or the ability to perform fixed and predictable tasks such as adhering to correct pre-flight procedures for an aircraft pilot. Lastly, in further support of the influence of context,

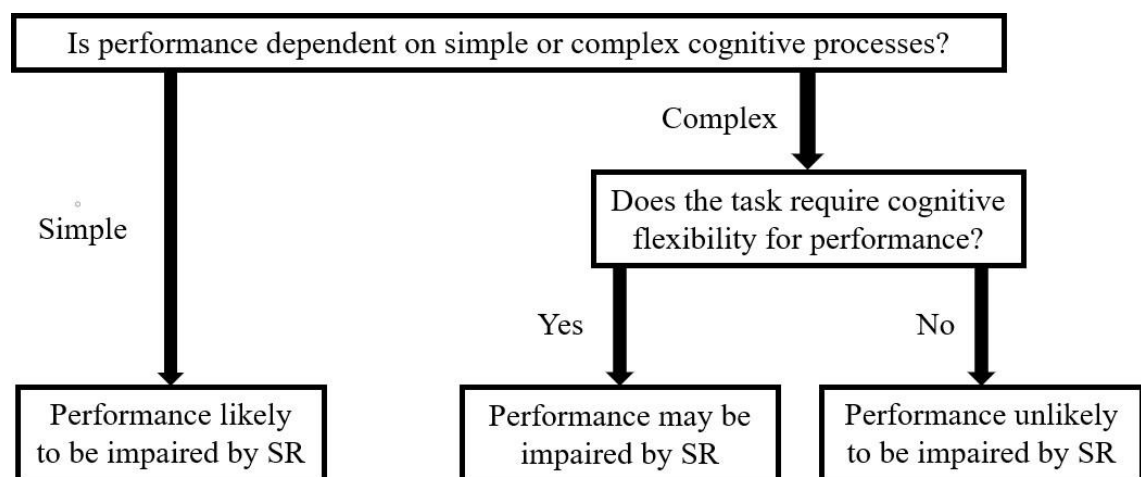
engagement and motivation, Englund and colleagues (Englund et al., 1985) noted in discussing the (statistically insignificant) trend of performance improvement on a complex ‘air defense’ game and the lack of performance loss on a reading efficiency following SR, that “competition and interest, each a motivating factor, influenced both psychomotor and cognitive task performance” (p. 84).

We separated *HSS* and *HSF* based on whether the task demanded *cognitive flexibility*; that is, the ability to shift attention to new, more relevant information, or to adapt to a changing task dynamic (Ionescu, 2012; Scott, 1962). Here we found that performance on a multitask test (Sallinen et al., 2008) and error rate within a task embedded within a marksmanship context (C. D. Smith et al., 2019), both of which required cognitive flexibility, were negatively affected by SR. The latter is particularly notable as error rate in a simpler adaptation of the same task, which didn’t require participants to adapt to the changing meaning of different colour targets throughout, was not negatively affected by the same conditions of SR. When considering the two studies that did *not* find a negative effect of SR on HSF outcomes, one used an adapted version of the aforementioned oil refinery distillation task with potential sensitivity issues, and the other one had a *moderate-to-high* risk of bias, with the authors (Hartzler et al., 2015) themselves stating that “practice effects were evident” (p. 24). The ability to flexibly shift attention and adapt to changing dynamics is of obvious ecological importance, particularly for safety-critical workers handling emergency situations. For example, aircraft pilots are presented with a multitude of information from dials, outputs, and air-traffic controllers when in an emergency situation (i.e., engine failure) and must be able to rapidly shift their focus to the most important interface to gather the most relevant information for the resolution of their current circumstance. Pilots must then be able to adapt to their new situation (flying an aircraft without the engine) and adjust their approach accordingly. Further experimental work is clearly required to understand how the cognitive flexibility of ECPs’ is affected by SR and how this impacts high-demand tasks within their workplace, given (a) the importance of cognitive flexibility particularly within emergency scenarios, (b) the increased prevalence of SR in ECPs versus the general population, and (c) the studies detailed within this review outlining the effects of SR on the cognitive flexibility of ECPs.

### **2.5.2. Strengths and Limitations**

It is accepted that the classification of performance tests in this review can be considered coarse. For example, standardised cognitive tests within the *HSS* category can be

attempting to test primarily inhibition, working memory, decision-making, executive functioning, *complex attention*, *cognitive throughput* and so on. These are regularly discussed as distinct cognitive outcomes with distinguishable underlying neural processes. Although previous meta-analyses have demonstrated differences in effect sizes among different *complex* cognitive domains (Lim & Dinges, 2010; Lowe et al., 2017), the most tangible distinctions regarding effects of sleep loss on cognitive performance appear to be: (a) the extent to which performance is dependent on sustained, simple attention (Harrison & Horne, 2000; Lim & Dinges, 2008, 2010; Lowe et al., 2017), and more recently, and (b) whether cognitive flexibility is prioritised over cognitive stability for performance (Honn et al., 2019; Whitney et al., 2015; Whitney et al., 2019) (spawning the rationale behind the classification used). This information is of direct practical use to members of safety-critical industries, elite athletes and coaches, and other individuals in occupations with cognitive demands spanning beyond the norm. The separation of tasks into cognitive domains limits applicability because tasks that ECPs engage in are complex by nature and require significant contributions from multiple domains simultaneously. By separating tasks as we have in the current review, we provide a simple framework that applies to *real-world* tasks in a host of occupations with large cognitive demands, but that seemingly distinguishes between tasks in which performance is likely or unlikely to be affected by SR (Figure 2-3).



**Figure 2-3** Proposed framework explaining the likelihood of sleep restriction affecting cognitive performance for Elite Cognitive Performers.

The current review had particularly stringent eligibility criteria. Studies that investigated sleep restriction and cognitive performance among ECPs, but that allowed significant variance (or lacked reporting) of participant sleep onset & wake times, or testing times,

were removed. Additionally, multiple studies included in this review presented additional measurements taken and statistical comparisons made that were not included due to these reasons (Hartzler et al., 2015; Haslam, 1985). By removing these studies and measurements we likely lost some degree of ecological validity as SR in practice is often accompanied by a shift in sleep onset and wake times (i.e., following transmeridian travel, change from day-shifts to night-shifts etc.) resulting in circadian desynchrony, and performance at different and often rapidly changing times of day is necessary (i.e., rotating shift schedules). The effect of rhythmic fluctuations of performance related to circadian rhythm, as well as the desynchronisation of circadian rhythm likely to arise from large variance in sleep onset and wake time, is non-trivial (Blatter & Cajochen, 2007; Carrier & Monk, 2000; Mollicone et al., 2010; Van Dongen & Dinges, 2000). Therefore, without controlling for changes based on when the participants were sleeping or when they were being tested, it would be incredibly difficult to discern whether differences in performance were due to the changes in sleep quantity or these circadian factors. Controlling for these factors allows us to more confidently conclude that any performance decrements observed were due to SR and *not* other influences. In short, this review only examined the effects directly related to change in sleep quantity, and that other features commonly experienced with SR such as shifts in sleep periods are likely to further exacerbate the performance impairments discussed in this review; hence, the findings of this review should be considered conservative and a *best case outcome* for how moderate sleep loss impacts task performance for Elite Cognitive Performers in the real-world.

Despite the stringency of the eligibility criteria, there was still a surprisingly small number of articles that met the inclusion criteria of the review, given the exhaustive nature of our systematic search. In particular, there was a dearth of research investigating the role of SR on cognitive performance among elite athletes. Using the criteria for defining and quantifying expertise as outlined by Swann et al. (2015), *semi-elite* athletes were the top level participants tested among the included studies. Some of the other studies utilised interns and junior medical residents (Saxena & George, 2005; Schlosser et al., 2012) and military personnel either within their first few years of service or within the process of completing specialised programs such as a naval preflight training program (Hartzler et al., 2015; C. D. Smith et al., 2019), and may not be representative of more experienced individuals who (a) have more experience performing while fatigued, and (b) have greater expertise on occupation-specific tasks. This distinction is highly important as more expert

individuals tend to utilise different cognitive strategies compared to their less skilled individuals when performing tasks within their field of expertise, such as different gaze and fixation strategies (Mann et al., 2007) and decision making processes (Salas et al., 2009). Hence, it is possible that experts and novices may be differentially affected by sleep loss on a particular cognitively demanding task. Ideally, further experimental work which directly investigates the potential for expertise to moderate the effect of SR on cognitive performance would elucidate such a possibility. However, recruiting such an array of participants within a specific area can prove difficult. Additionally, as demonstrated by Sallinen and colleagues (Sallinen et al., 2004), selecting a performance outcome within the context of one's area of cognitive expertise that is also sensitive enough to show performance deficits following SR provides another layer of difficulty. Still, such experimental work could be extremely beneficial in (a) understanding the relationships between sleep loss, cognitive performance, and cognitive expertise, and (b) further improving our overall understanding on how SR affects the task performance of ECPs.

### **2.5.3. Future Directions**

One area where it is both relatively easy to evaluate cognitive and occupational performance among individuals with a vast array of skill level is esports. Here, we believe that research on elite esports athletes may be able to provide insight into the moderating effect of cognitive expertise on performance loss resulting from SR. Esports refers to the competitive (and for some, professional) play of commercially available video games, with esports athletes being referred to as “cognitive athletes” due to the cognitive expertise that they possess (Campbell et al., 2018). Many esports games often adopt the Elo rating system, allowing for expertise to be quantified on a continuum and the digital nature of game play facilitates the collection of large amounts of relevant performance. In addition to being an exemplar test population, esports athletes also share many similarities with many ECPs with respect to work environment and the enhanced cognitive skills required by both for optimal performance (Smithies et al., 2020). Future research on the effects of SR on esports athletes could thus provide applicability to ECPs in general, furthering our understanding of how elite cognitive performers are affected by sleep loss. Moreover, as esports athletes can be considered ECPs themselves and that their shared commonalities with traditional athletes likely leading to higher-than-normal

prevalence of SR, the results of the current review are of great relevance to this population.

Sleep restriction presents as one of many factors which may adversely affect performance on complex, cognitively demanding tasks. In addition to circadian factors mentioned earlier, sleep inertia, referring to the grogginess and degraded performance immediately following wake, is of high relevance to individuals performing tasks at night or those working extended shifts and are able to sleep on the job but simultaneously may be required to respond to complex emergency situations at a moment's notice (i.e., night-shift medical workers, pilots, air traffic controllers, emergency responders). Extended periods of wakefulness and time on task (particularly for boring, monotonous tasks) can also further contribute to fatigue related performance impairment within the workplace (Caldwell et al., 2019), and are important considerations for safety-critical workers and other elite cognitive performers (i.e., athletes, esports athletes). As aforementioned, the context surrounding a given task (i.e. the presence of external motivating factors) is an important consideration in addition to the nature and demands of the task itself. Lastly, the extension of sleep quantity beyond what is habitually obtained has shown positive effects on cognitive performance outcomes for high-level collegiate athletes measured both through standardised cognitive tests and through outcomes directly related to their expertise (Mah et al., 2011), and may resemble a fruitful strategy to improve performance on demanding tasks for Elite Cognitive Performers overall.

#### **2.5.4. Conclusion**

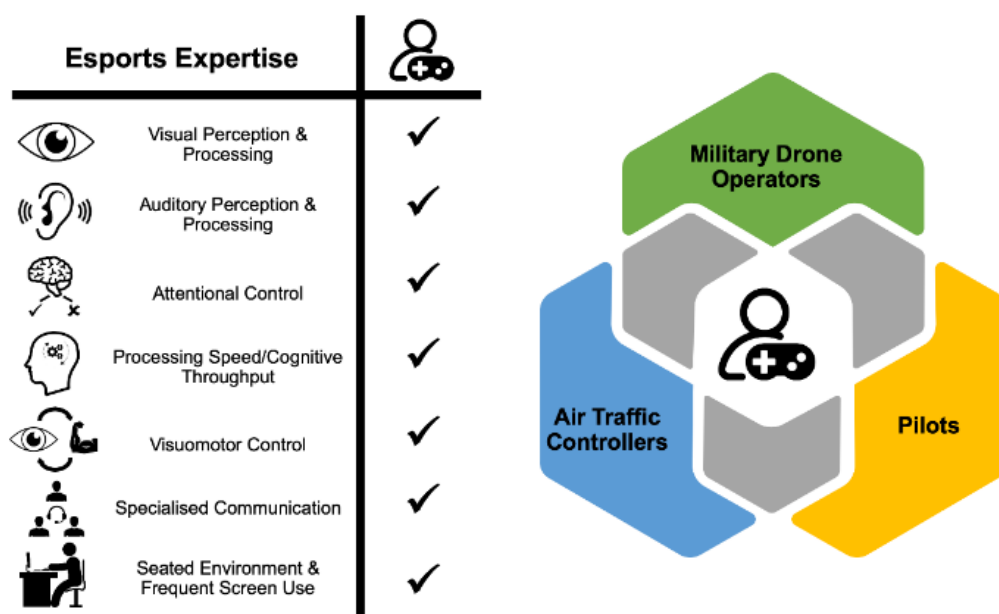
In summary, the current review demonstrates that the performance of ECPs is more negatively affected on simple cognitive tests and monotonous occupation-specific tasks, where simple attentional capabilities are instrumental to task success, over more complex cognitive tasks; however, performance may be more affected on complex tasks when adaptation to changing goal-oriented information and a shifting of attention (i.e., cognitive flexibility). Further research is required particularly when using tasks demanding cognitive flexibility as there is little and conflicting evidence on the effect of SR on the performance of such tasks. Lastly, we believe that esports presents as a fruitful medium to explore the effects of sleep loss on *Elite Cognitive Performers*, potentially uncovering moderating roles of expertise and providing applicability to many industries and occupations.

## 2.6. Linking Section

### 2.6.1. Esport Athletes are ECPs

The systematic review presented in the above chapter provides insight into how sleep loss (in the form of SR) could impact the performance of esport athletes. No prior research had experimentally determined how sleep loss impacts esports performance. The great cognitive demands of esports have led some researchers to refer to esport athletes as *cognitive athletes*. (Campbell et al., 2018). In order to gain insight into how sleep loss may impact esport performance and justify the experimental investigation of this phenomenon, I explored prior literature on an *Elite Cognitive Performer (ECP)* population. This population was chosen as they also are specifically required to perform cognitively demanding tasks with critical outcomes (i.e., requiring timely responses for success/ avoidance of failure). Furthermore, I explored performance both on laboratory based standardised cognitive tests and on occupation-specific cognitively demanding tasks, to provide insight into the effects of sleep loss on performance outside of a highly controlled environment and circumstances.

While the similarities between esport athletes and ECPs (as well as the fact that esport athletes themselves can be considered ECPs) was discussed in this review, it was elaborated on in much greater detail in my 2020 article *Life After Esports: A Grand Field Challenge* (Smithies et al., 2020). The purpose of this article was to outline post-career employment difficulties of esport athletes, suggesting that better connections should be established between the world of esports and occupations which could best benefit from the unique attributes that esport athletes possess. I uncovered these occupations using a systematic approach (querying the Occupational Information Network or *O\*net*), and found that the three occupations which best aligned with esports in terms of cognitive and environmental demands were military drone operators, aircraft pilots, and air traffic controllers; three ECP occupations. A figure from this article (presented here as Figure 2-4) highlights this relationship.



**Figure 2-4** Skills and experience of which esports athletes possess (left) and the three occupations which best match these skills and experience (right). Figure is from Smithies et al. (2020).

Additionally, it would appear that esports athletes exhibit the same improved cognitive performance that ECPs seemingly tend to exhibit within laboratory-testing. While there is limited evidence specifically for professional esports athletes, there is a now large body of evidence (summarised by Bediou et al. (2018), Bediou et al. (2023), and Toth et al. (2020)) demonstrating how frequent players of video games commonly played as esports outperform non-gamers in laboratory-based cognitive tests.

It is also pertinent to highlight that esports athletes may be (like the ECP populations included in the review) more susceptible for experiencing sleep restriction. This is as not only do esports athletes face most of same factors that traditional sport athletes, but they also possess some risk factors which are unique. These include the use of blue-light emitting monitors (which may negatively impact/ shorten sleep through suppression of endogenous melatonin secretion), the highly stimulating nature of the games commonly played as esports, and a ‘culture’ of late-night training and play. These factors are discussed in greater detail in **section 1.5**.

### 2.6.2. Implications for Experimental Research

The systematic review outlined in this chapter helped determine the cognitive domains of interest for my future inquiries into sleep loss and performance of esports players (**Chapter 7**). Firstly, the review found (in corroboration with reviews and meta-analyses in broader populations and under broader conditions of sleep loss; (Lim & Dinges, 2010; Lowe et al., 2017; Pilcher & Huffcutt, 1996) that performance on low-salience tasks (LS; defined as simple tasks with no distractors, limited decision-making, and requiring simple, timely responses) was consistently hindered by SR. I argue that this has limited relevance to in-game esports performance. Esports are highly engaging, almost always provide multiple sources of information to attend to or avoid distraction from, and require rapid and frequent decision making. Nonetheless, given the demonstrated sensitivity of LS tests to sleep loss, combined with the largely universal use of PVT to assess performance changes in sleep research, I decided to include one (PVT) in my experimental sleep loss study.

The review also suggested that performance on more complex tasks is likely more sensitive to sleep loss when cognitive flexibility is required (i.e., HSF tasks; task-switching tests, multitask tests, and tests where the nature of targets could change unpredictably throughout the test). HSF tests/ tasks appear much more relevant to esports than LS tests/ tasks, given the multitude of information sources that players must switch their attention between in most esports. Task-switching seems especially relevant, as studies consistently show task-switching ability to improve with the play of video games commonly played as esports (Nuyens et al., 2019; Toth et al., 2020).

In light of this, I sought to include a task-switching paradigm in addition to the PVT in my experimental sleep loss study. Within the “Inquisit” neuropsychological test platform (Millisecond Software, 2016), there are two unpredictable (i.e., the participant is unaware when the cue will switch) task-switching paradigms; the Color-Shape Task, and the Category Switch Task (CST). An unpredictable task-switching paradigm was sought (as opposed to a test with predictable switches) as it seemingly better reflects the unpredictable nature of player vs. opponent interactions within esports. Of these tests, the CST was shown to result in lower residual variance (i.e., variance unrelated to task-switching ability) than the Color-Shape Task, in a single administration setting (Friedman et al., 2008) using Switch Costs (SC) as the outcome measure. Hence, I sought to incorporate the (shortened) CST into the experimental sleep loss study. However, there were outstanding questions with regards to the test-retest reliability of this measure, and

whether practice effects may be an issue to overcome with the use of the CST. The latter is particularly notable, given that my review included many papers presenting results that were highly likely biased by the presence of practice effects (called *training effects* within the review). These questions are addressed through my pilot study, which is outlined in the following chapter.

### **Chapter 3. Test-Retest Reliability and Practice Effects on a Shortened Version of the Category Switch Task – A Pilot Study**

This chapter is currently under review for publication in a peer-review journal:

**UNDER REVIEW: Smithies, T. D., Toth, A. J., Campbell, M. J. (2023).** Test-Retest Reliability and Practice Effects on a Shortened Version of the Category Switch Task – A Pilot Study.

Changes to the version submitted for publication for the purposes of this thesis are outlined below:

- Change in referencing style (article version is in numbered format).
- References to supplementary files are changed to the appropriate location within the appendix.
- Words emphasised using quotation marks were changed to be emphasised using italics, in line with the thesis format.
- The words *Figure* and *Table* in in-text references to figures was capitalised. Furthermore, figure/ table numbering convention was changed in line with the thesis format.
- Additional of a *linking section* for the purpose of thesis flow.
- Minor amendments have been made based on examiner correction suggestions.

### 3.1. Abstract

**Study Objectives:** Test-retest reliability and practice effects (PEs) have not been assessed for the Category Switch Task (CST), a task-switching paradigm readily available to researchers. This pilot study aimed to assess the test-retest reliability of the CST, and the presence of Pes among three short test-retest intervals (*same day*, *next day*, and *next week*).

**Methods:** Forty-eight participants completed a shortened CST twice. Test-retest *intervals* were either *same day*, *one day*, or *one week*. Test-retest reliability was assessed via Pearson's correlation and intraclass correlation coefficient. PEs were assessed using paired-samples t-tests, and the effect of *interval* was examined through a series of ANCOVAs.

**Results:** Single task, switch cost and mixing cost response time test-retest reliability was comparable to other task-switching paradigms, while reliability for switch and mixing cost accuracy was poor. Test-retest PEs were present for single task response time and accuracy, and mixing cost response time. Of these, PEs varied as a function of *interval* only for single task accuracy, where an interval of *one week* resulted in a smaller improvement compared to *one day*.

**Discussion:** The CST produces reliable values for single task RT, single task accuracy, switch cost RT, and mixing cost RT. Researchers should be aware that PEs may confound results in a test-retest design when single task RT, single task accuracy, or mixing cost RT are considered as outcome measures.

Keywords: reliability, practice effects, task-switching, test-retest, category switch

### 3.2. Introduction

Researchers often wish to use cognitive tests in test-retest research designs. If a test is to be used in this fashion, it is important to consider its *test-retest reliability* as more reliable tests will minimise risk of error and will thus be generally better to use in practice. Test-retest reliability refers to a measure of the consistency of scores (or the minimisation of error) obtained from an individual when a test is administered multiple times. If test-retest reliability is low due to measurement error, the risk of both Type I and II error increases (however, note that a lower test-retest reliability as a function of low between-subject variability may be beneficial in test-retest designs; see Hedge et al. (2018) for further discussion). Sources of *error* that can reduce test-retest reliability can be random, such as natural fluctuations in an individual's alertness, or systematic, such as improved performance attributable specifically to an intervention or to repeated engagement with the test (*practice effects* (PEs); McCaffrey et al. (2000)).

When considering these sources of error, the presence and magnitude of PEs can, and should, be measured when using test-retest designs to account for the extent to which an intervention truly affects the behavioural outcome being quantified. If improvement due to repeated test administrations is not accounted for, the improvement could be mistakenly attributed to a positive effect of an intervention. Concurrently, a lack of numerical score difference between test and retest sessions could be erroneously interpreted as a null finding, when, in fact, test-retest PEs are masking the deleterious effect of a given intervention on performance. This scenario has been demonstrated within the context of underdiagnosis of mild cognitive impairment (Duff et al., 2011; Elman et al., 2018) and cognitive impairment among women undertaking chemotherapy for breast cancer (Cerulla et al., 2019). Many factors are known to influence the magnitude of PEs, including task difficulty and, perhaps even more importantly, the time interval between administrations (*test-retest interval*). For example, shortening the test-retest *interval* generally leads to a larger PE, with its influence seemingly being task dependent (Bartels et al., 2010; Calamia et al., 2012).

Task-switching (or set shifting) is an executive function commonly studied as a measure of one's cognitive flexibility. Task-switching paradigms have been used in a variety of test-retest designs, including the study of the potential cognitive benefits of action video game play (Boot et al., 2008; Green et al., 2012), and the study of the deleterious effects of total sleep deprivation on cognition (Couyoumdjian et al., 2010; Slama et al., 2018). The test-retest reliability and magnitude of PEs for various outcome metrics of some task-

switching paradigms have been previously described (affective task-switching: Eckart et al. (2021); colour-shape: Paap and Sawi (2016); number-letter task-switching: Soveri et al. (2018); Timmer et al. (2018); linguistic task-switching: Timmer et al. (2018); colour-shape and linguistic task-switching: Segal et al. (2021)), specifically within university student populations and with test-retest intervals ranging from 1-6 weeks. In task-switching paradigms, two common outcomes of interest are switch cost scores (SC; referring to the cost of responding to a changing cue/task/ruleset (hereafter simply ‘cue’) compared to when the cue remains the same) and mixing costs (MC; referring to the cost associated with *knowing* the cue *could* potentially change in a block of trials vs. knowing the cue *will not* change).

A specific task-switching paradigm which has not been assessed with respect to its test-retest reliability is the Category Switch Task (CST). Originally described by (Mayr & Kliegl, 2000), the CST requires participants to categorise target words as either *living/non-living*, or *smaller/ bigger than a basketball*, depending on a cue presented with or slightly before the target. The CST has been used (with adaptations depending on the purpose of the study) in research on the affective response to task-switching (Van Dessel et al., 2020; Vermeylen et al., 2019) and effect of reward on task-switching behaviour (Braem, 2017), as well as to test the congruency of targets (Schneider, 2015) and the effect of varying stimulus onset asynchronies (Schneider & Logan, 2014). It is one of two unpredictable (i.e., the participant is unaware when the cue will switch) task-switching paradigms (along with the colour-shape task) available on the popular *Inquisit* neuropsychological test platform (Millisecond Software, 2016). The CST was demonstrated by Friedman et al. (2008) in a single administration design to provide lower residual variance (i.e., task impurity and measurement error) than not only other task-switching paradigms (colour-shape and number-letter), but of all eight executive functioning tasks explored (when using the outcome measure SC), within a factor analysis model.

To date, only one study (Chihiro et al., 2017) has explored whether the CST (translated from English to Japanese) is susceptible to PEs, noting SC response time (RT) reduced from 230.5ms to 98.4ms following eight consecutive test blocks. However, given the blocks were performed consecutively, it remains unclear if / how the time interval between test and retest impacts PEs on the CST. Friedman et al. (2008) reported the split-half internal consistency for SC RT on a single administration of the CST as  $r = 0.85$ ; however, this does not capture error unique to the administration at a certain time point

and thus is not reflective of test-retest reliability. Hence, both test-retest reliability and PEs for the CST when considering SCs have not yet been examined; however it is also important to note that there are other metrics are often explored following the use of task-switching paradigms, such as *single task* performance and MCs.

Overall, the purpose of this study is to (a) describe the test-retest reliability of CST outcome measures, (b) to assess the presence and magnitude of test-retest PEs on these CST outcome measures, and (c) to explore whether presence and magnitude of PEs varied as a function of test-retest interval (*same day*, *next day*, and *next week*). Regarding (b) and (c), we hypothesised that performance in all outcome measures would improve from test to re-test, and that shorter test-retest intervals (*same day*) would lead to greater improvement compared to longer intervals (*next week*).

### 3.3. Methods

#### 3.3.1. Participants

51 healthy adults provided written informed consent to participate in the study. We did not consider the data of three participants due to the incompleteness of the second test session, English nonproficiency, or near chance performance. Our final sample included 48 healthy adults (mean age  $25.98 \pm 6.61$  years (range: 19 – 53), 17 females; see appendix 3.1 for demographics by *interval group*). All participants reported to be free from neuropsychological or neuromuscular disorders. All procedures described were approved by the Education and Health Sciences Research Ethics Committee at the University of Limerick, in accordance with the Declaration of Helsinki.

Participants were randomly assigned to one of three *interval groups* ( $N = 16$  in each group). In all groups, participants completed two test sessions of the Category Switch Test (CST; described below). In the first group (*same day*), participants arrived at the laboratory in the morning between the hours of 09:00 and 12:00 to complete their first test session. Participants then arrived back at the lab between the hours of 12:00 and 18:30 the *same day* to complete their second test session (mean interval length = 4hrs 19min  $\pm$  78min). In the second group (*next day*), participants completed their first and second test sessions at any time between the hours of 09:00 and 18:30 one day apart, given the times were similar between days (mean interval length = 23hrs and 59min  $\pm$  42min), while in the third group (*next week*), participants completed their first and second test sessions at any time between the hours of 09:00 and 18:30 seven days apart, given the times were similar between days (mean interval length = 7days  $\pm$  71min).

The Category Switch Task (CST) used in this study was adapted from that originally described by (Mayr & Kliegl, 2000) and administered using Inquisit 5.0 software, by Millisecond<sup>TM</sup>. Specifically, the test was shortened to one single task test block per cue, and two mixed task test blocks (see the procedure section below), and the stimulus response mappings between the response keys were pseudorandomised *across* test sessions. The task was completed by all individuals on the same computer, using the same computer peripherals and a 27-inch monitor with a 144hz refresh rate.

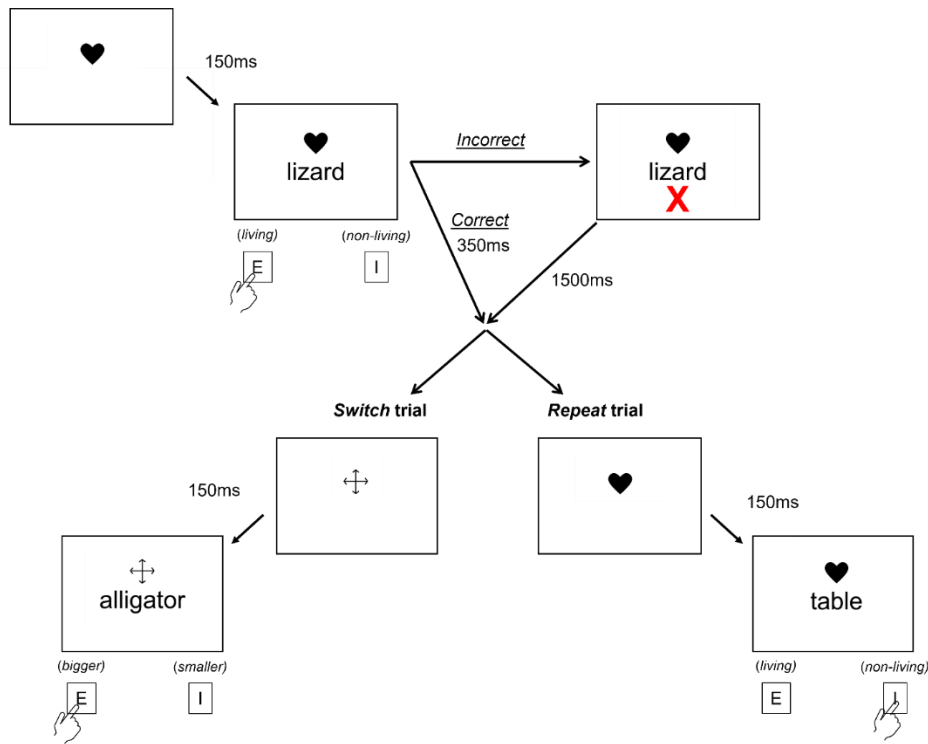
Within a CST trial, participants were presented with a word, which they categorised according to one of two distinct categorisation rules, determined by an image which served as a *cue*. Each word was randomly chosen from a set of sixteen words (Mayr & Kliegl, 2000), (see appendix 3.2 for the complete list). The first rule dictated whether a

given word resembled an item that was *living* or *non-living* (*living cue*), with this rule denoted by a *heart* image that appeared 150ms before the word and which remained present above the word until a response was inputted (cue-target interval or CTI = 150ms). A second rule dictated whether the word resembled an object that was bigger or smaller than a basketball (*size cue*), with this rule denoted by a *cross* image that also appeared with a CTI of 150ms. Participants inputted their responses using the *E* and *I* keys on the keyboard to classify the object as living/non-living or big/small respectively. There were four words that satisfied each living/size combination.

### 3.3.2. Procedure

Participants in all three groups completed two sessions of the CST, which each took approximately 15 minutes to complete. During each session, participants were presented with four test blocks (each 54 trials). A practice block (32 trials) preceded each of the first two blocks and allowed participants to familiarise themselves with the task. In the first block (single task living), the *living cue* was the only cue presented for all trials. In the second block (single task size), the *size cue* was the only cue presented for all trials.

In the third and fourth blocks, trials could be presented with either cue (mixed task), requiring participants to flexibly adapt their interpretation and categorisation of the word according to the cue provided in each trial. During this block, trials could be *switch trials*, in which the cue for the current trial differed from the previous trial, or *repeat trials*, whereby the cue for a current trial was identical to the previous trial. Participants used the cue provided to correctly classify each word using the *E* and *I* keys. Again, participants were presented a practice block, but this practice block consisted of a minimum of ten trials and continued until the correct response rate was 80%. Following the practice block, participants completed two test blocks of 54 trials. Within the mixed task blocks, cues were pseudorandomised such that there were an equal number of *switch* and *repeat* trials, with the constraint that there could not be more than four *switch trials* in succession. Figure 3-1 depicts both *switch* and *repeat trial* variations during the mixed task blocks of the CST.



**Figure 3-1** A visual representation of the third or ‘combined’ iteration of the category switch test. For each trial, a *living cue* or *size cue* is displayed alone for 150ms. After the 150ms, a word appears on the screen. The participant uses the ‘E’ or ‘I’ key to correctly categorise the word according to the rule associated with the cue. The inter-stimulus interval is 350ms for correct responses and 1500ms for incorrect responses. If the next trial has a different cue, it is a *switch trial*, whereas if the next trial has the same cue, it is a *repeat trial*.

Throughout all practice trials, visual error feedback was presented in the form of an *incorrect* message, with participants correcting the error before continuing. Within the testing trials, visual feedback for errors was provided however participants did not have to correct their error. There was a 350ms inter-stimulus interval (ISI) following correct responses, and a 1500ms ISI along with visual feedback (in the form of a *x*) for incorrect responses. Participants were told ‘try to minimize reaction time while avoiding making errors’.

The SRM between the *E* and *I* keys and *big*, *small*, *living* & *non-living* inputs was consistent within a given test session but was pseudorandomised *across* test sessions for a given participant and also among participants, such that 50% of the participants within each group had the mapping for one cue only changed, and 50% had the mapping for both cues changed.

### 3.3.3. Data Processing

The first two trials within each test block were classified as *buffer* trials for the given block (as per Paap and Sawi (2016)) and removed for any further analysis. Response times (RTs) beyond three standard deviations of the mean RT per participant per session were considered outliers and removed from subsequent analysis (2.00% of total responses removed as outliers).

Response time (RT; ms) and accuracy (Correct/Incorrect) were recorded for each trial within each test block. Performance on both single task blocks (*living cue* only, and *size cue* only) was pooled for analyses (collectively named *single task blocks*). For the mixed task, we explored two outcome measures commonly reported in task-switching literature; switch and mixing costs (SC and MC). We calculated SC as the difference in mean performance on switch trials and repeat trials, and considered MC as the difference in mean performance on repeat trials and trials within the *single task blocks* (see appendix 3.3 for further description of derived variables). In total, mean RT and response accuracy were considered for single task, SC, and MC, totalling six outcome measures.

### 3.3.4. Statistical Analysis - Test-Retest Reliability

The test-retest reliability of all outcome measures (irrespective of test-retest *interval*) was assessed using Pearson's  $r$  and intraclass correlation coefficient (ICC). Pearson's  $r$  demonstrates the strength and direction of linear relationship between test and re-test. We interpreted  $r < 0.60$  as *low* reliability,  $0.60 \leq r < 0.69$  as *marginal* reliability,  $0.70 \leq r < 0.79$  as *adequate* reliability,  $0.80 \leq r < 0.89$  as *high* reliability, and  $0.9 \leq r$  as *very high* reliability (Strauss et al., 2006). ICC further accounts for systematic error (i.e., PEs); we used ICC<sub>(2,1)</sub> as it is most appropriate when using ICC for test-retest reliability (Koo & Li, 2016). We interpreted ICC  $< 0.5$  as *poor* reliability,  $0.5 \leq \text{ICC} < 0.75$  as *moderate* reliability,  $0.75 \leq \text{ICC} \leq 0.9$  as *good* reliability, and  $0.9 < \text{ICC}$  as *excellent* reliability (Koo & Li, 2016). Lastly, we performed variance decomposition analysis on each outcome measure using the *psych* package (v 2.2.5; (Revelle, 2022)) in R (4.1.3; R Core Team (2022)).

### 3.3.5. Statistical Analysis - Practice Effects

Regarding practice effects (PEs), we hypothesised that PEs would be present for all CST outcome measures, and that the magnitude of PEs would be greater for shorter test-retest intervals.

To test if PEs were present across the entire sample, we pooled data from each interval group and compared scores at *test* against *retest* using a series of paired sample t-tests. Normality of difference scores (*retest* – *test*) was assessed through visual inspection and via the Shapiro-Wilk test; where the assumption of normality was violated, bootstrapping (Bias-corrected and accelerated (BCa), samples = 5000, 95%CI) was performed prior to the paired sample t-test. Differences between *test* and *retest* were considered significant if two-sided  $p < 0.05$ .

To test whether the magnitude of PEs varied across test-retest intervals, we ran a series of six bootstrapped one-way Analysis of Covariates (ANCOVAs). Bootstrapping (BCa, samples = 5000, 95%CI) was performed across all ANCOVAs due to performance at *retest* exhibiting non-normal distributions (inspected visually & using the Shapiro-Wilk test) in multiple groups for most outcome measures. In these ANCOVAs, we considered *interval* as the independent variable, with three levels (*same day*, *next day* & *next week*), and each performance variable at *retest* as a dependent variable. Performance at *test* was considered as a covariate. ANCOVA was chosen as it tends to produce less biased results when compared to other analytical methods factoring in baseline scores in a test-retest design (i.e., split-plot ANOVAs or change scores (Overall & Doyle, 1994; Stanley, 2022; Vickers & Altman, 2001). A main effect of *interval* ( $p < 0.05$ ) suggested that the magnitude of performance change varied as a result of test-retest interval. When this occurred, follow up Fishers LSD multiple comparisons were performed to determine which specific *interval groups* differed. Fishers LSD is an appropriate analysis given our design because it preserves type I error rate where three or fewer groups are tested (Hayter, 1986; Levin et al., 1994; Meier, 2006).

All analyses were performed using the IBM SPSS Statistics v28 (IBM Corp, 2021) software or R (4.1.3; R Core Team (2022)).

## 3.4. Results

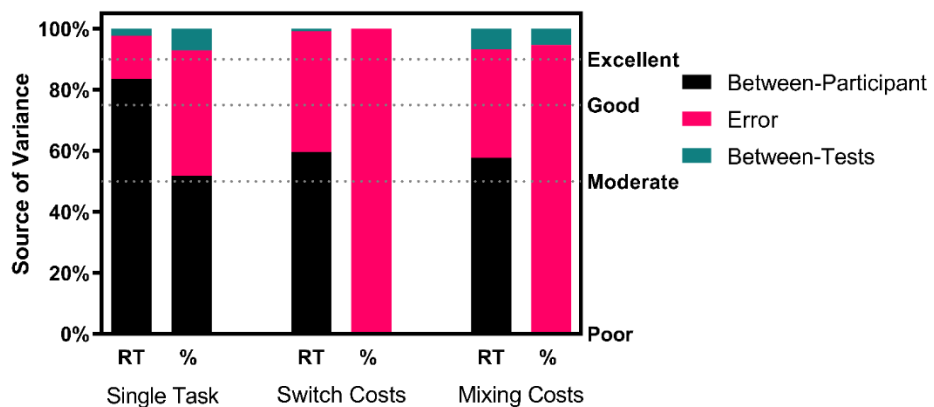
### 3.4.1. Test-Retest Reliability

Test-retest reliability was calculated for each of the six outcome measures. Reliability scores are shown in Table 3-1, and variance decomposition is shown in Figure 3-2.

**Table 3-1** Test-retest reliability results for outcome variables considered. Upper and lower 95% confidence intervals are given in squared brackets for ICC <sub>(2,1)</sub>.

Outcome Variable	Pearson's <i>r</i>	ICC <sub>(2,1)</sub>
<i>Single Task</i>		
<i>RT (ms)</i>	0.86*** (high)	0.84 [0.69, 0.91]*** (good)
<i>Accuracy (% correct)</i>	0.57*** (low)	0.52 [0.26, 0.70]*** (moderate)
<i>Switch Costs</i>		
<i>RT (<math>\Delta</math> ms)</i>	0.60*** (marginal)	0.60 [0.38, 0.75]*** (moderate)
<i>Accuracy (<math>\Delta</math> % correct)</i>	-0.08 (low)	0.00 [-0.28, 0.28] (poor)
<i>Mixing Costs</i>		
<i>RT (<math>\Delta</math> ms)</i>	0.66*** (marginal)	0.58 [0.32, 0.75]*** (moderate)
<i>Accuracy (<math>\Delta</math> % correct)</i>	-0.10 (low)	0.00 [-0.26, 0.27] (poor)

\*\*\* $p < 0.001$



**Figure 3-2** Stacked bar charts showing variance decomposition for the six outcome measures. As between-participant variance (black) is analogous to the reported ICC values, the thresholds used for defining reliability (poor, moderate, good, and excellent) are also provided.

### 3.4.2. Test-Retest PEs

All dependent variable difference scores were sufficiently normal besides mixing cost (MC) response time (RT), for which a bootstrapped paired samples t-test was performed. Paired-sample t-tests revealed a significant effect of test session (*test* vs. *retest*) for single task RT ( $\Delta RT = 30.52 \pm 10.04$ ;  $t(47) = 3.04$ ,  $p = 0.004$ , 95% CI [10.32, 50.72], Cohen's  $d = 0.44$ ), single task accuracy ( $\Delta \text{accuracy} = -1.46 \pm 0.46$ ;  $t(47) = -3.03$ ,  $p = 0.004$ , 95% CI [-2.43, -0.49], Cohen's  $d = -0.44$ ), and MC RT ( $\Delta RT = 73.29 \pm 23.00$ ;  $t(47) = 3.19$ ,  $p = 0.004$ , 95% CI [30.04, 121.30], Cohen's  $d = 0.46$ ), but not for SC RT ( $\Delta RT = 28.61 \pm 23.00$ ;  $p = 0.154$ ), SC accuracy ( $\Delta \text{accuracy} = 0.45 \pm 1.11$ ;  $p = 0.688$ ), or MC accuracy ( $\Delta \text{accuracy} = -1.53 \pm 0.83$ ;  $p = 0.072$ ).

### 3.4.3. Single Task

The covariate *performance at test* was significantly related to performance at *retest* for *single task* RT ( $F_{(1,44)} = 114.816$ ,  $p < 0.001$ ,  $\eta^2 = 0.723$ ). No main effect of the independent variable *interval group* was present ( $p > 0.05$ ), indicating that the change in performance from *test* to *retest* did not significantly vary as a function of test-retest interval.

The covariate *performance at test* was significantly related to performance at *retest* for *single task* accuracy ( $F_{(1,44)} = 26.990$ ,  $p < 0.001$ ,  $\eta^2 = 0.380$ ). A main effect of the independent variable *interval group* was present ( $F_{(2,44)} = 3.369$ ,  $p = 0.044$ ,  $\eta^2 = 0.133$ ), with multiple comparisons revealing significantly greater accuracy at retest ( $\Delta M = 2.21 \pm 0.86\%$ ,  $p = 0.014$ , BCa 95%CI [ 0.48, 3.94]) for the *next day interval group* ( $M = 97.34 \pm 0.45\%$ ) compared to the *next week interval group* ( $M = 95.13 \pm 0.68\%$ ).

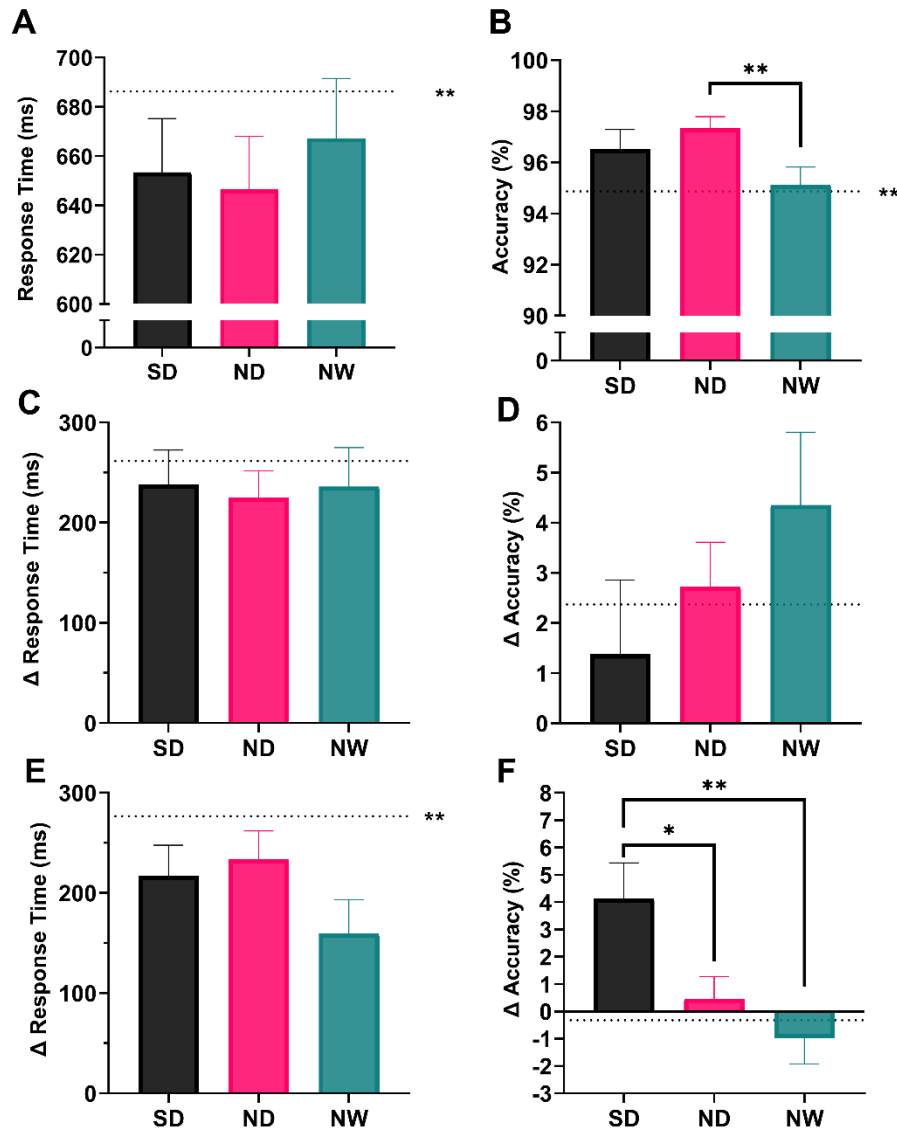
### 3.4.4. Switch Costs

The covariate *performance at test* was significantly related to performance at *retest* for SC RT ( $F_{(1,44)} = 24.64$ ,  $p < 0.001$ ,  $\eta^2 = 0.359$ ). No main effect of the independent variable *interval group* was present for SC RT ( $p > 0.05$ ). There was no main effect of *interval group* or significant relationship to *performance at test* for SC accuracy ( $p > 0.05$ ).

### 3.4.5. Mixing Costs

The covariate performance at *test* was significantly related to performance at *retest* for MC RT ( $F_{(1,44)} = 41.764$ ,  $p < 0.001$ ,  $\eta^2 = 0.487$ ). No main effect of the independent variable *interval group* was present ( $p > 0.05$ ) for MC RT.

The covariate performance at *test* was not significantly related to performance at *retest* for MC accuracy ( $p < 0.05$ ). A main effect of the independent variable *interval group* was present ( $F_{(2,44)} = 7.221$ ,  $p = 0.002$ ,  $\eta^2 = 0.247$ ), with multiple comparisons revealing that MC accuracy was significantly greater (i.e., a larger difference between *single task* accuracy and *repeat trial* accuracy in the *mixed task* blocks) in the *same day* ( $M = 4.13 \pm 0.97\%$ ) group, compared to both the *next day* ( $M = 0.46 \pm 0.81\%$ ,  $\Delta\% = 3.67 \pm 1.47\%$ ,  $p = 0.023$ , BCa 95%CI [ 0.95, 6.49]) and the *next week* ( $M = -0.97 \pm 0.96\%$ ;  $\Delta\% = 5.10 \pm 1.68\%$ ,  $p = 0.007$ , BCa 95%CI [1.89, 8.33]) groups. There was no significant difference between the *next day* and *next week* group.



**Figure 3-3** Bar charts depicting the ANCOVA adjusted mean ( $\pm$ SEM) retest values of the interval groups same day (SD), next day (ND), and next week (NW), adjusted for the ‘baseline’ test value (dotted line), for the six outcome variables of interest: **A** single task RT, **B** single task accuracy, **C** SC RT, **D** SC accuracy, **E** MC RT, and **F** MC accuracy. \*\* denotes  $p < 0.01$ , and \* denotes  $p < 0.05$ ; denotations next to the dotted line indicate a significant test-retest practice effect for the pooled data, and denotations above bars indicate a significant difference between different interval groups at retest.

### 3.5. Discussion

The current study aimed to uncover (a) test-retest reliability for various outcomes of a shortened Category Switch Task (CST), as well as (b) the presence and magnitude of test-retest practice effects (PEs) (c) among different test-retest intervals. Regarding (b) and (c), we hypothesised that performance on all outcome measures of interest within the CST would exhibit PEs, and that the magnitude of these effects would decrease as test-retest interval increased. We found greater test-retest reliability on the single task measure, compared to switch costs (SCs) and mixing costs (MCs), and on response time (RT) based outcome measures, compared to accuracy-based measures. We found evidence of test-retest PEs for all RT measures of the CST except SC, as well as for accuracy on the single task block. We did not find any evidence of PE magnitude difference between test-retest interval groups (*same day*, *next day* and *next week*) for any RT based measure; however, we found a *next day retest* to result in a greater single task accuracy improvement compared to a *next week* retest, and a *same day* retest to result in a greater MC increase compared to a *next day* or *next week* retest. Results and implications are discussed below.

#### 3.5.1. Test-Retest Reliability

Comparison between test-retest reliability score metrics found in the current study and other studies on task-switching paradigms can be found as appendix 3.4. Regarding single task RT, test-retest reliability was similar to or greater than those reported for single task components of other task-switching paradigms (Paap & Sawi, 2016; Segal et al., 2021 (data from Prior & Gollan, 2013)<sup>†</sup>; Soveri et al., 2018), and close to that expected ( $r = 0.9$ ) for a task with minimum 40 trials without consideration of day effects (Miller and Ulrich, 2013; cited in Paap & Sawi, 2016). Similarly, test-retest reliability of MC and SC was similar to or greater than most other task-switching paradigms. Only Eckart et al. (2021) reported an appreciably greater test-retest reliability for difference score measures, on an affective task-switching paradigm ( $r = 0.81 - 0.88$ ;  $ICC_{(2,1)} = 0.78 - 0.82$ ), however their design included many more trials per test session (48 switch trials & 192 repeat trials) than the current design (~52 switch trials & ~52 repeat trials). As Miller and Ulrich (2013) demonstrate using the individual differences in reaction time (IDRT) model, increasing the number of trials greatly increases test-retest reliability, particularly for difference score measures. This was also shown by Eckart et al. (2021), who in their own

<sup>†</sup> When referring to Segal et al., 2021 (data from Prior and Gollan (2013)), we refer only to their comparison between Test 1 and 3, which had a test-retest interval of one week, and not the comparison between Test 1 and 2 which immediately went from test to retest.

analysis noted smaller (though still adequate/ moderate) test-retest reliability ( $r = 0.74$ ;  $ICC_{(2,1)} = 0.66 - 0.69$ ) when only half the trials were included. Overall, the CST demonstrated test-retest reliability that was comparable-to or greater-than other task-switching paradigms for RT measures, boding favourably for its use both in single-test and test-retest design studies.

Test-retest reliability for accuracy measures was found to be much lower when compared to the test-retest reliability of RT measures from the same sub task type (single task, SC & MC). At least regarding SC, this finding corroborates work by Eckart et al. (2021), who found “consistently better psychometric properties for RT-based switch costs” (Eckart et al., 2021, p. 929) when compared to error based measures in their affective task-switching paradigm. While difference scores can often show poor reliability for individual differences despite remaining suitable for group level comparisons (Hedge et al., 2018; Paap & Sawi, 2016; Segal et al., 2021), test-retest reliability should at least reach statistical significance to be suitable for testing for group level differences (Paap & Sawi, 2016). Given non-significance of SC and MC accuracy test-retest reliability, we advocate against the consideration of SC and MC *accuracy* when using our shortened version of the CST. Note that *interval* effects are not responsible for the poor reliability of these measures, as the poor reliability remains when analysis is performed on individual *interval groups* (see appendix 3.5 for these results).

### 3.5.2. Test-Retest Practice Effects

We found evidence for test-retest PEs for both single task RT and accuracy, as well as MC RT, *but not* SC RT, and both SC and MC accuracy (Figure 3-3). While a lack of evidence for PEs in SC RT is contrary to our hypothesis, it is welcome, as it suggests that when SC RT (the primary outcome measure in most studies using task-switching paradigms) is considered as an outcome measure of the CST in a test-retest design, PEs are unlikely to systematically bias results. Regarding the effects of test-retest PE on SC and MC RT, previous analyses on task-switching paradigms have produced conflicting findings. Eckart et al. (2021) found reduced switch trial RTs but not repeat trial RTs at retest (*interval* = 11 – 17 days), resulting in a SC improvement of ~29 - ~41ms, while conversely Segal et al., 2021 (data from Prior and Gollan (2013)) found MC but not SC to improve with practice on a colour-switch task. Of course, differences in the specific paradigm used and its administration (number of trials, extent of prior practice) can all play a role in PE presence and magnitude. Our analysis suggests that those seeking to use

the CST in a test-retest design and aiming to analyse single task RT and/ or accuracy, or MC RT, should consider PE mitigation strategies such as extensive pre-testing practice (potentially with use of alternate forms; Beglinger et al. (2005)), counterbalancing (Greenwald, 1976), or using the performance of control groups (i.e., standardized regression-based methodology; McSweeny et al. (1993)).

### **3.5.3. Effect of ‘Interval’**

We found the magnitude of PEs to vary as a function of *test-retest interval* for single task and MC accuracy measures, however not for RT measures, or SC accuracy. Regarding single task accuracy, the direction of this effect corroborated with our hypothesis; a larger test-retest *interval* resulted in smaller PEs compared to a shorter interval (though this only reached significance when comparing to the *same day* group) (see Figure 3-3). A lack of effect of *interval* on RT measures, while contrary to our hypothesis, does corroborate with previous research finding no effect of *interval* on cognitive tests when only short-term (> 2 weeks) *intervals* were examined (Farahat et al., 2003; Salthouse & Tucker-Drob, 2008). Overall, our results suggest that researchers looking to use RT outcome measures on the CST should not be concerned about differences in *interval* confounding a test-retest design by producing different PE magnitudes, so long as *interval* remains between a few hours and one week.

### **3.5.4. Limitations**

Firstly, although all included participants were proficient English speakers, we did not explicitly check whether English was the first language of participants. Secondly, we did not check whether participants only spoke English, or whether they were bilingual. This is relevant as there is evidence that bilingual individuals possess enhanced task-switching abilities that persist beyond language-switching (Declerck et al., 2017; Prior & Gollan, 2013; Prior & Macwhinney, 2010; Weissberger et al., 2012), though this domain-generalizability is highly disputed (de Bruin et al., 2015; Paap & Greenberg, 2013; Paap et al., 2015; Paap et al., 2017). Larger studies allowing more control over participants should look to capture or control for bilingualism among groups. Thirdly, we examined RT and accuracy as separate entities, as is commonplace among studies using task-switching paradigms. However, we note that alternative approaches which have combined RT and accuracy measures have been developed and proposed for SC in task-switching paradigms, with comparable (and sometimes superior) reliability and validity compared

to RT alone (Hughes et al., 2014). These approaches have the additional advantage of accounting for potential speed-accuracy trade-offs which may be present. Fourthly regarding the assessment of PEs, we note that only one retest bout was performed. Performance improvement may be observable following many more than two repetitions of cognitive test administrations (i.e. Watson et al. (1994). It is possible that should more testing bouts have been performed, group differences compared to baseline may have been observable for measures such as SC RT or MC accuracy. Our test-retest design did not allow such performance improvement to be captured. Lastly, while the sample tested in this *pilot* study was less than 50 individuals (Hopkins, 2000), the sample size used was comparable to/ greater than those used in many of the studies used as comparisons (N = 47, Eckart et al., 2021; N = 34, Soveri et al., 2018; N = 53, Timmer et al., 2018).

### **3.5.5. Conclusions**

In conclusion, in the current article we demonstrate the shortened CST to produce reliable values for the outcome measures single task RT, single task accuracy, switch cost RT, and mixing cost RT. However, researchers should be cautious when seeking to analyse SC and MC accuracy based measures of the CST in their experimental designs. Of the above outcome measures recommended for use, researchers should be aware that without mitigation strategies, test-retest practice effects could confound results when single task RT, single task accuracy, or mixing cost RT are considered as outcome measures.

### 3.5.6. Linking Section

The investigation into the test-retest reliability and practice effects of a shortened CST task, presented in this chapter, both augmented and provided confidence in my use of the task as a high-salience flexible (HSF) task in my experimental sleep loss study. Firstly, all RT based outcome measures (single task, SC and MC) from the shortened CST presented with test-retest reliability that was either similar to or superior than other comparable task-switching paradigms. However when observing accuracy outcome measures, test-retest reliability was poor and insignificant for SC and MC. As stated by Paap and Sawi (2016), test-retest reliability should be significant to be suitable for use in testing group differences (see also Dorrian et al. (2004) for discussion about test-retest reliability specifically in sleep research). Hence, this study informed us that SC and MC accuracy were not reliable outcome measures that were suitable for use. Secondly, the study informed us that PEs may be a potential source of bias within the test-retest interval (1-week) I planned to use in the experimental sleep loss study. Given that recruitment of a specialised population was necessary for my experimental sleep loss study (with many individuals likely residing outside of Limerick and unable to attend one or multiple prior training sessions), prior training to asymptote on the CST was not a feasible option to overcome practice effects. I also note prior test-retest studies using task-switching paradigms that exhibit evidence of PEs, despite attempts to train to asymptote before study commencement (i.e., Couyoumdjian et al., 2010). Statistical approaches to overcome PEs (i.e., McSweeney et al., 1993) presented as limiting and overall unappealing. Hence, it was decided that counterbalancing was the ideal approach to overcome PE issues that may arise from some CST outcome measures in the experimental sleep loss study.

At this stage, I have identified the cognitive tests that I will use in the experimental sleep loss study (PVT and CST) to assess the performance of esports players on *low salience* and *high salience flexible* tasks (as per **Chapter 2**). The additional benefit to using the CST is the *single task* component of the test, would be considered a *high-salience stable* task using the categorisation rules of **Chapter 2**, and hence through using these two tests, I am covering all three categories outlined in Figure 2-3. However beyond cognitive performance, the current thesis aimed to explore how sleep loss would impact in-game esports performance. The following two chapters discuss the target esports examined within the thesis; *Rocket League*. **Chapter 4** aims to provide an introduction to the esports

for the esports *naïve* reader, while **Chapter 5** discusses performance and rank indicators (and hence, outcome metrics of interest) within this esports.

## **Chapter 4.      An introduction to Rocket League**

As mentioned in the introduction, a primary outcome of the current thesis is to explore if, how, and how much, sleep loss specifically impacts esports performance. Esports refers not to specifically one game, but to any video game played competitively in an organised manner. A recent compendium identified 1,007 video games which could be considered as an esports (Independent Electronic Sports, 2023). Within the esports umbrella is great diversity in-game dynamics, strategies, and specific human attributes more influential toward in-game success.

Importantly however, there is great diversity in how practical different esports are for use in an experimental research setting, if one wants to consider and measure performance *in-game*. Regarding use practicality, a first important factor is the game length where short and predictable is ideal, to allow for multiple matches within a given test session which spanning across a predictable (and hence, manageable) period of time. A second important factor is game popularity; an otherwise suitable esports is not worth consideration if there are not sufficient players of that esports within the community to render a study feasible. A third factor is in-game data availability, which varies between esports for various reasons, such as fear of “cheaters” using available data to create tools which provide an unfair in-game advantage (Reitman et al., 2019). A final factor is whether an esports includes competitive play as individuals (1v1), as opposed to as a team; this is as the inclusion of teammates greatly complicates analyses due to requiring the consideration not only of interactions between opponents but also of interactions between teammates (Ofoghi et al., 2013). Measuring gameplay within an 1v1 environment also removes the factor of in-game roles, which could complicate the interpretation of any performance data collected.

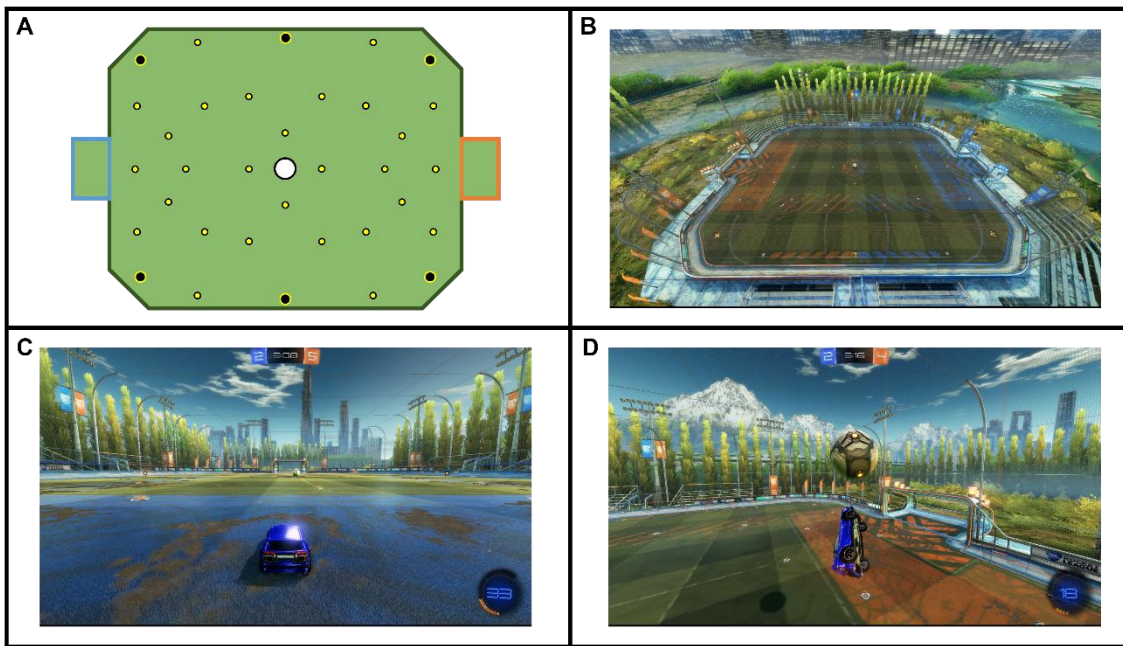
With a large playerbase (averaging ~90 million players per month, Active Player (2023), short and predictable match lengths, data-availability, and competitive play as individuals (1v1), *Rocket League* presented as an ideal esports with which to perform experimental research. Hence, the primary aim of this chapter is to provide the reader with an introduction to the esports Rocket League. The secondary aim of this chapter is to provide information regarding the process of obtaining in-game metrics within Rocket League, and an overview of the types of metrics that are obtained and used frequently in the analysis of competitive/ professional Rocket League. The hope is that following this chapter, the reader will have sufficient knowledge of Rocket League and in-game Rocket League metrics to facilitate an understanding of my analysis of performance and rank

indicators (PIs and RIs respectively) within 1v1 Rocket League, which is outlined in **Chapter 5**.

## 4.1. Rocket League

Rocket League, released in 2015 by the video game developer Psyonix LLC (now owned by Epic Games) can be described as a *vehicular soccer video game* (Smithies, Campbell, et al., 2021). Conceptually, it is one of the easiest esports to grasp for individuals new to esports. At a competitive level, teams consist of either one (1v1), two (2v2) or three (3v3) players. Rocket League can be played using both controller (i.e., Dualshock 4 or Xbox One controller) or a keyboard and mouse as input modalities.

In Rocket League, each player controls a single vehicle, viewed from a *third-person* camera angle (of which players have a large degree of control and customisation over). Players are free to use vehicles with one of six *hitboxes*. Vehicles of different hitboxes have minor differences in how they manoeuvre and interact with the ball in-game (i.e., length, width, height, centre of mass location, and turning circle of the vehicle). There is one spherical ball in play, which may spin when it hits the ground but which does not curve in the air (i.e., no Magnus effect) due to spin imparted on the ball. Teams must use their vehicles to prevent the ball from entering their goal (shaped similarly to a soccer goal) while simultaneously trying to hit the ball into their opponent's goal. The goals are situated on opposite ends of the *map*, which is a playing surface that is standardised in terms of length, width, and shape (rectangular, with soft edges), and are completely enclosed by walls and a ceiling (see Figure 4-1). Players are able to drive on these surfaces if desired. At the start of each match and following a goal, a *kickoff* commences; this places a member of each team at an identical distance to the ball, with a three second counter occurring before game (re)commencement.



**Figure 4-1.** **A** A simplified birds-eye view of a standardised Rocket League map. Blue and orange rectangles denote goal locations. The white circle denotes the location of the ball at the start of the match or immediately following a goal. Yellow circles (black outline) denote the location of *small pads*, and black circles (yellow outline) denote the location of *big pads*. **B** A top-down view of a Rocket League map (“Beckwith Park”), allowing view of the map walls and ceiling. **C** Players view of a kickoff about to commence. Kickoffs place a member of each team at equal distance to the ball, and allow players to move at an identical time. **D** A players view mid-game; in this instance, within an aerial-based attacking play. In **C** and **D**, the current score and time left can be seen in the top of the screen, and the player’s current boost total can be seen in the bottom-right of the screen; this information is present to players throughout the entirety of a Rocket League match.

Throughout the game, players may *drive*, *jump*, *double jump*, and *dodge* (a game mechanic which flips the car in a specified direction) to move and strike the ball. Players may also use *boost*, a finite resource collected at locations on the playing surface of the map, which provides acceleration of one’s vehicle (up to an inbuilt maximum speed). By using boost and pointing the nose of their vehicle away from the ground, players are also able to fly into the air. Lastly, players are able to remove their opponents from the map for three seconds by driving into them at ~95-100% of maximum speed and at a certain range of angles (this action is called a *demo*). For a further overview, I guide the reader to a 12-minute introductory video for Rocket League (Pilkin, 2022).

For the purposes of my experimental research, I focussed on 1v1 Rocket League. While most professional Rocket League is 3v3, there is a considerable (and growing) amount of competitive and professional Rocket League which is played as 1v1. In 2022 and 2023, 1v1 formed a major part of the Gamers8 Rocket League event, which possessed a 2 million USD prize pool on both occasions (Liquipedia, 2022, 2023). Elsewhere, it is estimated (using data from rlduels.gg (2023), provided as appendix 4.1) that as of June 22 2023, >€750,000<sup>†</sup> has been earned as prize money through 1v1 Rocket League.

## 4.2. In-Game Rocket League Metric Categories

As mentioned, Rocket League is uniquely positioned for research purposes in part because in-game data are readily available. In-game data can be made accessible using ballchasing.com, which is a large data repository for Rocket League match replays. The use of ballchasing.com is universal within the Rocket League gaming community, with over 90 million Rocket League match replays publicly uploaded to the repository as of 30/05/2023; this is inclusive of official Rocket League tournament and World Championship matches. One of the key features of ballchasing.com as a data repository is its Application Processing Interface (API) applied to all uploaded replays, which strips in-game data stored within the replay, and presents them as metrics generally considered relevant within the Rocket League community. Further, ballchasing.com subsequently allows for the downloading of these in-game metrics as .csv files. Essentially, this means that any individual can obtain a complete summary of any of the >90million Rocket League matches publicly uploaded to ballchasing.com (alongside any matches they upload privately), presented as 72 data points, per individual player, per match. Clearly, this is of extreme benefit if one seeks to explore in-game performance, either for an individual matches or for trends across many thousands of matches.

In **Chapter 5**, I utilise the ballchasing.com API to obtain in-game metrics for 21,588 Rocket League matches. Following data processing and steps (outlined in **Chapter 5**), I obtained 28 in-game metrics (26 when considering difference scores, i.e., differences between player and opponent), for which analysis was undertaken from. These metrics are divided into four categories for the purpose of **Chapter 5**; offense/ defence, boost, movement, and positioning. While a brief explanation of all 28 metrics is provided in

<sup>†</sup>I considered “Winnings” values presented by rlduels.gg as USD. While rlduels.gg inputs prize earnings in the currency they are presented, the vast majority of 1v1 showmatches or events with prize money provide potential earnings in USD – however, we accept as a limitation of this estimation that not all values included were necessarily provided in USD.

appendix 4.2., a brief description of these four categories is provided here to assist the reader's understanding of metrics in **Chapters 5, 7, and 8**.

#### **4.2.1. Offense/ Defence Metrics:**

These metrics relate to specific in-game events (saves, shots, and demos) rather than information regarding boost, movement, or positioning. These metrics are often the first considered and discussed by players and analysts alike. Live information regarding shot and save count is available to players and spectators throughout any Rocket League match. These are akin to metrics such as shots taken, saves made, tackles made, passes made etc. within soccer.

#### **4.2.2. Boost Related Metrics:**

Boost is a finite resource present in Rocket League which, when used, allows for rapid car acceleration and greatly assists in the ability to get high in the air. Boost can be collected from big pads (6 locations, which provide 100 boost and respawn every 10 seconds) or small pads (28 locations, which provide 12 boost and respawn every 3 seconds), which are in standardised locations on each map. A player can only have a maximum of 100 boost at any given time. Given that (a) access to boost is competed for between players and (b) boost is necessary to fly in the air and provides significantly greater speed and steering control, a player's ability to control their access and use of boost is considered highly beneficial to overall performance.

#### **4.2.3. Movement Related Metrics:**

Players have an incredibly large degree of control over the movement of their vehicles in Rocket League. On the ground, they can drive forward, reverse and turn in any direction. Just through driving, players can reach a maximum speed of 1400uu/s (unreal units per second; unreal units are the distance unit used within game), however through using boost, players may reach a maximum speed of 2300uu/s. If a player travels between 2200uu/s to 2300uu/s (~95-100% of maximum speed), they are considered to be *supersonic*. When supersonic, players will not lose speed even if boost is not inputted, provided they do not hit anything or turn beyond a certain angle. Being supersonic also allows players to demo opponents. Additionally, players may use *powerslide*, an ability which allows the vehicle to drift and thus have a tighter turning radius; using powerslide alongside boost can allow for different and precise turning movements. Powerslide also

allows players to maintain forward momentum while landing if a players vehicle is not facing in the exact direction of their movement. Players are also able to jump, and if within 1.25-1.45 seconds of a first jump (depending on the size of the jump), double jump. If a jump occurs while the players vehicle is in the air, the player may use the jump as a dodge (flip), which can be performed in a multitude of directions. Dodges can be used for hitting the ball or gaining speed beyond the maximum allowable through driving only (i.e., without boost), as well as for more sophisticated in-game *mechanics*, such as a *wave dash* and a *flip reset*. Lastly, players can fly in the air by *boosting* while aiming the front of their vehicle away from the ground. While in the air, players can control their vehicle in all three axis of motion (pitch, roll, and yaw).

Movement based metrics in Rocket League could draw analogy to metrics regarding a players movements on a soccer pitch (as measured by global positioning systems/ GPS), such as total distance covered, and proportions of matches spent walking/ jogging/ sprinting for example.

#### **4.2.4. Positioning Related Metrics:**

Like boost and vehicular movement control, optimal positioning is considered to be crucial to success in Rocket League. Unlike conceptually similar invasion ball traditional sports like soccer or hockey, Rocket League does not have set positions and roles due to team and map size (see Pilkin (2022) for elaboration), even when played as 2v2 or 3v3, let alone for 1v1 competition. Hence, players must play the roles of both an attacker and defended depending on the situation, and position one's vehicle accordingly. Positioning based metrics could again draw analogy to metrics taken by a Global Positioning System (GPS) in soccer for example, such as time spent in the penalty area, or time spent in front of/ goalside of the ball.

## **Chapter 5. A Random Forest approach to identify metrics that best predict match outcome and player ranking in the esports Rocket League**

This chapter has been published in a modified format in *Scientific Reports*:

Smithies, T. D., Campbell, M. J., Ramsbottom, N., & Toth, A. J. (2021). A Random Forest approach to identify metrics that best predict match outcome and player ranking in the esports Rocket League. *Scientific reports*, 11(1), 1-12. <https://doi.org/10.1038/s41598-021-98879-9>

Changes to the abovementioned publication for the purposes of this thesis are outlined below:

- The original article was written in the format Introduction → Results → Discussion → Methods, as per journal guidelines. To ensure consistency with other chapters and to facilitate readability, the order of sections has been changed to Introduction → Methods → Results → Discussion. As such, the first few paragraphs in the Methods section (up to the *Data Processing* subheading) was originally within the Results section, and was moved to ensure coherency within the chapter.
- Change in referencing style (article version is in numbered format).
- References to supplementary files are changed to the appropriate location within the appendix, or to an OSF online repository link for supplementary data.
- Words emphasised using quotation marks were changed to be emphasised using italics, in line with the thesis format.
- The words *Figure* and *Table* in in-text references to figures was capitalised. Furthermore, figure/ table numbering convention was changed in line with the thesis format.
- Addition of a *linking section* for the purpose of thesis flow.
- Minor amendments have been made based on examiner correction suggestions.

## 5.1. Abstract

Notational analysis is a popular tool for understanding what constitutes optimal performance in traditional sports. However, this approach has been seldom used in esports. The popular esports *Rocket League* is an ideal candidate for notational analysis due to the availability of an online repository containing data from millions of matches. The purpose of this study was to use Random Forest models to identify in-match metrics that predicted match outcome (performance indicators or *PIs*) and/or in-game player rank (rank indicators or *RIs*). We evaluated match data from 21,588 *Rocket League* matches involving players from four different ranks. Upon identifying goal difference (GD) as a suitable outcome measure for *Rocket League* match performance, Random Forest models were used alongside accompanying variable importance methods to identify metrics that were *PIs* or *RIs*. We found *shots taken*, *shots conceded*, *saves made*, and *time spent goalside of the ball* to be the most important *PIs*, and *time spent at supersonic speed*, *time spent on the ground*, *shots conceded* and *time spent goalside of the ball* to be the most important *RIs*. This work is the first to use Random Forest learning algorithms to highlight the most critical *PIs* and *RIs* in a prominent esports.

## 5.2. Introduction

The popularity of esports (competitive organised video game play) has grown rapidly over the past ten years to the point where viewership now rivals that in many traditional sports. In fact, it has been estimated that over one billion individuals viewed esports content in 2020 (Ahn et al., 2020). This rapid rise in interest in esports has led to increasing professionalisation, investment and attention towards optimising performance among the top players with the ultimate goal of individual or team success. However, extremely little research exists to date exploring what constitutes optimal performance within various esports. Until the factors that determine optimal performance are understood within a given sport (or any activity more broadly), it is very difficult to create and implement effective and efficient strategies towards achieving optimal performance.

For most tasks, optimizing performance is often predicated on the identification of performance indicators (PIs; individual variables that predict the overall outcome of a match or performance). A very popular approach to identifying PIs is a *notational approach*. Notational analyses is the study of patterns within a match/ contest/ competition/ performance that lead to a successful overall outcome (Hughes & Bartlett, 2002) and can uncover the components most important for match outcome. In traditional sports, identifying the PIs most important for successful task performance helps players and coaches to better direct focus to those key components to accelerate learning and, ultimately, improve performance. Thus, in traditional sport research, many have employed a notational approach to identify PIs in Australian Rules Football (Robertson et al., 2016), basketball (García et al., 2013; Leicht et al., 2017), ice hockey (Gu et al., 2016), rugby league (Whitehead et al., 2020; Woods et al., 2017), and rugby union (Bennett et al., 2020; Bennett et al., 2019; Bishop & Barnes, 2013; Hughes et al., 2017; Mosey & Mitchell, 2020; Vaz et al., 2010).

By using notational analysis to understand the components of an activity that are most important to success, one can direct their attention to those components to accelerate learning and ultimately improve performance. An example of a training method that could benefit from this understanding is Variable Priority Training (VPT), in which individuals complete a task with focused attention specifically towards improving key PIs within the task (Boot et al., 2010). VPT has been demonstrated to enhance learning in video game

contexts when compared to Fixed Priority Training (FPT: focussing on all aspects of a task) (Boot et al., 2010; Voss et al., 2012).

In light of the evidence above, notational analyses may stand to benefit esports. Notational analysis could be used in a similar way to that in the traditional sport examples mentioned above to find the most important PIs within a given game to focus on, resulting in more efficient training and use of techniques such as VPT to improve esports performance. However, little research has explored this topic in esports to date. One recent study has started to identify the important PI's for differentiating expertise in the First Person Shooter (FPS) esports, CS:GO, which has informed commercially available training software (Toth, Ramsbottom, et al., 2021), while two others have identified PIs in Multiplayer Online Battle Arena (MOBA) esports (Novak et al., 2020; Xia et al., 2017). The lack of notational analysis and subsequent analysis in esports is surprising given that esports appear ideal for such analyses as they are played digitally, with the ability to store in-game metrics directly for any game. However similarly to *traditional sports*, esports are extremely diverse in-game mechanics, objectives, equipment, and team size and structure, meaning that PIs from one esports are unlikely to be relevant to another. Additionally, in-match data can be difficult to obtain as they are often not made available by game development companies.

One such esports whereby performance data are readily available, making it an ideal candidate for notational analyses, is Rocket League. Rocket League is a *vehicular soccer video game* released in 2015 by Psyonix. In Rocket League, players each control a rocket-powered vehicle with the goal of hitting a large ball into a goal that is similar to a football/soccer goal, while simultaneously defending their own goal. The popularity of Rocket League has rapidly escalated since it became free-to-play on September 27 2020, with its peak concurrent player count of 1.85 million surpassing the popular esports mainstay, CS:GO, by more than 500,000 (Hindi, 2020; Moore, 2020). Alongside this high concurrent player count, Rocket League has reported ~90 million monthly users every month since November 2020 (Active Player, 2021b), approximately triple that received for CS:GO (Active Player, 2021a). Additionally, Rocket League has a thriving esports scene, with competing teams from top esports organisations such as Team Liquid, G2 esports, and NRG esports, and with ~12million USD won through Rocket League competition (As of 12/03/2021; Esports Earnings (2021)). Overall, its popularity, the drive for optimising player performance at the top levels and the wealth of freely and

readily available match data position Rocket League as an ideal candidate for notational analysis and the identification of the PIs that predict performance outcomes in this esports.

In Rocket League, players can save match replays and upload them to *ballchasing.com*, which in turn makes over 65 in-match metrics publicly available. As of May 5, 2021, there are over 24.5 million match replays available on the *ballchasing.com* online repository, across matches of various formats and with players of various ranks, freely available for anyone to download. Such volume of readily available match data is unheralded in esports and in traditional sports.

Previous research has employed the use of general linear mixed effects models (Novak et al., 2020) or solitary classification and regression trees (CARTs) (Xia et al., 2017) for PI identification within esports. While these methods have their benefits, one superior approach that has yet to be adopted in esports is the use of Random Forest models (Breiman, 2001). Random Forests are a machine learning ensemble algorithm and refer to an ensemble of CARTs each trained using a unique bootstrapped data set and random selection of splitting predictor features. Each case in the original data set is then run through all CARTs in the forest for which it was not part of the training process (and hence is *out-of-bag* or *OOB* for these CARTs), and the mean (for a regression model) or modal (for a classification model) response is considered the overall response of the model for that case.

Random Forests are a superior option to linear or logistic models and solitary CARTs for the current data and objectives for many reasons. Firstly, Random Forests can incorporate non-linear effects, and are superior to alternate methods at modelling complex interactions when the interactions are not, or cannot be, pre-specified (Cutler et al., 2007). This is ideal given the exploratory nature of PI identification in esports research and the unknown properties of the metrics included in model creation. Moreover, Random Forests have no distributional assumptions for predictor or response variables and are thus resistant to bias from non-parametric data, skewed data, and even nominal data, and perform exceptionally well even when many predictors are weak (or *noise*) (Breiman, 2001; Cutler et al., 2007; Díaz-Uriarte & Alvarez de Andrés, 2006). Moreover, the fact that Random Forests are an amalgamation of many CARTs using a bootstrapped data samples and a random selection of predictor variables for node splitting per tree, they inherently provide much greater predictive ability and reduce propensity for overfitting when compared to the CART method alone (Breiman, 2001; Siroky, 2009), making them

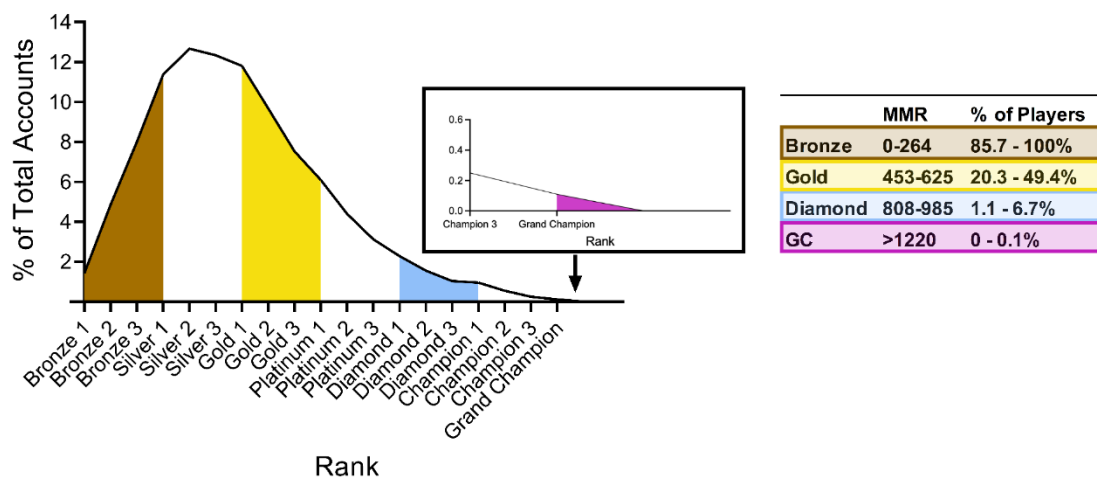
suitable for large datasets. Given the above advantages over existing methods and that Random Forest have been used previously to identify PIs within traditional sports (Bennett et al., 2020; Bennett et al., 2019; Mosey & Mitchell, 2020; Whitehead et al., 2020; Woods et al., 2017) they are arguably the most optimal method to identify PIs in Rocket League and esports more broadly.

By leveraging the immense amount of freely available match data in Rocket League and utilising the state-of-the-art notational approach of Random Forest machine learning modelling, the purpose of this study is to identify metrics that predict performance (PIs) and expertise (RIs) within the esports, Rocket League. Specifically, we aimed to first identify a suitable match outcome measure that could capture more information than provided by binary win vs. loss. We then aimed to identify in-match metrics that best predict our match outcome measure, across a variety of player ability levels. Finally, we aimed to also identify in-match metrics that best predict the ability level of the players within matches themselves.

### 5.2.1. Methods

While Rocket League can be played either individually (1v1) or in teams of two or three, the analyses of multiplayer competition requires consideration of the interactions between teammates, which is a necessary factor for team sports/esports and can greatly complicate analyses (Ofoghi et al., 2013). Therefore, this study focused solely on 1v1 Rocket League, which benefits from the fact that match metrics in this format are a direct result of player actions or interactions between player and opponent.

The data from four Rocket League rank groupings were considered for our analyses: Bronze, Gold, Diamond, and Grand Champion (GC). These rank groups were chosen to allow for the capture of a broad range of ability levels while simultaneously creating clear distinctions between each rank group (see Figure 5-1). Ranks within Rocket League correspond to a player's matchmaking rating (MMR). A player's MMR increases after every win and decreases after every loss, with the magnitude of the increase/decrease determined by the difference between players' MMR before the match.



**Figure 5-1** A density plot showing the distribution of accounts within the Rocket League rank system. Colour shaded areas correspond to the skill brackets, and associated MMRs, considered for the current study. This distribution is as per season 14 of Rocket League, which was the season at the time of the most recent match used in the analysis.

Data from 33,854 total matches were downloaded from *ballchasing.com* (<http://www.ballchasing.com>), a repository of Rocket League match replays and statistics, on 16/12/2020. In addition to downloading all the data for all Bronze (4,111 matches) and GC matches (9,743 matches), we downloaded all the data for the most recent 10,000 Gold and Diamond rank matches respectively. Data were gathered from

matches prior to September 29, 2020, and this was done for two reasons. Firstly, an update to the game with an accompanied rank redistribution saw additional ranks added after this date. Secondly, this update hindered the ability for *ballchasing.com* to recognise the ranking of players within a match. These issues have since been resolved, however were such during our data collection and analysis that we did not include match data from after September 29, 2020. These data were downloaded directly from the public domain, are freely available to all individuals, and results are completely de-identified. Further, all General Data Protection Regulations (GDPR) have been fulfilled.

### 5.2.2. Data Processing

Using the website's inbuilt filters and replay group function, match statistics were downloaded as a .csv file. Each match file contained general descriptions of the match (i.e., map, player names, cars used) as well as 65 columns corresponding to data describing the performance for 65 in-match metrics (potential performance (PIs) and rank (RIs) indicators) (<https://osf.io/z2fjg/> contains an anonymised sample file directly from *ballchasing.com*).

From here, many processing steps were undertaken to result in the final 28 *raw-score* metrics and 26 *difference-score* metrics included in the Random Forests analyses (see Table 5-1). We have provided a brief description of these steps below, however the reader is directed to appendix 5.1 where we provide a detailed description of these steps, allowing for reproduction.

First, we calculated match length using metrics provided, and used this to normalise all metrics that were not already presented as a percentage of match length to the average length of a rocket league match (360 seconds). Second, we removed all *draws* in the data, as well as matches that did not exceed 150 seconds duration to avoid overestimation of time normalised data. Next, we recalculated average speed using these time measures, and used metrics provided to calculate the metric *True boost wastage*. True Boost Wastage represents the proportion of *boost* used when a player is already travelling at max or near max speed. It is generally considered a measure of poor *boost* use, or wasted boost (Rocket Sledge, 2019; SquishyMuffinz, 2020). Appendix 4.2 contains descriptions for *boost*, true boost wastage and all other metrics are described in greater detail.

From here, we calculated *difference-scores* for each metric (the difference between a given player and their opponent's metric values). This was done in light of evidence that

difference-scores can provide superior predictive ability compared to *raw-score* metrics in a Random Forest analysis of PIs in Rugby Union (Bennett et al., 2019). We then maximised independence of data by removing all games besides the most recent ten from a given player, and de-identified the data. Penultimately, we ensured that no metrics could be combined to entirely explain the variance of another included metric. Lastly, shots conceded difference and demos taken difference were removed, as these metrics mirrored shots taken difference and demos inflicted difference metrics respectively (see appendix 4.2 & 5.1).

Following the above processing steps, 28 raw-score predictor metrics and 26 difference-score predictor metrics were retained per match. *Raw-score metrics* and *difference-score metrics* were split into two in separate dataset files and metrics in each file were divided into four categories, offense/defence metrics, boost metrics, player movement metrics and player positioning metrics (see Table 5-1).

**Table 5-1** Predictor metrics obtained through *ballchasing.com* and subsequent processing. Metrics are time normalised to an average match length (360 seconds) unless provided as a percentage of total time in the original dataset, Metrics are expressed both as *raw-score* and *difference-score* except those denoted by a †, which are *raw-score* only.

Offense/Defense	Boost	Movement	Positioning
Shots taken	Boost used	Average speed	Time spent on the ground
Shots conceded†	Average boost reserve	Time spent at <i>slow speed</i>	Time spent high in the air
Demos inflicted	Total boost collected	Time spent at <i>supersonic speed</i>	Time spent goalside of the ball
Demos taken†	Count boost collected from big pads	Average duration for a powerslide	Time spent in the defensive third
	Count boost collected from small pads	Instances of powerslides	Time spent in the offensive third
	Total boost stolen		
	Count boost stolen from big pads		
	Count boost stolen from small pads		
	True boost wastage (%)		
	Total boost overfill collected		
	Total boost overfill stolen		
	Time spent at 100 boost		
	Time spent at 0 boost		

### 5.2.3. Analysis 1: Identifying a continuous outcome measure

Upon identifying the relevant matches and metrics to carry forward for analyses, and in line with the first aim of this study, we determined a continuous *match outcome* metric that could be reasonably substituted for the binary win vs. loss outcome measure while providing additional information regarding the severity of a win or loss. To do this, the *in-game score difference* (IGSD) and *goal difference* (GD) metrics were considered as candidates. Point-beserial correlations were conducted between the candidate measures and the dichotomous *Win vs. Loss* (WL) metric across all rank groups and with matches from all ranks combined. Additionally, we explored the accuracy of the two candidate metrics in separating WL, using zero as the cut-off.

### 5.2.4. Analysis 2: Obtaining Performance Indicators (PIs)

Our second objective was to identify the metrics that best predicted our match outcome measure (GD) within matches across individual rank groupings, and within matches across all included ranks combined (PIs). To address this objective, individual Random Forest regression models were created each for matches within given ranks (i.e., Bronze matches only) and for all matches, regardless of rank. Two models were created per rank (and with all matches combined); one using raw-score metrics and one using difference-score metrics. Random Forest regression models were created using the statistical software, R: A Language and Environment for Statistical Computing (Vienna, Austria).

In addition to the steps taken in data processing to remove metrics that, when combined, could entirely account for the variance of another metric, multicollinearity was assessed for each dataset using qr-matrix decomposition ( $p < .05$ ) in the rfUtilities package in R (Evans & Murphy, 2018). Average speed within the model with GC matches only was identified as multicollinear and was subsequently removed from further analyses.

Random Forest models were then created using the randomForest package in R (Liaw & Wiener, 2002). The sole purpose of these models was to determine the optimal value of *ntree* for each model (amount of CARTs within the Random Forest model). The optimal *ntree* was the number under 1,000 that gave the lowest mean square error of GD, provided the mean square error in the number of trees surrounding this number was also stable. Mean square error was measured using out-of-bag (OOB) data; that is, using only matches that were not involved in the creation of a given tree within the forest. A maximum of 1,000 trees was chosen as it was likely that this would be sufficient to produce highly

predictive models if this was possible given the data (default is 500) while simultaneously balancing computational speed. The default mtry value was used, due to evidence that the default values provided within the RandomForest package perform well (Liaw & Wiener, 2002), and that this number does not tend to affect the performance of the model greatly (Breiman, 2001; Díaz-Uriarte & Alvarez de Andrés, 2006).

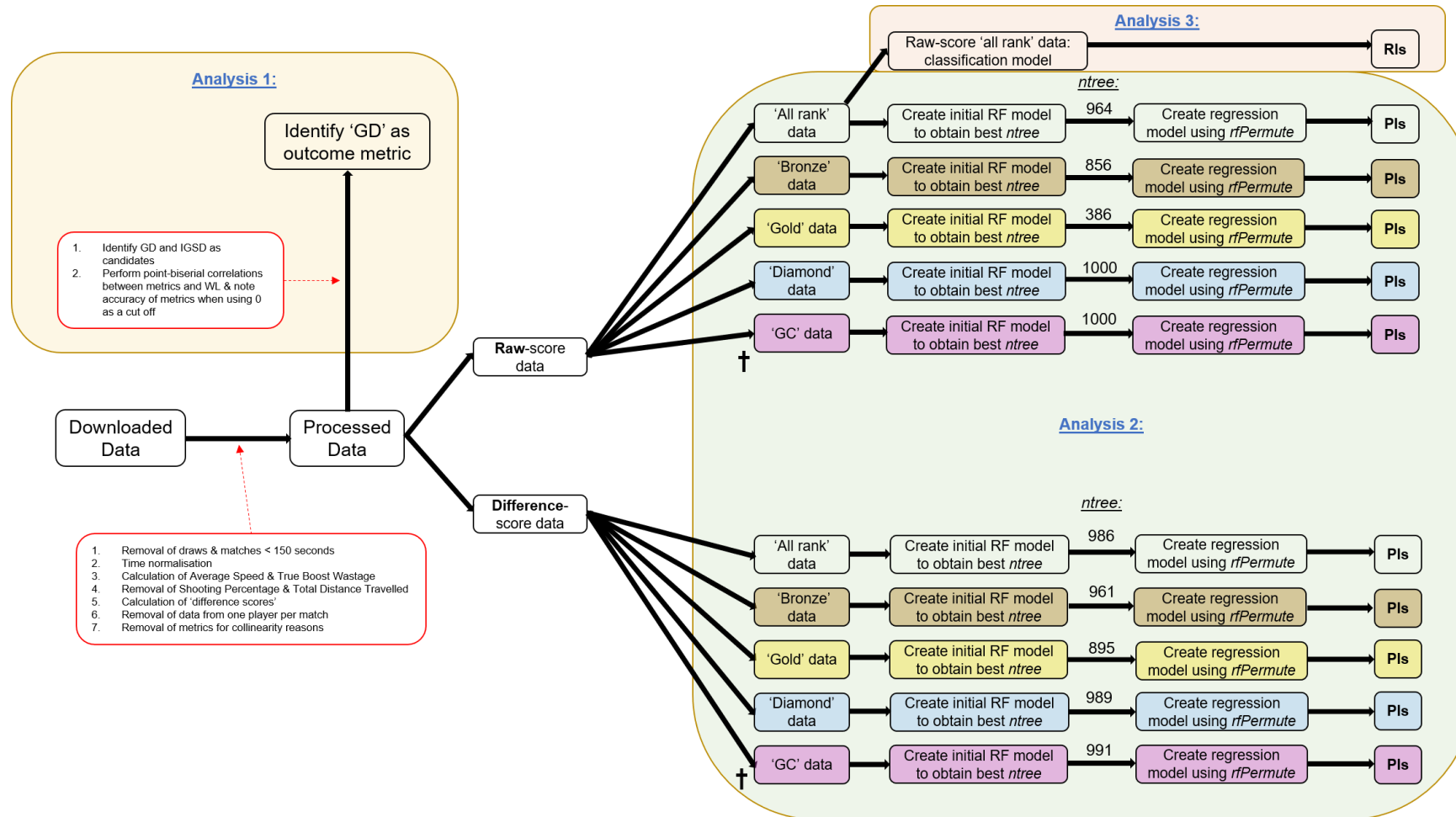
Using the optimal ntree, new Random Forest models were then created using the rfPermute package in R (Archer, 2020). As well as making a Random Forest model, the rfPermute package provides significance values for metric importance. The percentage increase in mean square error (%incMSE) observed when a metric is permuted compared to when no metrics are permuted was used as the measure of metric importance score for each metric. %incMSE was chosen over Mean Decrease in Impurity (Gini), as Gini has shown to be biased when the scale that features are measured on varies (Strobl et al., 2007). To obtain a significance value, rfPermute additionally permutes the outcome metric (GD) a specified number of times, so that there is to be no relationship between any predictor metric and GD. Significance values are obtained per predictor metric each time GD is permuted, forming a *null distribution* of importance scores per predictor metric. P-values are then calculated from the fraction of metric importance scores within this *null distribution* that are greater than the metric importance score obtained when GD was not permuted, with  $p < .05$  being considered a significant metric.

### **5.2.5. Analysis 3: Obtaining indicators of in-game rank (RIs)**

The third objective of this research was to identify the metrics that were able to predict the rank of players within a match regardless of match outcome (i.e., win vs. loss, IGSD & GD). To do so, a Random Forest *classification* model was created in R using data from all included ranks. Unlike a regression model, which provides a numerical outcome prediction, a Random Forest classification model provides a categorical prediction. Feature dependence was explored in the same manner as in Analysis 2. For metric importance, GD was permuted 50 times.

Raw-score Mean Decrease in Accuracy (MDA) was chosen as the measure of metric importance over Mean Decrease in Impurity (Gini) and normalised MDA, for the same reasons as mentioned for the regression models and %incMSE. A Random Forest classification model was only created using raw-score metrics because difference-score metrics should always tend to approach 0 when not considering match result.

A flowchart outlining the methods for this study can be found in Figure 5-2.



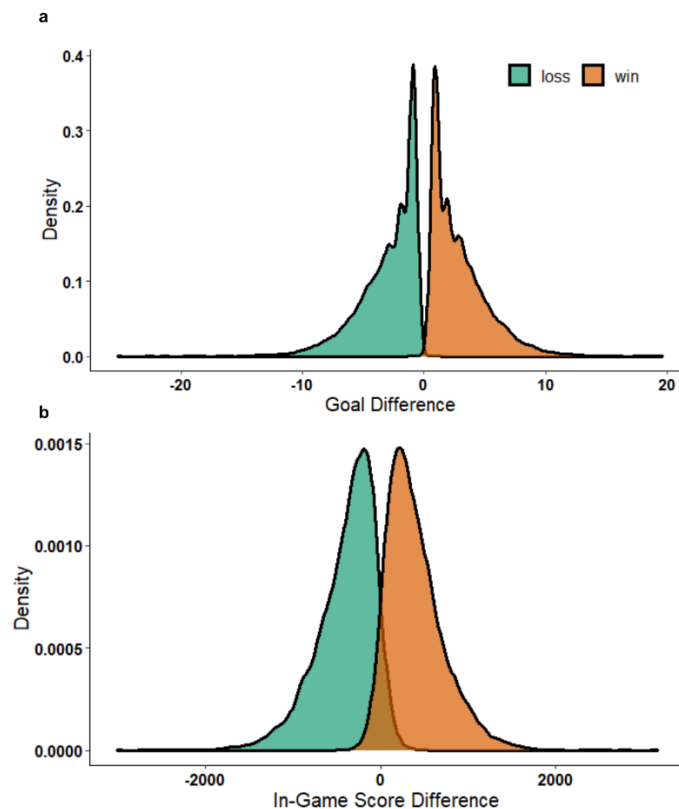
**Figure 5-2** Flowchart depicting the methods of the current study. The three outlined analyses are labelled in blue. The † highlights where average speed was removed due to multicollinearity.

## 5.3. Results

### 5.3.1. Obtaining a continuous performance outcome measure

In line with the first aim of this study, we determined a continuous match outcome metric that could be reasonably substituted for the binary win vs. loss (WL) outcome measure while providing additional information regarding the severity of a win or loss. To do this, the *in-game score difference* (IGSD) and *goal difference* (GD) metrics were considered as candidates.

We conducted point-biserial correlations that tested the association between each metric and WL across all matches and across matches within each specific rank group. All point-biserial correlations demonstrated large ( $r_{pb} > 0.70$ ) significant ( $p < .001$ ) associations, however GD yielded larger association with WL for matches within each rank and when matches for all ranks were combined (Bronze:  $r = 0.77$ , Gold:  $r = 0.80$ , Diamond:  $r = 0.79$ , GC:  $r = 0.78$ , all ranks:  $r = 0.79$ ) compared to IGSD (Bronze:  $r = 0.76$ , Gold:  $r = 0.78$ , Diamond:  $r = 0.77$ , GC:  $r = 0.75$ , all ranks:  $r = 0.77$ ). Finally, we noted that when using zero as a cut-off for IGSD and GD (positive scores corresponding to *win*, and negative scores corresponding to *loss*), IGSD correctly identified wins 93.56% of the time, and losses 93.70% of the time, while GD correctly identified wins and losses 99.94% of the time. Figure 5-3 displays the distribution of the data from all skill brackets combined using a density plot (default bandwidth).



**Figure 5-3** Density plots showcasing **a**: the distributions of goal difference and **b**: in-game score difference as a function of win vs. loss.

Through these analyses, we demonstrate that GD and IGSD are both appropriate continuous variables for game outcome. However, due to the superior association of GD with WL across all matches and matches within each rank group, GD was used as the performance outcome measure in subsequent analyses.

### 5.3.2. Obtaining indicators of performance (PIs)

Random Forest regression models were created using the raw-score metrics (player metrics, not accounting for opponent) and difference-score metrics (player metrics accounting for opponent). These models were created for 1v1 Rocket League matches occurring in Bronze rank (lowest in-game rank; 2,527 matches), Gold rank (7,226 matches), Diamond rank (7,193 matches) and Grand Champion (GC) rank (highest in-game rank; 4,642 matches), as well as in all matches regardless of rank (21,588 matches). The match outcome variable for these regression models was the goal difference (GD)

between players within a match. These models were used to identify the in-game metrics that best predicted in match outcome and could thus be described as PIs for Rocket League.

All of the models created were highly predictive of GD ( $R^2 > 0.7$ ). As can be seen in Table 5-2, models created using the difference-scores were better able to predict match outcome compared to models using raw-scores in all cases.

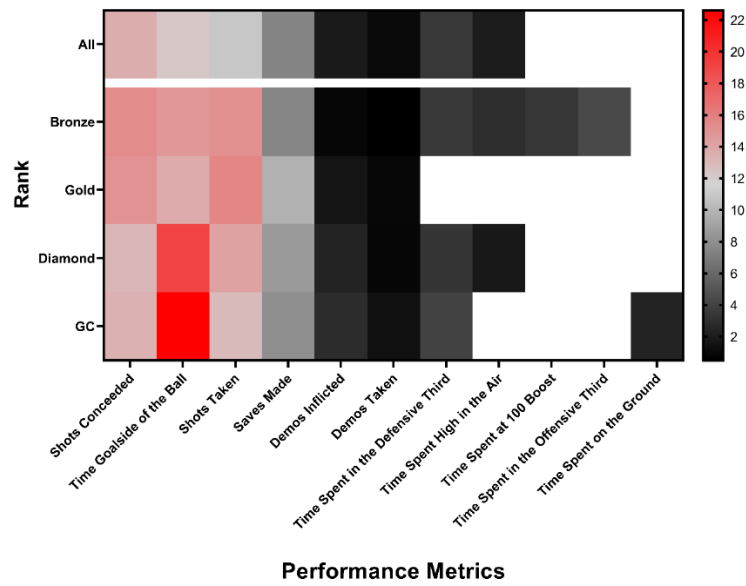
**Table 5-2**  $R^2$  (mean of squared residuals) of the Random Forest models created using the raw and difference-score metrics for each ranks and for all ranks combined

	Bronze	Gold	Diamond	GC	All
Raw-score	0.793 (3.91)	0.741 (3.55)	0.725 (3.66)	0.713 (4.06)	0.747 (3.61)
Difference-score	0.841 (2.99)	0.823 (2.43)	0.823 (2.35)	0.816 (2.59)	0.839 (2.29)

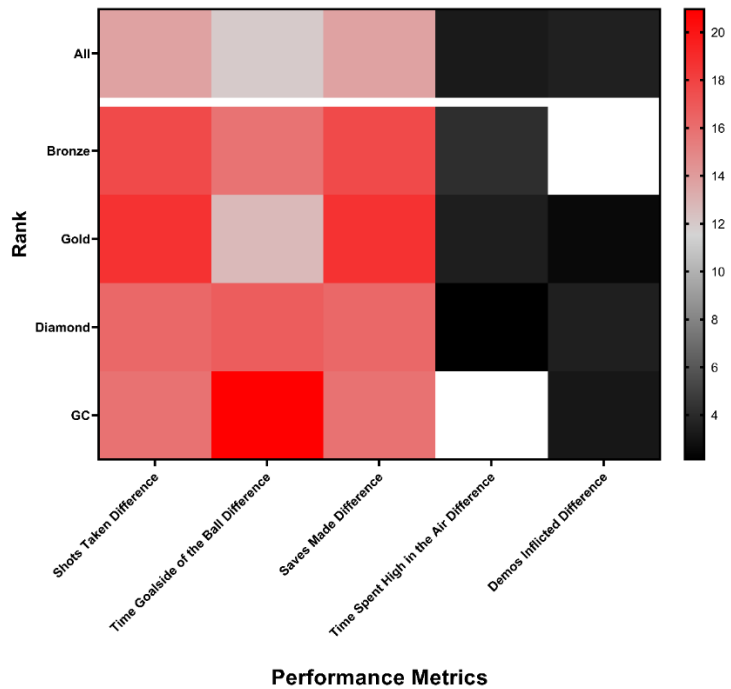
#### 5.3.2.1. Raw-Score Models

In all Random Forest regression models using raw-score metrics, the following metrics led to a significant ( $p < .05$ ) increase in mean square error (MSE) when permuted, and hence were identified as PIs: *shots taken*, *shots conceded*, *time spent goalside of the ball*, *saves made*, *demos taken*, and *demos inflicted*. Figure 5-4a shows the relative contribution that each PI metric made to the total MSE increase when all PIs were included together for matches within each rank category, as well as for matches across all ranks combined, for the raw-score models. For matches in the Bronze rank, Gold rank, and when all ranks are combined (i.e. all matches without considering player rank), *shots taken* and *shots conceded* were more important than *time spent goalside of the ball*, whereas for matches in the Diamond rank and GC rank, *time spent goalside of the ball* was more important than *shots taken*, and *shots conceded*.

**a**



**b**



**Figure 5-4** Heat map displaying the percentage of the total increase in MSE that can be found when a metric is permuted individually compared to the sum of increase in MSE for all metrics when permuted individually. Only metrics that were significantly important for predicting GD within each raw-score and difference-score model are presented. White squares represent metrics that were not significant for the rank they are assigned to. **a**: results from raw-score regression models, and **b**: results from difference-score regression models.

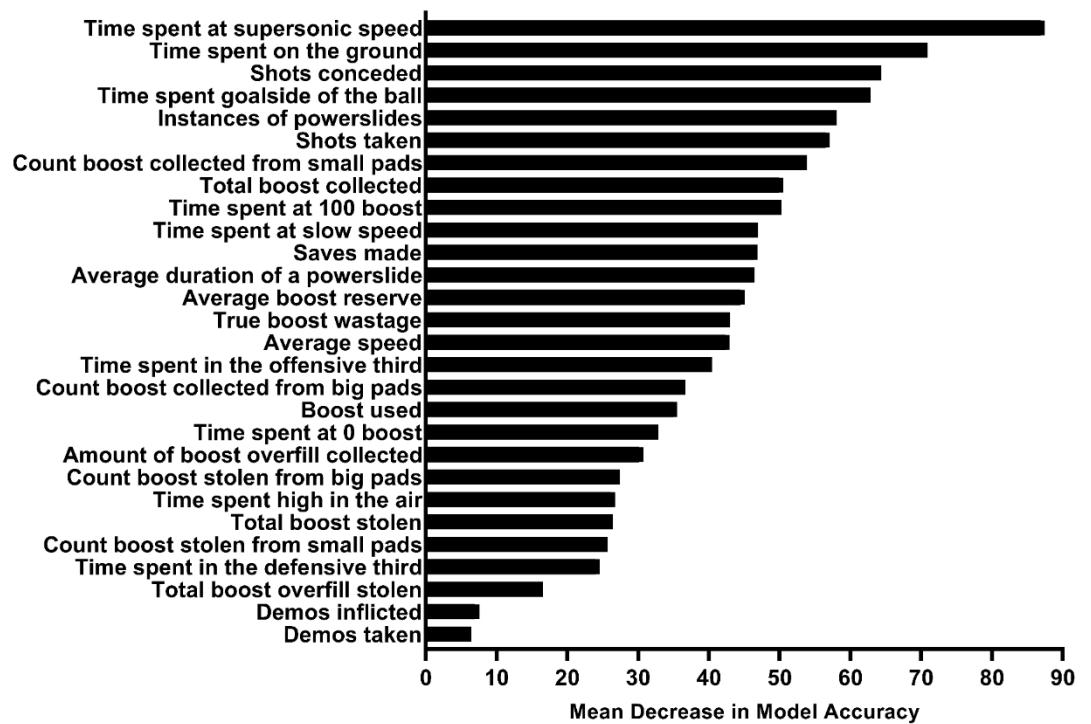
### 5.3.2.2. Difference-Score Models

In all Random Forest regression models using difference-score metrics, the following metrics led to a significant ( $p < .05$ ) increase in MSE when permuted, and were hence classified as PIs: *shots taken difference*, *time spent goalside of the ball difference*, and *saves made difference*. Figure 5-4b shows the relative contribution that each PI makes to the total MSE increase when all PIs are included together for matches within each rank category, as well as for matches across all ranks combined, for the difference-score models. For matches in the Bronze rank, Gold rank, and when all ranks are considered, *shots taken difference* and *saves made difference* were more important than *time spent goalside of the ball difference*, whereas for matches in the Diamond rank and GC rank, *time spent goalside of the ball difference* was more important than *shots taken difference* and *saves made difference*. *Saves made difference* was also more important than *time spent goalside of the ball difference* for matches in the Gold rank and when all ranks are considered.

### 5.3.3. Obtaining indicators of in-game rank (RIs)

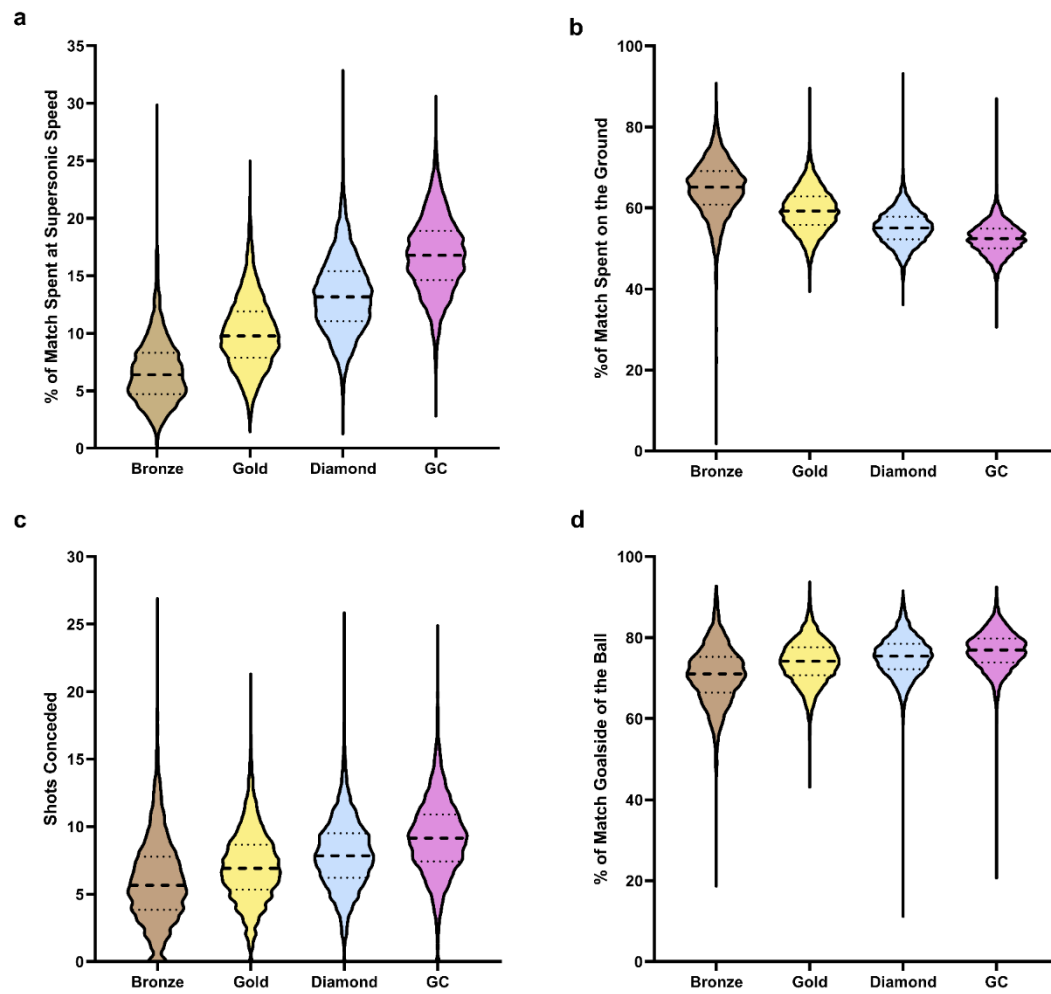
The Random Forest classification model correctly classified the rank of players within 1764 of 2527 Bronze matches (69.81%), 5394 of 7226 Gold matches (74.65%), 5098 of 7193 Diamond matches (70.87%), and 3417 of 4642 GC matches (73.61%), resulting in an overall out-of-bag (OOB) accuracy of 72.6%.

All metrics were found to significantly decrease the accuracy of the model when permuted ( $p < .05$ ), and so were deemed RIs in Rocket League (Figure 5-5).



**Figure 5-5** Metrics found to be of significant importance to the classification model created to predict the ranks of individuals playing 1v1 Rocket League, ordered by the increase in mean decrease in accuracy experienced within the model when each metric was permuted.

Overall, time spent at supersonic speed, time spent on the ground, shots conceded, and time spent goalside of the ball were the four RIs most important to the Random Forest model for correctly classifying data according to the rank of the players within the match. Violin plots showing the means and distributions of these four RIs across included ranks are displayed in Figure 5-6.



**Figure 5-6** Violin plots displaying the means and distributions, within each rank, of the four most important features for predicting rank, **a:** *Time spent at supersonic speed*, **b:** *Time spent on the ground*, **c:** *Shots conceded*, **d:** *Time spent goalside of the ball*.

## 5.4. Discussion

In this work, we used a Random Forest machine learning algorithm analysis to identify key performance metrics that predicted expertise (RIs) and match outcome (PIs) for the first time in a prominent esports, Rocket League. Specifically, we first aimed to identify a continuous match *outcome* metric that provided more information on in-game performance than the binary win vs. loss measure. Goal difference (GD) was identified as this suitable match outcome metric. Secondly, we aimed to identify metrics that are significantly important to influencing our outcome metric, GD (labelled as *PIs*), both across matches played by players of specific ranks and across matches played by those of all ranks. Hence, these PIs differentiate good and poor performance *within* a given rank. Here, we found specific PIs that are important to GD across all ranks, as well as PIs that are only important to GD in matches with players of specific ranks. Thirdly, we aimed to identify metrics that best predicted player expertise or rank within matches (labelled as *RIs*). These RIs differentiate *between* players of different ranks. All metrics were significantly important to the classification of rank in our Random Forest classification model. Importantly, we show for the first time the order of importance that each metric has for the prediction of rank within our model, with *time at supersonic speed*, *time spent on the ground*, *shots conceded*, and *time spent goalside of the ball* being the four that decreased the performance of the model the greatest when permuted. The following discusses the implications of these findings.

Firstly, our finding that difference-score metrics lead to better match outcome prediction compared to raw-score metrics corroborates previous literature in rugby union (Bennett et al., 2019). Models incorporating difference-score metrics for any rank were able to account for over 80% of the variance in GD between two players in a given match. This highlights the utility of the in-game statistics obtained from the online repository *ballchasing.com* for Rocket League and the utility of Random Forest models for predicting performance within Rocket League.

Focussing on PIs within the difference-score models, when compared to a rank-matched opponent, taking more shots, making more saves, and spending more time goalside of the ball all appear to be beneficial for success in Rocket League matches, regardless of one's rank. The difference in time spent goalside of the ball was found to be most important within higher ranked matches (Diamond & GC), suggesting that as the quality of players increases, so does the relative importance of maintaining one's positioning between the

ball and one's own goal, compared to simply taking more shots or making more saves. This could be due to the greater ability of higher ranked players to swiftly and accurately shoot a ball into a goal left unattended due to an opponent's poor positioning. This also may suggest that higher ranked players may generally be well served to adopt a *safer* playstyle, reducing the number of high-risk attacking plays such as *air dribbles* (referring to when a player achieves many controlled touches on the ball while both the player and ball are in the air) that if failed, might leave them positioned in front of the ball. For GC rank players, this is further supported by the finding that increasing time spent high in the air (a necessity for *air dribbles*) did not predict performance, whereas this was a PI of match outcome in all other ranks.

When considering the application this new awareness of important PIs, Rocket League players of all ranks can leverage Variable Priority Training (VPT), which has already been demonstrated to be superior to Fixed Priority Training (FPT) (Boot et al., 2010), to actively focus on improving performance on the key metrics that are actually shown to be important for match outcome. Based on the results, a player might specifically work to improve spending more time goalside of the ball than their opponent during their matches. Our results also suggest that lower ranked players (Bronze and Gold) could harness VPT by monitoring the shots they take relative to their opponent during matches and focussing on skills that facilitate improvement on this PI. This has been discussed by professional Rocket League coaches previously as a beneficial strategy as similarly lower-ranked opponents are less likely to save shots regardless of quality (Virge, 2020). Inflicting more demos on your opponent than they do on you also appears to provide a performance benefit in matches for all ranks except those in the Bronze rank group. A demo (short for demolition) is achieved when one player drives into an opposing player at supersonic speed at the correct angle and removes them from the field for three seconds before the opposing player respawns in one of two prespecified locations in their defensive third. The fact that this metric was not found to be a PI for Bronze level matches may be due to the fact that Bronze players may not possess the skills to capitalise on the three second advantage awarded by a demo to score, whereas higher ranked players may be better able to use demos to score or prevent goals.

The metrics that best predicted differences between ranks (RIs) were not necessarily predictive of performance when two rank-matched players play against one another (i.e., within rank). For example, the percentage of time that a player spent at supersonic speed was the most important RI (Figure 4-6), whereas this metric it did not significantly

improve the ability of regression models to predict the outcome of a match within a given rank group. The fact that *time at supersonic speed* was found to be a RI and not a PI may be due to the fact that playing at higher ranks requires one to have the ability to play at near maximum speed for longer durations so as to match the speed of the opponent in case they were to attack at maximum speed. However, once both players are able to do so, attacking at supersonic speed does not provide additional benefit within the match between two similarly ranked players. This explanation can also be applied to the PI *number of powerslides*, which, when permuted from the model, led to a large decrease in accuracy (which demonstrates its high importance) in the classification model predicting rank. Powerslides are a difficult manoeuvre that provide the opportunity to maintain speed when landing on the ground and turning sharply, however powerslide turns can be difficult to control. Higher skilled players appear to use this mechanic more often to achieve greater control of their car, however when players are of similar rank, powersliding more or less than an opponent within a match does not appear to provide an advantage. Taken together, higher rank players show better control over the movement of their car and are able to play a greater proportion of their matches at high speed. However, within rank-matched matches, this metric does not predict match outcome. Therefore, our findings suggest that while focussing on game speed and car movement may not provide immediate benefit to the outcome within matches, these PIs are important to develop as they may facilitate one's improvement in overall expertise over time.

#### **5.4.1. Significance**

While the identification of PIs to predict match outcome and in-game ranking within Rocket League provides new knowledge regarding how Rocket League players and coaches may structure training programs, the results from this analysis are also foundational for future experimental work utilising esports as a performance arena. Esports have been identified as a promising new avenue to study expertise (Campbell et al., 2018), due to their data rich nature, continuous and accurate skill rating systems (Elo), and the naturally controlled, laboratory like environment that esports are typically engaged in. More recently, esports have been identified as an ideal framework for exploring whether task expertise moderates task performance deficit experienced from sleep loss (Smithies, Toth, et al., 2021), with applications spanning beyond esports due to the shared work environment and cognitive skills required between esports and pilots,

air traffic controllers, and military drone operators for example (Smithies et al., 2020). This framework could be extended to study how ability level on a task moderates the effect of a given intervention on task performance.

1v1 Rocket League in particular is an ideal esports to use as a performance task in an experimental setting as it has a short & predictable match length (5 - 10 minutes), allowing for many trials within an experimental setting, a simpler experimental design than other esports due to the ability for one to play individually, and experimenters can easily access player rank and in-match metrics. The results of our analysis specifically inform as to the important PIs of interest when evaluating the efficacy of an experimental intervention on Rocket League performance. A reduction of the outcome variable, GD, alongside key PIs such as *difference in shots taken*, *difference in saves made*, and *difference in time spent goalside of the ball*, would represent a negative effect on the intervention on task performance. Interestingly, a reduction in *time spent at supersonic speed* or *instances of powerslides* following an intervention, but a maintenance of performance, could suggest an adaptation by players to simplify their play style to maintain performance following an intervention.

This is the first study to use Random Forest models to identify PIs within an esports. Random Forests are robust to data of any distribution from a large number of features (regardless of if many are actually predictive of the outcome or not) and can ascertain non-linear effects and complex interactions without prior specification. Thus, Random Forests present as a valuable tool for notational analysis within esports, which is in its infancy and has limited prior information available on potential PIs for various games and genres. Random Forests for notational analysis in esports could be used to explore what predictor metrics are most important for match outcome in other genres, such as FPS's and MOBA's.

#### **5.4.2. Limitations and Future Research**

When considering the power of Random Forests as a notational analysis, one limitation is that feature importance measures from Random Forest models can show bias when features are correlated (Hooker & Mentch, 2019; Strobl et al., 2008). To mitigate this, where the variance of one predictor metric could be entirely explained by one or more other metrics, these additional metrics were removed, and multicollinearity was assessed for each model with multicollinear metrics being removed. Additionally, features shown

to be important for game outcome or skill within each model showed no greater correlation with other features compared to those not found to be important (correlation matrices for all models can be found in appendix 5.2). Future research should consider feature importance measures such as permutation conditional on remaining features (Strobl et al., 2008), *leave-one-covariate-out* (Lei et al., 2018), and *permute and relearn* (Hooker & Mentch, 2019) to address correlated features, however given the large amount of data and extra computational resources required for these methods, they were not feasible here.

In this study, we chose to exclusively explore 1v1 Rocket League. While identical in game mechanics, positioning and decision making vary between 1v1, 2v2, and 3v3 formats of Rocket League. Hence, PIs and RIs for team-based Rocket League may be different to 1v1. However, this analysis would have been greatly complicated if we additionally included team-based Rocket League, as interactions between teammates would have to be considered, further complicating analysis (Ofoghi et al., 2013). Interestingly, 1v1 is considered by many professional Rocket League players (i.e., *Flakes*) to be the best way to improve in Rocket League overall due to affording players more time to interact with the ball compared to other formats. Hence, the PIs and RIs here can provide great benefit for all Rocket League players and coaches, even if improvement specifically in 1v1 Rocket League is not the primary goal. However, future research should attempt to use similar analysis methods to those described here to identify the PIs and RIs for 2v2 or 3v3 Rocket League.

### 5.4.3. Conclusions

In summary, this study is the first to use Random Forest models to identify PIs and RIs that could predict match outcome and rank respectively across over 20,000 matches in the rapidly emerging esport of Rocket League. Overall, spending more time goalside of the ball, *taking more shots*, *conceding less shots*, and *making more saves*, were all identified as beneficial for in-match performance across all ranked matches. All metrics were found to be significantly important (and thus, RIs) for a Random Forest model's ability to predict player rank, and we have classified the order of importance of these metrics using our model. Interestingly, we found that *time spent at supersonic speed*, *time spent on the ground*, *shots conceded*, and *time spent goalside of the ball* were the most important RIs. This type of analysis can provide useful insight to Rocket League players and coaches regarding the structuring of VPT programs to improve match success of in-

game rank. The findings from our analysis also provides researchers with key metrics to consider if using Rocket League as a performance task in experimental research.

## 5.5. Linking chapter

The analysis outlined in this chapter served several purposes for the thesis. Firstly, I demonstrated that *Goal Difference (GD)*, referring to the time-normalised difference in goals scored between players, was the most ideal continuous outcome metric for overall in-game performance. Secondly, the in-game performance indicator (PI) metrics outlined within the current study provide further metrics to explore when considering the effect of experimentally induced sleep loss on in-game Rocket League performance. Such analysis could elucidate whether sleep loss induces in-game *playstyle* changes within Rocket League. I decided to use difference score PIs as outcome variables of interest within exploratory analyses in the experimental sleep loss study, owing to models created using difference-score metrics consistently outperforming those made using raw-score metrics. This included *shots taken difference*, *time spent goalside of the ball difference*, and *saves made difference*, however also included *time spent high in the air difference* and *demos inflicted difference*, owing to these metrics being significantly important within the majority of ranks and when all ranks were combined.

## **Chapter 6. Key methodologies for the experimental sleep loss study**

The current chapter serves to outline some key methodologies relevant to the experimental sleep loss research, which is disseminated in the following chapter. The first two sections pertain to the use of actigraphy (specifically the Readiband<sup>TM</sup> (v5) by Fatigue Science) as the primary sleep-measurement method of choice within this study. Firstly, a brief outline of common sleep measurement methods is provided, with this outline narrowing to a specific description of the Readiband device. Secondly, my novel approach used to manage missing actigraphy derived sleep data is outlined, along with a demonstration of its efficacy. The third section of this chapter will specifically discuss the use of Mixed Effect Models (MEMs) within the analytical approach. MEMs are becoming increasingly common within sleep research for good reasons, including the inclusion of all data despite sources of dependence (this can be more than one source, and can be complex in nature), robustness to missing or unbalanced data, and the ability to manage categorical and continuous independent variables (or *fixed effects*) simultaneously. However, given the inherent increase in complexity associated with MEMs, along with the variation in implementation, it was necessary to provide a detailed explanation of how they were used in the research outlined in **Chapter 7**. I describe the best-practice guidelines followed in model selection and dissemination of results.

## **6.1. Using Readibands for Actigraphy-Derived Sleep Outcomes**

Sleep measurement was undertaken in the research outlined in **Chapter 7**, and the analysis described in **Section 6.2** of this chapter, using concurrent wrist-worn actigraphy (Readiband™ (v5)) and consensus sleep diary (CSD; Carney et al. (2012)). The following section outlines different methods of measuring individual nights of sleep within contemporary research, providing insight as to why these methods were chosen.

### **6.1.1. Polysomnography**

The measurement of individual nights of sleep for research purposes is normally undertaken using one of three approaches. On one end of the spectrum is polysomnography (or PSG), widely considered to be the gold standard of sleep measurement (Marino et al., 2013; Rundo & Downey, 2019). PSG (typically) employs the use of concurrent electroencephalography (EEG), electrooculography (EOG), chin and leg electromyography (EMG), pulse oximetry, nasal prongs, oronasal thermistors, respiratory inductance plethysmography, body position sensors, microphones and video recording. Naturally, such a set-up requires a dedicated sleep laboratory, with trained sleep researchers present to continuously monitor the various channels of information and use it to manually score sleep stages according to prespecified criteria; the most common of which are provided by the American Academy of Sleep Medicine (AASM; Troester et al. (2023)). Due to this, PSG tends to not only be expensive (estimated as \$1500-2000USD per night in the United States; Arnal et al. (2020)), but also may not be representative of an individual's natural sleep, due to a change in environment, discomfort arising from the instruments used, and stress. This disturbance to one's natural sleep may persist beyond a single day of habituation (Le Bon et al., 2001). Although there is continual and promising development of more practical derivatives of PSG for sleep measurement, which use automated sleep scoring and less instruments (generally EEG only or EEG and limited other instruments; i.e., Arnal et al. (2020); Myllymaa et al. (2016); Shambroom et al. (2012)), the use of these instruments remains mostly limited to single-night assessments or studies concerning sleep disorders.

### **6.1.2. Sleep Diaries**

On the opposite end of the spectrum are sleep diaries, the most commonly used being the Consensus Sleep Diary or CSD (Carney et al., 2012). Sleep diaries such as the CSD (the

gold standard of subjective sleep measurement) certainly have their benefits; they are extremely easy to administer, and present with little to no time or effort burden for participants. They are recommended by the AASM to be used concurrently with actigraphy (discussed in the following paragraph) for non-standard populations (Morgenthaler et al., 2007), and are required by some actigraphy devices in order to determine sleep onset and wake times. However, their subjective nature creates concerns regarding the reliability and agreement of the data obtained (Carney et al., 2012). Many sleep variables derived from sleep diaries present with consistent group-level biases when compared both to the gold standard PSG and to actigraphy; namely, a ~20 minute greater time in bed (TIB), ~15 to 55 minute greater total sleep time (TST), ~0 - 8% greater overall sleep efficiency (SE%), ~1 to 36 min smaller sleep onset latency (SOL), ~2 to 38 minute smaller wake after sleep onset (WASO), and a ~6-7-fold decrease in total number of awakenings (Kaplan et al., 2012; Lehrer et al., 2022; Matthews et al., 2018; McCall & McCall, 2012). Furthermore, the tendency to under or overestimate sleep variable quantities (& the severity of over or underestimation) compared to objective measures is highly variable between individuals, as shown by Moore et al. (2015) for Breast Cancer Survivors, and in **Section 6.2** for a young male population.

### **6.1.3. Actigraphy**

With PSG and sleep diaries at either end of the sleep measurement spectrum, sleep wearables or *actigraphy* devices provide a happy balance of objectivity and practicality, and hence are the default for field-based or longitudinal sleep measurement in research. Such devices are generally wrist-worn, and either exclusively or predominantly use tri-axial accelerometry to detect periods of movement or rest. This movement information is converted to sleep and wake data through pre-defined algorithms, which can be implemented both manually or automatically.

*Research-grade* actigraphy devices such as the Actigraph™ are the overall most widely used actigraphy devices in research. However, within contemporary research there has been a substantial increase in the use of alternative (and often commercially available) actigraphy devices, coinciding with validations of these devices within peer-reviewed literature (Evenson et al., 2015). Two major contemporary articles of this kind found many of these alternative actigraphy devices to be comparable to or outperform research grade actigraphy devices, when compared both to in-lab PSG and at-home EEG (Chinoy et al., 2021; Chinoy et al., 2022).

One of these devices is the Readiband™ (v5) wrist-worn activity monitor (Fatigue Science, Canada). When compared to at-home PSG, the Readiband has been shown to outperform the research-grade Actiwatch 2 (Philips Respironics), with no significant bias for TST, SE%, SOL and WASO (Chinoy et al., 2021). When compared to at-home EEG, the Readiband has been shown to provide superior specificity vs. the Actiwatch 2, and overall presents as a “viable option for sleep-wake tracking and with longer battery life (~30 days) compared with the other devices tested (~4-7 days)” (p. 512, Chinoy et al. (2022)). A separate study comparing both the Readiband and the research-grade ActiGraph found both actigraphy devices to be suitable for use when considering the sleep variables TST, time at sleep onset (TASO) and time at wake (TAW) (Dunican, Murray, et al., 2018). However, the authors encouraged exercising caution when interpreting the outcome variables SOL, WASO, and SE%, derived from *either* actigraphy device. The Readiband uses a proprietary sleep scoring algorithm which performs favourably (accuracy = 93% vs. PSG) compared to the commonly used Sadeh algorithm (91-93%; Sadeh et al. (1994)) and Cole-Kripke algorithm (88%; Cole et al. (1992)) on sleep data collected by the validated AMI-32 (Ambulatory Monitoring Inc) (Russell et al., 2010). This algorithm automatically scores all sleep and wake periods as well as bed-time, and as such, the Readiband does not possess an event marker. Sleep and wake scoring from this proprietary algorithm was assessed by an experienced researcher and cross-validated against consensus sleep diary measures.

The Readiband has high ( $ICC \geq 0.8$ ) inter-device reliability (including  $ICC = 0.99$  for total sleep time) and a mean inter-device difference of only two minutes per night of sleep (Driller et al., 2016). Readibands have been used in previously published sleep research for a variety of populations, including traditional and esports athletes (Bonnar et al., 2022; Dunican et al., 2023; S. Lee et al., 2021; Power et al., 2023; Smithies, Eastwood, et al., 2021), medical personnel (James et al., 2019; Min et al., 2023), pilots (Rocha & Silva, 2019), and military personnel (Edgar et al., 2023). For the reasons outlined above, the Readiband was used as the objective sleep measurement device within the study described in **Chapter 7**.

## 6.2. Managing Missing Actigraphy-Derived Sleep Data

### 6.2.1. Brief Background

One issue encountered within data collection for the study outlined in the following chapter was the existence of missing actigraphy-derived (Readiband) sleep data. 4.56% of total collected nights worth of actigraphy-derived sleep data were missing, while 3.64% of nights within three nights of test sessions (the range reported in **Chapter 7**) were also missing. The presence of missing actigraphy-derived sleep data was foreseen in the analytical approach, given how pervasive this issue is within actigraphy-based sleep research. This is demonstrated by three large-scale (>500 nights of data) actigraphy studies on different populations (healthy adults, Ustinov and Lichstein (2013); healthy women, Tworoger et al. (2005); children and adolescents, Acebo et al. (1999)) reporting rates of missing data between 14-28%.

There are five strategies generally employed to deal with missing actigraphy-derived sleep data (I note that published articles have documented methods to deal with missing epochs of actigraphy data, rather than missing sleep data for an entire night (Fuster-Garcia et al., 2013; Jang et al., 2020; Smith et al., 2021); For a detailed summary of missing data methods for summary and epoch actigraphy data, see Di et al. (2022)). These are listwise deletion, use of summary statistics, simple imputation, multiple imputation, and the use of analytical approaches which are robust to missing data (i.e., models that use maximum-likelihood estimation, such as MEMs). The pros and cons for each strategy, along with the scenarios which warrant a specific strategy, are beyond the scope of this section.

In this section, I outline a novel simple imputation approach (named *Diary ± Individual Bias*) and compare its agreement to other simple imputation strategies (adapted to the data collected). A simple imputation strategy was chosen for use as it was required to have a specific estimated value (as opposed to MI or maximum-likelihood approaches which do not provide one specific value) provided for any missing actigraphy data within the *critical* nights in my experimental design (the three nights prior to each test session), for the purposes of Figure 7-4. Also given that the rate of missingness was relatively small, the computational and theoretical complexity is not particularly warranted (Sainani, 2015).

## 6.2.2. Testing Approach

I considered data collected as part of the study described in the following chapter; however, I only considered participants who provided written informed consent and who did *not* have any missing data, so as to avoid bias risk. This resulted in data from 21 healthy young males ( $20.48 \pm 2.50$  y/o) being considered, providing 282 days of concurrent actigraphy-derived (Readiband) and Consensus Sleep Diary (CSD; (Carney et al., 2012)) data. The number of days with concurrent Readiband and CSD data varied between participants (as a function of days collected, not as a result of missing days); Table 6-1 shows the distribution of participants by number of days worth of data available. Data collection was approved by the Education and Health Sciences Research Ethics Committee (2021\_06\_13\_EHS) and conducted in accordance with The Declaration of Helsinki.

**Table 6-1** Number of participants with N amount of days available

Days Available	14	13	12
N	10	10	1

### 6.2.2.1. Simple Imputation Approaches Tested

#### 6.2.2.1.1. Proximity Imputation

This approach has been described by Bjorvatn et al. (2006), Bjorvatn et al. (2007), Forberg et al. (2010) and Saksvik et al. (2011). For a day of missing actigraphy data, this approach takes the mean actigraphy-derived value for the previous and following day as the imputed value, unless one of which is not available (i.e., when the first or last day within a date-range is missing), in which the value of the remaining available day was directly imputed as the replacement value. If three consecutive days are missing, the mean of the previous and following days around this three-day missing block were imputed for all consecutive missing days.

#### 6.2.2.1.2. Hot-Deck Imputation

This approach has been used by Rigney et al. (2015). Hot-deck imputation is a process whereby missing data is replaced with a value from an observation (*donor*) which exhibits similar characteristics. Characteristics that determine donor suitability are researcher-determined. When multiple donors are available, a donor is randomly selected from the pool of donors (for more detail, see Myers, 2011).

I considered two Hot-Deck approaches; one in which donor values were from the same individual in which the data is imputed for (*Within-participant or WP hot-deck*) and the other in which the donor values are from other participants (*Between-participant or BP hot-deck*). For within-participant hot-deck, potential donor values were categorised only by whether the data was from a weekday or weekend. For between-participant hot-deck, donor values were determined by the following characteristics (in order of importance; weekday/ weekend, overall Pittsburgh Sleep Quality Index (PSQI) score, self-report mean sleep duration, age).

#### **6.2.2.1.3. Diary Only**

This approach refers to the replacement of a missing actigraphy-derived sleep variable with the equivalent variable obtained from the same individuals' CSD on the corresponding day.

#### **6.2.2.1.4. Diary $\pm$ Individual Bias**

This approach refers to imputation using the corresponding CSD value (as per above), however also factoring in the mean difference between the participants CSD value and the actigraphy derived value for all other days in which both values were available. For example, if a participant overestimated TST by an average of 30 minutes, 30 minutes was taken off the CSD value subsequently used for imputation.

### **6.2.3. Imputation and Testing Procedure**

I tested the performance of the five approaches on three actigraphy-derived sleep variables:

*Time at Sleep Onset (TASO; hh:min)* The time of day in which the first epoch of sleep occurs in a nighttime sleep period

*Time at Wake (TAW; hh:min)* Time of day following the last epoch of sleep occurs in a nighttime sleep period, following by a prolonged period of wake

*Total Sleep Time (TST; mins):* Amount of time between TASO and TAW, minus any time spent awake (i.e., Wake After Sleep Onset). I also included napping periods (periods of sleep outside of a nighttime sleep period, with naps occurring before 12:00 added to the previous night, and naps occurring after 12:00 added to the upcoming night).

An imputed value was obtained for each day and for each participant, using all five imputation approaches. However, some approaches used data available from previous or following days from the same participant. For the data collected in **Chapter 7**, the minimum amount of days available for a given participant was 10, and the maximum (with at least one day of missing data) was 13. Thus, separate imputed value for each day, participant, and approach, were obtained for when participants had 13, 12, 11, and 10 days of other days worth of data available. For 13 days of available data, only the data from ten participants (140 days of total data) was available. Once an imputed value was obtained each day, participant, and approach, I removed one day's worth of data from each participant, such that each participant had 13 days of concurrent actigraphy-derived and CSD data available. This also allowed me to include participants with only 13 days available, increasing the days of data available ( $N = 20$ , 260 days of total data). From these data, I again obtained an imputed value for each day, participant, and approach. This process was iterated until the included participants had 11 days of concurrent actigraphy-derived and CSD data available ( $N = 21$ , 231 days of total data). This allowed for the testing of imputation approaches across the range of missingness present within the data in **Chapter 7**.

#### 6.2.4. Analysis

The following analyses were performed to compare sleep variables for all imputation approaches and across all amounts of data available per participant. Alpha was set to  $p < 0.05$  (two-tailed) for all analyses.

Agreement was assessed using two measures. Firstly, I calculated Absolute Percentage Error (APE; as per Stone et al. (2020)), according to the formula below:

$$APE = \frac{|Imputed\ Value - Readiband\ Value|}{Readiband\ Value} \times 100$$

To explore whether mean APE differed among imputation methods, a series of pairwise least-squared comparisons were made, with Satterthwaite's approximation for degrees of freedom, and using Tukey's HSD to correct for familywise error rate. Imputation approaches were inputted as fixed effects, with random intercepts for both participant and the participant day combination, to account for correlation between imputed values attributable to the specific day measured and to the specific individual.

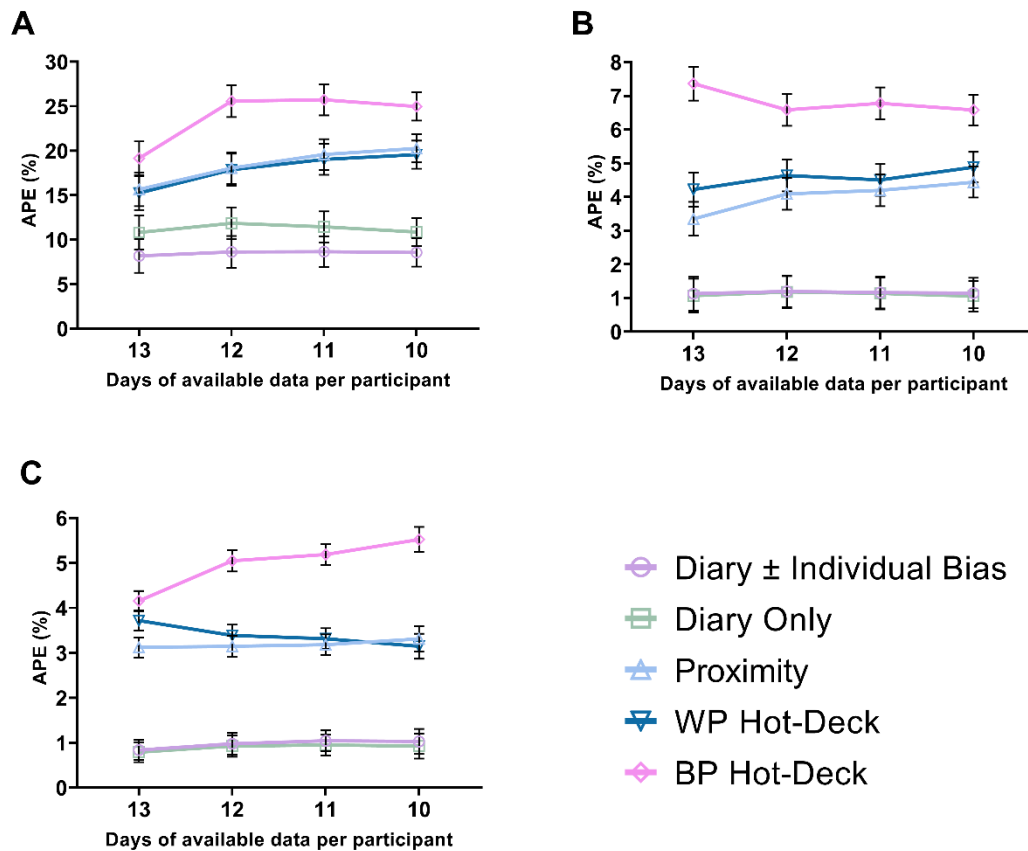
Secondly, I assessed agreement using a mixed-effects Limits of Agreement (LoA) analysis with accompanying Bland-Altman plots (Parker et al., 2020). LoA analysis is a common method of measuring agreement (particularly in actigraphy-based research) due to its easy interpretability. While I did not specify a clinically acceptable difference threshold *a priori*, I considered the imputation methods with the narrowest confidence limits to demonstrate the best agreement.

Lastly, I tested for presence of bias (tendency for imputation method to under or overestimate actigraphy-derived values) through a series of pairwise least-squared comparisons of mean derived values for each imputation approach, against actigraphy-derived values. Source of sleep variable values (imputation approaches and actigraphy-derived values) were inputted as fixed effects, with identical random effects to that mentioned for APE. Again, Satterthwaite's approximation for degrees of freedom was used. As only comparisons between each approach and actigraphy-derived values were relevant (as opposed to comparisons between all imputation approaches), Dunnett's test was used to correct for familywise error rate (Dunnett, 1955).

## **6.2.5. Outcomes and Discussion**

### **6.2.5.1. Absolute Percentage Error**

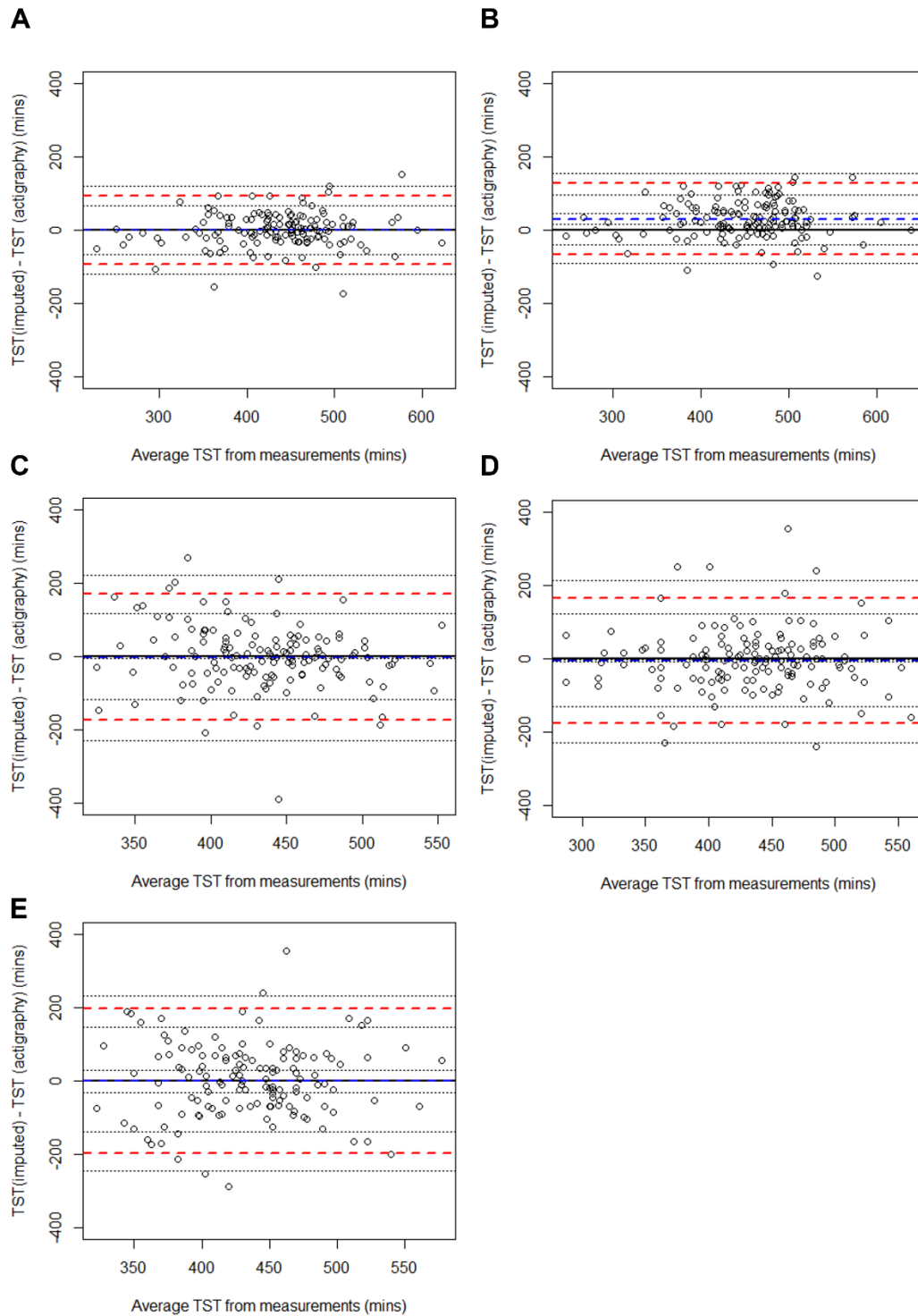
The following results are with reference to all levels of available data per participant (10 – 13 days) and for all sleep variables considered (TST, TASO & TAW). Firstly, the *BP Hot Deck* approach yielded a significantly larger APE compared to all other approaches ( $p < 0.001$ ), except for TST with 13 days available, where APE was significantly larger only when compared to the diary-based approaches, and TAW with 13 days available, where APE was significantly larger than all approaches except for *WP Hot Deck*. Secondly, *WP Hot Deck* and *Proximity Approach* did not significantly differ in APE ( $p > 0.05$ ), but yielded a significantly larger APE than diary based approaches ( $p < 0.05$ ) using all sleep variables with all numbers of days available. Lastly, APE did not significantly differ between diary-based approaches. Figure 6-1 shows the APE for each approach as across all days of available data per participant; Figures showing the APE for all sleep variables within individual amounts of data available per participant can be found as appendix 6.1.



**Figure 6-1** Line graphs depicting the mean absolute percentage error (APE) of imputed values, obtained using the five outlined imputation approaches, for **A** TST, **B** TASO, and **C** TAW.

#### 6.2.5.2. Limits of Agreement

Limits of Agreement (LoA) were consistently and considerably narrower for the two diary-based imputation approaches. Of the diary approaches, limits were equidistant from 0 for the *Diary  $\pm$  individual bias* approach, but not for the *Diary Only* approach. Figure 6-2 displays Bland-Altman plots for imputed vs. actual TST values, when 13 days of data were available per participant. Replicated figures for other sleep variables with 13 days of data available, as well as all sleep variables with 10 days of data available, can be found as appendix 6.2.



**Figure 6-2.** Bland-Altman Plots displaying the agreement between actigraphy-derived Total Sleep Time (TST) values, and values derived from the five considered imputation approaches (with 13 days available per participant): **A** *Diary  $\pm$  Individual Bias*, **B** *Diary Only*, **C** *Proximity Imputation*, **D** *WP Hot Deck*, and **E** *BP Hot Deck*. Blue dashed line represent the mean for TST (imputed) minus TST (actigraphy-derived). Red dashed lines represent 95% Confidence Limits. Black dotted lines represent the 95% bootstrap confidence intervals for these values.

### 6.2.5.3. Group Level Bias

For all levels of days of data available per participant, the *Diary Only* approach was the only approach with a mean TST significantly different to actigraphy-derived TST, overestimating TST by 29.76 to 31.84 minutes ( $p < 0.001$ ). The only other approach which resulted in a significant group-level difference from actigraphy-derived values was the *BP Hot Deck* approach when 12 days of data per participant were available, for which it underestimated TST by 19.31 minutes ( $p = 0.01$ ). Table 6-2 provides the mean values for actigraphy-derived and imputed sleep variables, assessments of bias, and limits of agreement, for analysis with 13 days of data available per participant; replicated table with 10 days of data available per participant can be found as appendix 6.3.

### 6.2.5.4. Discussion

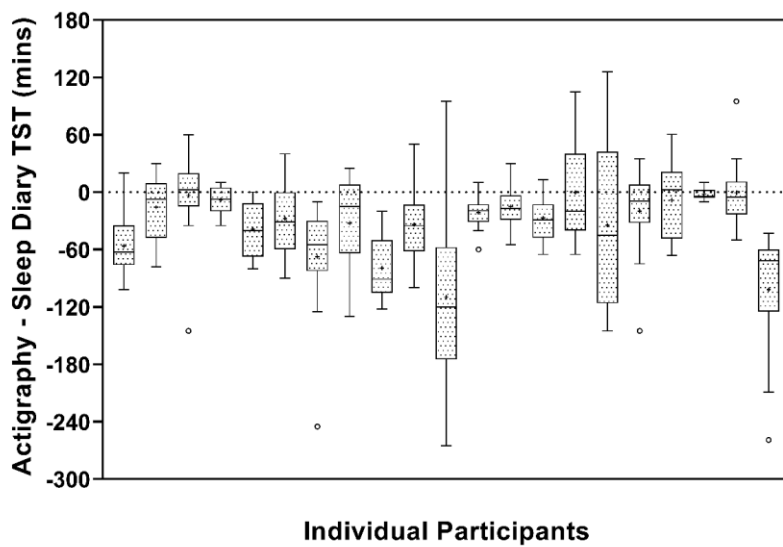
This section outlined analyses performed to uncover an optimal simple imputation approach for imputing missing actigraphy-derived sleep variable data. I assessed a method which used values from day/s immediately before and/ or following the missing day (*Proximity Approach*), and two *hot-deck* approaches, based on previous use in studies within the current scientific literature. I also assessed two approaches which utilised concurrent sleep diary derived data; direct diary imputation (*Diary Only*) and the use of diary values factoring in one's tendencies to over or underestimate sleep variable values when compared to actigraphy (*Diary  $\pm$  individual bias*).

Both the diary approaches demonstrated much greater agreement with all actigraphy-derived sleep variable values than the other approaches, as outlined both by lower absolute percentage error (APE) values, and narrower limits of agreement. Agreement between both diary-based approaches remained indistinguishable throughout the range of days available per participant assessed. However, when assessing bias, I found a consistent tendency for the *Diary Only* approach to overestimate TST by ~30min. Comparatively, the *Diary  $\pm$  Individual Bias* approach did not yield any group-level bias. Hence, the *Diary  $\pm$  Individual Bias* approach presents as the superior approach for simple imputation of actigraphy derived TST, while remaining equivalent to a diary-only approach for TASO and TAW. Due to this, the *Diary  $\pm$  Individual Bias* was chosen as the simple imputation approach for the study outlined in **Chapter 7**.

**Table 6-2** Mean values for actigraphy-derived and imputed sleep variables, assessments of bias, and limits of agreement, for analysis with 13 days of data available per participant.

Method	TST ( $\pm$ SE)	Overall Bias ( $\pm$ SE)	t (p)	Cohen's d	Lower LoA	Upper LoA
<b>Total Sleep Time (TST)</b>						
Readiband data	435 $\pm$ 9 min					
Diary $\pm$ Bias	435 $\pm$ 9 min	0	0 (1.00)	0	-1 hr 33 min	1 hr 33 min
Diary	465 $\pm$ 9 min	30 $\pm$ 7 min	4.19 (<0.001)***	0.84	-1 hr 8 min	2 hr 8 min
Proximity	433 $\pm$ 9 min	-2 $\pm$ 7 min	-0.28 (0.99)	-0.06	-2 hr 55 min	2 hr 51 min
WP Hot-Deck	430 $\pm$ 9 min	-5 $\pm$ 7 min	-0.70 (0.89)	-0.14	-2 hr 56 min	2 hr 46 min
BP Hot-Deck	436 $\pm$ 9 min	1 $\pm$ 7 min	0.11 (1.00)	0.02	-3 hr 16 min	3 hr 18 min
<b>Time at Sleep Onset (TASO)</b>						
Readiband data	25:31 $\pm$ 18min					
Diary $\pm$ Bias	25:31 $\pm$ 18min	0	0 (1.00)	0	-54 min	54 min
Diary	25:24 $\pm$ 18min	-7 $\pm$ 8 min	-0.82 (0.83)	-0.12	-59 min	45 min
Proximity	25:33 $\pm$ 18min	2 $\pm$ 8 min	0.25 (0.99)	0.04	-2 hr 15 min	2 hr 19 min
WP Hot-Deck	25:31 $\pm$ 18min	1 $\pm$ 8 min	0.07 (1.00)	0.01	-3 hr 7 min	3 hr 8 min
BP Hot-Deck	25:34 $\pm$ 18min	3 $\pm$ 8 min	0.35 (0.98)	0.05	-4 hr 42 min	4 hr 48 min
<b>Time at Wake (TAW)</b>						
Readiband data	33:13 $\pm$ 15min					
Diary $\pm$ Bias	33:13 $\pm$ 15min	0	0 (1.00)	0	-48 min	48 min
Diary	33:10 $\pm$ 15min	-4 $\pm$ 7 min	-0.50 (0.95)	-0.06	-52 min	45 min
Proximity	33:13 $\pm$ 15min	-1 $\pm$ 7 min	-0.08 (1.00)	-0.01	-2 hr 41 min	2 hr 40 min
WP Hot-Deck	33:07 $\pm$ 15min	-6 $\pm$ 7 min	-0.81 (0.84)	-0.10	-3 hr 18 min	3 hr 6 min
BP Hot-Deck	33:13 $\pm$ 15min	0 $\pm$ 7 min	0.03 (1.00)	0	-3 hr 22 min	3 hr 22 min

An interesting point to note is that while the *Dairy Only* approach overestimated actigraphy-derived TST by ~30min, which is consistent with previous literature from other samples, there were large individual differences in TST over/under estimation. This is demonstrated in Figure 6-3, and again is consistent with previous literature in other samples (i.e., breast cancer survivors; Moore et al., 2015), highlighting the utility of considering bias *within participant*.



**Figure 6-3.** Boxplots showing individual tendencies to over/ underestimate total sleep time (TST), in comparison to actigraphy-derived values. Each box represents an individual. Values below the dotted horizontal line resemble a tendency to overestimate TST. \* indicate outliers as determined by Tukey’s method (Tukey, 1977).

Obtaining a simple imputation approach with strong agreement does not entirely render the approach faultless. It is noted that simple imputation reduce variance estimates as they do not factor in any uncertainty into the estimation of missing values (Sainani, 2015); this is an accepted limitation of my approach. Secondly, if data is not missing-completely-at-random (MCAR), simple imputation methods can induce bias. I do not identify any reason for actigraphy-derived sleep data to be missing which is not detectable from other data collected (i.e., not-missing-at-random, or NMAR). A Little’s MCAR test (Little, 1988) was used to determine that the actigraphy-derived sleep data within **Chapter 7** could be considered MCAR, and as such, my *diary ± individual bias* approach was used to impute such data.

### 6.3. A Description and Discussion of Mixed Effect Models

**Chapter 7** includes the analysis of *multilevel* data; that is, rows of data which are likely to be correlated by a factor not considered as an independent variable (or *fixed effects*). Multilevel data is extremely common in cognitive testing, which typically involve taking many (>50) trials per given test session from many participants, who are further grouped (intervention vs. control for example). In this instance, trials taken from a certain individual are correlated to other trials from the same individual. Extending the cognitive testing example, often participants will be required to respond to a set of  $n$  items (i.e., one of ten possible stimuli), with multiple exposures of each item per test session; multiple trials with the same item are also likely to be correlated. Similar issues of dependence are extremely common in longitudinal research, with multiple measurements of a participant across a period of time. In these scenarios, performing subsequent statistical analysis which does not consider the *dependence* attributable to multiple trials from either the same participant/ item is sometimes called *psuedoreplication*, and can result in inflation of Type I error rate (Judd et al., 2009; Judd et al., 2012).

An increasingly popular way of dealing with multilevel data is through use of mixed-effect models (MEMs). MEMs are regression models built using all available rows of data. Hence, there is no need for aggregation<sup>†</sup> (i.e., *mean* response across a test session for each participant), even when data possesses correlation between rows not attributable to fixed effects (independent variables). This is facilitated by the inclusion of correlating factors as *random effects* within the created model. Essentially, random effects allow for the *explicit* modelling of variance components, including these otherwise problematic dependencies. Random effects can be included as *random intercepts*, which allow the intercept (zero-point) of a created model to vary as a function of the random effect. For example, say a model is to be built to predict response time within a population at two time-points; once when rested, and once when sleep deprived. If *participant* is included as a random intercept, the model allows the response speed to vary between participants, however this variance will be irrespective of the time-point. Random effects can also be included as *random slopes*, which allow magnitude of one or more fixed effects (or even their interaction) to vary as a function of the random effect. Following the earlier example, the inclusion of a *time-point by participant random slope* allows the effect of *time-point* to vary between participants. A thorough description of MEMs are beyond the scope of this thesis, however the reader is guided to descriptions by Brown (2021). MEMs with a

<sup>†</sup> I note that MEMs can also provide enhanced power when compared to other methods requiring aggregation; Aarts et al., 2014

continuous outcome variable are called linear mixed models, while MEMs with a categorical outcome variable are called generalised linear mixed models.

Regarding the use of MEMs, the fields of experimental psychology and ecology are trailblazers both for the use, and improvement of use, of MEMs (i.e., Baayen et al. (2008); Barr et al. (2013); Bolker et al. (2009); Brehm and Alday (2022); Jaeger (2008); Matuschek et al. (2017); Meteyard and Davies (2020)). However, MEMs have begun to feature very prominently both in sleep science and sport science journals within the past decade (i.e., Basner et al., 2017; Dunican, Higgins, et al., 2018; Dunican et al., 2019; Flaa et al., 2021; Honn & Van Dongen, 2023; Knufinke et al., 2018; Lastella et al., 2015; Rigney et al., 2015; Sargent et al., 2014; Smith et al., 2021; Smithies, Eastwood, et al., 2021 as a few of many examples). Unfortunately however, many of these articles suffer from the same flaws as in many psychology papers as outlined by Brehm and Alday (2022) and Meteyard and Davies (2020); namely, a severe lack of information regarding fixed and random effect selection and structure, and insufficient information regarding contrast coding (the numerical coding of categorical fixed effects). In particular, many studies report using *intercept-only* models without justification, which can severely inflate risk of Type I error when not justified by the data (Barr et al., 2013; Judd et al., 2009; Judd et al., 2012; Matuschek et al., 2017). To avoid these concerns, I have provided a brief summary and justification of the modelling choices made regarding my MEM use in **Chapter 7**:

**Fixed Effect Selection:** Models included fixed effects that were directly relevant to the specific hypothesis being tested, as well as all possible interactions (i.e., full factorial). I used treatment coding for all categorical fixed effects. Treatment coding refers to the coding of the two-levels of a dichotomous variable as 0 and 1 respectively (Brehm & Alday, 2022). Treatment coding was considered the intuitive option as there are sensible baseline levels within each fixed effect considered.

**Random Effect Selection (Including Structure Selection):** Random effect structure was selected using a data driven approach outlined by Matuschek et al. (2017) for each model created. This approach begun by building a MEM with a maximal random effect structure. Following this, a model was created simplifying the random effect structure by the smallest possible amount. These two models were compared using a likelihood ratio test (LRT; goodness-of-fit measure). If  $\alpha_{LRT} > 0.2$ , a new model is made, which further simplified the random effect structure by the smallest possible amount; this model was

subsequently compared to the previous model created. This process was continued until a model is created for which  $\alpha_{LRT} \leq 0.2$ . At that point, the model with the simplest random effect structure *and* for which  $\alpha_{LRT} > 0.2$  when compared to the model it was simplified from, was selected. Our procedure used here differed from that outlined by Matuschek et al. (2017) only when the model to be selected was singular (i.e., overfitted; as detected by using the default convergence control settings in the *lme4* package in R (Bates et al., 2015). Singular models were not interpreted as they can produce conservative fixed effect estimates and can result in inappropriate/ inaccurate results from inferential analysis procedures (Bates et al., 2015). Hence if a selected model was singular (/overfitted), then the next most complex model which was not singular was selected, and if all models prior were singular, random effect simplification continued until the next most complex model that was not singular was identified, and this model was subsequently selected.

**Assumption Checking:** Assumptions for linear mixed models include the normality of conditional residuals (i.e., differences between observed and fitted values should follow a normal distribution) and of random effects, constant variance (homoscedasticity), influential observations (/outliers), multicollinearity (I note that multicollinearity was not an issue for any MEM within **Chapter 7** due to no possibility of fixed effect correlation), linearity of simple/ main effects, as well as sensibility of overall model fit. Once a model was selected (using the abovementioned procedures), these assumptions were visually examined using the *model\_check()* function within the *performance* package (0.10.4; Lüdtke et al., 2021). Assumptions for binomial generalised linear mixed models are checked additionally using the simulation-based approach from the *DHARMa* package (0.4.6; Hartig, 2022) and include sensibility of overall model fit, the uniformity of scaled conditional residuals (i.e., differences between scaled observed and fitted values should follow a uniform distribution), normality of random effects, constant variance (homoscedasticity), influential observations (/outliers), and multicollinearity. The outputs from these packages (and hence, the assumption checking process) for Psychomotor Vigilance Task (PVT) Response Speed, Lapse likelihood, and raw reaction time, within the study described in **Chapter 7**, are provided as an example in appendix 6.4. Transformations performed to satisfy the assumptions listed above are specified in the results section of **Chapter 7** where applicable.

**Degrees of Freedom Estimation & Significance Testing:** The calculation of *p*-values can be performed in multiple ways within linear mixed models. Luke (2017) demonstrated that the use of Satterthwaite or Kenward-Roger approximations for degrees

of freedom produced the most robust significance tests, while other approaches can be anticonservative. Hence, I used the Satterthwaite approximation of degrees of freedom. Degrees of Freedom cannot be estimated using these methods in generalised linear mixed models, and as such, Wald tests are used on z-scores produced within the model to determine the significance of fixed effects included (Wald, 1943).

**Model Reporting:** The reporting of model selection steps and details of the final model selected was undertaken using a table format provided by a best practice guide (Meteyard & Davies, 2020). These are provided for each model created as appendix 6.5.

## **Chapter 7. Don't lose sleep over esports: exploring how total sleep deprivation affects the cognitive and in-game performance of rocket league players**

This chapter is currently under review for publication in a peer-review journal:

**UNDER REVIEW: Smithies, T. D.,** Toth, A. J., Campbell, M. J. (2023). Don't lose sleep over esports: exploring how total sleep deprivation effects the cognitive and in-game performance of rocket league players.

Changes to the version submitted for publication for the purposes of this thesis are outlined below:

- Change in referencing style (article version is in numbered format).
- References to supplementary files are changed to the appropriate location within the appendix.
- Words emphasised using quotation marks were changed to be emphasised using italics, in line with the thesis format.
- The words *Figure* and *Table* in in-text references to figures was capitalised. Furthermore, figure/ table numbering convention was changed in line with the thesis format.
- References to **previous chapters**, instead of supplementary files, for information presented within previous chapters.
- Minor amendments have been made based on examiner correction suggestions.
-

## 7.1. Abstract

**Study Objectives:** It is presumed by many that acute sleep loss results in degraded in-game esports (competitive, organised video game play) performance. However, this has not been experimentally investigated to date. The objective of the current experiment was to elucidate whether ~29hrs of total sleep deprivation impacts in-game performance for the popular esports *Rocket League*.

**Methods:** Twenty skill-matched pairs (N = 40 total) were recruited. Within each pair, one participant was assigned to an intervention group (TSD), while the other was assigned to a control group (CON). Two test sessions occurred; one while both participants were rested, and the other while the CON participant was rested, but the TSD participant was sleep deprived.

**Results:** Following total sleep deprivation, TSD participants reported higher Karolinska Sleepiness Scale-measured subjective sleepiness, and lower subjective alertness and motivation, as well as worsened PVT response speed (~50msec) and ~5 times greater PVT lapse incidence, and worsened response speed on a two-choice categorisation task (~40msec) ( $p < 0.05$  for all). However, overall in-game *Rocket League* performance (goal differential or ‘GD’) did not worsen due to total sleep deprivation ( $\Delta GD = 0.23 \pm 0.34$ ,  $p = 0.50$ ). Exploratory analyses on performance indicators suggest a potential shift toward a simpler and safer strategy following sleep deprivation.

**Conclusions:** Following a bout of ~29hrs total sleep deprivation, and in spite of increased subjective sleepiness, decreased subjective alertness and motivation, and decreased performance on the PVT and single task component of the category switch task, overall in-game *Rocket League* performance remained unaffected. This presents as a promising finding given high potential for acute pre-competition sleep disturbance in esports, though habitual sleep remains as a concern for esports athletes.

**Keywords:** *esports, performance, sleep deprivation, cognitive, task-switching, PVT, Rocket League*

**Statement of Significance:** Esports are quite comfortably the fastest growing competitive activity worldwide. The work presented is the first experimental study exploring how a bout of sleep loss impacts in-game performance in any esports. We found ~29hrs acute total sleep deprivation to have no impact on in-game outcome. This presents as a positive finding for

esport athletes and coaches alike, but certainly does not absolve sleep from being an impactful human factor within esports. Future studies should explore other esports with characteristics (i.e., longer bouts of sustained attention, such as Multiplayer Online Battle Arena or MOBA esports) purportedly sensitive to sleep loss, to see if the impact of sleep loss on esports performance is specific or agnostic to esport genre.

## **7.2. Introduction:**

Esports are by far the fastest growing competitive and high-performance activity worldwide. Defined as the competitive play of video games through the medium of cyberspace (Campbell et al., 2018), esports are a key part of the gaming industry, which has a projected market value of €375 billion in 2023 (Statista, 2023). The value of esports as an industry can be largely attributed to the size of its audience and public engagement. With viewership estimates exceeding one billion individuals in 2020 (Ahn et al., 2020) (and growing yearly), esports is and continues to be an enticing arena for investment. As a result, companies such as Xfinity, Kraft Group, PepsiCo, FTX, Red Bull, Coca-Cola, BMW, Nike, Asos, Ralph Lauren, and DC Comics (to name a few), over 319 traditional sporting teams (Code Red Esports, 2017), as well as even government organisations such as the U.S. Army (Nicholson, 2021) and Air Force, are either heavily invested in one or more esports teams, or own an esports team themselves. In response to this prolific interest in esports, there is an ever-increasing interest in understanding the human factors which influence esports competition performance in order to maximise the success of players.

One frequently highlighted human factor is sleep, or more specifically, the disturbance/loss of sleep experienced by players. Previously published literature has cited sleep loss in esports as a cause for concern specifically due to potential adverse impact on in-game performance (Bonnar, Castine, et al., 2019; Bonnar, Lee, et al., 2019; Bonnar et al., 2022; Kemp et al., 2021; S. Lee et al., 2021; Sanz-Milone et al., 2021), a sentiment shared with some esports athletes themselves (Baumann et al., 2022; Rudolf et al., 2020). Habitually, on average, professional esports athletes obtain a similar amount of sleep to others in their demographic (mid-late teenagers/ young adults, mostly male) (Bonnar et al., 2022; Gomes et al., 2021; S. Lee et al., 2021; Moen et al., 2022). However, esports athletes are characterised by incredibly late sleep onset (01:30 – 05:00) and wake (09:00 – 12:00) times on average, though large cultural/ regional group level differences have been noted (S. Lee et al., 2021). Sleep efficiency has also been cited as a concern, with a large longitudinal study of 1,243 nights of habitual esports player sleep data reporting a mean sleep efficiency of only 67.7% (Moen et al., 2022). Additional concern regarding the habitual sleep of esports players is warranted, given that multiple studies report mean insomnia severity index values at or beyond the cut-off for insomnia (Bonnar et al., 2022; S. Lee et al., 2021) and mean Pittsburgh Sleep Quality Index (PSQI) values well beyond the cut-off for poor sleep quality in this population (Gomes et al., 2021; Sanz-Milone et

al., 2021), somewhat mirroring the poor PSQI assessed sleep quality prevalent within traditional sport athletes (Doherty et al., 2021).

Furthermore, while elite esports athletes share many of the well-cited risk factors of sub-optimal sleep of traditional-sport athletes (i.e., pre-competition arousal/ anxiety, post-competition arousal, caffeine use, travel/ jet-lag), there are further risk factors uniquely associated with esports. Firstly, as esports are played through blue-light emitting computer monitors, there is a propensity for evening or night-time play leading to melatonin suppression, which can increase in sleep latency and reduce in sleep quality/ quantity (Green et al., 2017; Schöllhorn et al., 2023). Secondly, video games played as *esports* are cognitively/ physiologically arousing by design (i.e., high-intensity gaming), and as such, evening or night competition can reduce sleep quality and quantity through heightened arousal (Higuchi et al., 2005; Roberts, Teo, & Warmington, 2019). Such risk-factors have already been identified as potential mechanisms underlying associations between various sleep problems (poorer sleep quality, lower total sleep time [TST], increased prevalence of insomnia) and gaming frequency/duration (Kemp et al., 2021). Tying these risk factors together is a *culture* among professional esports athletes which promotes (and seemingly necessitates) training and playing late at night and into the early hours of the morning (Bonnar, Lee, et al., 2019; Lee et al., 2020). Overall, despite mean TSTs that are generally comparable to their peers, esports athletes are exposed to a cocktail of factors which together appear particularly conducive to bouts of acute sleep loss.

As mentioned, a common reason given for why sleep loss should be a major concern for esports athletes is that it can lead to in-game performance decrements. In contrast to many traditional sports, esports performance is predicated largely on cognitive abilities rather than physical abilities, leading some researchers to refer to esports athletes as *cognitive athletes* (Campbell et al., 2018). Though specific cognitive demands differ between different esports (Dobrowolski et al., 2015; Toth, Conroy, et al., 2021), most esports titles (especially those considered *action video games*) require rapid perception, processing and integration of multisensory stimuli originating from various sources (taxing visuospatial working memory systems), alongside fast and accurate decision making and responses through a peripheral device (keyboard/ mouse/ controller). Convincing evidence can be found for the robust cognitive demands of esports by looking at the now large body of quasi-experimental and intervention studies demonstrating how exposure to video games commonly played as esports improve aspects of cognition, even when tested outside of the specific game's context (see Bediou et al., 2018; Bediou et al., 2023; Toth et al., 2020

for relevant reviews/ meta-analyses). These improvements remain present even when disentangled from general improvements in motor execution (Bediou et al., 2023). Robust evidence is present for such video games improving attentional capacities (particularly visual attention), information processing speed and accuracy, and cognitive flexibility (in particular, task-switching), speaking to the importance of such elements for gameplay success.

Simultaneously, it is understood that acute sleep loss (i.e., total sleep deprivation/ sleep restriction) degrades performance in these same aspects of cognition (Lim & Dinges, 2010; Lowe et al., 2017). Though effects tend to be larger and more robust for simple (i.e., Psychomotor Vigilance Task or PVT) rather than complex attentional tasks (Glennville et al., 1978; Lim & Dinges, 2010; Pilcher & Huffcutt, 1996; Smithies, Toth, et al., 2021), sleep loss protocols have found response times and accuracy to worsen for tasks taxing visual attention, information processing, working memory, decision making, and executive functioning. A specific aspect of executive functioning, cognitive flexibility, has been highlighted as a domain particularly susceptible to sleep loss (Harrison & Horne, 2000; Honn et al., 2019; Whitney et al., 2019). Given that task-switching ability (a primary component of cognitive flexibility (see Ionescu, 2012; Uddin, 2021) appears integral to esports performance, the degradation of task-switching performance through sleep loss has been previously highlighted as an avenue for sleep loss to impact esports performance (Toth et al., 2020).

Despite this logical link between sleep, cognition, and esports performance, there has been no formal investigation into the effects of experimentally induced sleep loss on esports performance. Moen et al. (2022) investigated associations between habitual sleep and in-game performance for CS:GO (a popular first-person shooter esports) as a secondary analysis, finding no effect of TST on performance; however this approach was uncontrolled and only based on habitual sleep, and hence was unlikely to capture any subtle effects of TST on in-game performance, should they have been present. A controlled, experimental approach appears warranted to elucidate what (if any) observable effect acute sleep loss may have on the cognitive and in-game performance of esports players. Implications of such an experiment may be large for esports athletes and organisations alike, who have great desire to optimise every human factor which may impact their in-game performance.

The purpose of the current study is to explore how a bout of acute sleep deprivation (~29 hours awake) affects the cognitive and in-game performance of esports players in the esports *Rocket League*. Rocket League is a popular *vehicular soccer video game*, which averages ~90 million active players per month (Active Player, 2023), and as an esports, ranks 10<sup>th</sup> for most prize money earned (Esports Earnings, 2023b). The weight of evidence linking sleep loss to decreased cognitive performance, combined with the substantial cognitive demands of esports, leads us to hypothesise that sleep deprivation will worsen both cognitive (specifically, vigilance and task-switching performance) and in-game performance. We also aim to explore if (and how) certain established in-game Rocket League performance indicators are affected by sleep deprivation.

### 7.3. Methods:

All procedures and data collection were approved by the Education and Health Sciences Research Ethics Committee (2021\_06\_13\_EHS) and conducted in accordance with The Declaration of Helsinki.

#### 7.3.1. Participants

##### 7.3.1.1. Sample

An *a priori* power analysis was conducted, based on the predicted model structure for the primary analysis exploring the effect of total sleep deprivation on our overall in-game outcome measure, *goal difference* (GD), and following simulation processes outlined by DeBruine and Barr (2021). The details of this power analysis and R script can be found as appendix 7.1, however in short, we used an estimated effect size equivalent to the average effect size from all cognitive domains within a prior meta-analysis on sleep loss and cognitive performance (Lim & Dinges, 2010), combined with estimated variance components obtained through analysis of large databases of Rocket League matches, and a predicted level of warranted random effect complexity (correlated random intercept and slope). Using an alpha of 0.05, this power analysis suggested that 19 player pairs were required to achieve a power of 0.8. Using this (and adding one for the sake of evenness in counterbalancing), we sought to recruit 40 participants within the current study, allowing for 20 pairs.

46 young (18-35 years) adults provided written informed consent to participate in study. However, due to protocol non-adherence and participant drop-outs, we obtained a final sample of 40 ( $19.88 \pm 2.07$  years, 1 female) participants (20 pairs). Initially, we sought for participants to fulfil the criteria of a “normal healthy sleeper” according to the “Revised Research Diagnostic Criteria for Defining Normal Sleeping Controls” (RRDC) (Beattie et al., 2015), using answers obtained through an eligibility questionnaire. However due to extreme difficulty recruiting participants who were Rocket League players and also fulfilled this criteria, this was relaxed such that participants were eligible if they (a) habitually slept for six or more hours per night, (b) had no history of diagnosed sleep disorders and (c) were not alcohol dependant, nor were habitual users of other recreational drugs (besides tobacco). A summary of the included population with reference to the RRDC criteria can be found as Table 7.1. Of particular note, one included participant self-reported a seemingly inverted sleep-wake pattern (06:00 bed time, 17:00 rise time).

**Table 7-1.** Summary of included participants with reference to the Revised Research Diagnostic Criteria for Defining Normal Sleeping Controls (Beattie et al., 2015).

Component	Aspect	How Answered	Criteria/Cutoff	Mean( $\pm$ SD)/ Median	N above/ below threshold	Range
Sleep quality	Sleep duration	PSQI - Question 4	$\geq 6$ hrs	7.53 $\pm$ 1.17 hrs	0	6 - 11 hrs
	Time in bed	PSQI - Question 1 and 3	$\leq 9$ hrs	8.64 $\pm$ 1.25 hrs	11	6.5 - 11 hrs
	Sleep continuity	PSQI - Component 4	SE $\geq 85\%$	87.62 $\pm$ 9.98 %	14	65 - 100 %
		PSQI - Question 2	SL $\leq 30$ min WASO & EMA $\leq 30$ min	23.44 $\pm$ 14.85 mins	8	5 - 60 mins
	Subjective sleep impression Associated daytime effects	PSQI - Derived from Questions 1, 2, and 4		46.81 $\pm$ 52.09 mins	18	0 - 172.5 mins
		PSQI - Component 1	score $\leq 2$	1 (median)	0	0 - 2
Sleep timing	Habitual bed times	PSQI - Component 7	score $\leq 2$	1 (median)	1	0 - 3
	Habitual rise times	PSQI - a	22:00 - 01:00	00:57 $\pm$ 91 mins	13	23:00 - 06:00
	Stability of sleep timing (bedtime)	PSQI - c	06:00 - 09:00	09:35 $\pm$ 114 mins	21	07:00 - 17:00
	Stability of sleep timing (wake)	On a normal week, how many days would your (a) bedtime and (b) wake time deviate from your average by more than 1hr	total $\leq 3$	2.38 $\pm$ 0.95	5	0 - 4.5
	Associated daytime effects	On a normal week, how many days would your (a) bedtime and (b) wake time deviate from your average by more than 1hr	total $\leq 3$	2.23 $\pm$ 1.32	7	0 - 5
Sleep disorders	Diurnal Preference	PSQI - component 7	score $\leq 2$ Eve $\leq 41$ 42 < Int > 58 Morn $\geq 59$	1 (median)	1	0 - 3
		Horne-Östberg Morningness-Eveningness Questionnaire			NA	29 Intermediate, 11 Evening
	Insomnia disorder	HSDQ - 1, 7, 10, 12, 13, 14, 15 & 21	$\leq 3.68$	2.02 $\pm$ 0.63	1	1 - 3.88
	Circadian rhythm sleep disorder	HSQD - 5, 10, 13, 26, 27, 30	$\leq 3.41$	2.11 $\pm$ 0.68	2	1 - 4
	Sleep apnea	HSDQ 3, 17, 18, 19	$\leq 2.87$	1.57 $\pm$ 0.43	0	1 - 2.75
	PLMS/RLS	HSDQ	$\leq 2.70$	1.80 $\pm$ 0.59	2	1 - 4.2
General health	Narcolepsy	SNS ((6*Q1 + 9*Q2 - 5*Q3 - 11*Q4 - 13*Q5) + 20)	$\geq 0$	21.63 $\pm$ 14.50	2	-3 - 60
	Parasomnia	HSDQ - 4, 16, 20, 22, 24, 31	$\leq 2.42$	1.21 $\pm$ 0.30	0	1 - 2.33
General health	Physical health	Diagnosed with an ongoing physical/neurological disorder/problem?	No	NA	0	
	Mental health	Diagnosed with an ongoing psychological disorder/problem?	No	NA	0	

Medication use	Current medication use (besides hormonal contraceptives)?	No	NA	5	3: antihistamines, 1: asthma medication, 1: indigestion medication
Substance abuse: >400mg caffeine per day	Do you habitually (i.e. most days) consume more than 400mg of caffeine (i.e. >4 coffees or >3 large energy drinks)	No	NA	1	NA
No habitual nicotine or recreational drug use	Do you habitually (i.e. most days) consume tobacco/nicotine or any other recreational drugs (other than alcohol)?	No	NA	4 (all tobacco)	NA
no alcohol dependency	FAST/ AUDIT	≥ AUDIT SCORE 15	For those with an AUDIT score (N = 11); 8.64±2.11	0	0 - 14

Critically, participants were required to be players of the video game *Rocket League*. Participants were required to provide their in-game ranking using the rank tracking website <https://rocketleague.tracker.network>. This website provides a constantly updated record of a given player's in-game rank and matchmaking ranking (MMR). A player's MMR is measured on a continuous scale and is indicative of expertise within the 1v1 game mode of Rocket League (Smithies, Campbell, et al., 2021) & **Chapter 5**. Through <https://rocketleague.tracker.network>, we obtained each participant's highest and lowest 1v1 MMR over the most recent three months and calculated the mean of these two values to represent the participant's current expertise level. Additionally, the rank tracking website provides the participant's current rank percentile (i.e., the percentile of the current playing population in which the participant's rank resides in). Lastly, we also obtained an estimate of the total number of hours that the participant had accumulated playing Rocket League. A description of the method used to obtain this estimate is provided as appendix 7.2.

Where the MMR of two participants differed by less than 150 (equivalent of 15-21 total win vs. loss disparity), they were paired with one another. Paired individuals would complete aspects of the study at the same time, and play against one another in Rocket League matches during the two test sessions. We note that individuals who habitually use tobacco ( $N = 4$ ) were paired with one another. For each pair, one member was randomly selected to partake in the overnight sleep deprivation protocol (TSD), while the other individual was assigned as control (CON) (described below), using an automated web-based randomisation service (Haahr, 2021).

### **7.3.2. Materials**

#### **7.3.2.1. Eligibility Questionnaires & Participant Demographics**

The eligibility questionnaire provided to each participant included the Pittsburgh Sleep Quality Index (PSQI), Holland Sleep Disorder Questionnaire (HSDQ), Swiss Narcolepsy Scale (SNS), Horne-Östberg Morningness Eveningness Questionnaire (MEQ), and the Fast Alcohol Screening Test (FAST) & Alcohol Use Disorders Identification Test (AUDIT).

#### **7.3.2.1.1. Pittsburgh Sleep Quality Index (PSQI):**

The ten-item PSQI is the most commonly used and gold standard self-report measurement of sleep quality (Buysse et al., 1989). The PSQI shows strong reliability and validity and is appropriate for use in clinical and non-clinical populations (Mollaveva et al., 2016).

#### **7.3.2.1.2. Holland Sleep Disorder Questionnaire (HSDQ):**

The HSDQ is a 32-item questionnaire designed to screen for multiple sleep disorders as defined by the International Classification of Sleep Disorders (ICSD-2, 2005) (Kerkhof et al., 2013). We used the HSDQ (cut-off: sensitivity, specificity) to screen for; insomnia (2.02: 0.82, 0.51), circadian rhythm sleep disorder (CRSD) (3.41: 0.81, 0.75), sleep disordered breathing (2.87: 0.86, 0.81), periodic limb movement disorder (PLMS)/restless leg syndrome (RLS) (2.70: 0.82, 0.77), and parasomnia (2.42: 0.90, 0.90), as per Beattie et al. (2015). The HSQD has been found to be one of only two *comprehensive* questionnaires which screens for multiple sleep disorders (Klingman et al., 2017).

#### **7.3.2.1.3. Swiss Narcolepsy Scale:**

The SNS is a five-item questionnaire designed to screen for narcolepsy (Sturzenegger & Bassetti, 2004). The SNS exhibits superior performance (sensitivity = 0.93, specificity = 0.88) to other self-report screening tools for narcolepsy (Sturzenegger et al., 2018).

#### **7.3.2.1.4. Horne-Östberg Morningness Eveningness Questionnaire (MEQ):**

The MEQ is a 19-item questionnaire used to assess diurnal preference (Horne & Östberg, 1976). The MEQ shows agreement with actigraphy-derived measures around sleep timing (Thun et al., 2012), and its validity has been demonstrated against many other subjective and objective measures of human circadian rhythm (see Panjeh et al. (2021), p. 235, for a summary).

#### **7.3.2.1.5. Fast Alcohol Screening Test (FAST) & Alcohol Use Disorders Identification Test (AUDIT):**

The FAST was used to identify participants who may be at risk of alcohol use disorder (Hodgson et al., 2002). If participants were FAST positive (total  $\geq 3$ ;  $N = 11$ ), the remaining questions of the AUDIT were administered. The AUDIT is the gold-standard self-report measure for alcohol use disorder screening (Reinert & Allen, 2007; Saunders et al., 1993). We used cut-off values recommended by the World Health Organisation (Babor et al., 2001) (0 – 7: Zone I/ low risk, 8 – 15: Zone II/ hazardous drinking, 16 – 19: Zone III/ harmful drinking,  $\geq 20$ : Zone III/ possible dependence).

### **7.3.2.2. Subjective Sleep Measurement**

#### **7.3.2.2.1. Consensus Sleep Diary**

The Consensus Sleep Diary (Core) was used to obtain subjective sleep measures throughout the protocol (Carney et al., 2012). The Consensus Sleep Diary was created through collaboration of a large number of field experts (Carney et al., 2012), and is the research gold-standard for subjective sleep measurement.

### **7.3.2.3. Objective Sleep Measurement**

#### **7.3.2.3.1. Actigraphy**

Sleep variables were objectively measured using the Readiband<sup>TM</sup> (v5) wrist-worn activity monitor (Fatigue Science, Canada). This device uses tri-axial accelerometry (sampling frequency = 16Hz) to record wrist acceleration data, which are used to calculate sleep and wake events through a proprietary algorithm. The Readiband has demonstrated superior performance (most notably, less bias on sleep summary measures and greater sleep/ wake specificity) than the research standard Actiwatch 2, both at-home and in lab (when compared to the *gold-standard* polysomnography (PSG)) (Chinoy et al., 2021; Chinoy et al., 2022). It has also been independently found to be suitable when recording measures of total sleep time (TST), time at sleep onset (TASO) and time at wake (TAW) (Dunican, Murray, et al., 2018). Finally, the Readiband has high (ICC  $\geq 0.8$ ) inter-device reliability (including ICC = 0.99 for total sleep time; Driller et al. (2016)), and has been used in sleep research for a variety of populations, including traditional and esports athletes (Bonnar et al., 2022; Dunican et al., 2023; S. Lee et al., 2021; Power et al., 2023; Smithies, Eastwood, et al., 2021), medical personnel (James et al., 2019; Min et al., 2023), pilots (Rocha & Silva, 2019), and military personnel (Edgar et al., 2023).

A single trained researcher downloaded and processed the Readiband data. Outcome measures considered were TST, TASO, and TAW. Daytime naps were included in TST, with naps occurring before 12:00 added to TST for the previous night, and naps occurring after 12:00 added to the TST for the upcoming night, as per Smithies, Eastwood, et al. (2021).

### **7.3.2.4. Subjective Sleepiness, Alertness & Motivation**

To capture subjective sleepiness, alertness, and motivation of participants throughout the experimental protocol, participants completed The Karolinska Sleepiness Scale (KSS) as well as Alertness & Motivation Visual Analog Scales (VAS). The KSS is a widely used

single-item measure of individual subjective sleepiness at a given time-point (i.e., situational sleepiness). The KSS is answered on a nine-point Likert scale, with 1 denoting *extremely alert*, and 9 denoting *very sleepy, great effort to keep awake, fighting sleep*. Levels of subjective alertness and motivation were assessed using slider scales, with values (between 0 and 100) hidden. The alertness VAS ranged from *sleepy* (0) to *alert* (100), and the motivation VAS ranges from *motivated* (0) to *unmotivated* (100), as per Mathew et al. (2021). Motivation VAS scores were subsequently reverse scored for analysis, such that higher scores resembled greater motivation.

#### **7.3.2.5. Cognitive Performance**

Cognitive performance during the experimental protocol was assessed using the psychomotor vigilance task (PVT) and Category Switch Task (CST).

The 10 minute PVT was used to assess each participant's vigilance and reaction time. Participants were required to respond *as fast as possible* to the appearance of a red stopwatch in the centre of their screen by pressing the *space bar* on their keyboard. The inter-stimulus interval was set at random between 2,000-10,000ms for each trial. Participants were provided with feedback on their response time following each response, as well as their mean response time at the end of the testing session. If a response was made without the presence of the red stopwatch, a visual error message was displayed on the screen before the next trial commenced. The 10-minute PVT is the gold-standard performance test for vigilance, and exhibits stable performance over repeated measures testing (Balkin et al., 2004; Basner & Dinges, 2011; Basner et al., 2017). A one-minute practice block is undertaken prior to the ten-minute testing block. The test block is executed straight after the practice block with no break or indication it has changed, as per Thomann et al. (2014). False starts were removed from the data prior to analysis, as were trials responded to in  $\leq 100\text{ms}$  (as per Basner and Dinges (2011)). Dependent variables considered were response speed (equivalent to  $1000 / \text{reaction time}_{(\text{msec})}$ , and hereafter denoted as *RS*), and lapses, defined as trials where reaction time  $\geq 500\text{ms}$  (alternatively,  $\text{RS} \leq 2$ ), as these two measures have shown to display the best conceptual and statistical properties (including robustness to extreme values) and sensitivity to sleep loss for the PVT (Basner & Dinges, 2011).

The CST assesses task-switching ability, a component of executive functioning requiring cognitive flexibility. A detailed description of the CST used in this study can be found in **Chapter 3**. In short, participants were required to categorise words that appeared on a

screen according to a categorisation rule denoted by a cue. In some test blocks (*single task*), the cue was constant, while in others (*mixed task*), the cue (and categorisation rule) switched between one of two cues at random, and participants had to adapt to the corresponding change in categorisation rule. Stimulus-response mapping (SRM) changed from the first to second test sessions in a consistent manner between all participants (for the *living cue* but not *size cue*). We did not consider Switch Cost (SC) or Mixing Cost (MC) error rate, as per **Chapter 3**.

#### **7.3.2.6. In-Game Rocket League Performance**

The primary aim of the current study was to assess if an acute sleep deprivation intervention influenced in-game performance on the esports *Rocket League*. In Rocket League, players use a rocket-powered vehicle to hit a large ball into an opposing goal, while simultaneously defending one's own goal (as per soccer or hockey). Rocket League is played competitively in teams of 1, 2, or 3 players; in this study, we solely investigated 1v1 matches of Rocket League.

Paired participants played against one another on a local area network (LAN) connection. Participants were able to use their own input device for gameplay, however a DualShock 4 and Xbox Elite Controller (series 2), as well as a gaming mouse, keyboard, and headphones were provided if necessary. All input devices were used with a wired connection.

Participants were asked to log into their own Rocket League account on Steam or Epic, however were provided an account if they were unable. Participants were free to use headphones for game sound and/ or play music through the duration of the Rocket League matches. Participants were free to use in-game settings of their choosing (i.e., controller settings, camera settings etc.). Once both participants within the pair were ready, they were afforded five minutes for a warm-up. Participants were free to warm up however they chose (i.e., free-play training, training packs, workshop maps), with the exception of playing an online match. Once the five minutes have elapsed, participants joined a LAN (local area network) match, which was created by the researcher. Prior to the gameplay commencing, participants were (a) asked to save replays of the matches (a feature allowed at the end of any match by all users), and (b) asked to perform to the best of their ability for the entirety of each match, aiming to score as many goals as possible while simultaneously preventing their opponent from scoring, regardless of the match score. Participants then played seven consecutive matches against their paired opponent (for one

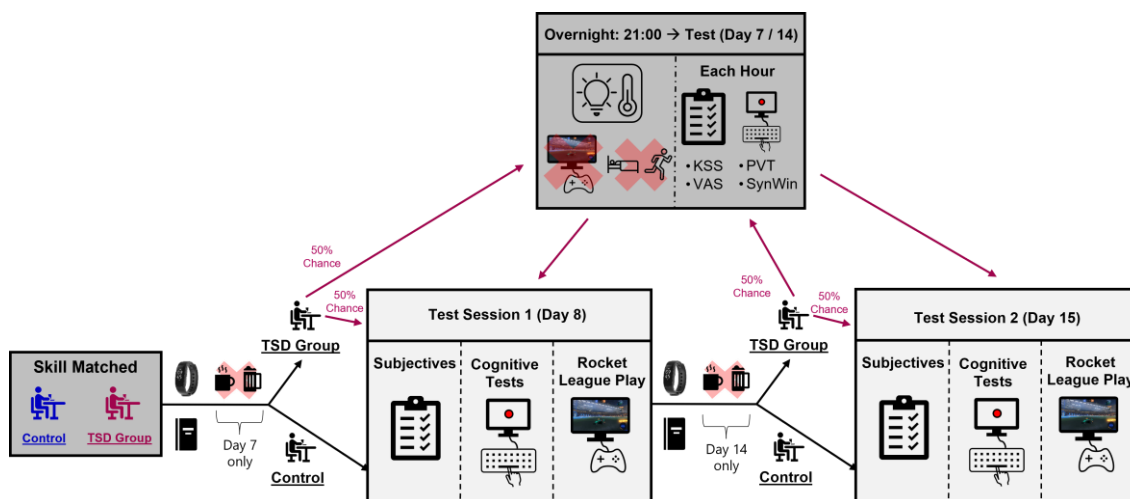
pair on one week, only six matches were played due to a participant needing to leave early). Short breaks (< 5 minutes) in between matches were afforded *ad libitum*.

Outcome measures were obtained via use of the ballchasing.com application programming interface (API) on saved match replays and processed as outlined by Smithies, Campbell, et al. (2021) **and in Chapter 5**.

### 7.3.3. Procedure

Figure 7-1 provides a visual outline of the study protocol. The study protocol lasted a minimum of 15-days per participant pair. On the first day (D1), participants were briefed on the study protocol, and provided their actigraphy device and consensus sleep diary. Participants determined a *target bed* (between 22:00 – 01:00+1) and *wake time* (between 06:00 – 09:00, to correspond with Beattie et al. (2015)). Participants were asked to adhere to their set *target bed* and *wake times* ( $\pm 1$ hr), particularly within the three days prior to each test session. Participants also agreed upon a *target gameplay amount* (hours) within the following week, of which they were asked to remain within  $\pm 20\%$  of (i.e., 80-120% of target hrs). *Target bed* and *wake times* and *target gameplay amount* remained consistent for individuals, but did not need to be consistent between participants within a given pair.

Participants were asked to synchronise their Readiband with their smartphone device upon waking to provide researchers access to their sleep data. Participants were reminded via text message to complete their sleep diary and synchronise their Readiband if they had not synchronised their Readiband by midday for all days in the protocol.



**Figure 7-1.** A timeline of the 15-day protocol for all participants within the protocol. The icons within this figure depict the following. Wristband = Readiband; book = Consensus

Sleep Diary; coffee mug and beer = caffeine and alcohol; clipboard = subjective measures. Computer with keyboard = computerised cognitive tests. Computer with controller = Rocket League gameplay; bed = sleeping; man running = strenuous activity; light and thermometer = light and temperature-controlled environment. TSD = participants within the sleep deprivation group; CON = participants within the control group.

#### **7.3.3.1. Test Session Protocol**

Following 12:00 on day seven (D7) and up until the upcoming test session (D8), participants were asked to refrain from consuming caffeine, alcohol, non-essential medication or any other drugs, as well as napping. Participants were also instructed to obtain a minimum of six hours of sleep in the upcoming night. On the eighth day of the protocol and between the hours of 11:30 to 15:00, participants attended the laboratory for *test session 1*. Sleep diary information were collected from each participant, and a new sleep diary was provided. Participants self-reported the amount of Rocket League gameplay undertaken in the previous seven days, and indicated their adherence to the abovementioned procedures.

Following this, the participants completed (in order) the PVT, CST, KSS, alertness & motivation VAS, and played their set of Rocket League matches against one another. Following completion of the Rocket League matches, participants again completed the KSS and alertness & motivation VAS. All procedures described were collected using gaming computers, comprising of a 27-inch monitor with a 144Hz refresh rate. All measures except for those in the Rocket League performance section were taken using identical input devices (Logitech Pro mouse, keyboard, and headphones). Following the set of Rocket League matches, both participants were asked “on a scale from 0 [not at all] to 10 [extremely], how much do you feel like fatigue affected your in-game performance?”, and “do you think the other participant had completed the overnight sleep deprivation protocol?”.

Following *test session 1*; the protocol was repeated, such that participants wore their Readiband and completed the sleep diary daily, played the agreed upon amount of Rocket League, adhered to the target bed and wake times within three days of the upcoming test session, and avoided consuming caffeine or alcohol, taking medication or drugs, or napping, within 24hrs of the upcoming *test session 2*. For 80% of pairs, the second test

session was exactly seven days following the test session. However due to participant availability, the timespan between tests was 14 days for three pairs, and 37 days for one pair.

### **7.3.3.2. Total Sleep Deprivation Protocol**

Within each pair, the protocol for the participant assigned to the control condition (CON) was exactly as described above. For the participant assigned to the TSD condition, one of the two *test sessions* (and the week prior) was exactly as described, however for the other *test session*, the participant completed the total sleep deprivation protocol prior. The *test session* for which the prior sleep deprivation was administered was counterbalanced (i.e., ten participants were sleep deprived prior to first and second *test session* respectively). Participants were (a) aware that they may be asked to complete the total sleep deprivation protocol prior to either of the test sessions, (b) unaware if they or their paired opponent were in the CON or TSD group, and (c) told at least three days in advance whether they were required to complete the total sleep deprivation protocol prior to the upcoming test session.

For the total sleep deprivation protocol, the participant arrived to the laboratory at 21:00 the night before the *test session*. The following day, the participant would remain in the laboratory until 30 minutes before the start of the *test session*. Participants were free to engage in activities of their choosing, except for strenuous exercise, or playing video games using the same input modality (i.e., keyboard or controller) they used to play Rocket League with. From 22:00 onwards, participants completed a 5-minute PVT and the SynWin multitask (Elsmore, 1994) on the hour each hour. The results of these tests are not within the scope of this article.

The light (~425 lux) and temperature ( $21 \pm 2^\circ\text{C}$ ) in the laboratory environment remained constant throughout this time and during all laboratory sessions. Each participant was supervised throughout the duration of the sleep deprivation protocol to ensure wakefulness. Water, fruit, low-sugar snacks, and caffeine-free hot beverages (i.e., peppermint tea) were available to participants *ad libitum* throughout the sleep deprivation protocol. Another standardised meal (toast with peanut butter and honey, fruit, and fruit juice) was provided at 08:00 the following morning. In the 30 minutes prior to the test session, participants left the laboratory, and were supervised on a walk; this was to simulate a walk to the laboratory, as would occur if they were the CON participant. This

protocol resulted in an average of  $28.78 \pm 1.22$  hours between last wake time and the start of Rocket League gameplay.

#### **7.3.4. Statistical Analysis**

Statistical analyses were performed using R: A language and Environment for Statistical Computing (Vienna, Austria) and/ or IBM SPSS Statistics v26 (Armonk, N.Y.) software. Alpha was set to  $p < 0.05$  (two-tailed) for all analyses. Variance measures ( $\pm$ ) are presented as standard error unless explicitly specified.

##### **7.3.4.1. Participant Pairs & Rank**

Means and standard deviations are provided for participant MMR. The relationship between time spent playing Rocket League (hours) and in-game expertise (MMR) was assessed through a simple linear regression, which was subsequently used to impute missing *hours played* for individuals from whom we could not obtain a confident estimate.

##### **7.3.4.2. Protocol Adherence**

Means and standard deviations are provided for each individual's TASO and TAW within the three days preceding either test session. Additionally, means and standard deviations are provided for TST, both for the night before each test session (TST[1]) and the two nights (combined) prior to TST[1] (TST[2-3]), within each *group x session* combination. Independent-sample t-tests and paired-sample t-tests were used to assess by-group and by-session group differences, respectively. Nonparametric equivalents (Mann-Whitney U and Wilcoxon Signed-Rank Test) were used when values within one or more groups were significantly non-normal (Shapiro-Wilk test,  $p < 0.05$ ).

Rate of adherence to target Rocket League gameplay was expressed as a percentage. Additionally, the mean and standard deviation for the proportion of target gameplay achieved was calculated for the entire sample, and for each *group x session* combination. By-group and by-session differences in target RL gameplay achieved were assessed identically to that described for TST above.

##### **7.3.4.3. Cognitive Performance**

Cognitive Performance Measures (PVT & CST) were assessed using Mixed Effect Models (MEMs). All MEMs were created using the *lme4* package in R (1.1-31; Bates et

al. (2015)). Random effect structures were determined using a backward-selection approach as outlined by Matuschek et al. (2017) **and described in chapter 6.3**, deviating only to avoid selection of singular models or models which did not converge. If a selected model was singular, then the next most complex model which was not singular was selected, and if all models prior were singular, random effect simplification continued until the next most complex model that was not singular was identified, and this model was subsequently selected. Within the selection process, random slopes were considered for each fixed effect (and their interaction) that vary within a given random effect (Barr, 2013). Once the most appropriate random effects structure was identified, mixed-model assumptions (see **Chapter 6.3**) were visually examined using the *performance* package (0.10.4; Lüdtke et al. (2021)), and *DHARMa* package (0.4.6; Hartig (2022)). For fixed effects, degrees of freedom were estimated using the Satterthwaite method to allow for significance testing of fixed effects, while the Wald method was used for confidence interval estimation. Additionally, Wald tests were used to determine fixed effect significance for any MEMs created with categorical outcomes (binomial generalised linear mixed effects models; i.e., for PVT lapses). We used *treatment coding* for all categorical *fixed effects*. Treatment coding refers to the coding of the two-levels of a dichotomous variable as 0 and 1 respectively (Brehm & Alday, 2022). Treatment coding was considered the intuitive option as there are sensible *baseline* levels (always coded as 0) within each fixed effect considered. Details regarding both model selection (including specific model selection decisions made) and the details of the final model selected are provided within table layouts based on a best practice guideline (Meteyard & Davies, 2020) and are provided as appendix 6.5.

For PVT measures, MEMs were created for RS and lapses, respectively. We note a deviation from normality observed at low RS values, however given that (a) this is representative of an expected phenomenon (lapses), (b) RS is an already transformed outcome measure which satisfies MEM assumptions substantially better than raw RT, and (c) RS is considered alongside lapses to be the best outcome variable to use for the PVT (Basner & Dinges, 2011), RS was retained as the outcome measure within the model created. The models created for RS and lapses included the between-participant fixed effect *group* (CON vs. TSD) and the within-participant fixed effect *session* (baseline vs. experimental), as well as their interaction, with participant considered as a random effect.

For CST, we planned on using  $RT_{(msec)}$  as an outcome measure for Single Task, SC and MC, however MEM assumptions were not met using this outcome measure. When RTs

were converted to RS using the same transformation used for PVT ( $RS = 1000 / \text{reaction time}_{(\text{msec})}$ ), MEM assumptions were satisfied. Hence, MEMs were created for Single Task RS and error rate, Switch Cost RS, and Mixing Cost RS. For Single Task models, they included the between-participant fixed effect *group* (CON vs. TSD) and the within-participant fixed effect *session* (baseline vs. experimental), as well as their interaction, while for the SC and MC models, *group*, *session*, and *trial type* (for SC, this was switch vs. repeat; for MC, this was repeat vs. single task), as well as their interactions (i.e., full factorial), were included as fixed effects. For all of these models, *participants*, *word* (i.e., the specific word displayed), *cue* (living or size), and a *cue* by *word* interaction, were considered as random effects.

#### **7.3.4.4. Subjective Measures**

By-group and by-session differences in KSS, Alertness VAS and Motivation VAS were analysed using independent/ paired sample t-tests or nonparametric equivalents, as per protocol adherence measures. We note that data from one pair are missing due to a technical error (N pairs = 19). Participant's self report of how much fatigue affected their in-game performance was also analysed in an identical manner.

#### **7.3.4.5. Rocket League Performance**

The primary aim of the study was to test the null hypothesis that TSD would not affect our in-game outcome variable, GD. To test this, we created a MEM with *session* (baseline vs. experimental) as a fixed effect, and *pair* as a random effect.

Irrespective of the result of the above analysis, we sought to conduct exploratory analysis on whether TSD impacted any performance indicators in Rocket League. We built five separate MEMs with identical fixed and random effects to that above, to predict the following outcome measures: *Shots Taken Difference*, *Time Spent Goalside of the Ball Difference*, *Saves Made Difference*, *Time Spent High in the Air Difference*, and *Demos Inflicted Difference*. These five PIs were chosen as they were the five PIs shown to predict game performance in 1v1 Rocket League when all in-game ranks are considered (Smithies, Campbell, et al., 2021), **see Chapter 5**. All metrics were calculated as the value of the TSD participant minus the value of the CON participant within each pair. As this analysis is exploratory, we did not conduct any familywise error rate adjustment; however, we do not make claims based on the results of these analyses, instead using

them as ways to highlight potential effects of TSD on in-game strategy to be explored in future research.

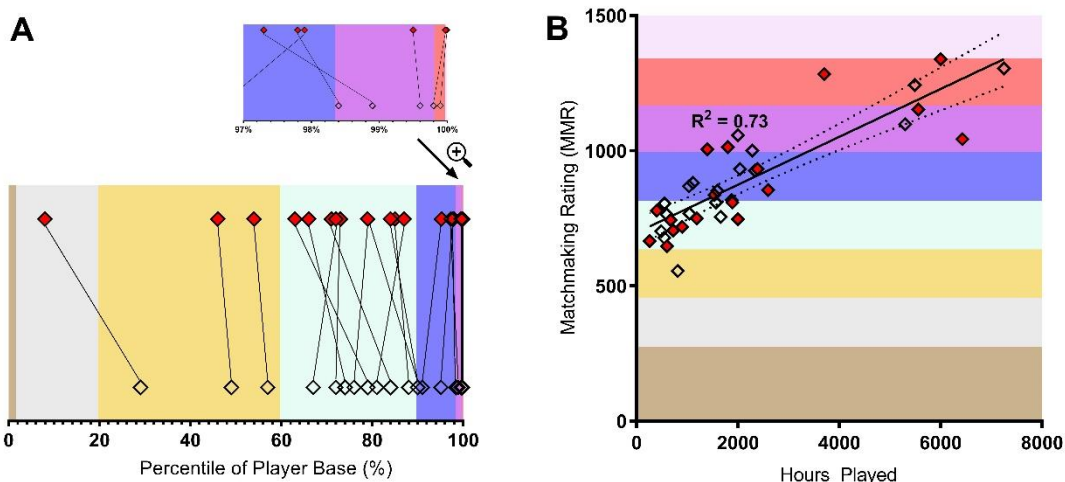
## 7.4. Results:

### 7.4.1. Participant Pairs & Rank

The mean MMR for participants in the study was 874.01 (SD = 203.59), corresponding to the top 20.62% of the overall playerbase. A simple linear regression was conducted to examine the relationship between *hours played* (from participants where a reliable value estimate could be obtained, N = 37) and MMR. This model was significant,  $F(1,35) = 93.48$ ,  $p < 0.001$ ,  $R^2 = 0.73$ , with *hours played* explaining 73% of the variance of participant 1v1 Rocket League Expertise. The equation for the regression model can be found below

$$MMR = 0.08853 \times \text{Hours Played} + 696.8$$

This equation was used to predict missing values for *hours played*, however for two of the three participants it predicted a negative value. Hence for these participants, *hours played* was conservatively estimated as zero. After including these participants, mean *hours played* among the sample was 2014.15hrs (SD = 1881.30hrs), or ~85 days. Figure 7-2A shows the rank distribution of the participants and their paired opponent, and Figure 7-2B shows the abovementioned relationship between *Hours Played* and *MMR*.



**Figure 7-2 A** Rank distribution and pairing of included players. Clear diamonds resemble CON participants, and red diamonds resemble TSD participants. Pairs are denoted by lines joining participants. The x-axis denotes the participants in-game MMR (a proxy for expertise) relative to the esports overall player base at the time of recruitment, such that lower values resemble a lower-ranked player and vice versa (i.e., 99% denotes a player

in the top 1% of players). To better visualize pairs in the top 3% of ranked 1v1 Rocket League players, a magnified display is depicted above the main graph. **B** Relationship between estimated total hours of Rocket League played (x-axis) and player expertise (y-axis). Dashed lines represent 95% CI for line fitted using the linear regression equation. For both A and B, colours represent the in-game rank of the participants (in order from bottom to top; bronze, silver, gold, platinum, diamond, champion, grand champion, supersonic legend).

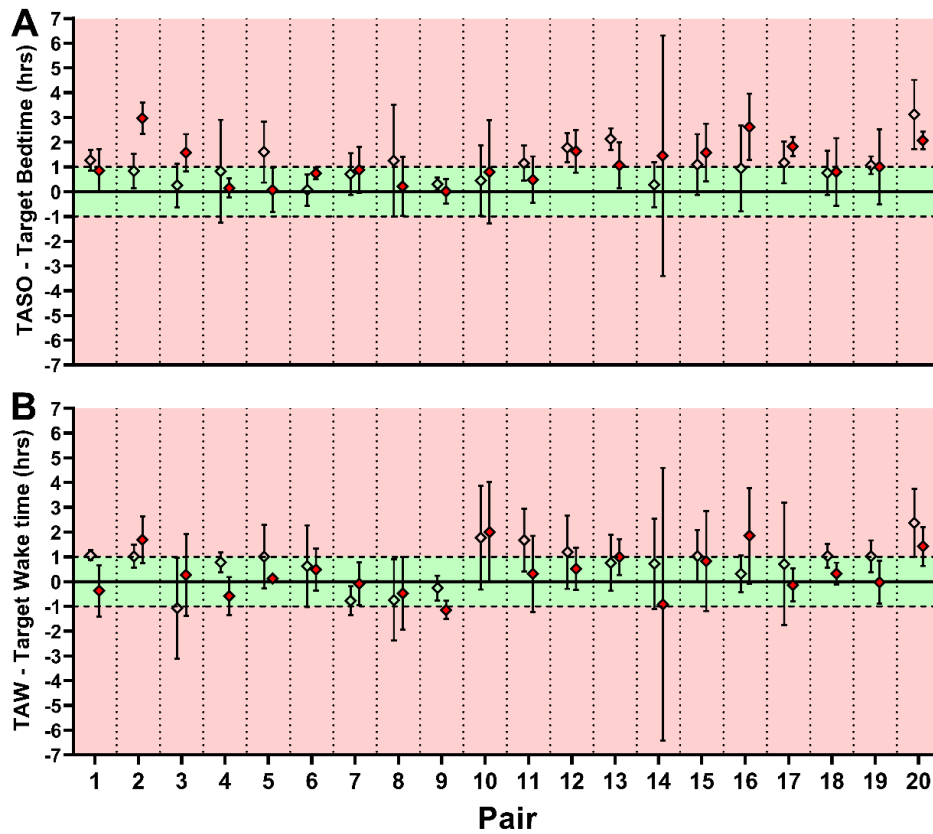
## 7.4.2. Protocol Adherence

### 7.4.2.1. Subjective & Objective Sleep Data

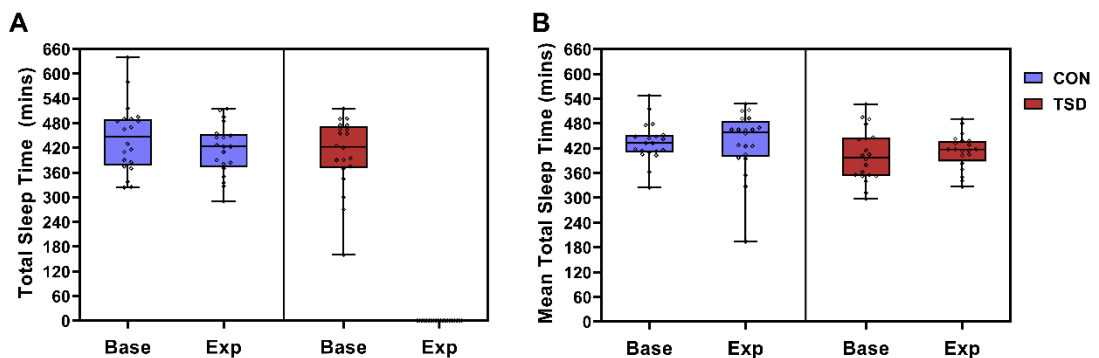
For sleep data within 3 nights of each test session, 3.64% of days of actigraphy derived sleep data were missing or unusable, while 4.55% of days of Consensus Sleep Diary (CSD) data were missing or unusable. A missing value analysis using Little's MCAR test (Little, 1988) was not significant ( $\chi^2 = 7.32$ ,  $p = 0.50$ ), suggesting the data can be treated as MCAR and as such, missing actigraphy-derived sleep data were imputed using a simple imputation method described in **Chapter 6.2**.

Actigraphy-derived TASO and TAW, in comparison to participant-defined target bed and wake times, are shown for each participant (and pair) in Figure 7-3. 48.2% of nights within three days of a test session had a TASO within one hour of the individuals target bedtime (mean difference = 1.09hrs (SD = 1.39)), while 54.84% of TAW values were within one hour of the individuals target wake time (mean difference = 0.55hrs (SD = 1.62)).

Mean TST (including naps) the night before each test session (TST[1], 7-4A), as well as the mean TST for two nights prior (TST[2-3], 7-4B), are shown for each condition in Figure 7-4. All between- and within-participant comparisons for TST[1] and TST[2,3] were not significant ( $p < 0.05$ ) with the exception of those involving TST[1] values for TSD on the experimental *session*. In other words, for the three nights preceding test sessions, the only observable difference in TST was as a direct result of the sleep deprivation protocol.



**Figure 7-3** Mean ( $\pm$ SD) discrepancy between **A** target bedtime and TASO, and **B** target wake-time and TAW, for each participant within each pair. Participants in CON are denoted by clear diamonds, while TSD participants are denoted by the red diamonds. The green band denotes TASO or TAW within 1hr of the target bed/ wake time, while the red area denotes TASO or TAW outside of that range.



**Figure 7-4** Box and whisker (min  $\rightarrow$  max) plots showing the group mean TSTs for CON and TSD **A** the night before test sessions, and **B** the mean of the two nights prior to that shown in **A**.

#### 7.4.2.2. Caffeine, Alcohol, Napping

No participants reported caffeine or alcohol use within 24-hours of any test session. One participant (CON) reported medication use within 24-hours of test sessions; a daily asthma medication on the morning of both test sessions, and one dosage of cough medication the night before the experimental test session. Two participants (one CON, one TSD) napped (45-60mins) within 24-hours of the baseline test session, as both self-reported and corroborated through actigraphy.

#### 7.4.2.3. Weekly Rocket League Play

Participants remained within  $\pm 20\%$  of their target gameplay prior to testing for 56.25% of test sessions ( $M = 89.44\%$  of target hours,  $SD = 38.02\%$ ). No significant differences were found between % target gameplay achieved prior to *baseline* and *experimental* test sessions for either group ( $p < 0.05$ ).

### 7.4.3. Cognitive Performance

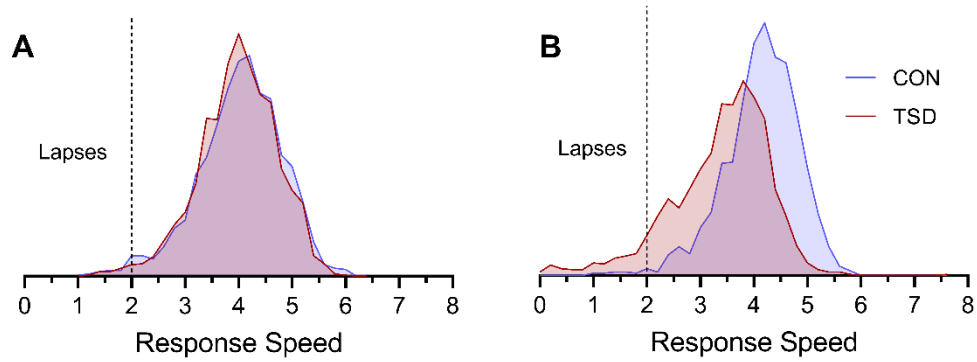
Model selection process and details of the final MEMs (as per Meteyard and Davies (2020)) can be found as appendix 6.5.

#### 7.4.3.1. Psychomotor Vigilance Task

Distributions of PVT Response time across *groups* and *sessions* are shown in Figure 7-5.

For PVT response time, neither *group* nor *session* alone significantly contributed to the model ( $p > .05$ ), however a significant *condition* by *session* interaction was present ( $b = -0.72 \pm 0.11$ , 95% CI  $[-0.94, -0.51]$ ,  $t(1, 37.38) = -6.58$ ,  $p < .001$ ), such that being in the TSD group on the experimental day resulted in a mean reaction time worsening of 48.61msec.

For PVT lapses, neither *group* nor *session* alone significantly contributed to the model ( $p > 0.05$ ), however a significant *condition* by *session* interaction was present ( $b = 2.38 \pm 0.40$ , 95% CI  $[1.56, 3.16]$ ,  $z = 5.91$ ,  $p < 0.001$ ), such that being in the TSD group on the experimental day resulted in 4.91 times more lapses occurring.



**Figure 7-5** Frequency distribution of RS ( $1000/RT_{(msec)}$ ) for participants in the **A** CON group and **B** TSD group. I direct the reader to **B**, and note both a leftward shift and leftward skew of the response distribution in the experimental test session for the TSD participants, consistent with previous literature (i.e., Figure 4, Grant et al. (2017)) and demonstrative of both the broadband decrease in RS and increase in lapses observed (trials to the left of the dotted line).

#### 7.4.3.2. Category Switch Task

For performance on the Single Task component of the CST, both RT (msec) and error rate were examined. RT was transformed to RS ( $1000/RT_{(msec)}$ ), as per PVT analysis) to fulfil MEM assumptions. For RS, neither *group* nor *session* alone significantly contributed to the model ( $p > .05$ ), however a significant *condition* by *session* interaction was present ( $b = 0.10 \pm 0.40$ , 95%CI [-0.19, -0.01],  $t(1, 37.90) = -2.27$ ,  $p = .029$ ), such that being in the TSD group on the experimental day resulted in a 42.19msec increase in mean reaction time. No fixed effect, nor interaction were significant within the model predicting errors in the Single Task component of the CST ( $p > 0.05$ ).

For SC RS, only *trial type* significantly contributed to the model ( $b = 0.24 \pm 0.40$ , 95%CI [-0.30, -0.19],  $t(1, 37.83) = -8.40$ ,  $p < 0.001$ ), corresponding to a switch cost of 102.19msec. No other main effects or interactions were significant ( $p < 0.05$ ). For MC RS, there was a simple main effect for *trial type* ( $b = -0.27 \pm 0.05$ , 95% CI [-0.36, -0.18],  $t(1, 52.78) = -5.79$ ,  $p < 0.001$ ), corresponding to a mixing cost of 83.45msec. There was also a significant *condition* by *session* interaction present ( $b = -0.10 \pm 0.04$ , 95% CI [-0.19, -0.01],  $t(1, 37.88) = -2.27$ ,  $p = 0.029$ ), such that according to the model, being in the TSD group on the experimental day resulted in a 42.21 msec increase in mean reaction time

for single task trials (identical to that in the single task only model). No other main effects or interactions were significant ( $p < 0.05$ ).

#### **7.4.4. Subjective Sleepiness, Alertness and Motivation**

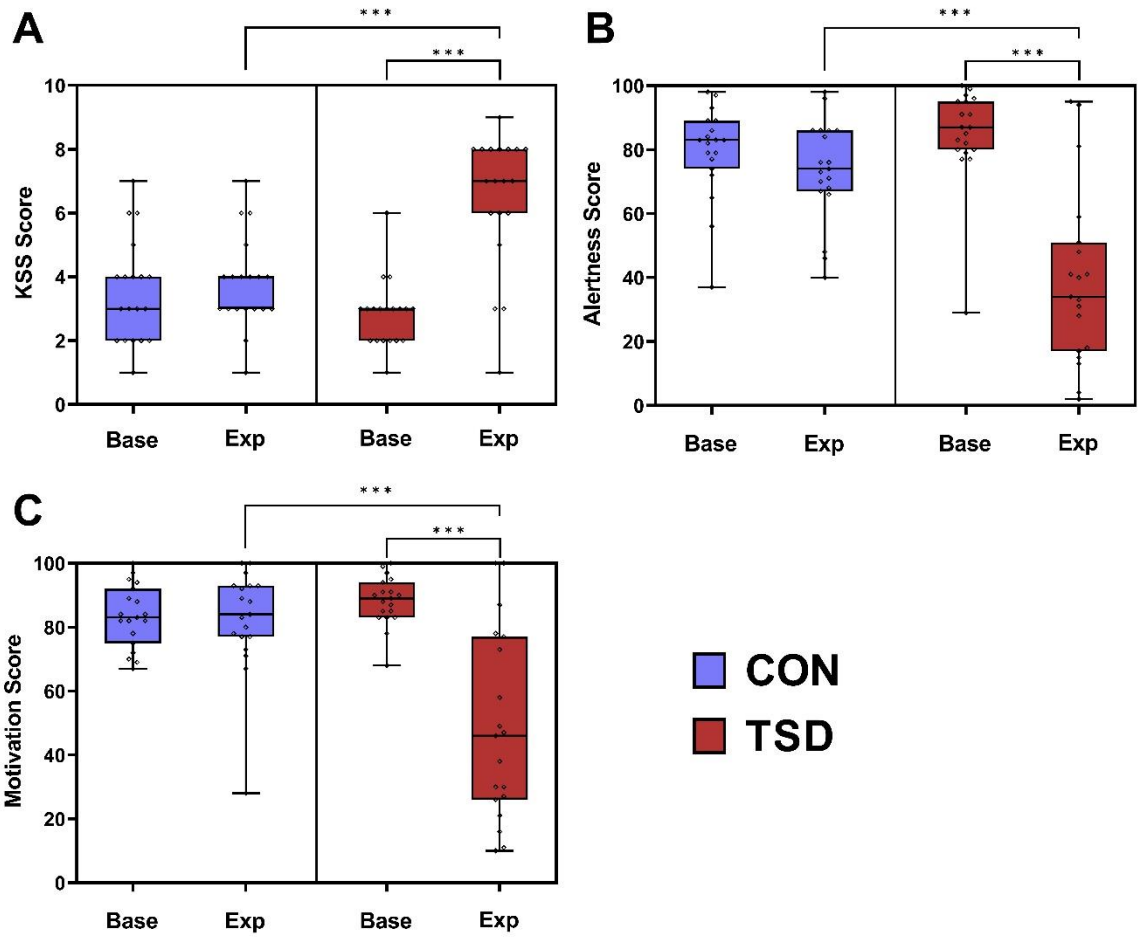
KSS, alertness VAS and motivation VAS scores for CON and TSD within both the baseline and experimental test sessions are shown in Figure 7-6.

KSS scores for CON participants did not differ between baseline ( $3.53 \pm 0.37$ ) and experimental ( $3.79 \pm 0.33$ ) sessions ( $p = 0.367$ ). Additionally, KSS scores obtained in the baseline test session did not differ between CON and TSD participants ( $p = 0.191$ ). However, significant differences were found between the KSS scores of TSD participants at baseline ( $2.84 \pm 0.24$ ) vs. experimental ( $6.47 \pm 0.49$ ) sessions ( $Z = -3.74$ ,  $p < 0.001$ ) and between CON and TSD participants on the experimental session ( $Z = -3.66$ ,  $p < 0.001$ ).

Alertness VAS scores for CON participants did not change between baseline ( $79.26 \pm 3.32$ ) and experimental ( $73.53 \pm 3.64$ ) sessions ( $p = 0.184$ ). Additionally, alertness VAS scores obtained in the baseline test session did not differ between CON and TSD participants ( $p = 0.130$ ). However, significant differences were found between the alertness VAS scores of TSD participants at baseline ( $84.74 \pm 3.56$ ) vs. experimental ( $39.21 \pm 6.30$ ) sessions ( $Z = -3.82$ ,  $p < 0.001$ ) and between CON and TSD participants on the experimental session ( $t(36) = 4.72$ ,  $p < 0.001$ , Hedges  $g = 1.50$ ).

Motivation VAS scores for CON participants did not change between baseline ( $83.32 \pm 2.25$ ) and experimental ( $82.26 \pm 3.76$ ) sessions ( $p = 0.948$ ). Additionally, alertness VAS scores obtained in the baseline test session did not differ between CON and TSD participants ( $p = 0.094$ ). However, significant differences were found between the alertness VAS scores of TSD participants at baseline ( $88.21 \pm 1.75$ ) vs. experimental ( $48.63 \pm 6.76$ ) sessions ( $t(18) = 6.75$ ,  $p < 0.001$ , Hedges  $g = 1.48$ ) and between CON and TSD participants on the experimental session ( $Z = -3.29$ ,  $p < 0.001$ ).

In summary, KSS, alertness VAS, or motivation VAS scores did not change except as a direct result of the sleep deprivation protocol. Within 10 minutes of the Rocket League matches commencing, participants who had undertaken the sleep deprivation protocol reported higher subjective sleepiness, and lower subjective alertness and motivation.



**Figure 7-6** Box and whisker (min → max) plots showing **A** KSS scores, **B** Alertness VAS scores, and **C** Motivation VAS scores (reverse scored) for CON and TSD participants in the baseline and experimental sessions.

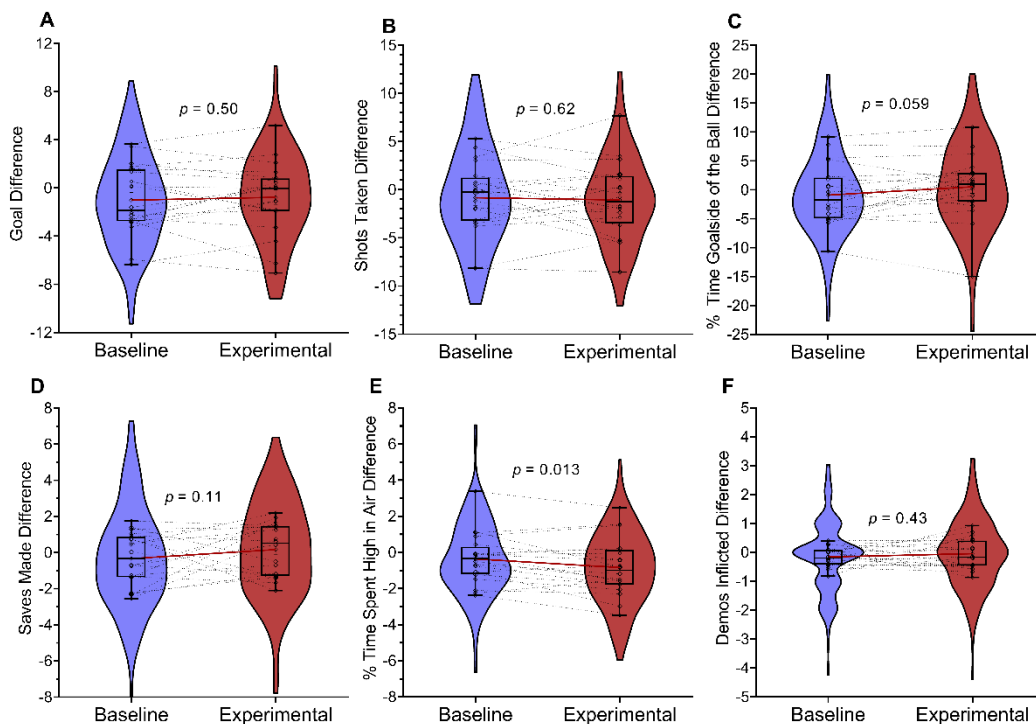
#### 7.4.5. In-Game Performance

To assess the effect of the TSD protocol on overall in-game performance, a model was created with GD as the outcome variable, session (baseline vs. experimental) as a fixed effect, and with a by-pair random intercept.

By-pair random intercept standard deviation was 2.61, and residual standard deviation was 2.83. The model intercept (corresponding to GD for the baseline session) was  $-1.01 \pm 0.63$ , which was not significantly different to 0 ( $t(1, 22.05) = -1.61, p = 0.12$ ), suggesting that neither group of participants were significantly better than the other at baseline. The effect of day (i.e., change from baseline to experimental) was not significant ( $\Delta GD = 0.23 \pm 0.34, t(1, 258.02) = 0.68, p = 0.498$ ), suggesting that the sleep deprivation protocol did not significantly impact GD. Figure 7-7A shows the distribution of GD for baseline and experimental sessions.

Similar models were created to assess whether the TSD protocol impacted any of the top five PIs within 1v1 Rocket League. The only PI to significantly change from baseline to test session was *Time Spent High in the Air Difference* ( $p = 0.013$ ) (Figure 7-7E), such that (compared to their opponent) the TSD individual spent  $0.48 \pm 0.19\%$  less time in the air in the experimental session, compared to the control session.

Lastly, participant's self-reported ratings for how much fatigue affected their in-game performance did not change between baseline ( $1.85 \pm 0.39$ ) and experimental ( $2.40 \pm 0.46$ ) sessions ( $p = 0.251$ ). Additionally, ratings obtained in the baseline test session did not differ between CON and TSD participants ( $p = 0.779$ ). However, significant differences were found for TSD participants at baseline ( $1.70 \pm 0.44$ ) vs. experimental ( $4.75 \pm 0.50$ ) sessions ( $Z = -3.33$ ,  $p < 0.001$ ) and between CON and TSD participants on the experimental session ( $Z = -2.96$ ,  $p = 0.003$ ). Participants in the CON group correctly guessed when their TSD group opponents were rested (*baseline*) 80% of the time, and when TSD group opponents were sleep deprived (*experimental*) 95% of the time. Participants in the TSD group correctly guessed that their CON group opponents were rested 85% of the time.



**Figure 7-7** Violin plots displaying the distribution of **A** GD and **B-F** exploratory PIs for *baseline* and *experimental* sessions, across all 279 matches. Box and whisker (min → max) plots inside the violin plots resemble the distribution of *mean* outcome variables across a test session for a given pair ( $N = 20$  for each box and whisker plot). Diamonds

represent pair means for each session, with pair means connected via the dotted lines. The solid red line represents the estimated mean $\pm$ SE from each model used for analysis.

## 7.5. Discussion:

The current study aimed to establish whether an acute bout of total sleep deprivation (TSD) decreased in-game performance in the popular esports *Rocket League*. We recruited 40 Rocket League players, pairing them based on expertise level, with half of the participants completing two test sessions while rested (CON) and the other half (TSD) completing one test session while rested (baseline), and the other test session following ~29 hours of TSD (experimental). Following this bout of TSD, we found these individuals to respond ~50msec slower and lapse (responses >500msec after stimulus onset) almost five times more often on the Psychomotor Vigilance Task (PVT). They also responded slower on the two-choice component of the Category Switch Task (CST), however, error rate on this component of the CST as well as Switch Cost and Mixing Cost (SC and MC respectively; measures of task-switching ability) response speeds were unchanged. Additionally, immediately (~10min) before Rocket League play, participants reported higher subjective sleepiness and lower subjective alertness and motivation when sleep deprived, when compared both to their own scores when well rested and compared to their paired opponents for Rocket League gameplay. Despite the cognitive impairments observed as a result of TSD, as well as the fact that participants felt that fatigue affected their in-game performance more following TSD, we did not find evidence that TSD impacted game outcome in Rocket League matches. The implications of our findings are discussed.

While we hypothesised that TSD would negatively impact our in-game esports outcome measure (GD), in line with the sentiment of previous articles (Bonnar, Castine, et al., 2019; Bonnar, Lee, et al., 2019; Bonnar et al., 2022; S. Lee et al., 2021; Sanz-Milone et al., 2021) and some esports athletes themselves (i.e., Baumann et al., 2022; Rudolf et al., 2020), we can identify (at least) four rational arguments for why such an effect was not found in the current study. Firstly, we note that not all aspects of cognitive performance are equally affected by sleep loss (Lim & Dinges, 2010; Lowe et al., 2017; Smithies, Toth, et al., 2021; Wickens et al., 2015), with the general trend being that as task complexity increases (for which, esports would be considered particularly complex), the magnitude of measurable adverse effect of sleep loss decreases (Harrison & Horne, 2000). Secondly (however relatedly), motivation (both intrinsic and extrinsic) appears to play an important role in the maintenance of performance (top-down mechanisms) in spite of sleep loss (Massar, Lim, & Huettel, 2019). As stated by Massar and Colleagues (p. 2), “In conditions in which incentives are high to perform, e.g., in military emergency situations, people

may be able to maintain performance. However, situations that do not contain significant extrinsic incentives may fail to generate sufficient motivation—and thus lead to reduced performance.” Our study involved the play of an esports highly familiar (and judging by total hours played, highly enjoyed) by participants in a highly competitive (and hence, motivating) environment. It is understood that task-specific factors can influence the degree to which performance occurs, specifically through promoting/ dissuading motivation. Regarding task complexity for example, Harrison and Horne (2000) (p.g. 236) state “the prevailing view in SD [sleep deprivation] research is that high-level complex skills are relatively unaffected by SD because of the interest they generate and the implicit encouragement for participants to apply compensatory effort to overcome their sleepiness”. Not only is Rocket League a highly cognitively complex activity, the play of Rocket League in the current study was undertaken in a set of circumstances which lends itself to compensatory mechanisms being activated. Thirdly, very repetitive tasks (often labelled monotonous) may experience greater performance loss due to persistent use of a very specific brain circuitry (Hudson et al., 2020), while tasks with greater stimulus/ response diversity (for which Rocket League very much fits) may not experience this effect. Fourthly (and again, relatedly), we note that the *time-on-task* effect (or vigilance decrement), which is accelerated and exaggerated by sleep loss (Doran et al., 2001), may not have been a factor within Rocket League gameplay. Rocket League matches are only ~6-7 minutes in length, and allow ~10 second breaks between each goal (occurring every ~40 seconds in the current study), allowing for frequent brief rest opportunities. This is not consistent among all esports. For example, major multiplayer online battle arena (MOBA) esports such as DOTA2 and LoL (the first and fourth largest esports by prize money earned; Esports Earnings (2023b), have average match lengths of ~20-30minutes (but can extend to >90 minutes) with very limited and unpredictable rest break opportunities. Lastly, while esports performance appears to be largely predicated on cognitive performance, there are a myriad of other factors involved, such as mood, biomechanics related factors, *playstyle* (individual differences in in-game abilities and strategy preferences) and interactions between competitor’s playstyles. Such factors could increase performance variation, confounding any expected effects of sleep loss.

With regards to the last argument, the authors argue that even if there are many extraneous factors in play, it is very unlikely that these factors would have completely nullified the effects of TSD on GD in the current study. A power analysis conducted *a priori* provided an estimated power of 0.829 for our analysis on GD, using estimated effect size and

variance measures, and an estimated MEM design. We observed variance that was larger than predicted (random intercept SD = 2.61 [predicted = 1.94], residual SD = 2.83 [predicted = 1.83]), however also found that our data did not (according to the procedure outlined by Matuschek et al. (2017)) warrant a model including a session by participant random slope. Retaining the estimated mean effect of TSD as a GD change of 1.218 from the *a priori* power analysis and including the variance measures and model structure from the results, an updated power analysis suggested that if the predicted mean effect magnitude was accurate, the power to detect it would have been 0.958 (the R script for this reanalysis can be found as appendix 7.3). Hence, we argue that our underestimation of variance is highly unlikely to be the root cause of the inability to reject the null hypothesis that TSD has no impact on in-game Rocket League performance, and that if an effect of ~29hrs of TSD on GD exists, the magnitude of this effect is most likely substantially smaller than anticipated.

It should be explicitly stated that our results do not suggest that sleep is a human factor to be disregarded within the world of esports. Sufficient sleep health is imperative for physical and mental wellbeing (Itani et al., 2017), and plays an instrumental role in memory consolidation (Stickgold, 2005; Walker & Stickgold, 2004); these are all critical factors when considering the everyday life of esports athletes and the downstream effects of sleep on competition performance. Furthermore, we certainly do not suggest that an acute bout of sleep deprivation does not impact alertness or cognitive performance, as our measures for such (as well as many decades of research; see Lim and Dinges (2010) for meta-analyses) are mostly in direct conflict with such a notion. What our results do suggest however, is that an acute bout of ~29 hours of sleep deprivation is unlikely to impact in-game esports performance to any measurable degree. This perhaps provides a positive message to esports players and coaches; that a night of poor sleep immediately prior competition is unlikely to adversely impact in-game performance. This message has high importance given some traditional athletes often experience sleep disturbances the night prior to competition (Juliff et al., 2015), with scholars suggesting that these disturbances are equally likely for esports athletes (Bonnar, Castine, et al., 2019).

In addition to the in-game outcome measure (GD), we gathered in-game data pertaining to a myriad of game-specific factors (i.e., offense/ defense, boost, movement & positioning; see Smithies, Campbell, et al. (2021) & **Chapter 5**) and explored whether established performance indicators (PIs) in 1v1 Rocket League (Smithies, Campbell, et al., 2021) varied as a function of sleep deprivation. Through this exploratory analysis, we

identified the PI *time spent high in the air difference* (TSD minus CON) to lower quite substantially (0.48%; average time spent high in the air across all matches in our sample = 3.13%) between *baseline* and *experimental* sessions. We also noted numerical increases in the *time spent goalside of the ball difference*, though with greater overall uncertainty as indicated by  $p = 0.059$ . These two PI's were specifically discussed by Smithies, Campbell, et al. (2021) as potentially indicating *safer* or *riskier* playstyles, with greater time *goalside of the ball* and less time *high in the air* resembling a *safer* overall playstyle. However, changes in these specific PIs could alternatively be argued to resemble an individual adopting an *easier* in-game strategy, as voyages *high in the air* typically involve much more difficult and precise movements, while staying grounded and *goalside of the ball* and relying primarily on counterattacking may present as a strategy requiring comparatively less effort than complicated attacking approaches. Sleep deprivation resulting in either (or both) *safer* or *simpler* decision making have theoretical support. While sleep loss is generally considered to result in *riskier* decision making (Satterfield & Killgore, 2019; Womack et al., 2013), decision making tasks (i.e., the Balloon Analog Risk Task or BART) typically show safer strategy employment following sleep deprivation (when 48hrs or less) (Killgore, 2007; Killgore et al., 2008), a trend mimicked by subjective risk-taking propensity following sleep deprivation of 48hrs or less (Chaumet et al., 2009; Killgore, 2007; Killgore et al., 2008). Interestingly, when discussing why sleep deprivation leads to a safer strategy on the BART but not other decision making tasks (i.e., Iowa Gambling Task), Satterfield and Killgore (2019) note that riskier decisions on the BART are also more effortful, and that “sleep deprived individuals appear to be less willing to expend effort to engage in risky activities.” (p. 353). Work by Engle-Friedman and colleagues (Engle-Friedman et al., 2010; Engle-Friedman et al., 2003) can be looked to for additional support for the notion that TSD evoked a *simpler* strategy among our participants. Despite these hypotheses regarding our observed PI differences, we emphasise the exploratory nature of this analysis and emphasise the need for more formal testing before claims can be substantiated. Nonetheless, we note this as an interesting line of future enquiry, especially given participants tended to *feel* (self-report) that fatigue affected their in-game performance following TSD.

Along with in-game performance, we also examined how sleep loss impacted the cognitive performance of esports players using the PVT and CST. As expected, overall response speed worsened, and the likelihood of lapses increased substantially (~5 times)

following TSD, indicating that players' vigilance was impaired. In the single task component of the CST, response speed slowing was similar in magnitude to that observed in the PVT, which was somewhat surprising given previous research (i.e., Smithies, Toth, et al., 2021) suggested the level of impairment decreases as task complexity increases so long as cognitive flexibility is not introduced as a task requirement. Even more surprisingly, we found no evidence for decreases in task-switching ability (measured both by SC and MC response time), contrasting findings by Couyoumdjian et al. (2010). However, these findings were consistent with Nakashima et al. (2018), who noted no change in SC reaction time following TSD, and somewhat in agreement with Slama et al. (2018) who noted a change in SC accuracy but not SC reaction time following TSD. Overall, these contrasting findings may be a result of subtle task-characteristics (i.e., types of stimuli, stimulus-response mappings, interstimulus intervals, frequency of task-switches) which warrant further investigation. The ability to rapidly switch ones' attention between multiple information sources appears integral to esports performance (in the context of Rocket League, this could be the switching of visual attention between the ball, the opponent, the players vehicle and the players boost meter), as demonstrated by improvements in task-switching ability coinciding with the play of action video games (Nuyens et al., 2019; Toth et al., 2020). Although the results of our study suggest that this ability may be unaffected by acute sleep loss for esports athletes, we interpret with caution given mixed findings in the literature.

### **7.5.1. Limitations**

We outline several limitations regarding the presented experimental study. Firstly, despite our best efforts, we had only one female participant in our final sample of 40, resulting in a clear sex imbalance. Similar difficulties recruiting female esports players have been previously noted (Bonnar et al., 2022; Ratan et al., 2015). We also note that this large gender imbalance is (regrettably) reflective of *elite* esports demographics (with estimates of only 5% of professional esports athletes being female; Hilbert (2019)), a disparity actively highlighted in many articles (Darvin et al., 2021; Taylor & Stout, 2020). Secondly, we note a lesser degree of control over factors such as participant demographics, sleep, and weekly gameplay than desirable. We note the extreme difficulty in recruiting participants sufficiently experienced with Rocket League while also fulfilling the somewhat strict criteria for healthy sleeping participants outlined by Beattie et al. (2015). Due to this, affordances were made to the inclusion criteria and as such, our participant pool included some individuals screening at risk for sleep disorders, and one

participant with habitual caffeine use > 400mg. We found that many participants reported great difficulty maintaining a regular sleep and wake time in particular, as resulting in some participants experiencing less than desired TST (i.e. ~3hrs) in the 1-3 nights prior to testing, even when meant to be *well rested* (see Figure 7-4). This is noteworthy and we would encourage future research to explore sleep/ wake variability within habitual esports playing populations, potentially through use of the sleep regularity index (SRI; Phillips et al. (2017)), as done recently in elite team sport athletes by Halson et al. (2022). Continuing the topic of sleep and wake times, implementing somewhat standardised target bed (22:00 – 01:00) and wake times (06:00 – 09:00) that were outside of those habitually experienced by some participants may have affected the sleep experienced by some participants within the three nights prior to test sessions, potentially leading to the higher-than-desired variability in TST on these days within control conditions. However, we note that this research design decision was made to better comply with the RRDC criteria (Beattie et al., 2015) and for logistical reasons (i.e. some participants habitual wake times could interfere with test session availability). We also caution that the results of the current study may have limited applicability to other esports besides Rocket League. Although most major esports share a lot of similarities (fast and accurate responses to rapidly changing stimuli executed through fast and precise fine motor movements, complex interactions with other individuals, use of computer peripheries, seated environment etc.), their diversity has resulted in some observed differences in the relevant importance of specific cognitive abilities (Dobrowolski et al., 2015; Toth, Conroy, et al., 2021), and hence potentially, diversity in the impact of acute sleep loss. Also as previously mentioned, Rocket League has short match lengths with frequent break opportunities when compared to other esports, which may lend to a lesser ability of the *time-on-task* effect (which sleep loss accelerates and exaggerates) to negatively impact performance. We note however that the short and predictable match lengths within Rocket League are also one of the key characteristics which make it a feasible esports to conduct experimental research on (as it allows for multiple and consistent amounts of trials per test session (Smithies, Campbell, et al., 2021; see **Chapter 5**)). Nonetheless, generalising the results of the current study to other esports should be done with caution. Lastly, we note that our subjective measure of motivation (motivation VAS; as per Mathew et al. (2021)) was not suitably timed or worded to appropriately capture participants motivation to perform within the Rocket League gameplay. Had this item been implemented immediately following either the warm-up provided or in between the matches played, it

may have been able to shed light on whether this mechanism may have played a role in performance preservation in spite of sleep loss.

### **7.5.2. Conclusion**

Overall, the results of our study suggest that an acute bout of sleep loss (~29hrs TSD) does not adversely impact in-game Rocket League performance, despite degrading vigilance and attentional capabilities as measured by both subjective and objective instruments. Our findings suggest that efforts may be better placed optimising day-to-day sleep health, as opposed to austere avoidance of sleep loss immediately prior to competition.

**Chapter 8.      The BART effect: Rocket League players  
appear to play both simpler and safer when sleep deprived**

“There’s a 4:30 in the morning now?” – Bart Simpson (The Simpsons; S6, E1)

## 8.1. Introduction:

In the previous chapter, I examined the effects of ~29hours of total sleep deprivation (TSD) on in-game performance, in the esport Rocket League. While I did not observe an effect of the TSD protocol on match outcome, I performed exploratory analysis to see whether any in-game performance indicators (PIs; as identified by Smithies, Campbell, et al. (2021) and in **Chapter 5**) differed as a result of TSD.

Through this analysis, I observed a 15.34% decrease in the PI *Time Spent High in the Air Difference* (from 3.13% of match to 2.64% of match), and (a less certain,  $p = 0.059$ ) 1.89% increase in the PI *Time Spent Goalside of the Ball Difference*. In the discussion of this chapter, I stated that these changes in PI values could be reminiscent of a shift toward a *safer* (vs. riskier) in-game strategy, or alternatively a *simpler* (vs. more complex) strategy (though I was careful to avoid making strong claims about this evidence given the exploratory nature of the analyses). Evidence from previous literature suggesting either (or both) of these changes to be feasible due to sleep loss is discussed below.

Regarding a *simpler* playstyle, there is a substantial body of literature suggesting that under conditions of sleep deprivation, tasks with high cognitive demands are perceived as more effortful (see Massar, Lim and Huettel (2019) for an overview). A 2003 article outlines multiple experiments in which sleep deprived university students, when provided a choice between easier and more difficult math questions, tended to choose easier questions than when well rested or when vs. well rested counterparts, with the authors concluding that “These studies demonstrate that sleep loss results in the choice of low-effort behaviour that helps maintain accurate responding.” (p. 113, Engle-Friedman et al. (2003)). Another study with competitive adolescent ice-skaters found that those who slept less perceived relevant skating-specific manoeuvres as more difficult, while those with greater sleep disturbances (awakening count & wake after sleep onset) were more likely to choose easier manoeuvres to perform (Engle-Friedman et al., 2010). These are two nice examples suggesting that when given options between *easier* and *more difficult* alternatives, sleep loss drives individuals toward the easier alternative, likely due to an increase in perceived task effort demands when sleep deprived (Massar, Lim, & Huettel, 2019).

Regarding a *safer* playstyle, the evidence base is substantially more conflicting. While sleep deprivation of 48hrs or less has been demonstrated to result in self-reported reductions in risk-taking (Chaumet et al., 2009; Killgore, 2007; Killgore et al., 2008), the

current body of experimental work suggests that sleep deprivation tends to lead to riskier decision making in practice (for reviews, see Satterfield and Killgore (2019) and Womack et al. (2013)). However, there is one decision making paradigm called the Balloon Analog Risk Task (BART) which typically shows sleep deprived individuals adopting *safer* strategies compared to well rested counterparts (Killgore, 2007; Killgore et al., 2008). In discussing why sleep deprived individuals adopt safer strategies on the BART but riskier strategies on other decision making paradigms (like the Iowa Gambling Task or IGT), Satterfield and Killgore (2019) propose that while riskier decisions require no more effort on the IGT, they do on the BART. This suggests that a *safe* strategy and *less effortful* strategy are equivalent on the BART. In a similar vein, I proposed in **Chapter 7** that spending less time *high in the air* (and to a lesser extent, more time *goalside of the ball*) is both *safer* and *simpler* in 1v1 Rocket League.

However, I note that the link between the mentioned PIs and both *safe vs. risky* and *simple vs. complex* was established by one author (TDS), albeit with extensive familiarity with the esports (~2,800 hours played) and in-game metrics. In order to instil more confidence in the interpretation of PIs and their relationship with in-game strategy, I see great value in gaining the opinion of *field experts* within the given esports; that is, former professional players, coaches, analysts, and/ or casters who possess significant knowledge and experience of Rocket League and its in-game metrics.

Hence, the current study aimed to establish a stronger understanding of which Rocket League in-game metrics best differentiate both *safe vs. risky* and *simple vs. complex* playstyles. Using this understanding, I then aimed to explore whether ~29 hours of TSD resulted in a playstyle perceived to be more *safe*, more *simple*, or both. I hypothesised that there would be a large amount of overlap between in-game metrics that distinguish *playstyle risk* and *playstyle complexity*, and as such, the TSD protocol would lead to changes in playstyle perceived as *safe* and *simple*.

## 8.2. Methods

All procedures and data collection was approved by the Education and Health Sciences Research Ethics Committee (2021\_06\_13\_EHS) and conducted in accordance with The Declaration of Helsinki.

### 8.2.1. Participants

Eleven individuals were contacted to participate in the current study, to which nine responded, and seven provided informed consent and participated. Participants were sought on the basis of their prior playing/ coaching/ casting/ analyst experience, and were contacted through email or through social media (i.e., Twitter, Discord). Each participant had significant experience playing, coaching, analysing or casting Rocket League, and expressed familiarity with the in-game metrics provided through ballchasing.com, and hence could be considered *field experts*. The expertise of each participating individual is given below:

Participant 1: Two-years of experience casting professional level Rocket League (predominantly 1v1; >500 matches casted). Top <~2% (of total player base) player.

Participant 2: Two-years of coaching Rocket League (predominantly 3v3) focussing on individuals between the top ~0.33-5% of total playerbase. Has achieved top 0.2% of playerbase in 1v1 Rocket League.

Participant 3: Retired professional Rocket League player (7 years, >\$80,000USD winnings). Current professional coach.

Participant 4: Former coach of world championship winning Rocket League team, current analyst.

Participant 5: Retired professional Rocket League player (3 years, >\$10,000USD winnings). Four-years of casting professional level Rocket League (including 1v1).

Participant 6: Current professional Rocket League player (3 years, ~\$200,000USD winnings), international tournament winner. One year of coaching experience at all levels. Frequently ranked number 1 in 1v1 Rocket League.

Participant 7: Seven years of high-level Rocket League gameplay, including being top 0.2% of playerbase in 1v1 Rocket League since 2021, scrimmaging at a semi-professional

level, and substituting professionally. Experience coaching at international collegiate level, and casting/ analyst roles at a national level.

### 8.2.2. Instruments & Procedure

Participants completed a survey instrument via. Qualtrics. A copy of this survey can be found as appendix 8.1.

The survey instrument included two questions. For each question, participants were provided all 26 in-game Rocket League metrics provided by ballchasing.com (see appendix 4.2, and note the removal of Shots Conceded and Demos Taken as difference scores were to be used within the analysis), along with the capacity to rank-order them according to the question. The two questions requested were worded in the following manner:

The following question is with regards to Rocket League player(s) within a 1v1 match, who are between 500MMR (i.e. gold) to 1300MMR (i.e. high GC3/ low SSL [supersonic legend, the highest rank in Rocket League as of 22/09/2020]).

From the options below, please rank the in-game metrics that you feel would best discriminate between a player playing a strategy which is [question 1; 'safer' or 'riskier', question 2; 'more simple' or 'more complex']. **Note that we only require the top 10**, with 1 being the metric that best discriminates. **Feel free to ignore the ordering below 10**

### 8.2.3. Data Processing and Statistical Analysis

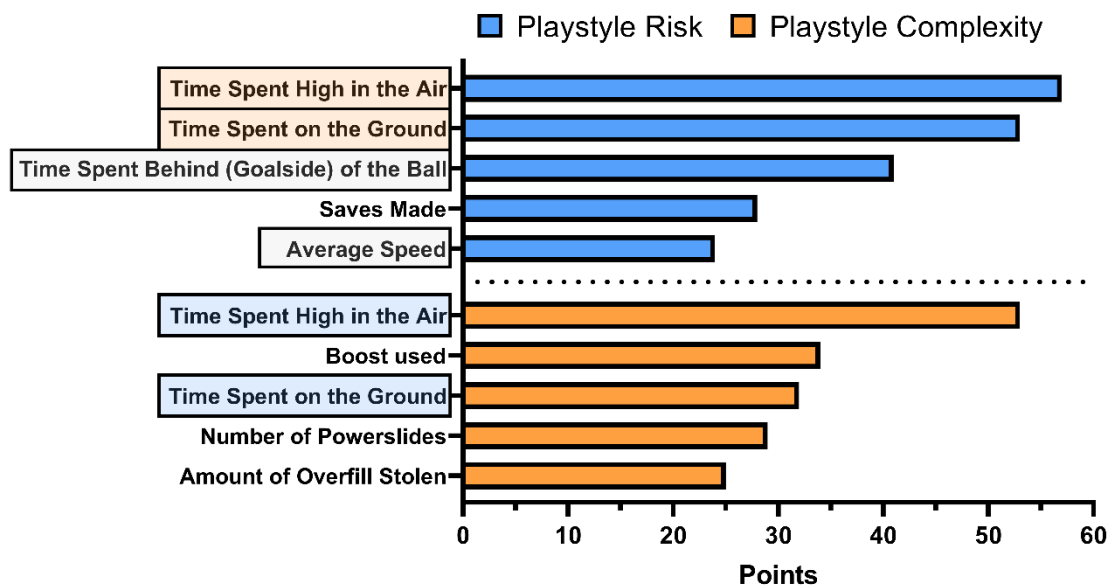
For each question (called *playstyle risk* and *playstyle complexity* from here onwards), metrics chosen by each individual were assigned points according to their ranking, in a reverse scored manner (such that a metric ranked number 1 was given 10 points, and a metric ranked 10 was assigned 1 point). The five metrics which received the most points (per question) were considered the metrics that best distinguish (a) *playstyle risk* and (b) *playstyle complexity*.

The mean difference scores (TSD participant – CON participant, see **Chapter 5&7**) obtained per pair and per test session for (a) *playstyle risk* and (b) *playstyle complexity* metrics, were subsequently inputted into two separate repeated-measures multivariate analyses of variance (MANOVAs), with session (baseline vs. experimental) as the independent variable. This analysis was performed using the MANOVA function in IBM SPSS Statistics v28 (Armonk, N.Y.) software. MANOVA uses all inputted dependent variables to create a canonically derived dependent variable (called *playstyle variable* hereafter) which maximally discriminates between groups in the independent variable (in my case, baseline and experimental test sessions). It is an appropriate analytical technique to analyses conceptually interrelated variables together (as opposed to conceptually unrelated variables; Huberty and Morris (1992)), as is present in the current study. I provided descriptive information and performed a follow-up analysis on these *playstyle variables* (a process called descriptive discriminant analysis or DDA) to examine differences between sessions, using paired-samples t-tests (as per Enders, 2003). Data and syntax used in this analysis can be found at <https://osf.io/z2fjg/>. Alpha was set as  $p < 0.05$  unless otherwise specified.

## 8.3. Results

### 8.3.1. Metrics Relevant to Playstyles

The metrics that best distinguish *playstyle safety* and *playstyle complexity* in 1v1 Rocket League according to *field experts* are shown in Figure 8-1. I note a considerable overlap between metrics considered to best distinguish between the two playstyles, consistent with the notion that a simple playstyle is analogous to a safe playstyle in 1v1 Rocket League.



**Figure 8-1.** The five in-game metrics that received the most points, corresponding to *field expert* determined ability to distinguish *playstyle risk* and *playstyle complexity* in 1v1 Rocket League. Metrics labels highlighted in blue or orange are metrics that are also a top five distinguisher of the other playstyle, which metrics labels highlighted in grey are metrics that are also a top ten distinguisher of the other playstyle

### 8.3.2. MANOVA Analysis

For *playstyle risk* metrics, a statistically significant MANOVA effect for session (baseline vs. experimental) was found (Hotelling's  $T^2 = 1.15$ ,  $F(5, 15) = 3.46$ ,  $p = 0.028$ ,  $\eta_p^2 = 0.54$ ), such that 54% of the variance in the *playstyle variable* could be accounted for by session (baseline vs. experimental). For *playstyle complexity* metrics, a statistically significant MANOVA effect for session (baseline vs. experimental) was found (Hotelling's  $T^2 =$

1.03,  $F(5, 15) = 3.09$ ,  $p = 0.041$ ,  $\eta_p^2 = 0.51$ ), such that 51% of the variance in the *playstyle variable* could be accounted for by session (baseline vs. experimental). Exploratory univariate post-hoc t-tests of PIs included in each MANOVA found only the PI *Time Spent High in the Air Difference* to significantly change from baseline to experimental sessions ( $\Delta$  value =  $-0.47 \pm 0.18$ ,  $F(1, 19) = 6.99$ ,  $p = 0.016$ , 95% CI  $[-0.84, -0.10]$ ,  $\eta_p^2 = 0.27$ ), corroborating findings in **Chapter 7** and suggesting that session alone (i.e., ~29hrs TSD) accounts for 27% of variance within this PI.

The eigenvalue, standardised discrimination function coefficients (SDFC), and correlation between these coefficients and the *playstyle variable* (structure coefficient) can be found in Table 8-1. The former gives a measure of the relative contribution from each metric to the linear equation which creates the *playstyle variable*, while the latter gives a measure of the *actual* relationship between the metrics and the *playstyle variable* (K. N. Smith et al., 2019). Metrics with a large |SDFC| but a small structure coefficient (i.e., *Time Spent On the Ground Difference* in both models, and *Overfill Stolen Difference* in the playstyle complexity model) are suppressor variables, meaning that while they have little to no relationship with the *playstyle variable* alone, they function to strengthen the relationship between other metrics and the *playstyle variable*.

Paired-sample t-tests were performed to test the magnitude of effect of session (baseline vs. experimental) on each *playstyle variable*, as per Enders (2003). A conservative alpha of  $p < 0.001$  was used for these analyses, due to differences in sampling distribution between univariate variables and canonically derived variables (Neufeld & Gardner, 1990). Nonetheless, I found significant between session (baseline  $\rightarrow$  experimental) effects on both the *risk playstyle variable* ( $\Delta$  value =  $1.48 \pm 0.32$ ,  $t(1, 19) = 4.68$ ,  $p = 0.0002$ , 95% CI  $[0.82, 2.14]$ ,  $g = 1.03$ ) and *complexity playstyle variable* ( $\Delta$  value =  $1.40 \pm 0.32$ ,  $t(1, 19) = 4.42$ ,  $p = 0.0003$ , 95% CI  $[0.74, 2.06]$ ,  $g = 0.97$ ).

**Table 8-1** Eigenvalues, canonical correlation, standard discrimination function coefficients (SDFC) and structure coefficients, for metrics (note that all are considered as difference scores) that distinguish *playstyle risk* and *playstyle complexity* in 1v1 Rocket League. The linear equation for the *playstyle variables* is given at the bottom of the table using raw discrimination function coefficients.

Playstyle Risk			Playstyle Complexity		
		Eigenvalue	Canonical Correlation		
		1.15	0.73		
In-Game Metrics	SDFC	Structure Coefficient	In-Game Metrics	SDFC	Structure Coefficient
A. Time Spent High in the Air	1.10	0.56	A. Time Spent High in the Air	0.74	0.60
B. Time Spent On the Ground	0.56	-0.04	B. Boost Used	1.40	0.43
C. Time Goalside of the Ball	-0.87	-0.39	C. Time Spent On the Ground	0.49	-0.04
D. Saves	-0.24	-0.26	D. Number of Powerslides	-0.26	-0.12
E. Average Speed	0.01	0.22	E. Overfill Stolen	-0.94	0.06
$Playstyle\ Variable = 1.95A + 0.32B - 0.36C - 0.19D - 0.0003E$			$Playstyle\ Variable = 1.31A + 0.05B + 0.28C - 0.03D - 0.02E$		

## 8.4. Discussion

The aim of the further analysis undertaken in this chapter was to follow-up on exploratory analysis undertaken in **Chapter 7**, which indicated that TSD may have led Rocket League players to employ a *safer* or *simpler* (or both) playstyle. Specifically, I sought the opinions of *field experts* to uncover the metrics perceived to best distinguish *playstyle risk* and *playstyle complexity* for 1v1 Rocket League players within the bounds of those studied in **Chapter 7**. Using these metrics in a multivariate approach, I found that ~29hr TSD could account for ~54% of variance in the canonically derived *risk playstyle variable* and ~51% of variance in the canonically derived *complexity playstyle variable*, with follow-up analyses showing both of these variables to be highly sensitive to sleep loss.

The first standout observation is that many of the metrics which the *field experts* felt best distinguished *playstyle risk* were the same that distinguished *playstyle complexity*. Of the five metrics determined to best distinguish between each playstyle, two were present in both; *Time Spent on the Ground*, and *Time Spent High in the Air*. This is notable when considering that (a) *Time Spent High in the Air Difference* was rated the best metric to distinguish both *playstyle risk* and *playstyle complexity*, *Time Spent High in the Air Difference* is a PI, with higher values being generally associated with better match outcome (**Chapter 5**), and that *Time Spent High in the Air Difference* was the only PI that changed as a result of ~29hrs of TSD, within the univariate exploratory analysis outlined in **Chapter 7**. I also note that two of the five (and six of the top ten) metrics that best distinguish *playstyle risk* were present in the top ten metrics that distinguish *playstyle complexity*. This provides weight to the argument that within the context of 1v1 Rocket League, a subjectively defined safe playstyle is (at least to a moderate degree) analogous to a subjectively defined simple playstyle, and is best demarked (if only considering a single metric) by *Time Spent High in the Air Difference*.

Additionally, MANOVAs produced by the five most agreed-upon distinguishing metrics for *playstyle risk* and *complexity* both produced *playstyle variables* that varied by a similar amount as a function of ~29hr TSD (~29% and ~26% respectively), and for which, values obtained for baseline and experimental sessions were highly significantly different. It appears that the slightly larger variance value for the *risk playstyle variable* was driven primarily by the inclusion of the *Time Goalside of the Ball Difference* metric (having the second largest structure coefficient), which was the equal sixth best discriminator of *playstyle complexity* according to the *field experts*, and was highlighted

in **Chapter 7** as a metric that likely resembles *both* a safer and simple playstyle. Conversely, the second largest contributor to the *complexity playstyle variable* was *Boost Used Difference*; a metric which only received five points from *field experts* and was rated 18 out of 26 in-game metrics for determining *playstyle risk*. This could suggest that ~29hrs TSD leads to a *simpler playstyle* employed, which mostly (but not completely) equates to a *safer playstyle* within the context of 1v1 Rocket League. However, I stress the tentative nature of this interpretation given the non-significant univariate post-hoc t-test on this metric ( $p = 0.071$ ).

My analysis presented here appears to support the idea that ~29hr TSD led players to employ a *safer* playstyle in 1v1 Rocket League, however with the caveat that this *safer* playstyle was also a *simpler* playstyle. Drawing an analogy between playstyle complexity and effort required, my results are in line with studies by Killgore (2007) and Killgore et al. (2008) using the Balloon Analog Risk Task (BART), and observations from Satterfield and Killgore (2019) that risk-taking (while appearing to generally increase following acute sleep loss) actually decreases when greater risk is also associated with greater effort/complexity. Riskier plays in 1v1 Rocket League (best exemplified by taking the ball high in the air on attack) appear to be inherently more complex/ difficult, as shown by *field expert* opinion and supported by similar discriminability of the derived *playstyle variables*.

### 8.4.1. Limitations & Future Research

I stress that the analysis presented in this chapter is based on data discussed in **Chapter 7** and not new data, and as such, the analysis remains exploratory rather than confirmatory. Furthermore, I note that *playstyle variables* are not variables derived from a shared variance/ correlation (such as the output of a factor analysis for example), but rather are a variable that, taking five *field expert* chosen metrics, weight the metrics in such a way to maximise discrimination between rested and sleep deprived gameplay. This approach was chosen due to understanding that *safe* or *simple* playstyles may manifest in multiple different (and potentially uncorrelated) ways. In other words, there may be multiple ways to play *safely* or *more simple*, and these ways may not be correlated with one another. An alternative data-driven approach could have been to use factor analysis to determine playstyle variables derived from highly correlated variables, and performing paired-samples t-tests on the differences between sessions on these variables. However, sensible interpretation of derived variables from such an approach may prove difficult.

Here, this problem of interpretability is circumvented by directly asking *field experts* which metrics best distinguish *safe vs. risky* and *simple vs. complex* playstyles.

The analysis provided showcases that the differences observed following ~29hr TSD appear to not be based on random fluctuation in an individual in-game metric, but instead are more representative of a fundamental change in the way that Rocket League players played the game while sleep deprived. In other words, should differences found be actually attributable to chance, it would be that players just happened to play safer and simpler Rocket League when they were sleep deprived, rather than just happening to go high in the air less. This potential *safer and simpler approach when sleep deprived* idea (to which I coin ‘the BART Effect’) presents as an interesting line of enquiry for future research within Rocket League but also within other esports, perhaps considering varying levels of in-game expertise as well.



The ultimate aim of the work presented in this thesis was to illuminate how sleep loss impacts the in-game ability of esports athletes. This avenue of enquiry was explored for three key reasons. Firstly, the negative effects of sleep loss on performance within traditional sports has been a burgeoning research field since the 1980s, as athletes, coaches, and stakeholders seek to examine and understand the human factors which may contribute to success in their given sport. However, esports performance is more predicated on cognitive performance than virtually all traditional sports (Campbell et al., 2018), and sleep loss tends to cause larger and more robust deficits in cognitive rather than physical performance. Secondly, much of the seminal research regarding sleep and esports has discussed sleep loss with respect to its hypothesised detrimental effect on performance, due to the abovementioned assumptions (Bonnar, Castine, et al., 2019; Bonnar, Lee, et al., 2019; Bonnar et al., 2022; S. Lee et al., 2021; Sanz-Milone et al., 2021). Lastly and most importantly, the relationship between sleep loss and esports performance has not been studied within any capacity<sup>†</sup> beyond exploring associations between habitual sleep and in-game outcome measures in an uncontrolled setting and against unknown opponents (i.e., online *ranked* matches; Moen et al. (2022)).

A total sleep deprivation (TSD) study, detailed in **Chapter 7**, primarily addressed the aim of illuminating how sleep loss impacts the in-game ability of esports athletes. Within this study, I subjected twenty individuals to ~29hrs of experimentally controlled TSD, before playing seven matches of the esports *Rocket League* against well-rested peers of a similar *in-game* expertise level. Seven matches were also played between both players while well rested, allowing comparison of performance under such circumstances. Immediately prior to Rocket League matches taking place, participants (compared to when well-rested) reported increased subjective sleepiness, decreased subjective alertness and motivation, and decreased performance both on a *low-salience* vigilance taxing task (PVT) and a *high-salience-stable* single-cued component of the Category Switch Task (CST); all as a direct result of the TSD protocol. Despite this, I found no evidence of in-game Rocket League performance degrading as a function of the TSD bout which caused clear subjective and objective impairment on other measures.

<sup>†</sup> This statement could be argued as technically untrue if one was to consider the use of *Tetris score* as measure of esports performance; for which I note one study (Kariv et al., 2007) that used it as a general measure of the cognitive performance of physicians before and after a night shift.

## 9.1. Assessing Performance In-Situ

A feature which presents as a major strength of the work in this thesis is the unique approach I used to directly measure in-game performance. In my experimental protocol, I measured performance during a live competition, matching the environment and context experienced in actual esports competition (i.e., *in-situ*). Such an approach is largely (but not completely; see Fox et al. (2021) and Staunton et al. (2017) as sleep-related examples) avoided in studies of human factors contributing to team-based traditional sport performance, in favour of performance measurement within *proxies*. Such performance proxies are examined in environments/ with procedures including *affordances*, in an effort to reduce extraneous variables inducing high amounts of uncontrollable variance (see Araújo and Davids (2015) for further discussion). However, such affordances can result in task environments/ stimuli with varied (and often unclear) correspondence to environments/ stimuli experienced within esports competition. In simple terms, performance changes within proxies may not reflect performance changes during actual esports competition (*in-situ*). The esports performance assessment used in this thesis avoided this issue altogether, as performance was explicitly measured in-competition. Additionally, participants were unconstrained regarding preferences such as their in-game settings, input modalities, and even their ability to play music in the background during gameplay (a common practice for esports athletes). Essentially, efforts were made to assess performance within a testing environment and set of circumstances that bore a level of ecological validity<sup>†</sup> surpassing that which is generally present in traditional sport performance research.

Such environmental/ circumstantial factors bare additional relevance within sleep loss literature. How engaging the particular task is, or how motivated participants are to perform optimally, are known to play a non-trivial role in whether (or to what extent) sleep loss actually impacts performance. By measuring performance in a live competition environment, I avoided the potential of finding an effect that does not translate to live competition, where high levels of task engagement/ motivation (and hence, propensity for compensatory mechanisms) is inherently present.

<sup>†</sup> When ecological validity is mentioned throughout the discussion, it is with reference to Orne's definition of ecological validity; the generalisability of experimentally obtained findings to a real-world context, or to the context for which the results directly apply to. This is sometimes called representative design (Araújo and Davids, 2015), and is as opposed to Brunswik's original definition of ecological validity, being very specifically the correlation between a proximal cue and a distal object. Discussion around the distinction between these two definitions is provided by Kihlstrom (2021).

Not only did my esports performance analysis bare high levels of ecological validity, but by using an outcome variable which was practically analogous<sup>†</sup> to victory, the practical relevance of findings was maximised. After all, consideration of human factors within esports (from a performance perspective) stems from whether it influences match outcome. It can be argued that acute sleep loss immediately before competition is not a factor worth considering in practice (despite its demonstrated adverse effects on some tasks, and even if it changes *the way* esports athletes play) if ultimately, the end result does not change.

## 9.2. Why *didn't* sleep deprivation lead to worsened performance?

That I was not able to statistically detect a performance difference following ~29hrs TSD was surprising and contrary to my hypothesis. It also runs contrary to the expectations outlined in much of the previous sleep and esports literature (Bonnar, Castine, et al., 2019; Bonnar, Lee, et al., 2019; Bonnar et al., 2022; S. Lee et al., 2021; Sanz-Milone et al., 2021), as well as the assumptions of some esports athletes themselves (Baumann et al., 2022; Rudolf et al., 2020).

While the measurement of performance in live competition clearly presents with many benefits from an ecological validity perspective, there are valid reasons for its infrequent use in traditional sport research. Such concerns are outlined by Araújo and Davids, who note that the study of motor behaviours strictly within competitive performance environments is “clearly not possible, nor desirable, due to the presence of irrelevant idiosyncrasies of specific competitive events which might contaminate data” (Araújo & Davids, 2015, p. 269). Indeed, a similar sentiment was raised by a reviewer for an article I was also peer-reviewing, which, like I did, used in-game esports performance as their primary outcome metric. This reviewer raised the point that variability increases by using a design in which players play against other players, and raised the question to authors of whether they can be confident that effects are attributable to their given intervention and not random fluctuation.

A major strength of the analytical approach taken is that this very same question could be directly addressed. Using the observed levels of between-pair and residual variance, the appropriate level of random effect complexity (or variance components) that was justified by the data (according to procedures outlined by Matuschek et al. (2017)), and the estimated

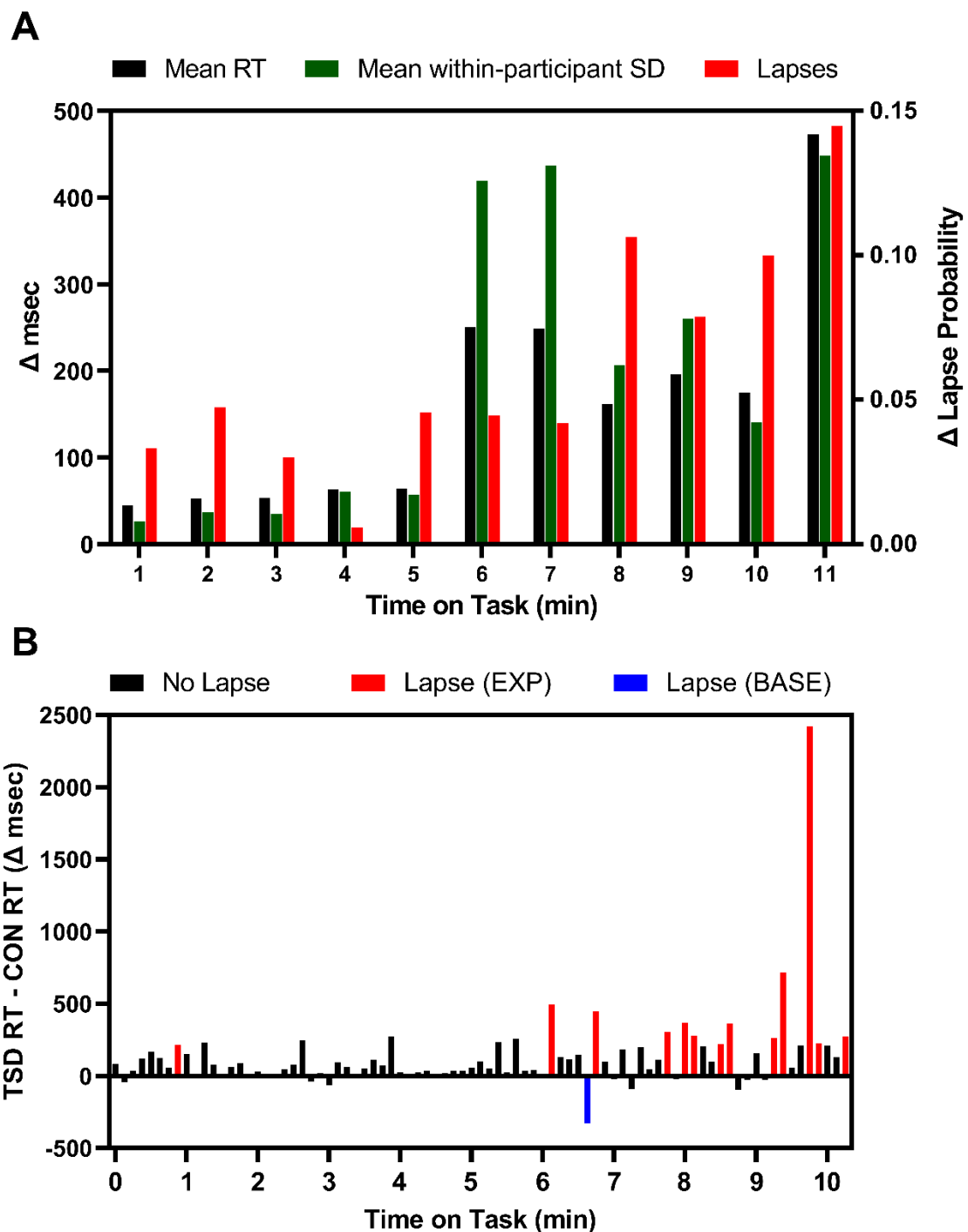
<sup>†</sup> While win vs. loss can be argued to be the ultimate game objective within Rocket League, the sign of GD always matches with win vs. loss except when a player forfeits while ahead, which was not allowed within the experimental work. This, combined with specific instructions provided to participants to maximise effort in scoring goals and preventing opponents score irrespective of current score within a given game, resulted in GD being analogous to victory while providing additional information on the closeness of the game.

magnitude of effect from my initial power analysis (a mean difference of 1.218 goals per match attributable to the ~29hrs of TSD), a power of 0.96 was observed. According to this power re-estimation, if the true effect of ~29hrs of TSD was in fact a mean decrease in GS of 1.218, it should be observed as a significant ( $p < 0.05$ ) effect ~96% of the time. This presents as quite convincing evidence that the null finding of sleep deprivation on in-game esports performance was not simply due to too much noise, and that the true effect of ~29hr TSD on performance, if present at all, is much smaller than anticipated.

If esports performance is predicated on cognitive factors which are consistently shown to be impacted by sleep loss, why *didn't* sleep loss cause performance deficits in my study? Although potential answers to this question are discussed somewhat in **Chapter 7**, it is interesting to revisit this question specifically with respect to mechanisms of sleep loss performance deficits, outlined in the **Section 1.6**; starting with the effects of sleep loss on the *time-on-task effect*.

The time-on-task effect refers to the increase in response time (and variability in response times) during a task requiring sustained attention (Dinges & Powell, 1988, 1989; Doran et al., 2001). It is understood that sleep loss expedites and exacerbates the time-on-task effect. This is pleasingly visible within the PVT data of participants who completed testing both when rested and following ~29hrs of sleep deprivation in my research (TSD participants), when looking at performance across the timespan of the PVT administrations. Figure 9A shows the difference in mean reaction time (msec), average *within-participant* standard deviation of reaction time (a measure of response variability), and lapse likelihood, between PVT administrations while sleep deprived and well rested, separated into one-minute epochs across the 11 minutes of the PVT (this included the first minute, which was excluded for analytical purposes in **Chapter 7**). Firstly, participant responses were slower and more variable, with higher lapse propensity, following sleep deprivation at all time points of the PVT, showing that the time-on-task effect does not explain the entirety of simple attentional capacity degradation under sleep loss *alone*. However, it is also extremely clear that the difference in reaction time, reaction time variability, and lapse probability, between sleep deprived and well rested individuals, was substantially greater during the latter half of the PVT, particularly following 5-6 minutes of test time. By isolating this comparison for a randomly chosen individual (Figure 9B), one can easily observe the increase in lapse frequency and performance variability of the sleep deprived compared to the rested individual following 6 minutes of test time. It is curious

to note that this particular individual actually improved in GD by 1.76 following TSD, compared to when rested, despite clear impairment on the PVT following TSD.



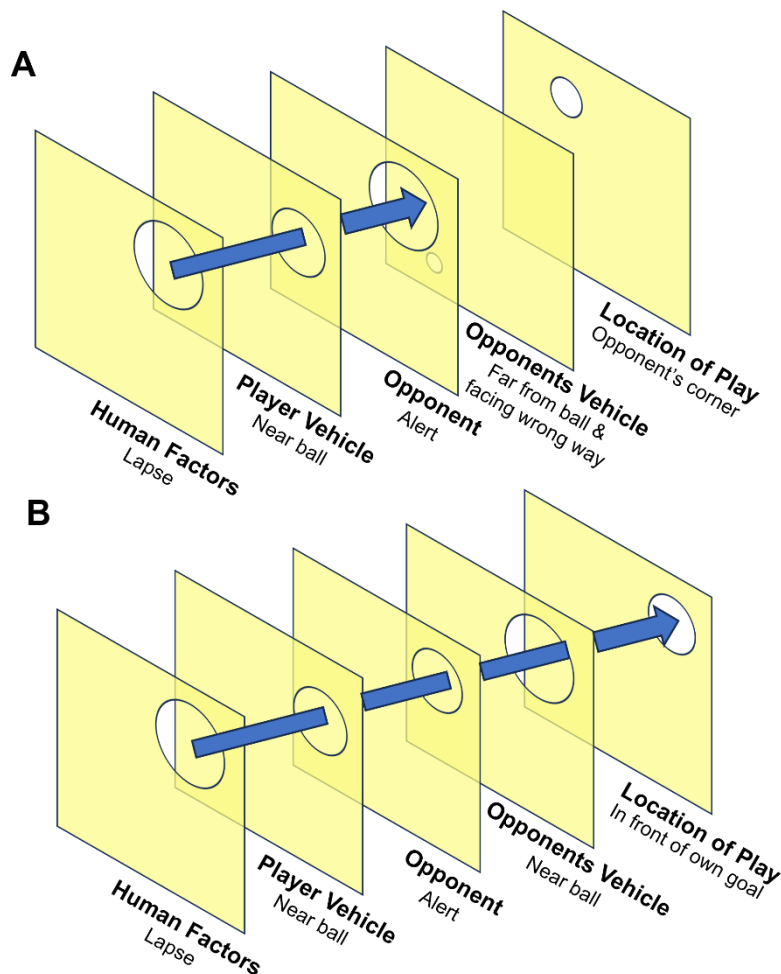
**Figure 9-1** A Bar chart showing the difference in mean reaction time (black), mean within-participant standard deviation or reaction time (i.e., green), and lapse probability (red), for TSD participants following sleep deprivation protocol and when well rested, within each one-minute epoch of the 10-min PVT (inclusive of the first minute of the PVT). Values above 0 resemble

greater mean values following sleep deprivation, compared to when rested. Black and green bars are measured on the left y-axis, and red bars are measured on the right y-axis. **B** Bar chart showing the difference in mean reaction time for each trial of the PVT for a randomly-selected TSD participant. Values above zero resemble a trial in which the reaction time was poorer following sleep deprivation, compared to when well rested. Red bars resemble trials which were a lapse within the session while sleep deprived, and blue bars resemble trials which were a lapse within the session when rested.

Contrast this 10-minute PVT to the 1v1 Rocket League matches, in which my esports specific outcome measures were obtained. The average Rocket League match within my sample spanned 6 minutes 42 seconds, with ~1-2 minute breaks in between matches. Ironically, this was one of the characteristics that made 1v1 Rocket League highly suitable as a target esports for experimental research; consistently short matches allowed for multiple matches (seven in my case), and hence multiple data points, per testing session. Moreover, goals occurred every ~40 seconds, with a ~10 second break afforded to each player in between each goal. These small breaks in task are far from trivial, considering that the effect of sleep loss on the time-on-task effect can be mitigated by task breaks (Ralph et al., 2017). With the average Rocket League match length close to matching the timespan in which performance was *not* greatly affected in the 10-minute PVT, and with regular breaks even within this short timespan, I posit that the non-trivial influence of sleep loss on time-on-task effects, likely bore little (if any) relevance to in-game performance in 1v1 Rocket League.

Although the expedition and exacerbation of the time-on-task effect due to sleep loss was a major contributor to simple attention deficits as seen on the PVT, it was not the entire story, as shown by the general increased lapse frequency across the entire timespan of the PVT following TSD. Hence, it is important to consider the potential outcome of an attentional lapse experienced during a Rocket League match. This may be best done so through the lens of Reason's *Swiss Cheese model of accident causation* (Reason, 2000). Within this model, each layer of the *cheese* resembles an error defence system, while each *hole* resembles a source of error (human or otherwise), with holes in all layers needing to line up for an incident to occur. Within this analogy, each layer resembles a set of circumstances, instead of a defence layer, in a similar manner to some adaptations of this model used to describe consequences of sleep loss in operational settings (i.e., Van Dongen et al. (2022)) and for motor vehicle accidents (Van Dongen, 2017). In the context of 1v1 Rocket League, the first layer could resemble the given

individual (human factors), while other layers could resemble the players vehicle (i.e., position, speed, orientation), opponent (human factors), opponents vehicle, the location of the ball with respect to the players, and the location of the ball with respect to the map. An attentional lapse would certainly resemble a hole in the first layer, but unless holes in the other layers align (i.e., critical events occur or circumstances exist at the same time as a lapse), the lapse would not result in a negative outcome. This model is conceptualised with respect to Rocket League in Figure 9-2.



**Figure 9-2** Reason's Swiss Cheese Model (Reason, 2000), conceptualised to 1v1 Rocket League. **A** resembles a hypothetical scenario in which an attentional lapse does not lead to an adverse match outcome, while **B** resembles a hypothetical scenario in which an attentional lapse does lead to an adverse match outcome (i.e., a goal conceded). In each panel, only the *holes* demonstrating the scenario also described are shown; in reality, there may be many holes at different positions within a given layer at a given period of time.

It is important to note that different esports may possess different likelihoods of attentional lapses impacting in-game outcomes. Again, this can be conceptualised by the number/ nature of layers and holes within layers. Future work using correlates of attentional lapses (as measured using EEG for example (Armanfard et al., 2016) or eye-tracking (McIntire et al., 2013)) time-synchronised with match replay files may even be able to attain the likelihood of lapses resulting in adverse in-game outcomes within Rocket League and other esports. Player expertise may also play a factor, and could be conceptualised by the widening of holes within the Swiss Cheese model. In summary, while lapses and *broadband* decreases in reaction time are impactful by nature on a task such as the PVT (which is maximally sensitive to them by design; Basner and Dinges (2011); Dorrian et al. (2004)), they may have only impacted in-game Rocket League performance under a particular set of (infrequently) aligning circumstances.

It is also worthwhile to consider the potential consequence of any in-game event affected by an attentional lapse. Regarding lapses in a motor vehicle context for example, they almost never will result in any adverse outcome when driving. However, they remain (rightly so) of the upmost concern with regards to motor vehicle safety, given the potentially fatal consequences of (infrequently occurring) fatigue-related incidents. Again contrasting this with 1v1 Rocket League, the absolute worst case scenario that could have occurred from a single event (i.e., lapse) in the Rocket League matches was a *two-goal turnaround* (a circumstance in which a certain goal for one player turned into a goal against that player). This event is rare in the context of Rocket League, however even then, I note that one two-goal turnaround is less than the average goal discrepancy observed within my sample of 279 matches (~3 goal discrepancy). In other words, the worst-case scenario for a single attentional lapse impacting in-game performance in 1v1 Rocket League is not a very large effect. Just like the earlier point regarding the likelihood of an adverse outcome from a lapse, I note that the potential consequence of an attentional lapse is almost certainly not uniform among esports genres. A useful traditional sport analogy for the differential potential impact of an attentional lapse between different esports could be the differential scope of impact for an attentional lapse at the starting line for a sprinter when compared to a marathon runner.

In summary regarding attentional lapses, I argue that they were extremely unlikely to play any role in the in-game performance of the esports players under sleep deprivation. I firstly note that the time-on-task effect, a major proponent of lapses, is extremely unlikely to be a factor within

1v1 Rocket League due to the short match length with frequent rest breaks. Secondly, I note that even if an attentional lapse were to occur, it would require many other factors to align to result in a negative in-game outcome. Lastly, I note that the absolute worst-case scenario of such an event is a notable, but not calamitous, two-goal swing in the outcome measure GD.

The above discussion regarding why the effects of sleep loss which are observable on the PVT, were unlikely to impact in-game Rocket League performance, are undertaken under the pretence that such effects would have translated from the dull and monotonous PVT assessment to the stimulating and motivating context of esports competition. However, participants were asked to play an esports (which are cognitively arousing/ engaging *by design*) in a competitive context (playing against an opponent in a series of seven matches). Such a context and its potential performance preserving properties are not seen as a limitation but instead a major strength of my performance assessment, as it matched the normal circumstance in which esports are played within over and above the conditions present within standard cognitive testing (and hence attained a greater level of ecological validity).

The impact of sleep loss on cognitive performance spans beyond simple attentional capacity, as outlined in **Chapter 2** and by a plethora of prior literature. However, compensatory mechanisms (of which task engagement and motivation are understood to encourage) function to largely preserve performance in much of the more *complex* cognitive tasks under what would be considered *mild* bouts (<36hrs TSD) of sleep loss (i.e., Horne & Pettitt, 1985). However, it is theorised that these compensatory mechanisms function to maintain *cognitive* stability at the expense of *cognitive flexibility* (Whitney et al., 2019). This appears highly relevant to an esports context, as aspects of cognitive flexibility have been outlined as particularly important within esports generally (Valls-Serrano et al., 2022). The specific impact of sleep loss on task-switching performance outlined by some prior research (Couyoumdjian et al., 2010; Slama et al., 2018), as well as the improved task-switching ability of action video game players (Nuyens et al., 2019; Toth et al., 2020), has led some researchers to suggest that sleep may affect esports performance through reducing task-switching ability (Toth et al., 2020).

However somewhat surprisingly, on my formal test of task-switching (the Category Switch Task), I found no evidence of ~29hrs TSD impacting task-switching performance. This is in contrast to Couyoumdjian et al. (2010), who found switch cost reaction time but not accuracy following one night of TSD, somewhat in agreement with Slama et al. (2018) who found switch cost (SC) accuracy but not SC RT to worsen following one night of TSD, and in agreement

with Nakashima et al. (2018), who found SC RT to be unaffected by 24hrs of TSD. Could the inability to find an effect of ~29hr TSD on task-switching ability be a feature of the esports player population, who may be better able to preserve their task-switching ability under bouts of sleep loss? Potential explanations for such a hypothesis could be that esports players frequently play while sleep restricted, and as such, may have trained their ability to maintain task-switching performance (seemingly an integral part of their gameplay) under such fatigue. However, this is unlikely given that previous work has shown stable trait-like performance impairment (i.e., level of performance impairment remains consistent over multiple bouts of sleep loss) in many previous cognitive tests (see Tkachenko & Dinges, 2018, for elaboration). Alternatively, as esports lends itself to sleep disturbances resulting in play under sleep-deprived fatigue, naturally tolerant individuals may self-select into esports, or alternatively, naturally vulnerable individuals may self-select out of esports; this follows a previous line of discussion regarding military/ aviation (Caldwell et al., 2005; Caldwell et al., 2012; Van Dongen & Belenky, 2009; Van Dongen, Caldwell, et al., 2011) and medical resident (Schlosser et al., 2012; Veasey et al., 2002) contexts. However, this is an unlikely explanation within the context of the presented work, as the population explored was not homogenously professional/ frequently competing esports athletes.

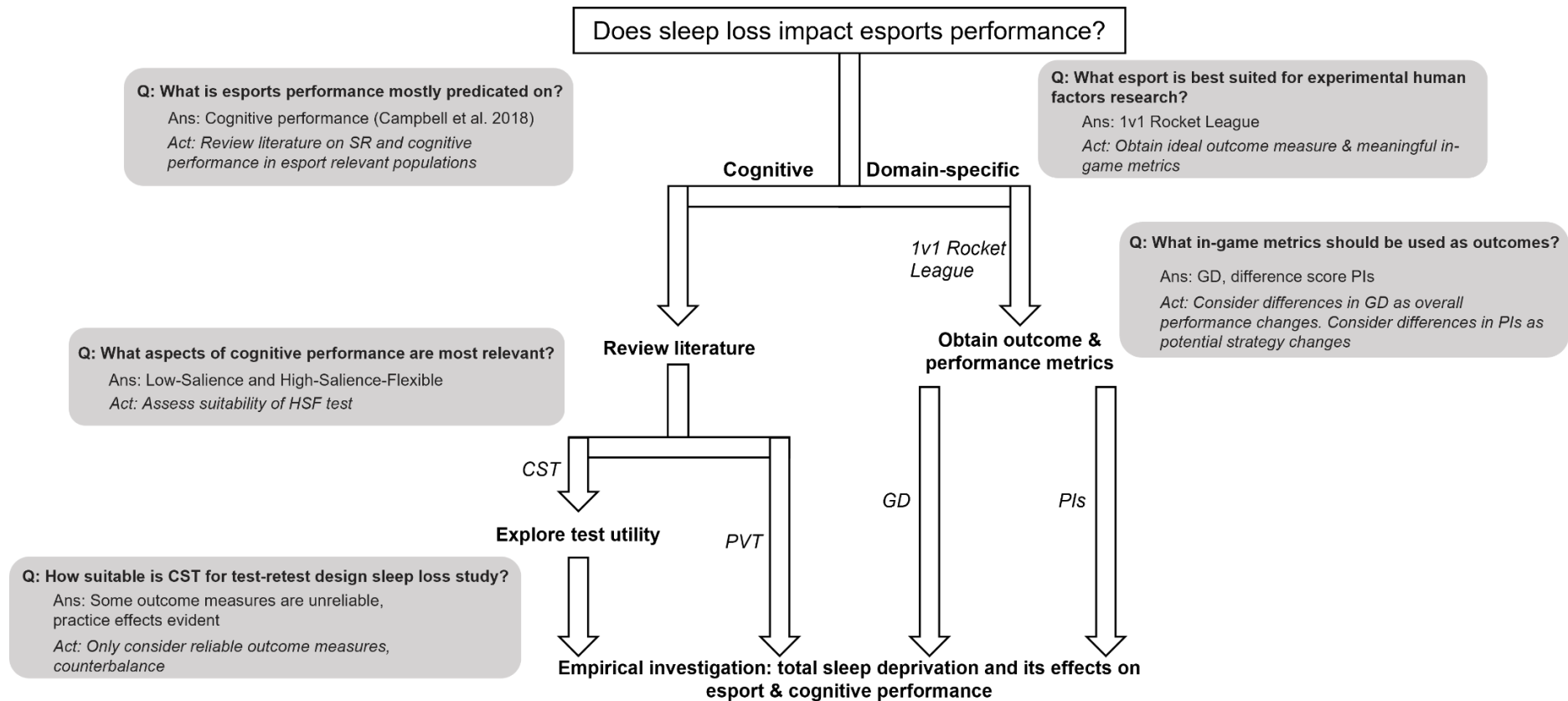
Instead, given the relative dearth of research exploring sleep loss and effects on task-switching, along with the diversity of test procedures (i.e., stimulus response mappings, cue-target intervals, nature of stimuli etc.), outcome measures, and findings, it is much more likely that these factors may explain some of the nuances and diversity of results within this area. All of the previously mentioned studies (Couyoumdjian et al., 2010; Nakashima et al., 2018; Slama et al., 2018), like ours, utilised a paradigm which taxes one's ability to switch between two or more categorisation rules in an unpredictable manner, while simultaneously using the same stimulus-response mappings (SRMs) between each cue, and responding to the same type of stimuli regardless of cue. Slama et al. (2018) (18 participants undertaking TSD) used a *cued match-to-sample* task with cues being *colour, shape, number, or outline*, Couyoumdjian et al. (2010) (54 participants undertaking TSD) used a *number-letter* paradigm, Nakashima et al. (2018) (12 participants undertaking TSD) used a *value-numerosity* paradigm, and I (20 participants undertaking TSD) used a *living-size* Category Switch Task. It is important to note that both Slama et al. (2018) and Couyoumdjian et al. (2010) demonstrated their effects to be distinguishable from downstream effects of arousal/ basic attentional processing. It appears plausible that sleep loss does lead to task-switching deficits, however it is apparent that there

are task/ procedural/ outcome measure factors which are not yet fully understood and which may play a large role in the aforementioned mixed findings. Ideally, staunch consistency in procedure would be extremely useful in providing a clearer picture as to the effects of TSD on task-switching ability, and as such, replication studies (particularly for the procedure and analyses of Slama et al. (2018) and Couyoumdjian et al. (2010) as they found significant effects) would be incredibly useful and encouraged. However, it is simultaneously recognised that the differences in procedures is, in part, attributed to the fact that task-switching ability measured on the task-switching paradigm used was not the sole outcome measure of interest in any of the four studies explored. Hence, a systematic collation of literature exploring sleep loss and task-switching (and cognitive flexibility), which has not been performed to date, may provide insight onto important task-specific factors, with implications both in the lab and beyond. This review is currently being undertaken by a Masters student at the University of Limerick, alongside supervision from Prof. Mark Campbell and I.

Irrespective of the potential reasons behind the null finding of TSD on task-switching performance, it stands to reason that if the cognitive task specifically designed to measure this component did not find effects, it is unlikely that task-switching ability deficits were to impact in-game performance to any measurable degree.

### **9.3. Strengths of the overall approach**

The work comprising this thesis was highly multidisciplinary, owing to the novel nature of the aims and infancy of esports science research more generally. As such, a wide range of topics and methodologies were used. There were strengths and practical implications which spawned from each individual piece of work but which are not directly related to the thesis as an overall body of work. These strengths and implications are discussed within each individual chapter, and as such, will not be rehashed here. Instead, this section will discuss the strength of the overall approach outlined in the thesis, and its potential utility for future work aiming to assess human factors that are relevant to esports performance. A flowchart of the overall approach is conceptualised in Figure 9-3, and this will be referred to in the subsequent paragraphs. This flowchart can be utilised and expanded on within future work.



**Figure 9-3** A flowchart describing the individual components of the thesis and their linkage to the overall research objective.

Delving into the left side of the flowchart first, the starting point for the body of work was the question of what human factors theoretically contribute most greatly to esports performance. This question was relevant, as it provided the specific element of performance which was most pertinent to explore, in addition to ecologically valid in-game measures. The consensus answer to this question within the scientific literature is cognitive factors (Campbell et al., 2018; Pedraza-Ramirez et al., 2020). As such, it was relevant to consider the influence of sleep loss on cognitive performance, with respect to how this could ultimately impact in-game performance. An abundance of original research had already been conducted regarding sleep loss and its effects on cognitive performance, however prominent previous systematic collations (Lim & Dinges, 2010; Lowe et al., 2017; Pilcher & Huffcutt, 1996; Wickens et al., 2015) had not been performed with specific regards to populations who engage in cognitively demanding tasks with critical outcomes in their occupation or area of expertise (*ECPs*). Furthermore, such systematic collations (with the exception of Wickens et al. (2015)) had tended to avoid examining ecologically relevant occupation specific (/ expertise relevant) task performance, instead focussing on standardised cognitive tests (Lim & Dinges, 2010; Lowe et al., 2017; Pilcher & Huffcutt, 1996). This is despite previous accounts of effects of sleep loss on standardised cognitive tests not necessarily translating to performance loss in task specific circumstances (a pattern supported by the review). Regarding this, I point toward a quote from a seminal narrative review by Harrison and Horne (2000) (p. 236); “Much of the SD research, as of this writing, has focused on cognitive processes that have little to do with the true nature of the job or normal working duties (e.g., serial reaction time, vigilance). Sometimes, the overall picture can be confusing, with findings showing no impairments for certain clinical skills and concurrent deterioration in psychological performance tasks of unknown relevance to these and other medical skills.” Overall, my review was particularly pertinent as it provided a summary of literature that was most relevant to the specific context of esports, both in terms of population and outcome measures.

The thorough and systematic nature of the review presents as an obvious strength of this piece of work. In particular, the use of a database combination with a demonstrated optimal sensitivity/ specificity trade-off (Bramer et al., 2017), along with a rigorous grey literature and backward snowballing procedure, ensured that no relevant body of work was missed. The usage of a field-relevant grey literature source (Defence Technical Information Centre/ DTIC) resulted in an otherwise missed article (Hartzler et al., 2015)

being included; as such, this review showcases the benefits of grey literature searching, particularly using grey literature sources with specific field-relevance. This review provided crucial insight into the nature of cognitive deficits we may expect to see within sleep deprived esports athletes. The first was of rudimentary attentional capacity while performing simple and monotonous (low-salience) tasks. Such tasks are not reflective of the complex, engaging, and stimulating environment that esports provide, but were included in the later experimental work nonetheless via the PVT. The review suggested that more complex tasks appeared to be more likely to be impacted by SR if they involved high levels of cognitive flexibility (formally tested using task-switching, reversal learning, or multitask tests), somewhat corroborating contemporary theories/ frameworks (Whitney et al., 2019). This presented as high relevance for an esports context, as both quasi-experimental and intervention studies have shown that exposure to video games commonly played as esports results in improved cognitive flexibility, and in particular, task-switching ability (Nuyens et al., 2019; Toth et al., 2020). Thus, my findings somewhat aligned with suggestions by Toth et al. (2020), that sleep loss may impact esports performance specifically through decreasing esports athlete's task-switching ability.

A pervasive issue in the realm of sleep literature is the usage of tests with untested properties or properties which are unsuitable for sleep deprivation designs. Two such properties are the test-retest reliability of the task, along with its propensity to result in practice effects within the test-retest timeframe of consideration (Dorrian et al., 2004). A major strength of the approach taken in this thesis is that these elements were explicitly tested through a pilot study (outlined in **Chapter 3**, and the next box on the left path of Figure 9-3). In doing so, I was able to identify outcome measures that were *not* suitable for use. This work also confirmed my suspicion regarding the potential issue practice effects would present that warranted addressing through experimental design (counterbalancing). Running a pilot study to attain test-retest reliability and the propensity of practice effects to bias results is essential where cognitive assays are considered for repeated measures assessment.

Switching over to the right side of Figure 9-3, the important first step was to uncover an esports competition that would be suitable for experimental testing. I believe that my thesis makes a strong case for 1v1 Rocket League being the default esports of choice for future research examining human factors in esports. This is for many reasons, which have been discussed many times already. However, one important consideration not yet thoroughly

discussed is game popularity. It is important that a sufficient number of players of a given esports are available, such that sufficiently powered experiments can take place. This is one of the major successes of the outlined work; forty Rocket League players were recruited specifically for an in-person total sleep deprivation study, within a relatively small city in a relatively small country. This, alongside the already mentioned short and predictable match lengths, data (and now, outcome measures and performance indicator) availability, and play as individuals, positions Rocket League as arguably the best starting point (without consideration of genre specific demands as discussed in the previous section) with regards to human factors research in esports.

A unique element of esports (when compared to most traditional sports) is computerised gameplay, resulting in in-game statistics/ metrics actually being recorded digitally by the game itself. This data can then subsequently be accessed using application processing interfaces (APIs), should they be available for the particular game. Unfortunately, data availability is highly variable between different esports titles. Rocket League is relatively unique in that in-game data are freely available and are readily updated to a replay database (ballchasing.com), which provides in-game statistics for over 90 million games (as per 30/05/2023). This large amount of data availability facilitated my use of a machine learning and feature selection approach (taking inspiration from similar previous work in traditional sports, (i.e., Bennett et al. (2020); Bennett et al. (2019); Bishop and Barnes (2013); García et al. (2013); Hughes et al. (2017); Leicht et al. (2017); Mosey and Mitchell (2020); Robertson et al. (2016); Vaz et al. (2010); Whitehead et al. (2020); Woods et al. (2017)) to extract information about certain gameplay styles or strategies that influences one's performance; information that can subsequently be used for analytical depth in esports performance research. **Chapter 5** provides such groundwork for Rocket League, while similar work in other esports provide performance indicator metrics as well (Bahrololloomi et al., 2023; Bialecki et al., 2023; Hitar-García et al., 2023; Hojaji et al., 2023; D. Lee et al., 2021; Novak et al., 2020; Xia et al., 2017). However, one issue that is pervasive within most esports is *meta-shifts* (changes in the dominant strategy/ set of strategies, or perceived optimal playstyle, within an esports; Kemp et al. (2020); Kokkinakis et al. (2021)) resulting from *game patches* (changes of game parameters introduced by game developers Chitayat et al. (2023)). These can be analogised to traditional sport are game-changing rule changes, such as the introduction of the three-point field goal in basketball (Jaguszewski, 2020), or castling in modern chess (Pratesi, 2008). However, while these occur on a highly frequent basis in most major

esports (i.e., approximately every two weeks in League of Legends; Sabtan et al., 2022), they have barely been a factor in Rocket League<sup>†</sup>, due to its stable and standardised game mechanics. Hence, the understanding of in-game factors leading to success in Rocket League remains far robust than most other esports over time; yet another strength of using Rocket League as a target sport for human factors research.

## 9.4. Limitations and Future Research

Again, it is extremely important to outline the novelty of the work outlined in the current thesis. This was the first experimental foray into a large, complicated, and highly nuanced area. As such, there are naturally many more questions raised than answers gained at the conclusion of the work. To avoid the risk of taking of too much space within the discussion, I have omitted most of these future avenues of inquiry here, but have instead provided a paragraph(/s) description of four of them within Appendix 9.1. Also, like the previous section, I have omitted the limitations and future research avenues which are highly specific to each chapter and, as they are discussed within the given chapter. Instead I will focus on three broader concepts here; generalisability, expertise, and forms of sleep loss.

### 9.4.1. Generalisability

Throughout this discussion, I have suggested that the effects of sleep loss on in-game performance may not be completely uniform between esports of different genres. Despite great diversity in game dynamics, esports are often discussed as one activity (to a much greater degree than traditional sports are). This is likely due to a combination of reasons, the first being the sheer infancy of esports research (Cranmer et al., 2021). The second reason is a relative ambiguity or line-blurring between esports of different *genres* (Apperley, 2006). Lastly, almost all esports use similar (if not identical) input (keyboard and mouse or controller) and output (computer monitor) devices, within a seated posture. A traditional sport analogy would be to consider all sports which involve running as the primary movement modality as identical; including sprinting, marathon running, rugby, netball, baseball, and squash. Following this analogy, we would certainly not consider all human factors (including relative demands of strength, aerobic/ anaerobic capacity,

<sup>†</sup>While some Rocket League patches have changed gameplay, very few have resulted in *meta-shifts*. Examples of such could be the addition of directional air roll in patch v1.17 (25/04/2016), addition of deadzone/ sensitivity customisation in patch v1.25 (07/12/2016), standardisation of car hitbox/ turning radii in patch v1.35 (05/07/2017), and the standardisation of maps in Competitive Season 6 (Legacy; 29/08/2017).

agility etc.) leading to optimal performance to be identical between these sports, however we can certainly find insight and applicability from research within one of these sports, for another one of these sports. In the same vein, we should not consider the effects of sleep loss to be exactly identical between esports of different genres. As discussed in **Chapter 7**, multiplayer online battle arena (MOBA) games such as League of Legends (LoL), which have match lengths spanning from 20-90 minutes and often without gameplay breaks, present as an esports in which the time-on-task effect is far more likely to play a role than other esports. Likewise, a first person shooter (FPS) game like Counter Strike: Global Offensive (CS:GO) is far more likely to have events in which fast responses are the primary determinant of success, compared to other esports.

This issue of generalisability remains relevant even within the discussion of Rocket League as an esports. I specifically explored 1v1 Rocket League, due to the potential procedural and analytical complexities that the inclusion of participants and team-based gameplay would introduce. However, the game mode which receives the greatest amount of attention as a competitive game mode is 3v3 Rocket League. The differences between 1v1 and 3v3 Rocket League, with respect to how sleep loss may impact performance, warrant consideration. Game specific differences (i.e., relative importance of positioning, certain in-game mechanics etc.) between 1v1 and 3v3 are frequently theorised however not formally explored within peer-review literature. Beyond gameplay specific differences, the adverse impact that sleep loss may have on leader-follower interactions (Barnes et al., 2016; Guarana & Barnes, 2017; Olsen et al., 2016) and specialised communication abilities (Banks et al., 2019; Harrison & Horne, 2000; Holding et al., 2019; Whitmore & Fisher, 1996) warrant consideration.

It is important to stress that the results outlined within this thesis with regards to sleep loss present as the most generalisable and applicable, probably to all esports contexts, to date. However, it is noted that the generalisability of my results to all esports contexts is not perfect, and many factors which are relevant to specific esports contexts could not be considered.

#### **9.4.2. Expertise**

Within the discussions of **Chapter 2** and **Chapter 5**, I discussed how esports (and specifically Rocket League) may be a useful window to explore how expertise moderates the effect of sleep loss on cognitively demanding task performance. This is due to the use of the Elo system (called MMR in Rocket League), which is a continuous and accurate

measure of in-game expertise, along with measures of match outcome and performance indicators outlined in **Chapter 5**. However, this was not a primary aim within the scope of the thesis, and as such, analysis with consideration to player expertise was not provided in **Chapter 7**. By including the MMR of the TSD player within each pair as a fixed effect and allowing for a *session by MMR* interaction within the models exploring GD and PIs, this idea could be explored using our data. I performed this analysis for curiosities sake, and included the model creation process and results tables (as per **Chapter 7**) as Appendix 9.2. In short, player MMR did not appear to influence the (lack of) relationship between the TSD protocol and GD, nor the relationship between the TSD protocol and any PI with the exception of *time spent goalside of the ball difference*, where a negative *session by MMR* interaction trended toward significance ( $b = -0.0064 \pm 0.0035$ ,  $t(1, 257.06) = -1.83$ ,  $p = 0.058$ ). This interaction would suggest that the extent to which *time spent goalside of the ball difference* increased following TSD (a trend towards significance,  $p = 0.059$ , that was replicated in this analysis,  $p = 0.058$ ) decreased as player MMR increased. In other words, there was a trend towards TSD impacting this PI more in less skilled players, compared to more skilled players.

I would encourage future use of esports as a tool to explore the moderating role of task expertise on sleep loss and its effects on performance. However, this question may be better approached by comparing two clearly defined and distinct expertise groups (i.e., novice vs. highly skilled). I note that such an approach has been previously undertaken to examine differential effects of neurostimulation during esports specific skill learning (Toth, Ramsbottom, et al., 2021).

### 9.4.3. Forms of Sleep Loss

The systematic review disseminated in **Chapter 2** specifically explored performance within the context of *sleep restriction (SR)*. This was because SR is the most ecologically relevant form of sleep loss for esports athletes. At the time the review was undertaken, the specific design of the sleep loss experimental research was not fleshed out, and the direction was leaning towards the undertaking of a SR protocol. However, as protocol drafting ensued (coinciding with the easing of COVID-19-related restrictions), it became increasingly clear that a SR protocol was not feasible. This was primarily due to participant burden and safety concerns. Practically all SR protocols which go beyond one night of mild SR (~5hr SO for example, which was highly unlikely to result in any effect) necessitate in-person monitoring of sleep, and restriction of participant movement (i.e.,

the avoidance of driving; this is typically realised by restricting participants completely to the laboratory) throughout the protocol. This was highly impractical considering (a) most of the included participants were university students and/ or travelling from outside of Limerick to complete the protocol, and (b) the absence of a purpose-built laboratory. Conversely, a performance impairment (as measured by PVT) equivalent to multiple days of SR (~7 days of 4hr SO and ~10 days of 6hr SO; see Figure 1-1, lining up where ~29hrs on the 0hr SO line corresponds to relative to the 4hr SO and 6hr SO lines with regards to PVT lapses) could be achieved through one night of TSD, and as such, TSD was chosen as the sleep loss modality.

It is important to note that while TSD and SR have some differential effects, these seem limited mostly to (a) a slower decrease in a subjective alertness relative to objective performance in SR compared to TSD (Banks et al., 2010; Belenky et al., 2003; Van Dongen et al., 2003), and (b) a longer recovery period (to return to baseline cognitive performance) in SR, compared to the equivalent performance impairment realised through TSD (Banks et al., 2010; Belenky et al., 2003). There is no evidence (to my knowledge) that outside of the dosage required, TSD and SR result in differential effects on an aspect of cognitive performance (i.e., no evidence that SR impacts performance within a cognitive domain that isn't impacted by TSD). This is exemplified by the similarity of results between a meta-analysis on the cognitive performance effects of TSD (Lim & Dinges, 2010) and SR (Lowe et al., 2017). Hence, I argue that the results from the review in **Chapter 2** and **Chapter 7** can be discussed together, despite focussing on different types of sleep loss.

## **9.5. Practical implications**

The results presented within this thesis primarily suggest that an acute sleep loss bout is unlikely to impact overall in-game esports performance to a measurable degree. This presents as somewhat of a positive message for esports coaches and players alike, who may be concerned about sleep loss experienced immediately before competition spelling the difference between victory and defeat in upcoming competitions (which can have significant potential financial ramifications). This is highly relevant considering that esports athletes face significant risk of experiencing disturbed sleep (Bonnar, Lee, et al., 2019; Lee et al., 2020; S. Lee et al., 2021), including the same disturbances frequently observed the night prior to competition for traditional sport athletes (Bonnar, Castine, et al., 2019; Juliff et al., 2015).

However, this implication is caveated by the fact that only one esports was explored, and the generalisability of the results presented here (while almost certainly being greater than any other work disseminated to date) to other esports remains unknown. Athletes, coaches, and individuals with a vested interest in optimising in-game esports performance outside of Rocket League, should interpret the relevance of the current findings at the same level that individuals outside of a basketball context should interpret findings relating sleep to basketball-related performance measures, such as those from Mah et al. (2011), Staunton et al. (2017), and Fox et al. (2021), for example.

Also, these results presented within this thesis do not absolve sleep as an important human factor within the world of esports. Sleep plays an instrumental role in learning (including the refinement of fine-motor skills, i.e., Walker et al. (2002)) and hence is important for the learning/ refinement of in-game skills and strategies. These skills/ strategies are the primary determinant of success in esports, and as such, are the ultimate currency for the esports athlete, who compete in a profession with remarkably low job security (Smithies et al., 2020). Given this importance of habitual sleep for esports athletes, combined with worrying poor sleep quality and behaviours previously outlined for esports athletes (Bonnar et al., 2022; Lee et al., 2020; S. Lee et al., 2021) and a relative reluctance/ dislike of sleep monitoring/ hygiene practices aiming to address such issues (Bonnar et al., 2023), I caution that my results do not downplay the importance of sleep for esports. As such, I encourage future research towards and employment of interventions aiming to improve habitual sleep outcomes for esports athletes, such as that outlined by Bonnar et al. (2022).

## **9.6. Conclusions**

Esports are by far the fastest growing competitive activity worldwide. Successful esports athletes can receive significant earnings (135 players have earned over €1million as of June 20, 2023; Esports Earnings (2023a)) by playing the video game they love, making esports an attractive career path for many. However, esports as a career is characterised by extremely short career spans (Ward & Harmon, 2019) and minimal job security (Smithies et al., 2020). These factors, alongside demands from sponsors and stakeholders, result in an extreme drive for in-game performance maximisation. Esports is characterised by a high relative importance of cognitive factors, leading some researchers to refer to esports athletes as *cognitive athletes* (Campbell et al., 2018). This, combined with the wealth of literature linking sleep loss to worsened cognitive performance, has led many researchers

and esports athletes alike to hypothesise that sleep loss leads to worsened in-game performance.

In this thesis, I tested this for the first time in an experimental design, using 1v1 Rocket League. I show that ~29hr TSD (equivalent to ~7 days of 4hr TIB or ~10 days of 6hr TIB, considering lapse propensity on the PVT (Van Dongen et al., 2003)) does not impact in-game performance in 1v1 Rocket League, despite potentially causing strategy changes observed using in-game performance indicators (which were identified using a machine learning notational analysis approach). I conclude that acute sleep loss immediately prior to competition may not be of primary concern for esports athletes, though caution this interpretation with the observation that different esports can vary in factors (for example, length of gametime without a break) which may influence the relative impact of sleep loss on performance.

## Chapter 10. References

- Aarts, E., Verhage, M., Veenvliet, J. V., Dolan, C. V., & van der Sluis, S. (2014). A solution to dependency: using multilevel analysis to accommodate nested data. *Nature Neuroscience*, 17(4), 491-496. <https://doi.org/10.1038/nn.3648>
- Acebo, C., Sadeh, A., Seifer, R., Tzischinsky, O., Wolfson, A. R., Hafer, A., & Carskadon, M. A. (1999). Estimating Sleep Patterns with Activity Monitoring in Children and Adolescents: How Many Nights Are Necessary for Reliable Measures? *Sleep*, 22(1), 95-103. <https://doi.org/10.1093/sleep/22.1.95>
- Active Player. (2021a). *CS: Go Live Player Count and Statistics*. Retrieved 01/03/2021 from <https://activeplayer.io/counter-strike-global-offensive/>
- Active Player. (2021b). *Rocket League Live Player Count and Statistics*. Activeplayer. Retrieved 01/03/2021 from <https://activeplayer.io/rocket-league/>
- Active Player. (2023). *Rocket League Live Player Count and Statistics*. Retrieved 16/04/2023 from <https://activeplayer.io/rocket-league/>
- Adams, R., & Ericsson, A. (2000). Introduction to cognitive processes of expert pilots. *Journal of Human Performance in Extreme Environments*, 5(1), 5. <https://doi.org/10.7771/2327-2937.1006>
- Ahn, J., Collis, W., & Jenny, S. (2020). The one billion dollar myth: Methods for sizing the massively undervalued esports revenue landscape. *International Journal of Esports*, 1(1), 1-19. <https://www.ijesports.org/article/15/html>
- Anothaisintawee, T., Reutrakul, S., Van Cauter, E., & Thakkinstian, A. (2016). Sleep disturbances compared to traditional risk factors for diabetes development: Systematic review and meta-analysis. *Sleep Medicine Reviews*, 30, 11-24. <https://doi.org/10.1016/j.smrv.2015.10.002>
- Apperley, T. H. (2006). Genre and game studies: Toward a critical approach to video game genres. *Simulation & Gaming*, 37(1), 6-23. <https://doi.org/10.1177/1046878105282278>
- Araújo, D., & Davids, K. (2015). Towards a theoretically-driven model of correspondence between behaviours in one context to another: Implications for studying sport performance. *International Journal of Sport Psychology*, 46(6), 745-757. <https://doi.org/10.7352/IJSP.2015.46.745>
- Arbula, S., Capizzi, M., Lombardo, N., & Vallesi, A. (2016). How Life Experience Shapes Cognitive Control Strategies: The Case of Air Traffic Control Training. *PLOS ONE*, 11(6), e0157731. <https://doi.org/10.1371/journal.pone.0157731>

- Archer, E. (2020). *rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics*. In <https://CRAN.R-project.org/package=rfPermute>
- Armanfard, N., Komeili, M., Reilly, J. P., & Pino, L. (2016, 15-18 May 2016). Vigilance lapse identification using sparse EEG electrode arrays. 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE),
- Arnal, P. J., Thorey, V., Debellemanniere, E., Ballard, M. E., Bou Hernandez, A., Guillot, A., Jourde, H., Harris, M., Guillard, M., Van Beers, P., Chennaoui, M., & Sauvet, F. (2020). The Dreem Headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging. *Sleep*, 43(11), zsaa097. <https://doi.org/10.1093/sleep/zsaa097>
- Aserinsky, E., & Kleitman, N. (1953). Regularly Occurring Periods of Eye Motility, and Concomitant Phenomena, During Sleep. *Science*, 118(3062), 273-274. 10.1126/science.118.3062.273
- Axelsson, J., Ingre, M., Kecklund, G., Lekander, M., Wright, K. P., Jr., & Sundelin, T. (2020). Sleepiness as motivation: a potential mechanism for how sleep deprivation affects behavior. *Sleep*, 43(6), zsz291. <https://doi.org/10.1093/sleep/zsz291>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The Alcohol Use Disorders Identification Test: Guidelines for Use in Primary Care* (Second ed.). World Health Organization Geneva. [https://apps.who.int/iris/bitstream/handle/10665/67205/WHO\\_MSD\\_MSB\\_01.6a.pdf?sequence=1](https://apps.who.int/iris/bitstream/handle/10665/67205/WHO_MSD_MSB_01.6a.pdf?sequence=1)
- Bahrololloomi, F., Klonowski, F., Sauer, S., Horst, R., & Dörner, R. (2023). E-Sports Player Performance Metrics for Predicting the Outcome of League of Legends Matches Considering Player Roles. *SN Computer Science*, 4(3), 238. <https://doi.org/10.1007/s42979-022-01660-6>
- Baldassari, C. M., Mitchell, R. B., Schubert, C., & Rudnick, E. F. (2008). Pediatric obstructive sleep apnea and quality of life: A meta-analysis. *Otolaryngology–Head and Neck Surgery*, 138(3), 265-273. <https://doi.org/10.1016/j.otohns.2007.11.003>
- Balkin, T. J., Bliese, P. D., Belenky, G., Sing, H., Thorne, D. R., Thomas, M., Redmond, D. P., Russo, M., & Wesensten, N. J. (2004). Comparative utility of instruments

- for monitoring sleepiness-related performance decrements in the operational environment. *Journal of sleep research*, 13(3), 219-227. <https://doi.org/10.1111/j.1365-2869.2004.00407.x>
- Banks, S., & Dinges, D. F. (2007). Behavioral and Physiological Consequences of Sleep Restriction. *Journal of Clinical Sleep Medicine*, 3(5), 519-528. <https://doi.org/10.5664/jcsm.26918>
- Banks, S., Landon, L. B., Dorrian, J., Waggoner, L. B., Centofanti, S. A., Roma, P. G., & Van Dongen, H. P. A. (2019). Effects of fatigue on teams and their role in 24/7 operations. *Sleep Medicine Reviews*, 48, 101216. <https://doi.org/10.1016/j.smrv.2019.101216>
- Banks, S., Van Dongen, H. P. A., Maislin, G., & Dinges, D. F. (2010). Neurobehavioral dynamics following chronic sleep restriction: Dose-response effects of one night for recovery. *Sleep*, 33(8), 1013-1026. <https://doi.org/10.1093/sleep/33.8.1013>
- Barnes, C. M., Guarana, C. L., Nauman, S., & Kong, D. T. (2016). Too tired to inspire or be inspired: Sleep deprivation and charismatic leadership. *Journal of Applied Psychology*, 101(8), 1191-1199. <https://doi.org/10.1037/apl0000123>
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, 4, 328. <https://doi.org/10.3389/fpsyg.2013.00328>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. *BMC neuroscience*, 11(118). <https://doi.org/10.1186/1471-2202-11-118>
- Basner, M., & Dinges, D. F. (2011). Maximizing Sensitivity of the Psychomotor Vigilance Test (PVT) to Sleep Loss. *Sleep*, 34(5), 581-591. <https://doi.org/10.1093/sleep/34.5.581>
- Basner, M., Hermosillo, E., Nasrini, J., McGuire, S., Saxena, S., Moore, T. M., Gur, R. C., & Dinges, D. F. (2017). Repeated administration effects on Psychomotor Vigilance Test performance. *Sleep*, 41(1), 1-6. <https://doi.org/10.1093/sleep/zsx187>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>

- Baumann, A., Mentzoni, R. A., Erevik, E. K., & Pallesen, S. (2022). A qualitative study on Norwegian esports students' sleep, nutritional and physical activity habits and the link to health and performance. *International Journal of Esports*, 1(1), 1-13. <https://www.ijesports.org/article/88/html>
- Beattie, L., Espie, C. A., Kyle, S. D., & Biello, S. M. (2015). How are normal sleeping controls selected? A systematic review of cross-sectional insomnia studies and a standardized method to select healthy controls for sleep research. *Sleep Medicine*, 16(6), 669-677. <https://doi.org/10.1016/j.sleep.2015.01.010>
- Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, 144, 77-110. <https://doi.org/10.1037/bul0000130>
- Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2023). Effects of action video game play on cognitive skills: A meta-analysis. *Technology, Mind, and Behavior*, 4(1), 1-21. <https://doi.org/10.1037/tmb0000102>
- Beebe, D. W., Di Francesco, M. W., Tlustos, S. J., McNally, K. A., & Holland, S. K. (2009). Preliminary fMRI findings in experimentally sleep-restricted adolescents engaged in a working memory task. *Behavioral and Brain Functions*, 5. <https://doi.org/10.1186/1744-9081-5-9>
- Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J., Fastenau, P. S., & Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology*, 20(4), 517-529. <https://doi.org/10.1016/j.acn.2004.12.003>
- Belenky, G., Krueger, G. P., Balkin, T. J., Headley, D. B., & Solick, R. E. (1987). *Effects of continuous operations (CONOPS) on soldier and unit performance: Review of the literature and strategies for sustaining the soldier in CONOPS* (WRAIR-BB-87-1). (WALTER REED ARMY INST OF RESEARCH WASHINGTON DC DEPT OF BEHAVIORAL BIOLOGY).
- Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., Russo, M. B., & Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *Journal of sleep research*, 12(1), 1-12. <https://doi.org/10.1046/j.1365-2869.2003.00337.x>
- Bennett, M., Bezodis, N. E., Shearer, D. A., & Kilduff, L. P. (2020). Predicting performance at the group-phase and knockout-phase of the 2015 Rugby World

- Cup. *European Journal of Sport Science*, 1-9.  
<https://doi.org/10.1080/17461391.2020.1743764>
- Bennett, M., Bezodis, N. E., Shearer, D. A., Locke, D., & Kilduff, L. P. (2019). Descriptive conversion of performance indicators in rugby union. *Journal of Science and Medicine in Sport*, 22(3), 330-334.  
<https://doi.org/10.1016/j.jsams.2018.08.008>
- Bialecki, A., Gajewski, J., Bialecki, P., Phatak, A., & Memmert, D. (2023). Determinants of victory in Esports - StarCraft II. *Multimedia Tools and Applications*, 82(7), 11099-11115. <https://doi.org/10.1007/s11042-022-13373-2>
- Bills, A. G. (1931). Blocking: A New Principle of Mental Fatigue. *The American Journal of Psychology*, 43(2), 230-245. <https://doi.org/10.2307/1414771>
- Bishop, L., & Barnes, A. (2013). Performance indicators that discriminate winning and losing in the knockout stages of the 2011 Rugby World Cup. *International Journal of Performance Analysis in Sport*, 13(1), 149-159.  
<https://doi.org/10.1080/24748668.2013.11868638>
- Bjerner, B. (1949). *Alpha depression and lowered pulse rate during delayed actions in a serial reaction test: A study in sleep deprivation* (Vol. 19). Haeggströms.
- Bjorvatn, B., Stangenes, K., Øyane, N., Forberg, K., Lowden, A., Holsten, F., & Åkerstedt, T. (2006). Subjective and Objective Measures of Adaptation and Readaptation to Night Work on an Oil Rig in the North Sea. *Sleep*, 29(6), 821-829. <https://doi.org/10.1093/sleep/29.6.821>
- Bjorvatn, B., Stangenes, K., Øyane, N., Forberg, K., Lowden, A., Holsten, F., & Åkerstedt, T. (2007). Randomized placebo-controlled field study of the effects of bright light and melatonin in adaptation to night work. *Scandinavian Journal of Work, Environment & Health*, 33(3), 204-214.  
<http://www.jstor.org/stable/40967644>
- Blatter, K., & Cajochen, C. (2007). Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior*, 90(2), 196-208. <https://doi.org/10.1016/j.physbeh.2006.09.009>
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127-135.  
<https://doi.org/10.1016/j.tree.2008.10.008>

- Bonnar, D., Castine, B., Kakoschke, N., & Sharp, G. (2019). Sleep and performance in Eathletes: for the win! *Sleep Health*, 5(6), 647-650. <https://doi.org/10.1016/j.sleh.2019.06.007>
- Bonnar, D., Hwu, M., Lee, S., Gradisar, M., Suh, S., & Kahn, M. (2023). The Influence of Coaches and Support Staff on the Sleep Habits of Esports Athletes Competing at Professional and Semiprofessional Level. *Journal of Electronic Gaming and Esports*, 1(1), jege.2022-0023. <https://doi.org/10.1123/jege.2022-0023>
- Bonnar, D., Lee, S., Gradisar, M., & Suh, S. (2019). Risk Factors and Sleep Intervention Considerations in Esports: A Review and Practical Guide. *Sleep Med Res*, 10(2), 59-66. <https://doi.org/10.17241/smr.2019.00479>
- Bonnar, D., Lee, S., Roane, B. M., Blum, D. J., Kahn, M., Jang, E., Dunican, I. C., Gradisar, M., & Suh, S. (2022). Evaluation of a Brief Sleep Intervention Designed to Improve the Sleep, Mood, and Cognitive Performance of Esports Athletes. *International Journal of Environmental Research and Public Health*, 19(7), 4146. <https://doi.org/10.3390/ijerph19074146>
- Bonnet, M. H., & Arand, D. L. (2003). Clinical effects of sleep fragmentation versus sleep deprivation. *Sleep Medicine Reviews*, 7(4), 297-310. <https://doi.org/10.1053/smr.2001.0245>
- 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. <https://doi.org/10.1016/j.actpsy.2010.09.005>
- Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychologica*, 129(3), 387-398. <https://doi.org/10.1016/j.actpsy.2008.09.005>
- Borbély, A. A. (1982). A two process model of sleep regulation. *Human Neurobiology*, 1, 195-204.
- Braem, S. (2017). Conditioning task switching behavior. *Cognition*, 166, 272-276. <https://doi.org/10.1016/j.cognition.2017.05.037>
- Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews*, 6(1). <https://doi.org/10.1186/s13643-017-0644-y>

- Brehm, L., & Alday, P. M. (2022). Contrast coding choices in a decade of mixed models. *Journal of Memory and Language*, 125, 104334. <https://doi.org/10.1016/j.jml.2022.104334>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920960351. <https://doi.org/10.1177/2515245920960351>
- Burgess, H. J., Legasto, C. S., Fogg, L. F., & Smith, M. R. (2013). Can small shifts in circadian phase affect performance? *Applied Ergonomics*, 44(1), 109-111. <https://doi.org/10.1016/j.apergo.2012.05.007>
- Buyse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, 28(2), 193-213. [https://doi.org/10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4)
- Calamia, M., Markon, K., & Tranel, D. (2012). Scoring Higher the Second Time Around: Meta-Analyses of Practice Effects in Neuropsychological Assessment. *The Clinical Neuropsychologist*, 26(4), 543-570. <https://doi.org/10.1080/13854046.2012.680913>
- Caldwell, J. A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease*, 3(2), 85-96. <https://doi.org/10.1016/j.tmaid.2004.07.008>
- Caldwell, J. A., Caldwell, J. L., Thompson, L. A., & Lieberman, H. R. (2019). Fatigue and its management in the workplace. *Neuroscience and Biobehavioral Reviews*, 96, 272-289. <https://doi.org/10.1016/j.neubiorev.2018.10.024>
- Caldwell, J. A., Mu, Q., Smith, J. K., Mishory, A., Caldwell, J. L., Peters, G., Brown, D. L., & George, M. S. (2005). Are Individual Differences in Fatigue Vulnerability Related to Baseline Differences in Cortical Activation? *Behavioral Neuroscience*, 119(3), 694-707. <https://doi.org/10.1037/0715-7044.119.3.694>
- Caldwell, J. L., Chandler, J. F., & Hartzler, B. M. (2012). Battling fatigue in aviation: recent advancements in research and practice. *Journal of Medical Sciences*, 32(2), 47-56. [https://doi.org/10.6136/JMS.2012.32\(2\).047](https://doi.org/10.6136/JMS.2012.32(2).047)
- Campbell, M. J., Toth, A. J., Moran, A. P., Kowal, M., & Exton, C. (2018). eSports: A new window on neurocognitive expertise? In S. Marcora & M. Sarkar (Eds.), *Progress in Brain Research* (Vol. 240, pp. 161-174). Elsevier. <https://doi.org/10.1016/bs.pbr.2018.09.006>

- Capaldi, V. F., Balkin, T. J., & Mysliwiec, V. (2019). Optimizing Sleep in the Military Challenges and Opportunities. *Chest*, 155(1), 215-226. <https://doi.org/10.1016/j.chest.2018.08.1061>
- Cappuccio, F. P., D'Elia, L., Strazzullo, P., & Miller, M. A. (2010). Sleep Duration and All-Cause Mortality: A Systematic Review and Meta-Analysis of Prospective Studies. *Sleep*, 33(5), 585-592. <https://doi.org/10.1093/sleep/33.5.585>
- Cappuccio, F. P., Taggart, F. M., Kandala, N.-B., Currie, A., Peile, E., Stranges, S., & Miller, M. A. (2008). Meta-Analysis of Short Sleep Duration and Obesity in Children and Adults. *Sleep*, 31(5), 619-626. <https://doi.org/10.1093/sleep/31.5.619>
- Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The Consensus Sleep Diary: Standardizing Prospective Sleep Self-Monitoring. *Sleep*, 35(2), 287-302. <https://doi.org/10.5665/sleep.1642>
- Carrier, J., & Monk, T. H. (2000). Circadian rhythms of performance: new trends. *Chronobiology International*, 17(6), 719-732. <https://doi.org/10.1081/CBI-100102108>
- Cerulla, N., Arcusa, À., Navarro, J.-B., de la Osa, N., Garolera, M., Enero, C., Chico, G., & Fernández-Morales, L. (2019). Cognitive impairment following chemotherapy for breast cancer: The impact of practice effect on results. *Journal of Clinical and Experimental Neuropsychology*, 41(3), 290-299. <https://doi.org/10.1080/13803395.2018.1546381>
- Chaumet, G., Taillard, J., Sagaspe, P., Pagani, M., Dinges, D. F., Pavy-Le-Traon, A., Bareille, M.-P., Rascol, O., & Philip, P. (2009). Confinement and Sleep Deprivation Effects on Propensity to Take Risks. *Aviation, Space, and Environmental Medicine*, 80(2), 73-80. <https://doi.org/10.3357/ASEM.2366.2009>
- Cheng, Y. H., Roach, G. D., & Petrilli, R. M. A. (2014). Current and future directions in clinical fatigue management: An update for emergency medicine practitioners. *Emergency Medicine Australasia*, 26(6), 640-644. <https://doi.org/10.1111/1742-6723.12319>
- Chihiro, A., Yousuke, K., Yul-Wan, S., & Seiji, O. (2017). The effects of short-term practice on executive functions. *Tohoku psychological folia*, 75, 15-21. <https://cir.nii.ac.jp/crid/1050001202738712320>
- Chinoy, E. D., Cuellar, J. A., Huwa, K. E., Jameson, J. T., Watson, C. H., Bessman, S. C., Hirsch, D. A., Cooper, A. D., Drummond, S. P. A., & Markwald, R. R. (2021).

- Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*, 44(5). <https://doi.org/10.1093/sleep/zsaa291>
- Chinoy, E. D., Cuellar, J. A., Jameson, J. T., & Markwald, R. R. (2022). Performance of Four Commercial Wearable Sleep-Tracking Devices Tested Under Unrestricted Conditions at Home in Healthy Young Adults. *Nature and Science of Sleep*, 14, 493-516. <https://doi.org/10.2147/nss.S348795>
- Chitayat, A. P., Block, F., Walker, J., & Drachen, A. (2023). Beyond the Meta: Leveraging Game Design Parameters for Patch-Agnostic Esport Analytics. *arXiv preprint arXiv:2305.18477*. <https://doi.org/10.48550/arXiv.2305.18477>
- Chiu, H.-Y., Lee, H.-C., Chen, P.-Y., Lai, Y.-F., & Tu, Y.-K. (2018). Associations between sleep duration and suicidality in adolescents: A systematic review and dose-response meta-analysis. *Sleep Medicine Reviews*, 42, 119-126. <https://doi.org/10.1016/j.smrv.2018.07.003>
- Chuah, Y. L., Venkatraman, V., Dinges, D. F., & Chee, M. W. (2006). The neural basis of interindividual variability in inhibitory efficiency after sleep deprivation. *Journal of Neuroscience*, 26(27), 7156-7162. <https://doi.org/10.1523/JNEUROSCI.0906-06.2006>
- Coates, D., & Parshakov, P. (2016). Team vs. individual tournaments: evidence from prize structure in eSports. *Higher School of Economics Research Paper No. WP BRP*, 138.
- Code Red Esports. (2017). *Sports teams in Esports by Redeye* Google Sheets. [https://docs.google.com/spreadsheets/d/1qkr79b8V3GR2Uk2ZYCzeccm7oWWeNV72TZRKOL\\_Khbg/edit#gid=0](https://docs.google.com/spreadsheets/d/1qkr79b8V3GR2Uk2ZYCzeccm7oWWeNV72TZRKOL_Khbg/edit#gid=0)
- Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic Sleep/Wake Identification From Wrist Activity. *Sleep*, 15(5), 461-469. <https://doi.org/10.1093/sleep/15.5.461>
- Couyoumdjian, A., Sdoia, S., Tempesta, D., Curcio, G., Rastellini, E., De Gennaro, L., & Ferrara, M. (2010). The effects of sleep and sleep deprivation on task-switching performance. *Journal of sleep research*, 19(1 Pt 1), 64-70. <https://doi.org/10.1111/j.1365-2869.2009.00774.x>
- Cranmer, E. E., Han, D.-I. D., van Gisbergen, M., & Jung, T. (2021). Esports matrix: Structuring the esports research agenda. *Computers in Human Behavior*, 117, 106671. <https://doi.org/10.1016/j.chb.2020.106671>

- Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random Forests For Classification In Ecology. *Ecology*, 88(11), 2783-2792. <https://doi.org/10.1890/07-0539.1>
- da Silva, A. A., de Mello, R. G. B., Schaan, C. W., Fuchs, F., D. , Redline, S., & Fuchs, S., C. . (2016). Sleep duration and mortality in the elderly: a systematic review with meta-analysis. *BMJ Open*, 6(2), e008119. <https://doi.org/10.1136/bmjopen-2015-008119>
- Daaloul, H., Souissi, N., & Davenne, D. (2019). Effects of Napping on Alertness, Cognitive, and Physical Outcomes of Karate Athletes. *Medicine and science in sports and exercise*, 51(2), 338-345. <https://doi.org/10.1249/MSS.0000000000001786>
- Darvin, L., Holden, J., Wells, J., & Baker, T. (2021). Breaking the glass monitor: examining the underrepresentation of women in esports environments. *Sport Management Review*, 24(3), 475-499. <https://doi.org/10.1080/14413523.2021.1891746>
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: an example of publication bias? *Psychological Science*, 26(1), 99-107. <https://doi.org/10.1177/0956797614557866>
- de Bruin, E. J., van Run, C., Staaks, J., & Meijer, A. M. (2017). Effects of sleep manipulation on cognitive functioning of adolescents: a systematic review. *Sleep Medicine Reviews*, 32, 45-57. <https://doi.org/10.1016/j.smrv.2016.02.006>
- de Vet, H. C. W., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, 59(10), 1033-1039. <https://doi.org/10.1016/j.jclinepi.2005.10.015>
- DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science*, 4(1), 1-15. <https://doi.org/10.1177/2515245920965119>
- Declerck, M., Grainger, J., Koch, I., & Philipp, A. M. (2017). Is language control just a form of executive control? Evidence for overlapping processes in language switching and task switching. *Journal of Memory and Language*, 95, 138-145. <https://doi.org/10.1016/j.jml.2017.03.005>
- Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, 9(4), 673-690. [https://doi.org/10.1016/0013-4694\(57\)90088-3](https://doi.org/10.1016/0013-4694(57)90088-3)

- Dewald, J. F., Meijer, A. M., Oort, F. J., Kerkhof, G. A., & Bögels, S. M. (2010). The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. *Sleep Medicine Reviews*, 14(3), 179-189. <https://doi.org/10.1016/j.smr.2009.10.004>
- Di, J., Demanuele, C., Kettermann, A., Karahanoglu, F. I., Cappelleri, J. C., Potter, A., Bury, D., Cedarbaum, J. M., & Byrom, B. (2022). Considerations to address missing data when deriving clinical trial endpoints from digital health technologies. *Contemporary Clinical Trials*, 113, 106661. <https://doi.org/10.1016/j.cct.2021.106661>
- Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1), 3. <https://doi.org/10.1186/1471-2105-7-3>
- Dijk, D.-J., Duffy, J. F., & Czeisler, C. A. (1992). Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *Journal of sleep research*, 1(2), 112-117. <https://doi.org/10.1111/j.1365-2869.1992.tb00021.x>
- Dinges, D. F., & Kribbs, N. B. (1991). Performing while sleepy: Effects of experimentally-induced sleepiness. In T. H. Monk (Ed.), *Sleep, sleepiness and performance*. (pp. 97-128). John Wiley & Sons.
- Dinges, D. F., & Powell, J. (1988). Sleepiness is more than lapsing. *Sleep Research*, 17, 84.
- Dinges, D. F., & Powell, J. (1989). Sleepiness impairs optimum response capability. *Sleep Research*, 18, 366.
- Dobrowolski, P., Hanusz, K., Sobczyk, B., Skorko, M., & Wiatrow, A. (2015). Cognitive enhancement in video game players: The role of video game genre. *Computers in Human Behavior*, 44, 59-63. <https://doi.org/10.1016/j.chb.2014.11.051>
- Doherty, R., Madigan, S. M., Nevill, A., Warrington, G., & Ellis, J. G. (2021). The Sleep and Recovery Practices of Athletes. *Nutrients*, 13(4), 1330. <https://doi.org/10.3390/nu13041330>
- Doran, S. M., Van Dongen, H. P., & Dinges, D. F. (2001). Sustained attention performance during sleep deprivation: evidence of state instability. *Archives Italiennes de Biologie*, 139(3), 253-267. <https://doi.org/10.4449/aib.v139i3.503>
- Dorrian, J., Rogers, N. L., & Dinges, D. F. (2004). Psychomotor vigilance performance: Neurocognitive assay sensitive to sleep loss. In C. A. Kushida (Ed.), *Sleep deprivation* (1 ed., Vol. 193, pp. 39-70). CRC Press.

- Driller, M. W., McQuillan, J., & O'Donnell, S. (2016). Inter-device reliability of an automatic-scoring actigraph for measuring sleep in healthy adults. *Sleep Science*, 9(3), 198-201. <https://doi.org/10.1016/j.slsci.2016.08.003>
- Drummond, S. P. A., Brown, G. G., Gillin, J. C., Stricker, J. L., Wong, E. C., & Buxton, R. B. (2000). Altered brain response to verbal learning following sleep deprivation. *Nature*, 403(6770), 655-657. <https://doi.org/10.1038/35001068>
- Drummond, S. P. A., Brown, G. G., Salamat, J. S., & Gillin, J. C. (2004). Increasing task difficulty facilitates the cerebral compensatory response to total sleep deprivation. *Sleep*, 27(3), 445-451. <https://doi.org/10.1093/sleep/27.3.445>
- Drummond, S. P. A., Meloy, M., Yanagi, M. A., Orff, H. J., & Brown, G. G. (2005). Compensatory recruitment after sleep deprivation and the relationship with performance. *Psychiatry Research: Neuroimaging*, 140(3), 211-223. <https://doi.org/10.1016/j.psychresns.2005.06.007>
- Duff, K., Lyketsos, C. G., Beglinger, L. J., Chelune, G., Moser, D. J., Arndt, S., Schultz, S. K., Paulsen, J. S., Petersen, R. C., & McCaffrey, R. J. (2011). Practice effects predict cognitive outcome in amnesic mild cognitive impairment. *The American Journal of Geriatric Psychiatry*, 19(11), 932-939. <https://doi.org/10.1097/JGP.0b013e318209dd3a>
- Dunican, I. C., Higgins, C. C., Jones, M. J., Clarke, M. W., Murray, K., Dawson, B., Caldwell, J. A., Halson, S. L., & Eastwood, P. R. (2018). Caffeine use in a Super Rugby game and its relationship to post-game sleep. *European Journal of Sport Science*, 18(4), 513-523. <https://doi.org/10.1080/17461391.2018.1433238>
- Dunican, I. C., Higgins, C. C., Murray, K., Jones, M. J., Dawson, B., Caldwell, J. A., Halson, S. L., & Eastwood, P. R. (2019). Sleep Patterns and Alertness in an Elite Super Rugby Team During a Game Week. *Journal of human kinetics*, 67, 111-121. <https://doi.org/10.2478/hukin-2018-0088>
- Dunican, I. C., Murray, K., Slater, J. A., Maddison, K. J., Jones, M. J., Dawson, B., Straker, L. M., Caldwell, J. A., Halson, S. L., & Eastwood, P. R. (2018). Laboratory and home comparison of wrist-activity monitors and polysomnography in middle-aged adults. *Sleep and Biological Rhythms*, 16(1), 85-97. <https://doi.org/10.1007/s41105-017-0130-x>
- Dunican, I. C., Perry, E., Maissey, G., Nesci, E., & Roberts, S. (2023). Sleep hygiene education for improving sleep in ultra-marathon swimmers: Guidance for coaches and swimmers. *International Journal of Sports Science & Coaching*, 1018-1024. <https://doi.org/10.1177/17479541231152196>

- Dunnett, C. W. (1955). A Multiple Comparison Procedure for Comparing Several Treatments with a Control. *Journal of the American Statistical Association*, 50(272), 1096-1121. <https://doi.org/10.1080/01621459.1955.10501294>
- Durmer, J. S., & Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology*, 25(1), 117-129. <https://doi.org/10.1055/s-2005-867080>
- Eckart, C., Kraft, D., & Fiebach, C. J. (2021). Internal consistency and test–retest reliability of an affective task-switching paradigm. *Emotion*, 21(5), 921-931. <http://doi.org/10.1037/emo0000972>
- Edgar, D. T., Beaven, C. M., Gill, N. D., Zaslona, J. L., & Driller, M. W. (2023). Operation early-bird: Investigating altered light exposure in military barracks on sleep and performance—a placebo-controlled study. *Journal of sleep research*, 32(4), e13837. <https://doi.org/10.1111/jsr.13837>
- Elman, J. A., Jak, A. J., Panizzon, M. S., Tu, X. M., Chen, T., Reynolds, C. A., Gustavson, D. E., Franz, C. E., Hatton, S. N., Jacobson, K. C., Toomey, R., McKenzie, R., Xian, H., Lyons, M. J., & Kremen, W. S. (2018). Underdiagnosis of mild cognitive impairment: A consequence of ignoring practice effects. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 372-381. <http://doi.org/10.1016/j.dadm.2018.04.003>
- Elsmore, T. F. (1994). SYNWORK1: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments, & Computers*, 26(4), 421-426. <https://doi.org/10.3758/BF03204659>
- Enders, C. K. (2003). Performing Multivariate Group Comparisons Following a Statistically Significant MANOVA. *Measurement and Evaluation in Counseling and Development*, 36(1), 40-56. <http://doi.org/10.1080/07481756.2003.12069079>
- Engle-Friedman, M., Palencar, V., & Riela, S. (2010). Sleep and effort in adolescent athletes. *J Child Health Care*, 14(2), 131-141. <https://doi.org/10.1177/1367493510362129>
- Engle-Friedman, M., Riela, S., Golan, R., Ventuneac, A. M., Davis, C. M., Jefferson, A. D., & Major, D. (2003). The effect of sleep loss on next day effort. *Journal of sleep research*, 12(2), 113-124. <https://doi.org/10.1046/j.1365-2869.2003.00351.x>
- Englund, C. E., Ryman, D. H., Naitoh, P., & Hodgdon, J. A. (1985). Cognitive performance during successive sustained physical work episodes. *Behavior Research Methods, Instruments & Computers*, 17(1), 75-85. <http://doi.org/10.3758/BF03200899>

- Esports Earnings. (2021). *Top Games Awarding Prize Money*. Esports Earnings. Retrieved 12/03/2021 from <https://www.esportsearnings.com/players>
- Esports Earnings. (2023a). *Top 200 Highest Overall Earnings*. Esports Earnings. Retrieved 20/06/2023 from <https://www.esportsearnings.com/players/highest-earnings-top-200>
- Esports Earnings. (2023b). *Top Games Awarding Prize Money*. Retrieved 01/09/2023 from <https://www.esportsearnings.com/games>
- Evans, J. S., & Murphy, M. A. (2018). *rfUtilities*. In <https://cran.r-project.org/package=rfUtilities>
- Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), 159. <http://doi.org/10.1186/s12966-015-0314-1>
- Farahat, F. M., Rohlman, D. S., Storzbach, D., Ammerman, T., & Anger, W. K. (2003). Measures of Short-Term Test–Retest Reliability of Computerized Neurobehavioral Tests. *NeuroToxicology*, 24(4-5), 513-521. [http://doi.org/10.1016/S0161-813X\(03\)00079-2](http://doi.org/10.1016/S0161-813X(03)00079-2)
- Fatima, Y., Doi, S. A. R., & Mamun, A. A. (2016). Sleep quality and obesity in young subjects: a meta-analysis. *Obesity Reviews*, 17(11), 1154-1166. <https://doi.org/10.1111/obr.12444>
- Flaa, T. A., Bjorvatn, B., Pallesen, S., Røislien, J., Zakariassen, E., Harris, A., & Waage, S. (2021). Subjective and objective sleep among air ambulance personnel. *Chronobiology International*, 38(1), 129-139. <http://doi.org/10.1080/07420528.2020.1802288>
- Forberg, K., Waage, S., Moen, B., & Bjorvatn, B. (2010). Subjective and objective sleep and sleepiness among tunnel workers in an extreme and isolated environment: 10-h shifts, 21-day working period, at 78 degrees north. *Sleep Medicine*, 11(2), 185-190. <https://doi.org/10.1016/j.sleep.2009.07.015>
- Foster, R. G., & Kreitzman, L. (2014). The rhythms of life: what your body clock means to you! *Experimental Physiology*, 99(4), 599-606. <https://doi.org/10.1113/expphysiol.2012.071118>
- Fox, J. L., Stanton, R., Scanlan, A. T., Teramoto, M., & Sargent, C. (2021). The Association Between Sleep and In-Game Performance in Basketball Players. *International Journal of Sports Physiology and Performance*, 16(3), 333-341. <http://doi.org/10.1123/ijsp.2020-0025>

- Franken, P., Chollet, D., & Tafti, M. (2001). The Homeostatic Regulation of Sleep Need Is under Genetic Control. *The Journal of Neuroscience*, 21(8), 2610. [10.1523/JNEUROSCI.21-08-02610.2001](https://doi.org/10.1523/JNEUROSCI.21-08-02610.2001)
- Franks, R. R., Jr., & King, D. (2023). Esports and the Esports Athlete—Simply Misnomer Terms, or Are They the Real Deal? *Clinical Journal of Sport Medicine*, 33(2). <https://doi.org/10.1097/JSM.0000000000001120>
- Frestad, D., & Prescott, E. (2017). Vital Exhaustion and Coronary Heart Disease Risk: A Systematic Review and Meta-Analysis. *Psychosomatic Medicine*, 79(3). <http://doi.org/10.1097/PSY.0000000000000423>
- Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, 137(2), 201-225. <http://doi.org/10.1037/0096-3445.137.2.201>
- Fullagar, H. H. K., Skorski, S., Duffield, R., Hammes, D., Coutts, A. J., & Meyer, T. (2015). Sleep and Athletic Performance: The Effects of Sleep Loss on Exercise Performance, and Physiological and Cognitive Responses to Exercise. *Sports Medicine*, 45(2), 161-186. <http://doi.org/10.1007/s40279-014-0260-0>
- Fuster-Garcia, E., Juan-Albarracín, J., Bresó, A., & García-Gómez, J. M. (2013). Monitoring changes in daily actigraphy patterns of free-living patients. International Work Conference on Bioinformatics and Biomedical Engineering (IWBBIO),
- Gallicchio, L., & Kalesan, B. (2009). Sleep duration and mortality: a systematic review and meta-analysis. *Journal of sleep research*, 18(2), 148-158. <https://doi.org/10.1111/j.1365-2869.2008.00732.x>
- García, J., Ibáñez, S. J., Martínez De Santos, R., Leite, N., & Sampaio, J. (2013). Identifying Basketball Performance Indicators in Regular Season and Playoff Games. *Journal of human kinetics*, 36(1), 161-168. <http://doi.org/10.2478/hukin-2013-0016>
- García, J., & Murillo, C. (2020). Sports video games participation: what can we learn for esports? *Sport, Business and Management: An International Journal*, 10(2), 169-185. <http://doi.org/10.1108/SBM-01-2019-0006>
- Gillberg, M., & Åkerstedt, T. (1994). Sleep restriction and SWS-suppression: effects on daytime alertness and night-time recovery. *Journal of sleep research*, 3(3), 144-151. <http://doi.org/10.1111/j.1365-2869.1994.tb00121.x>

- Glenville, M., Broughton, R., Wing, A. M., & Wilkinson, R. T. (1978). Effects of Sleep Deprivation on Short Duration Performance Measures Compared to the Wilkinson Auditory Vigilance Task. *Sleep*, 1(2), 169-176. <https://doi.org/10.1093/sleep/1.2.169>
- Gomes, M. A., Narciso, F. V., de Mello, M. T., & Esteves, A. M. (2021). Identifying electronic-sport athletes' sleep-wake cycle characteristics. *Chronobiology International*, 38(7), 1002-1009. <https://doi.org/10.1080/07420528.2021.1903480>
- Goulart, J. B., Aitken, L. S., Siddiqui, S., Cuevas, M., Cardenas, J., Beathard, K. M., & Riechman, S. E. (2023). Nutrition, lifestyle, and cognitive performance in esports athletes. *Frontiers in Nutrition*, 10. <http://doi.org/10.3389/fnut.2023.1120303>
- Grace, A. A. (2000). The tonic/phasic model of dopamine system regulation and its implications for understanding alcohol and psychostimulant craving. *Addiction*, 95(8s2), S119-S128. <https://doi.org/10.1046/j.1360-0443.95.8s2.1.x>
- Grandou, C., Wallace, L., Fullagar, H. H. K., Duffield, R., & Burley, S. (2019). The Effects of Sleep Loss on Military Physical Performance. *Sports Medicine*, 49(8), 1159-1172. <http://doi.org/10.1007/s40279-019-01123-8>
- Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S. M., & Van Dongen, H. P. A. (2017). 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. *Behavior Research Methods*, 49(3), 1020-1029. <https://doi.org/10.3758/s13428-016-0763-8>
- Green, A., Cohen-Zion, M., Haim, A., & Dagan, Y. (2017). Evening light exposure to computer screens disrupts human sleep, biological rhythms, and attention abilities. *Chronobiology International*, 34(7), 855-865. <https://doi.org/10.1080/07420528.2017.1324878>
- Green, C. S., Sugarman, M. A., Medford, K., Klobusicky, E., & Bavelier, D. (2012). The effect of action video game experience on task-switching. *Computers in Human Behavior*, 28(3), 984-994. <http://doi.org/10.1016/j.chb.2011.12.020>
- Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, 83(2), 314-320. <http://doi.org/10.1037/0033-2909.83.2.314>
- Gu, W., Saaty, T. L., & Whitaker, R. (2016). Expert System for Ice Hockey Game Prediction: Data Mining with Human Judgment. *International Journal of Information Technology & Decision Making*, 15(04), 763-789. <http://doi.org/10.1142/S0219622016400022>

- Guarana, C. L., & Barnes, C. M. (2017). Lack of sleep and the development of leader-follower relationships over time. *Organizational Behavior and Human Decision Processes*, 141, 57-73. <https://doi.org/10.1016/j.obhdp.2017.04.003>
- Gupta, A., Huettner, D. P., & Dukewich, M. (2017). Comparative Effectiveness Review of Cooled Versus Pulsed Radiofrequency Ablation for the Treatment of Knee Osteoarthritis: A Systematic Review. *Pain Physician*, 20(3), 155-171. <https://www.painphysicianjournal.com/linkout?issn=&vol=20&page=155>
- Gupta, L., Morgan, K., & Gilchrist, S. (2017). Does elite sport degrade sleep quality? A systematic review. *Sports Medicine*, 47(7), 1317-1333. <http://doi.org/10.1007/s40279-016-0650-6>
- Haahr, M. (2021). *RANDOM.ORG: True Random Number Service*. Randomness and Integrity Services Ltd. Retrieved 20/12/2021 from <https://www.random.org>
- Hallmann, K., & Giel, T. (2018). eSports – Competitive sports or recreational activity? *Sport Management Review*, 21(1), 14-20. <https://doi.org/10.1016/j.smr.2017.07.011>
- Halson, S. L., Johnston, R. D., Piromalli, L., Lalor, B. J., Cormack, S., Roach, G. D., & Sargent, C. (2022). Sleep Regularity and Predictors of Sleep Efficiency and Sleep Duration in Elite Team Sport Athletes. *Sports Medicine - Open*, 8(1), 79. <https://doi.org/10.1186/s40798-022-00470-7>
- Hamari, J., & Sjöblom, M. (2017). What is eSports and why do people watch it? *Internet Research*, 27(2), 211-232. <http://doi.org/10.1108/IntR-04-2016-0085>
- Harris, L. M., Huang, X., Linthicum, K. P., Bryen, C. P., & Ribeiro, J. D. (2020). Sleep disturbances as risk factors for suicidal thoughts and behaviours: a meta-analysis of longitudinal studies. *Scientific Reports*, 10(1), 13888. <http://doi.org/10.1038/s41598-020-70866-6>
- Harrison, Y., & Horne, J. A. (1999). One Night of Sleep Loss Impairs Innovative Thinking and Flexible Decision Making. *Organizational Behavior and Human Decision Processes*, 78(2), 128-145. <https://doi.org/10.1006/obhd.1999.2827>
- Harrison, Y., & Horne, J. A. (2000). The impact of sleep deprivation on decision making: a review. *Journal of Experimental Psychology: Applied*, 6(3), 236-249. <https://doi.org/10.1037//1076-898x.6.3.236>
- Hartig, F. (2022). *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. In R package <https://CRAN.R-project.org/package=DHARMA>

- Hartzler, B. M., Chandler, J. F., Levin, C. B., & Turnmire, A. E. (2015). *Predicting performance during chronic sleep loss: identification of factors sensitive to individual fatigue resistance* (NAVAL MEDICAL RESEARCH UNIT (DAYTON) WRIGHT-PATTERSON AFB OH). N. M. R. U. D. W.-P. A. OH.
- Haslam, D. R. (1985). Sustained operations and military performance. *Behavior Research Methods, Instruments & Computers*, 17(1), 90-95. <http://doi.org/10.3758/BF03200901>
- Hayter, A. J. (1986). The Maximum Familywise Error Rate of Fisher's Least Significant Difference Test. *Journal of the American Statistical Association*, 81(396), 1000-1004. <http://doi.org/10.1080/01621459.1986.10478364>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166-1186. <http://doi.org/10.3758/s13428-017-0935-1>
- Higuchi, S., Motohashi, Y., Liu, Y., & Maeda, A. (2005). Effects of playing a computer game using a bright display on presleep physiological variables, sleep latency, slow wave sleep and REM sleep. *Journal of sleep research*, 14(3), 267-273. <https://doi.org/10.1111/j.1365-2869.2005.00463.x>
- Hilbert, J. (2019). *Gaming & gender: how inclusive are eSports?* The Sports Integrity Initiative. Retrieved 11/05/2023 from <https://www.sportsintegrityinitiative.com/gaming-gender-how-inclusive-are-esports/>
- Hilburn, B. (2004). Cognitive complexity in air traffic control: A literature review. *EEC note*, 4(04).
- Hindi, M. (2020). *Almost 2 million concurrent Rocket League players achieved in first FTP weekend.* Retrieved 05/02/2021 from <https://rocketleague.tracker.network/rocket-league/articles/almost-2-million-concurrent-rocket-leauge-players-in-first-ftp-weekend>
- Hirshkowitz, M., Whiton, K., Albert, S. M., Alessi, C., Bruni, O., DonCarlos, L., Hazen, N., Herman, J., Katz, E. S., Kheirandish-Gozal, L., Neubauer, D. N., O'Donnell, A. E., Ohayon, M., Peever, J., Rawding, R., Sachdeva, R. C., Setters, B., Vitiello, M. V., Ware, J. C., & Adams Hillard, P. J. (2015). National Sleep Foundation's sleep time duration recommendations: methodology and results summary. *Sleep Health*, 1(1), 40-43. <https://doi.org/10.1016/j.sleh.2014.12.010>

- Hitar-García, J. A., Morán-Fernández, L., & Bolón-Canedo, V. (2023). Machine Learning Methods for Predicting League of Legends Game Outcome. *IEEE Transactions on Games*, 15(2), 171-181. <http://doi.org/10.1109/TG.2022.3153086>
- Hodgson, R., Alwyn, T., John, B., Thom, B., & Smith, A. (2002). The Fast Alcohol Screening Test. *Alcohol and Alcoholism*, 37(1), 61-66. <https://doi.org/10.1093/alcalc/37.1.61>
- Hojaji, F., Toth, A. J., & Campbell, M. J. (2023). A Machine Learning Approach for Modeling and Analyzing of Driver Performance in Simulated Racing. *Artificial Intelligence and Cognitive Science*, Cham.
- Holden, J. T., Kaburakis, A., & Rodenberg, R. (2017). The Future Is Now: Esports Policy Considerations and Potential Litigation. *Journal of Legal Aspects of Sport*, 27(1), 46-78. <https://doi.org/10.1123/jlas.2016-0018> (78)
- Holding, B. C., Sundelin, T., Lekander, M., & Axelsson, J. (2019). Sleep deprivation and its effects on communication during individual and collaborative tasks. *Scientific Reports*, 9(1), 3131. <http://doi.org/10.1038/s41598-019-39271-6>
- Honn, K. A., Hinson, J. M., Whitney, P., & Van Dongen, H. P. A. (2019). Cognitive flexibility: A distinct element of performance impairment due to sleep deprivation. *Accident Analysis & Prevention*, 126, 191-197. <https://doi.org/10.1016/j.aap.2018.02.013>
- Honn, K. A., & Van Dongen, H. P. A. (2023). Paradoxical effects from stimulus density manipulation provide new insight into the impact of sleep deprivation on PVT performance. *SLEEP Advances*, 4(1), zpac045. <https://doi.org/10.1093/sleepadvances/zpac045>
- Hooker, G., & Mentch, L. (2019). Please Stop Permuting Features: An Explanation and Alternatives. *arXiv e-prints*, arXiv:1905.03151. <https://ui.adsabs.harvard.edu/abs/2019arXiv190503151H>
- Hopkins, W. G. (2000). Measures of Reliability in Sports Medicine and Science. *Sports Medicine*, 30(1), 1-15. <https://doi.org/10.2165/00007256-200030010-00001>
- Horne, J. A., & Minard, A. (1985). Sleep and sleepiness following a behaviourally 'active' day. *Ergonomics*, 28(3), 567-575. 10.1080/00140138508963171
- Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4, 97-110. <https://europepmc.org/article/med/1027738>

- Horne, J. A., & Pettitt, A. N. (1985). High incentive effects on vigilance performance during 72 hours of total sleep deprivation. *Acta Psychologica*, 58(2), 123-139. [https://doi.org/10.1016/0001-6918\(85\)90003-4](https://doi.org/10.1016/0001-6918(85)90003-4)
- How, J. M., Foo, S. C., Low, E., Wong, T. M., Vijayan, A., Siew, M. G., & Kanapathy, R. (1994). Effects of sleep deprivation on performance of Naval seamen: I. Total sleep deprivation on performance. *Annals of the Academy of Medicine, Singapore*, 23(5), 669-675. <http://europepmc.org/abstract/MED/7847745>
- Huberty, C. J., & Morris, J. D. (1992). Multivariate analysis versus multiple univariate analyses. In A. E. Kazdin (Ed.), *Methodological issues & strategies in clinical research*. (pp. 351-365). American Psychological Association. <https://doi.org/10.1037/10109-030>
- Hudson, A. N., Van Dongen, H. P. A., & Honn, K. A. (2020). Sleep deprivation, vigilant attention, and brain function: a review. *Neuropsychopharmacology*, 45(1), 21-30. <https://doi.org/10.1038/s41386-019-0432-6>
- Hughes, A., Barnes, A., Churchill, S. M., & Stone, J. A. (2017). Performance indicators that discriminate winning and losing in elite men's and women's Rugby Union. *International Journal of Performance Analysis in Sport*, 17(4), 534-544. <https://doi.org/10.1080/24748668.2017.1366759>
- Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, 20(10), 739-754. <https://doi.org/10.1080/026404102320675602>
- Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods*, 46(3), 702-721. <https://doi.org/10.3758/s13428-013-0411-5>
- Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W. F., Miller, J. C., & Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation Space and Environmental Medicine*, 75(3), A44-A53. <https://www.ingentaconnect.com/content/asma/asem/2004/00000075/a00103s1/art00005>
- IBM Corp. (2021). *IBM SPSS Statistics for Windows, Version 28.0*. In IBM Corp.
- Independent Electronic Sports. (2023). Catalogue of Competitive Video Games. In Independent Electronic Sports (Ed.), *Electronic Sports Compendium* (pp. 3-36). Independent Electronic Sports. <https://consulting.independent-esports.com/wp->

- Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, 30(2), 190-200. <https://doi.org/10.1016/j.newideapsych.2011.11.001>
- Irwin, M. R., Olmstead, R., & Carroll, J. E. (2016). Sleep Disturbance, Sleep Duration, and Inflammation: A Systematic Review and Meta-Analysis of Cohort Studies and Experimental Sleep Deprivation. *Biological Psychiatry*, 80(1), 40-52. <https://doi.org/10.1016/j.biopsych.2015.05.014>
- Itani, O., Jike, M., Watanabe, N., & Kaneita, Y. (2017). Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression. *Sleep Medicine*, 32, 246-256. <https://doi.org/10.1016/j.sleep.2016.08.006>
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446. <https://doi.org/10.1016/j.jml.2007.11.007>
- Jaguszewski, M. (2020). Increasing role of three-point field goals in National Basketball Association. *Trends in Sport Sciences*, 27(1), 5-11. <https://doi.org/10.23829/TSS.2020.27.1-1>
- James, L., Smart, D., Odom-Maryon, T., Honn, K. A., & Rowan, S. (2019). Sleep deprivation in Air National Guard medical personnel responding to simulated disaster-training exercises. *Military Psychology*, 31(2), 138-146. <https://doi.org/10.1080/08995605.2019.1565909>
- Janelle, C. M., & Hillman, C. (2003). Expert performance in sport: Current Perspectives and Critical Issues. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (pp. 19-47). Human Kinetics.
- Jang, J.-H., Choi, J., Roh, H. W., Son, S. J., Hong, C. H., Kim, E. Y., Kim, T. Y., & Yoon, D. (2020). Deep Learning Approach for Imputation of Missing Values in Actigraphy Data: Algorithm Development Study. *JMIR Mhealth Uhealth*, 8(7), e16113. <https://doi.org/10.2196/16113>
- Jang, W., & Byon, K. K. (2020). Antecedents of esports gameplay intention: Genre as a moderator. *Computers in Human Behavior*, 109, 106336. <https://doi.org/10.1016/j.chb.2020.106336>
- Jenny, S. E., Manning, R. D., Keiper, M. C., & Olrich, T. W. (2017). Virtual(ly) Athletes: Where eSports Fit Within the Definition of “Sport”. *Quest*, 69(1), 1-18. <https://doi.org/10.1080/00336297.2016.1144517>

- Jewett, M. E., Dijk, D.-J., Kronauer, R. E., & Dinges, D. F. (1999). Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep*, 22(2), 171-179. <https://doi.org/10.1093/sleep/22.2.171>
- Johnson, K. G., & Johnson, D. C. (2010). Frequency of Sleep Apnea in Stroke and TIA Patients: A Meta-analysis. *Journal of Clinical Sleep Medicine*, 06(02), 131-137. <https://doi.org/10.5664/jcsm.27760>
- Johnson, M. P., Duffy, J. F., Dijk, D. J., Ronda, J. M., Dyal, C. M., & Czeisler, C. A. (1992). Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature. *Journal of sleep research*, 1(1), 24-29. <https://doi.org/10.1111/j.1365-2869.1992.tb00004.x>
- Jonasson, K., & Thiborg, J. (2010). Electronic sport and its impact on future sport. *Sport in Society*, 13(2), 287-299. <https://doi.org/10.1080/17430430903522996>
- Judd, C. M., McClelland, G. H., & Ryan, C. S. (2009). *Data analysis: A model comparison approach*, 2nd ed. Routledge/Taylor & Francis Group.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54-69. <https://doi.org/10.1037/a0028347>
- Juliff, L. E., Halson, S. L., & Peiffer, J. J. (2015). Understanding sleep disturbance in athletes prior to important competitions. *Journal of Science and Medicine in Sport*, 18(1), 13-18. <https://doi.org/10.1016/j.jsams.2014.02.007>
- Kahn, M., Fridenson, S., Lerer, R., Bar-Haim, Y., & Sadeh, A. (2014). Effects of one night of induced night-wakings versus sleep restriction on sustained attention and mood: a pilot study. *Sleep Medicine*, 15(7), 825-832. <https://doi.org/10.1016/j.sleep.2014.03.016>
- Kaplan, K. A., Talbot, L. S., Gruber, J., & Harvey, A. G. (2012). Evaluating sleep in bipolar disorder: comparison between actigraphy, polysomnography, and sleep diary. *Bipolar Disorders*, 14(8), 870-879. <https://doi.org/10.1111/bdi.12021>
- Kariv, D., Offer, I., Notzer, N., & Dolev, E. (2007). Israeli Physicians' Performance after Sleep Deprived Situations - Differences among Male and Female Hospital Residents. *Studies on Ethno-Medicine*, 1(1), 59-70. <https://doi.org/10.1080/09735070.2007.11886302>
- Kemp, C., Pienaar, P. R., & Rae, D. E. (2020). Brace yourselves: esports is coming. *South African Journal of Sports Medicine*, 32(1), 1-2. <https://doi.org/10.17159/2078-516X/2020/v32i1a7596>

- Kemp, C., Pienaar, P. R., Rosslee, D. T., Lipinska, G., Roden, L. C., & Rae, D. E. (2021). Sleep in Habitual Adult Video Gamers: A Systematic Review. *Frontiers in Neuroscience*, 15, 1-15. <https://doi.org/10.3389/fnins.2021.781351>
- Kerkhof, G. A., Geuke, M. E. H., Brouwer, A., Rijsman, R. M., Schimsheimer, R. J., & Van Kasteel, V. (2013). Holland Sleep Disorders Questionnaire: a new sleep disorders questionnaire based on the International Classification of Sleep Disorders-2. *Journal of sleep research*, 22(1), 104-107. <https://doi.org/10.1111/j.1365-2869.2012.01041.x>
- Kihlstrom, J. F. (2021). Ecological Validity and “Ecological Validity”. *Perspectives on Psychological Science*, 16(2), 466-471. <https://doi.org/10.1177/1745691620966791>
- Killgore, W. D. S. (2007). Effects of sleep deprivation and morningness-eveningness traits on risk-taking. *Psychological Reports*, 100(2), 613-626. <https://doi.org/10.2466/pr0.100.2.613-626>
- Killgore, W. D. S., Grugle, N. L., Killgore, D. B., Leavitt, B. P., Watlington, G. I., McNair, S., & Balkin, T. J. (2008). Restoration of Risk-Propensity During Sleep Deprivation: Caffeine, Dextroamphetamine, and Modafinil. *Aviation, Space, and Environmental Medicine*, 79(9), 867-874. <https://doi.org/10.3357/ASEM.2259.2008>
- Kleitman, N. (1923). STUDIES ON THE PHYSIOLOGY OF SLEEP. *American Journal of Physiology-Legacy Content*, 66(1), 67-92. <https://doi.org/10.1152/ajplegacy.1923.66.1.67>
- Klingman, K. J., Jungquist, C. R., & Perlis, M. L. (2017). Questionnaires that screen for multiple sleep disorders. *Sleep Medicine Reviews*, 32, 37-44. <https://doi.org/10.1016/j.smrv.2016.02.004>
- Knight, B. (2022, 06/05/2022). *The Most Valuable Esports Companies 2022*. Forbes. Retrieved 20/06/2023 from <https://www.forbes.com/sites/brettknight/2022/05/06/the-most-valuable-esports-companies-2022/>
- Knufinke, M., Nieuwenhuys, A., Geurts, S. A. E., Møst, E. I. S., Maase, K., Moen, M. H., Coenen, A. M. L., & Kompier, M. A. J. (2018). Train hard, sleep well? Perceived training load, sleep quantity and sleep stage distribution in elite level athletes. *Journal of Science and Medicine in Sport*, 21(4), 427-432. <https://doi.org/10.1016/j.jsams.2017.07.003>

- Kokkinakis, A., York, P., Patra, M., Robertson, J., Kirman, B., Coates, A., Pedrassoli Chitayat, A., Demediuk, S. P., Drachen, A., & Hook, J. D. (2021). Metagaming and metagames in Esports. *International Journal of Esports*. <https://www.ijesports.org/article/51/html>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Krause, A. J., Ben Simon, E., Mander, B. A., Greer, S. M., Saletin, J. M., Goldstein-Piekarski, A. N., & Walker, M. P. (2017). The sleep-deprived human brain. *Nature Reviews Neuroscience*, 18(7), 404-418. <https://doi.org/10.1038/nrn.2017.55>
- Krueger, J. M., Rector, D. M., Roy, S., Van Dongen, H. P. A., Belenky, G., & Panksepp, J. (2008). Sleep as a fundamental property of neuronal assemblies. *Nature Reviews Neuroscience*, 9(12), 910-919. <https://doi.org/10.1038/nrn2521>
- Kunorozva, L., Rae, D. E., & Roden, L. C. (2017). Chronotype distribution in professional rugby players: Evidence for the environment hypothesis? *Chronobiology International*, 34(6), 762-772. <https://doi.org/10.1080/07420528.2017.1322600>
- Lastella, M., Roach, G. D., Halson, S. L., & Sargent, C. (2015). Sleep/wake behaviours of elite athletes from individual and team sports. *European Journal of Sport Science*, 15(2), 94-100. <https://doi.org/10.1080/17461391.2014.932016>
- Lawrence-Sidebottom, D., Hinson, J. M., Whitney, P., Van Dongen, H. P. A., & Honn, K. A. (2020). Reversal learning deficits during sleep deprivation: investigating the role of information acquisition failures. *Chronobiology International*, 37(9-10), 1445-1451. <https://doi.org/10.1080/07420528.2020.1819306>
- Le Bon, O., Staner, L., Hoffmann, G., Dramaix, M., San Sebastian, I., Murphy, J. R., Kentos, M., Pelc, I., & Linkowski, P. (2001). The first-night effect may last more than one night. *Journal of Psychiatric Research*, 35(3), 165-172. [https://doi.org/10.1016/S0022-3956\(01\)00019-X](https://doi.org/10.1016/S0022-3956(01)00019-X)
- Lee, D., Kim, M.-J., & Ahn, C. W. (2021). Predicting combat outcomes and optimizing armies in StarCraft II by deep learning. *Expert Systems with Applications*, 185, 115592. <https://doi.org/10.1016/j.eswa.2021.115592>
- Lee, M. A. M., & Kleitman, N. (1923). STUDIES ON THE PHYSIOLOGY OF SLEEP. *American Journal of Physiology-Legacy Content*, 67(1), 141-152. <https://doi.org/10.1152/ajplegacy.1923.67.1.141>

- Lee, S., Bonnar, D., Kim, Y., Lee, Y., Lee, S., Gradisar, M., & Suh, S. (2020). Sleep Characteristics and Risk Factors of Korean Esports Athletes: An Exploratory Study. *Sleep Medicine Research*, 11(2), 77-87. <https://doi.org/10.17241/smr.2020.00773>
- Lee, S., Bonnar, D., Roane, B., Gradisar, M., Dunican, I. C., Lastella, M., Maissey, G., & Suh, S. (2021). Sleep Characteristics and Mood of Professional Esports Athletes: A Multi-National Study. *International Journal of Environmental Research and Public Health*, 18(2), 664. <https://doi.org/10.3390/ijerph18020664>
- Lehrer, H. M., Yao, Z., Krafty, R. T., Evans, M. A., Buysse, D. J., Kravitz, H. M., Matthews, K. A., Gold, E. B., Harlow, S. D., Samuelsson, L. B., & Hall, M. H. (2022). Comparing polysomnography, actigraphy, and sleep diary in the home environment: The Study of Women's Health Across the Nation (SWAN) Sleep Study. *SLEEP Advances*, 3(1), zpac001. <https://doi.org/10.1093/sleepadvances/zpac001>
- Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-Free Predictive Inference for Regression. *Journal of the American Statistical Association*, 113(523), 1094-1111. <https://doi.org/10.1080/01621459.2017.1307116>
- Leicht, A. S., Gomez, M. A., & Woods, C. T. (2017). Team Performance Indicators Explain Outcome during Women's Basketball Matches at the Olympic Games. *Sports*, 5(4), 96. <https://doi.org/10.3390/sports5040096>
- Levin, J. R., Serlin, R. C., & Seaman, M. A. (1994). A controlled, powerful multiple-comparison strategy for several situations. *Psychological Bulletin*, 115(1), 153-159. <https://doi.org/10.1037/0033-2909.115.1.153>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18-22. <https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf>
- Libedinsky, C., Massar, S. A. A., Ling, A., Chee, W., Huettel, S. A., & Chee, M. W. L. (2013). Sleep Deprivation Alters Effort Discounting but not Delay Discounting of Monetary Rewards. *Sleep*, 36(6), 899-904. <https://doi.org/10.5665/sleep.2720>
- Lim, J., & Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences*, 1129(1), 305-322. <http://doi.org/10.1196/annals.1417.002>
- Lim, J., & Dinges, D. F. (2010). A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychological Bulletin*, 136(3), 375-389. <https://doi.org/10.1037/a0018883>

- Liquipedia. (2022, 26/05/2023). *Gamers8 2022*. Liquipedia. Retrieved 22/03/2023 from <https://liquipedia.net/rocketleague/Gamers8/2022>
- Liquipedia. (2023, 26/05/2023). *Gamers8 2023*. Liquipedia. Retrieved 29/05/2023 from <https://liquipedia.net/rocketleague/Gamers8/2023>
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83(404), 1198-1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Liu, R. T., Steele, S. J., Hamilton, J. L., Do, Q. B. P., Furbish, K., Burke, T. A., Martinez, A. P., & Gerlus, N. (2020). Sleep and suicide: A systematic review and meta-analysis of longitudinal studies. *Clinical Psychology Review*, 81, 101895. <https://doi.org/10.1016/j.cpr.2020.101895>
- Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers*, 36(2), 339-346. <http://doi.org/10.3758/BF03195580>
- Lowe, C. J., Safati, A., & Hall, P. A. (2017). The neurocognitive consequences of sleep restriction: A meta-analytic review. *Neuroscience & Biobehavioral Reviews*, 80, 586-604. <https://doi.org/10.1016/j.neubiorev.2017.07.010>
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139. <https://doi.org/10.21105/joss.03139>
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494-1502. <http://doi.org/10.3758/s13428-016-0809-y>
- Lundahl, A., Kidwell, K. M., Van Dyk, T. R., & Nelson, T. D. (2015). A Meta-Analysis of the Effect of Experimental Sleep Restriction on Youth's Attention and Hyperactivity. *Developmental neuropsychology*, 40(3), 104-121. <http://doi.org/10.1080/87565641.2014.939183>
- Ma, N., Dinges, D. F., Basner, M., & Rao, H. Y. (2015). How Acute Total Sleep Loss Affects the Attending Brain: A Meta-Analysis of Neuroimaging Studies. *Sleep*, 38(2), 233-240. <http://doi.org/10.5665/sleep.4404>
- Mah, C. D., Mah, K. E., Kezirian, E. J., & Dement, W. C. (2011). The Effects of Sleep Extension on the Athletic Performance of Collegiate Basketball Players. *Sleep*, 34(7), 943-950. <http://doi.org/10.5665/SLEEP.1132>

- Mah, C. D., Sparks, A. J., Samaan, M. A., Souza, R. B., & Luke, A. (2019). Sleep restriction impairs maximal jump performance and joint coordination in elite athletes. *Journal of Sports Sciences*, 37(17), 1981-1988. <http://doi.org/10.1080/02640414.2019.1612504>
- Mann, D. T. Y., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: a meta-analysis. *Journal of sport & exercise psychology*, 29(4), 457-478. <http://doi.org/10.1123/jsep.29.4.457>
- Marino, M., Li, Y., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., Dulin, H., Berkman, L. F., & Buxton, O. M. (2013). Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography. *Sleep*, 36(11), 1747-1755. <http://doi.org/10.5665/sleep.3142>
- Marques, N. (2019). The Role of Breakthrough Technologies in the Growth of Esports. *IEEE Potentials*, 38(3), 24-26. <http://doi.org/10.1109/MPOT.2019.2893754>
- Massar, S. A. A., Lim, J., & Huettel, S. A. (2019). Sleep deprivation, effort allocation and performance. In H. P. A. Van Dongen, P. Whitney, J. M. Hinson, K. A. Honn, & M. W. L. Chee (Eds.), *Progress in Brain Research* (Vol. 246, pp. 1-26). Elsevier. <https://doi.org/10.1016/bs.pbr.2019.03.007>
- Massar, S. A. A., Lim, J., Sasmita, K., & Chee, M. W. L. (2019). Sleep deprivation increases the costs of attentional effort: Performance, preference and pupil size. *Neuropsychologia*, 123, 169-177. <https://doi.org/10.1016/j.neuropsychologia.2018.03.032>
- Mathew, G. M., Strayer, S. M., Bailey, D. S., Buzzell, K., Ness, K. M., Schade, M. M., Nahmod, N. G., Buxton, O. M., & Chang, A.-M. (2021). Changes in Subjective Motivation and Effort During Sleep Restriction Moderate Interindividual Differences in Attentional Performance in Healthy Young Men. *Nature and Science of Sleep*, 13, 1117-1136. <https://doi.org/10.2147/NSS.S294409>
- Matthews, K. A., Patel, S. R., Pantescio, E. J., Buysse, D. J., Kamarck, T. W., Lee, L., & Hall, M. H. (2018). Similarities and differences in estimates of sleep duration by polysomnography, actigraphy, diary, and self-reported habitual sleep in a community sample. *Sleep Health*, 4(1), 96-103. <https://doi.org/10.1016/j.sleh.2017.10.011>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305-315. <https://doi.org/10.1016/j.jml.2017.01.001>

- Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *Journal of experimental psychology. Learning, memory, and cognition*, 26(5), 1124-1140. <https://doi.org/10.1037//0278-7393.26.5.1124>
- McCaffrey, R. J., Duff, K., & Westervelt, H. J. (2000). *Practitioner's guide to evaluating change with neuropsychological assessment instruments*. Kluwer Academic Publishers.
- McCall, C., & McCall, W. V. (2012). Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. *Journal of sleep research*, 21(1), 122-127. <https://doi.org/10.1111/j.1365-2869.2011.00917.x>
- McCauley, P., Kalachev, L. V., Smith, A. D., Belenky, G., Dinges, D. F., & Van Dongen, H. P. A. (2009). A new mathematical model for the homeostatic effects of sleep loss on neurobehavioral performance. *Journal of Theoretical Biology*, 256(2), 227-239. <https://doi.org/10.1016/j.jtbi.2008.09.012>
- McIntire, L., McKinley, R. A., McIntire, J., Goodyear, C., & Nelson, J. (2013). Eye Metrics: An Alternative Vigilance Detector for Military Operators. *Military Psychology*, 25(5), 502-513. <http://doi.org/10.1037/mil0000011>
- McSweeney, A. J., Naugle, R. I., Chelune, G. J., & Lüders, H. (1993). "TScores for Change": An illustration of a regression approach to depicting change in clinical neuropsychology. *Clinical Neuropsychologist*, 7(3), 300-312. <http://doi.org/10.1080/13854049308401901>
- Meier, U. (2006). A note on the power of Fisher's least significant difference procedure. *Pharmaceutical Statistics*, 5(4), 253-263. <http://doi.org/10.1002/pst.210>
- Merrow, M., & Harrington, M. (2020). A functional context for heterogeneity of the circadian clock in cells. *PLOS Biology*, 18(10), e3000927. <http://doi.org/10.1371/journal.pbio.3000927>
- Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. <https://doi.org/10.1016/j.jml.2020.104092>
- Mikulincer, M., Babkoff, H., Caspy, T., & Sing, H. (1989). The effects of 72 hours of sleep loss on psychological variables. *British Journal of Psychology*, 80(2), 145-162. <https://doi.org/10.1111/j.2044-8295.1989.tb02309.x>
- Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review*, 20(5), 819-858. <http://doi.org/10.3758/s13423-013-0404-5>

- Miller, N. L., Shattuck, L. G., & Matsangas, P. (2007). *Fatigue in military operational environments: an annotated bibliography* N. P. S. M. CA.
- Millisecond Software. (2016). *Millisecond Test Library*. In <https://www.millisecond.com/download/library>
- Min, A., Hong, H. C., Son, S., & Scott, L. D. (2023). Overtime and alertness of rotating-shift nurses: An observational study using ecological momentary assessment. *Journal of Clinical Nursing*, 32(1-2), 199-207. <https://doi.org/10.1111/jocn.16218>
- Miner, B., & Lucey, B. P. (2021). Normal Human Sleep: An Overview. In M. H. Kryger, T. Roth, C. A. Goldstein, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine: Seventh Edition* (pp. 27-24). Elsevier.
- Moen, F., Vatn, M., Olsen, M., Haugan, J. A., & Skalicka, V. (2022). Sleep Characteristics in Esport Players and Associations With Game Performance: Residual Dynamic Structural Equation Modeling. *Frontiers in Sports and Active Living*, 3, 1-13. <https://doi.org/10.3389/fspor.2021.697535>
- Mollaveva, T., Thuraiajah, P., Burton, K., Mollaveva, S., Shapiro, C. M., & Colantonio, A. (2016). The Pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 25, 52-73. <https://doi.org/10.1016/j.smrv.2015.01.009>
- Mollicone, D. J., Van Dongen, H. P. A., Rogers, N. L., Banks, S., & Dinges, D. F. (2010). Time of day effects on neurobehavioral performance during chronic sleep restriction. *Aviation Space and Environmental Medicine*, 81(8), 735-744. <http://doi.org/10.3357/ASEM.2756.2010>
- Monk, T., Buysse, D., Reynolds Iii, C., Berga, S., Jarrett, D., Begley, A. M. Y., & Kupfer, D. (1997). Circadian rhythms in human performance and mood under constant conditions. *Journal of sleep research*, 6(1), 9-18. <https://doi.org/10.1046/j.1365-2869.1997.00023.x>
- Montaruli, A., Castelli, L., Mulè, A., Scurati, R., Esposito, F., Galasso, L., & Roveda, E. (2021). Biological rhythm and chronotype: new perspectives in health. *Biomolecules*, 11(4), 487. <https://doi.org/10.3390/biom11040487>
- Moore, B. (2020). *Rocket League surpasses CS: GO in all-time concurrent players*. sportskeeda. Retrieved 05/02/2021 from <https://www.sportskeeda.com/esports/rocket-league-surpasses-cs-go-all-time-concurrent->

[players#:~:text=Rocket%20League%20has%20surpassed%20CS,concurrent%20player%20base%20of%201%2C321%2C924](#)

- Moore, C. M., Schmiede, S. J., & Matthews, E. E. (2015). Actigraphy and Sleep Diary Measurements in Breast Cancer Survivors: Discrepancy in Selected Sleep Parameters. *Behavioral Sleep Medicine*, 13(6), 472-490. <http://doi.org/10.1080/15402002.2014.940108>
- Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., Brown, T., Chesson, A., Jr., Coleman, J., Lee-Chiong, T., Pancer, J., & Swick, T. J. (2007). Practice Parameters for the Use of Actigraphy in the Assessment of Sleep and Sleep Disorders: An Update for 2007. *Sleep*, 30(4), 519-529. <http://doi.org/10.1093/sleep/30.4.519>
- Mosey, T. J., & Mitchell, L. J. G. (2020). Key performance indicators in Australian sub-elite rugby union. *Journal of Science and Medicine in Sport*, 23(1), 35-40. <http://doi.org/10.1016/j.jsams.2019.08.014>
- Muck, R. A., Hudson, A. N., Honn, K. A., Gaddameedhi, S., & Van Dongen, H. P. A. (2022). Working around the Clock: Is a Person's Endogenous Circadian Timing for Optimal Neurobehavioral Functioning Inherently Task-Dependent? *Clocks & Sleep*, 4(1), 23-36. <http://doi.org/10.3390/clockssleep4010005>
- Mumaw, R. J., Swatzler, D., Roth, E. M., & Thomas, W. A. (1994). *Cognitive skill training for nuclear power plant operational decision making* <https://www.osti.gov/biblio/10161883>.
- Myllymaa, S., Muraja-Murro, A., Westeren-Punnonen, S., Hukkanen, T., Lappalainen, R., Mervaala, E., Töyräs, J., Sipilä, K., & Myllymaa, K. (2016). Assessment of the suitability of using a forehead EEG electrode set and chin EMG electrodes for sleep staging in polysomnography. *Journal of sleep research*, 25(6), 636-645. <https://doi.org/10.1111/jsr.12425>
- Nadeem, R., Molnar, J., Madbouly Essam, M., Nida, M., Aggarwal, S., Sajid, H., Naseem, J., & Loomba, R. (2013). Serum Inflammatory Markers in Obstructive Sleep Apnea: A Meta-Analysis. *Journal of Clinical Sleep Medicine*, 09(10), 1003-1012. <http://doi.org/10.5664/jcsm.3070>
- Naitoh, P., Englund, C. E., & Ryman, D. H. (1987). *Sustained Operations: Research Results* [Interim Report] (87-17). (Naval Medical Command National Capital Region). <https://apps.dtic.mil/sti/citations/ADA191794>.

- Nakashima, A., Bouak, F., Lam, Q., Smith, I., & Vartanian, O. (2018). Task switching following 24 h of total sleep deprivation: a functional MRI study. *NeuroReport*, 29(2), 123-127. <https://doi.org/10.1097/wnr.0000000000000934>
- Neufeld, R. W. J., & Gardner, R. C. (1990). Data aggregation in evaluating psychological constructs: Multivariate and logical deductive considerations. *Journal of Mathematical Psychology*, 34(3), 276-296. [https://doi.org/10.1016/0022-2496\(90\)90033-6](https://doi.org/10.1016/0022-2496(90)90033-6)
- Nicholson, J. (2021). *Dignitas takes flight with U.S. Air Force partnership*. esportsinsider. Retrieved 17/02/2022 from <https://esportsinsider.com/2021/07/dignitas-us-air-force-partnership/>
- Novak, A. R., Bennett, K. J. M., Pluss, M. A., & Fransen, J. (2020). Performance analysis in esports: modelling performance at the 2018 League of Legends World Championship. *International Journal of Sports Science & Coaching*, 15(5-6), 809-817. <http://doi.org/10.1177/1747954120932853>
- Nuyens, F. M., Kuss, D. J., Lopez-Fernandez, O., & Griffiths, M. D. (2019). The Empirical Analysis of Non-problematic Video Gaming and Cognitive Skills: A Systematic Review. *International Journal of Mental Health and Addiction*, 17(2), 389-414. <https://doi.org/10.1007/s11469-018-9946-0>
- O'Hare, D. (1997). Cognitive Ability Determinants of Elite Pilot Performance. *Human factors*, 39(4), 540-552. <http://doi.org/10.1518/001872097778668004>
- Odle-Dusseau, H. N., Bradley, J. L., & Pilcher, J. J. (2010). SUBJECTIVE PERCEPTIONS OF THE EFFECTS OF SUSTAINED PERFORMANCE UNDER SLEEP-DEPRIVATION CONDITIONS. *Chronobiology International*, 27(2), 318-333. <http://doi.org/10.3109/07420520903502226>
- Ofoghi, B., Zeleznikow, J., MacMahon, C., & Raab, M. (2013). Data Mining in Elite Sports: A Review and a Framework. *Measurement in Physical Education and Exercise Science*, 17(3), 171-186. <http://doi.org/10.1080/1091367X.2013.805137>
- Ohayon, M., Wickwire, E. M., Hirshkowitz, M., Albert, S. M., Avidan, A., Daly, F. J., Dauvilliers, Y., Ferri, R., Fung, C., Gozal, D., Hazen, N., Krystal, A., Lichstein, K., Mallampalli, M., Plazzi, G., Rawding, R., Scheer, F. A., Somers, V., & Vitiello, M. V. (2017). National Sleep Foundation's sleep quality recommendations: first report. *Sleep Health*, 3(1), 6-19. <https://doi.org/10.1016/j.sleh.2016.11.006>

- Olsen, O. K., Pallesen, S., Torsheim, T., & Espevik, R. (2016). The effect of sleep deprivation on leadership behaviour in military officers: an experimental study. *Journal of sleep research*, 25(6), 683-689. <https://doi.org/10.1111/jsr.12431>
- Olympics.com. (2022, 08/08/2022). *Commonwealth esports championships 2022: India win bronze in Dota 2 - know all medal winners*. Olympics.com. Retrieved 20/06/2023 from <https://olympics.com/en/news/commonwealth-esports-championships-birmingham-2022-india-medal-winners>
- Olympics.com. (2023, 20/06/2023). *What you need to know about exhibition show matches at the Olympic Esports Week 2023*. Olympics.com. Retrieved 20/06/2023 from <https://olympics.com/en/news/everything-to-know-about-exhibition-show-matches-olympic-esports-week>
- Overall, J. E., & Doyle, S. R. (1994). Implications of chance baseline differences in repeated measurement designs. *Journal of Biopharmaceutical Statistics*, 4(2), 199-216. <http://doi.org/10.1080/10543409408835083>
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232-258. <http://doi.org/10.1016/j.cogpsych.2012.12.002>
- Paap, K. R., Johnson, H. A., & Sawi, O. M. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, 69, 265-278. <http://doi.org/10.1016/j.cortex.2015.04.014>
- Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, 29(2), 89-112. <http://doi.org/10.1080/20445911.2016.1248436>
- Paap, K. R., & Sawi, O. M. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, 274, 81-93. <http://doi.org/10.1016/j.jneumeth.2016.10.002>
- Page, M., Moher, D., Bossuyt, P., Boutron, I., Hoffman, T., Mulrow, C., Shamseer, L., Tetzlaff, J., Akl, E., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., . . . McKenzie, J. (2020). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372. <http://doi.org/10.31222/osf.io/jb4dx>

- Panjeh, S., Pompeia, S., Archer, S. N., Pedrazzoli, M., von Schantz, M., & Cogo-Moreira, H. (2021). What are we measuring with the morningness–eveningness questionnaire? Exploratory factor analysis across four samples from two countries. *Chronobiology International*, 38(2), 234-247. <https://doi.org/10.1080/07420528.2020.1815758>
- Parker, R. A., Scott, C., Inácio, V., & Stevens, N. T. (2020). Using multiple agreement methods for continuous repeated measures data: a tutorial for practitioners. *BMC Medical Research Methodology*, 20(1), 154. <http://doi.org/10.1186/s12874-020-01022-x>
- Patel, V. L., Kaufman, D. R., & Magder, S. A. (1996). The acquisition of medical expertise in complex dynamic environments. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (1 ed., pp. 127-165). Lawrence Erlbaum Associates, Inc.
- Paton, D., & Flin, R. (1999). Disaster stress: an emergency management perspective. *Disaster Prevention and Management*, 8(4), 261-267. <http://doi.org/10.1108/09653569910283897>
- 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. <http://doi.org/10.1080/1750984X.2020.1723122>
- Pelayo, R., & Dement, W. C. (2021). History of Sleep Physiology and Medicine. In M. H. Kryger, T. Roth, C. A. Goldstein, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine: Seventh Edition* (pp. 3-15). Elsevier.
- Phillips, A. J. K., Clerx, W. M., O'Brien, C. S., Sano, A., Barger, L. K., Picard, R. W., Lockley, S. W., Klerman, E. B., & Czeisler, C. A. (2017). Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Scientific Reports*, 7(1), 3216. <https://doi.org/10.1038/s41598-017-03171-4>
- Pigeon, W. R., Pinguat, M., & Conner, K. (2012). Meta-analysis of sleep disturbance and suicidal thoughts and behaviors. *The Journal of Clinical Psychiatry*, 73(9), e1160-1167. <http://doi.org/10.4088/JCP.11r07586>
- Pilcher, J. J., & Huffcutt, A. I. (1996). Effects of Sleep Deprivation on Performance: A Meta-Analysis. *Sleep*, 19(4), 318-326. <https://doi.org/10.1093/sleep/19.4.318>
- Pilkin, W. (2022, 20/04/2022). *An Introduction to Rocket League in 11 Minutes*. YouTube. <https://www.youtube.com/watch?v=Os7TTRtF-HI>

- Power, C. J., Fox, J. L., Teramoto, M., & Scanlan, A. T. (2023). Sleep Patterns Fluctuate Following Training and Games across the Season in a Semi-Professional, Female Basketball Team. *Brain Sciences*, 13(2), 238. <https://doi.org/10.3390/brainsci13020238>
- Pratesi, F. (2008). Chess Theory—Its Structure And Evolution. In H. Holländer & U. Schädler (Eds.), *Scacchia Ludus* (pp. 479-501).
- Prior, A., & Gollan, T. H. (2013). The elusive link between language control and executive control: A case of limited transfer. *Journal of Cognitive Psychology*, 25(5), 622-645. <http://doi.org/10.1080/20445911.2013.821993>
- Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13(2), 253-262. <http://doi.org/10.1017/S1366728909990526>
- R Core Team. (2022). *R*. In (Version 4.1.3) <https://www.r-project.org/>
- Rae, D. E., Stephenson, K. J., & Roden, L. C. (2015). Factors to consider when assessing diurnal variation in sports performance: the influence of chronotype and habitual training time-of-day. *European Journal of Applied Physiology*, 115(6), 1339-1349. <https://doi.org/10.1007/s00421-015-3109-9>
- Ralph, B. C. W., Onderwater, K., Thomson, D. R., & Smilek, D. (2017). Disrupting monotony while increasing demand: benefits of rest and intervening tasks on vigilance. *Psychological Research*, 81(2), 432-444. <http://doi.org/10.1007/s00426-016-0752-7>
- Ratan, R. A., Taylor, N., Hogan, J., Kennedy, T., & Williams, D. (2015). Stand by your man: An examination of gender disparity in League of Legends. *Games and Culture: A Journal of Interactive Media*, 10, 438-462. <https://doi.org/10.1177/1555412014567228>
- Reason, J. (2000). Human error: models and management. *BMJ*, 320(7237), 768-770. <http://doi.org/10.1136/bmj.320.7237.768>
- Reinert, D. F., & Allen, J. P. (2007). The Alcohol Use Disorders Identification Test: An Update of Research Findings. *Alcohol: Clinical and Experimental Research*, 31(2), 185-199. <https://doi.org/10.1111/j.1530-0277.2006.00295.x>
- Reitman, J. G., Anderson-Coto, M. J., Wu, M., Lee, J. S., & Steinkuehler, C. (2019). Esports Research: A Literature Review. *Games and Culture*, 15(1), 32-50. <http://doi.org/10.1177/1555412019840892>

- Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. In (Version 2.2.5) R. <https://cran.r-project.org/web/packages/psych/index.html>
- Reyner, L. A., & Horne, J. A. (2013). Sleep restriction and serving accuracy in performance tennis players, and effects of caffeine. *Physiology and Behavior*, 120, 93-96. <http://doi.org/10.1016/j.physbeh.2013.07.002>
- Reynolds, A. C., & Banks, S. (2010). Total sleep deprivation, chronic sleep restriction and sleep disruption. In G. A. Kerkhof & H. P. A. v. Dongen (Eds.), *Progress in Brain Research* (Vol. 185, pp. 91-103). Elsevier. <https://doi.org/10.1016/B978-0-444-53702-7.00006-3>
- Ribeiro, T., Almeida, V., Calapez, A., Matsuoka, H., & Yamashita, R. (2023). Esports and Olympic Games: a cross-cultural exploration of the player support behaviour towards the Olympics. *International Journal of Sports Marketing and Sponsorship*, ahead-of-print(ahead-of-print). <http://doi.org/10.1108/IJSMS-12-2022-0215>
- Rigney, G., Blunden, S., Maher, C., Dollman, J., Parvazian, S., Matricciani, L., & Olds, T. (2015). Can a school-based sleep education programme improve sleep knowledge, hygiene and behaviours using a randomised controlled trial. *Sleep Medicine*, 16(6), 736-745. <https://doi.org/10.1016/j.sleep.2015.02.534>
- rlduels.gg. (2023). *All Players*. rlduels.gg. Retrieved 22/06/2023 from <https://rlduels.gg/players>
- Roberts, S. S. H., Teo, W.-P., Aisbett, B., & Warmington, S. A. (2019). Extended Sleep Maintains Endurance Performance Better than Normal or Restricted Sleep. *Medicine & Science in Sports & Exercise*, 51(12), 2516-2523. <http://doi.org/10.1249/mss.0000000000002071>
- Roberts, S. S. H., Teo, W.-P., & Warmington, S. A. (2019). Effects of training and competition on the sleep of elite athletes: a systematic review and meta-analysis. *British Journal of Sports Medicine*, 53(8), 513-522. <https://doi.org/10.1136/bjsports-2018-099322>
- Robertson, S., Back, N., & Bartlett, J. D. (2016). Explaining match outcome in elite Australian Rules football using team performance indicators. *Journal of Sports Sciences*, 34(7), 637-644. <http://doi.org/10.1080/02640414.2015.1066026>
- Rocha, J., & Silva, J. (2019, October 9-11, 2019). *Pilots Performance and Flight Safety: The Case of Cognitive Fatigue in Unpressurized Aircraft Cabins* International

- Congress of the Iberoamerican Air Transportation Research Society, University of Beira Interior, Covilhã, Portugal. ISBN: 978-989-654-594-9
- Rocket Sledge. (2019). *How to stop wasting boost in Rocket League*. YouTube. <https://www.youtube.com/watch?app=desktop&v=ykC-ozFSROg>
- Roenneberg, T., Kantermann, T., Juda, M., Vetter, C., & Allebrandt, K. V. (2013). Light and the Human Circadian Clock. In A. Kramer & M. Mewes (Eds.), *Circadian Clocks* (pp. 311-331). Springer. [http://doi.org/10.1007/978-3-642-25950-0\\_13](http://doi.org/10.1007/978-3-642-25950-0_13)
- Roenneberg, T., Wirz-Justice, A., & Mewes, M. (2003). Life between Clocks: Daily Temporal Patterns of Human Chronotypes. *Journal of Biological Rhythms*, 18(1), 80-90. <http://doi.org/10.1177/0748730402239679>
- Rogers, W. P., Armstrong, N., Acheson, D., Covert, E., Feynman, R., Hotz, R., Kutyna, D., Ride, S., Rummel, R., & Sutter, J. (1986). *Presidential commission on the space shuttle Challenger accident*. Washington, DC
- Romdhani, M., Hammouda, O., Chaabouni, Y., Mahdouani, K., Driss, T., Chamari, K., & Souissi, N. (2019). Sleep deprivation affects post-lunch dip performances, biomarkers of muscle damage and antioxidant status. *Biology of Sport*, 36(1), 55-65. <http://doi.org/10.5114/biolSport.2018.78907>
- Rudolf, K., Bickmann, P., Froböse, I., Tholl, C., Wechsler, K., & Grieben, C. (2020). Demographics and Health Behavior of Video Game and eSports Players in Germany: The eSports Study 2019. *International Journal of Environmental Research and Public Health*, 17(6), 1870. <https://doi.org/10.3390/ijerph17061870>
- Rundo, J. V., & Downey, R. (2019). Polysomnography. In K. H. Levin & P. Chauvel (Eds.), *Handbook of Clinical Neurology* (Vol. 160, pp. 381-392). Elsevier. <https://doi.org/10.1016/B978-0-444-64032-1.00025-4>
- Russell, C., Caldwell, J., Arand, D., Myers, L., Wubbels, P., & Downs, H. (2010). *Validation of the fatigue science readiband actigraph and associated sleep/wake classification algorithms* (Archinoetics LLC).
- Sabtan, B., Cao, S., & Paul, N. (2022). Current practice and challenges in coaching Esports players: An interview study with league of legends professional team coaches. *Entertainment Computing*, 42, 100481. <https://doi.org/10.1016/j.entcom.2022.100481>
- Sadeh, A., Sharkey, M., & Carskadon, M. A. (1994). Activity-Based Sleep-Wake Identification: An Empirical Test of Methodological Issues. *Sleep*, 17(3), 201-207. <http://doi.org/10.1093/sleep/17.3.201>

- Sainani, K. L. (2015). Dealing With Missing Data. *PM&R*, 7(9), 990-994.  
<https://doi.org/10.1016/j.pmrj.2015.07.011>
- Saksvik, I. B., Bjorvatn, B., Harvey, A. G., Waage, S., Harris, A., & Pallesen, S. (2011). Adaptation and readaptation to different shift work schedules measured with sleep diary and actigraphy. *Journal of occupational health psychology*, 16(3), 331-344.  
<https://doi.org/10.1037/a0022770>
- Salas, E., Rosen, M. A., & DiazGranados, D. (2009). Expertise-Based Intuition and Decision Making in Organizations. *Journal of Management*, 36(4), 941-973.  
<http://doi.org/10.1177/0149206309350084>
- Sallinen, M., Härmä, M., Akila, R., Holm, A., Luukkonen, R., Mikola, H., Müller, K., & Virkkala, J. (2004). The effects of sleep debt and monotonous work on sleepiness and performance during a 12-h dayshift. *Journal of sleep research*, 13(4), 285-294. <http://doi.org/10.1111/j.1365-2869.2004.00425.x>
- Sallinen, M., Holm, A., Hiltunen, J., Hirvonen, K., Härmä, M., Koskelo, J., Letonsaari, M., Luukkonen, R., Virkkala, J., & Müller, K. (2008). Recovery of cognitive performance from sleep debt: Do a short rest pause and a single recovery night help? *Chronobiology International*, 25(2-3), 279-296.  
<http://doi.org/10.1080/07420520802107106>
- Salthouse, T. A., & Tucker-Drob, E. M. (2008). Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*, 22(6), 800-811.  
<http://doi.org/10.1037/a0013091>
- Saltzman, J. A., & Liechty, J. M. (2016). Family correlates of childhood binge eating: A systematic review. *Eating Behaviors*, 22, 62-71.  
<https://doi.org/10.1016/j.eatbeh.2016.03.027>
- Santhi, N., Horowitz, T. S., Duffy, J. F., & Czeisler, C. A. (2007). Acute Sleep Deprivation and Circadian Misalignment Associated with Transition onto the First Night of Work Impairs Visual Selective Attention. *PLOS ONE*, 2(11), e1233.  
<http://doi.org/10.1371/journal.pone.0001233>
- Sanz-Milone, V., Yoshinori, P., & Maculano Esteves, A. (2021). Sleep quality of professional e-Sports athletes (Counter Strike: Global Offensive). *International Journal of Esports*, 1(1), 1-10. <https://www.ijesports.org/article/45/html>
- Sargent, C., Lastella, M., Halson, S. L., & Roach, G. D. (2014). The impact of training schedules on the sleep and fatigue of elite athletes. *Chronobiology International*, 31(10), 1160-1168. <http://doi.org/10.3109/07420528.2014.957306>

- Satterfield, B. C., & Killgore, W. D. S. (2019). Sleep loss, executive function, and decision-making. In M. A. Grandner (Ed.), *Sleep and Health* (pp. 339-358). Academic Press. <https://doi.org/10.1016/B978-0-12-815373-4.00026-5>
- Saunders, J. B., Aasland, O. G., Babor, T. F., De La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction*, 88(6), 791-804. <https://doi.org/10.1111/j.1360-0443.1993.tb02093.x>
- Saxena, A. D., & George, C. F. (2005). Sleep and motor performance in on-call internal medicine residents. *Sleep*, 28(11), 1386-1391. <http://doi.org/10.1093/sleep/28.11.1386>
- Schlosser, K., Maschuw, K., Kupietz, E., Weyers, P., Schneider, R., Rothmund, M., Hassan, I., & Bartsch, D. K. (2012). Call-associated acute fatigue in surgical residents--subjective perception or objective fact? A cross-sectional observational study to examine the influence of fatigue on surgical performance. *World journal of surgery*, 36(10), 2276-2287. <http://doi.org/10.1007/s00268-012-1699-5>
- Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine*, 65(10), 611-621. <http://doi.org/10.1097/00001888-199010000-00001>
- Schneider, D. W. (2015). Isolating a mediated route for response congruency effects in task switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(1), 235-245. <http://doi.org/10.1037/xlm0000049>
- Schneider, D. W., & Logan, G. D. (2014). Modelling Response Selection in Task Switching: Testing the Contingent Encoding Assumption. *Quarterly Journal of Experimental Psychology*, 67(6), 1074-1095. <http://doi.org/10.1080/17470218.2013.843009>
- Schöllhorn, I., Stefani, O., Lucas, R. J., Spitschan, M., Slawik, H. C., & Cajochen, C. (2023). Melanopic irradiance defines the impact of evening display light on sleep latency, melatonin and alertness. *Communications Biology*, 6(1), 228. <https://doi.org/10.1038/s42003-023-04598-4>
- Scott, W. A. (1962). Cognitive Complexity and Cognitive Flexibility. *Sociometry*, 25(4), 405-414. <https://doi.org/10.2307/2785779>
- Segal, D., Prior, A., & Gollan, T. H. (2021). Do all switches cost the same? Reliability of language switching and mixing costs. *Journal of cognition*, 4(1), 3. <https://doi.org/10.5334/joc.140>

- Serfaty, D., MacMillan, J., Entin, E. E., & Entin, E. B. (1997). The decision-making expertise of battle commanders. In C. E. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 233-246). Lawrence Erlbaum Associates, Inc.,.
- Shambroom, J. R., Fábregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of sleep research*, 21(2), 221-230. <https://doi.org/10.1111/j.1365-2869.2011.00944.x>
- Shawa, N., Rae, D. E., & Roden, L. C. (2018). Impact of seasons on an individual's chronotype: current perspectives. *Nature and Science of Sleep*, 10, 345-354. <https://doi.org/10.2147/NSS.S158596>
- Shi, L., Chen, S.-J., Ma, M.-Y., Bao, Y.-P., Han, Y., Wang, Y.-M., Shi, J., Vitiello, M. V., & Lu, L. (2018). Sleep disturbances increase the risk of dementia: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 40, 4-16. <https://doi.org/10.1016/j.smrv.2017.06.010>
- Short, M. A., & Weber, N. (2018). Sleep duration and risk-taking in adolescents: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 41, 185-196. <https://doi.org/10.1016/j.smrv.2018.03.006>
- Siegel, J. M. (2021). Normal Human Sleep: An Overview. In M. H. Kryger, T. Roth, C. A. Goldstein, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine: Seventh Edition* (pp. 52-53). Elsevier.
- Signal, T. L., & Gander, P. H. (2007). Rapid counterclockwise shift rotation in air traffic control: Effects on sleep and night work. *Aviation Space and Environmental Medicine*, 78(9), 878-885.
- Siroky, D. S. (2009). Navigating Random Forests and related advances in algorithmic modeling. *Statistics Surveys*, 3, 147-163. <https://doi.org/10.1214/07-SS033>
- Slama, H., Chylinski, D. O., Deliens, G., Leproult, R., Schmitz, R., & Peigneux, P. (2018). Sleep Deprivation Triggers Cognitive Control Impairments in Task-Goal Switching. *Sleep*, 41(2), zsx200. <https://doi.org/10.1093/sleep/zsx200>
- Smith, C. D., Cooper, A. D., Merullo, D. J., Cohen, B. S., Heaton, K. J., Claro, P. J., & Smith, T. (2019). Sleep restriction and cognitive load affect performance on a simulated marksmanship task. *Journal of sleep research*, 28(3), e12637. <https://doi.org/10.1111/jsr.12637>
- Smith, K. N., Lamb, K. N., & Henson, R. K. (2019). Making Meaning out of MANOVA: The Need for Multivariate Post Hoc Testing in Gifted Education Research. *Gifted Child Quarterly*, 64(1), 41-55. <https://doi.org/10.1177/0016986219890352>

- Smith, M. G., Wusk, G. C., Nasrini, J., Baskin, P., Dinges, D. F., Roma, P. G., & Basner, M. (2021). Effects of six weeks of chronic sleep restriction with weekend recovery on cognitive performance and wellbeing in high-performing adults. *Sleep*, 44(8), zsab051. <https://doi.org/10.1093/sleep/zsab051>
- Smithies, T. D., Campbell, M. J., Ramsbottom, N., & Toth, A. J. (2021). A Random Forest approach to identify metrics that best predict match outcome and player ranking in the esports Rocket League. *Scientific Reports*, 11(1), 19285. <https://doi.org/10.1038/s41598-021-98879-9>
- Smithies, T. D., Eastwood, P. R., Walsh, J., Murray, K., Markwick, W., & Dunican, I. C. (2021). Around the world in 16 days: the effect of long-distance transmeridian travel on the sleep habits and behaviours of a professional Super Rugby team. *Journal of Sports Sciences*, 1-7. <https://doi.org/10.1080/02640414.2021.1947617>
- Smithies, T. D., Toth, A. J., Conroy, E., Ramsbottom, N., Kowal, M., & Campbell, M. J. (2020). Life After Esports: A Grand Field Challenge. *Frontiers in Psychology*, 11(883). <https://doi.org/10.3389/fpsyg.2020.00883>
- Smithies, T. D., Toth, A. J., Dunican, I. C., Caldwell, J. A., Kowal, M., & Campbell, M. J. (2021). The effect of sleep restriction on cognitive performance in elite cognitive performers: a systematic review. *Sleep*, 44(7), zsab008. <https://doi.org/10.1093/sleep/zsab008>
- Soveri, A., Lehtonen, M., Karlsson, L. C., Lukasik, K., Antfolk, J., & Laine, M. (2018). Test–retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology: Adult*, 25(2), 155-165. <https://doi.org/10.1080/23279095.2016.1263795>
- SquishyMuffinz. (2020). *How to manage your boost like a pro player (tutorial) | the best tips for improving in rocket league*. YouTube. <https://www.youtube.com/watch?app=desktop&v=eK3DLp-Yjwc>
- Stanley, P. (2022). Analysis of covariance: A useful tool for the pharmacologist to reduce variation and improve precision using fewer animals. *British Journal of Pharmacology*, 179(14), 3645-3650. <https://doi.org/10.1111/bph.15855>
- Statista. (2023). *Video Games - Worldwide*. Statista. Retrieved 20/06/2023 from <https://www.statista.com/outlook/dmo/digital-media/video-games/worldwide?currency=EUR>
- Staunton, C., Gordon, B., Custovic, E., Stanger, J., & Kingsley, M. (2017). Sleep patterns and match performance in elite Australian basketball athletes. *Journal of Science*

- Stenson, A. R., Whitney, P., Hinson, J. M., Hansen, D. A., Lawrence-Sidebottom, D., Skeiky, L., Riedy, S. M., Kurinec, C. A., & Van Dongen, H. P. A. (2023). Effects of total sleep deprivation on components of top-down attentional control using a flexible attentional control task. *Journal of sleep research*, 32(2), e13744. <https://doi.org/10.1111/jsr.13744>
- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437, 1272-1278. <https://doi.org/10.1038/nature04286>
- Stone, J. D., Rentz, L. E., Forsey, J., Ramadan, J., Markwald, R. R., Finomore, V. S., Galster, S. M., Rezai, A., & Hagen, J. A. (2020). Evaluations of Commercial Sleep Technologies for Objective Monitoring During Routine Sleeping Conditions. *Nature and Science of Sleep*, 12, 821-842. <https://doi.org/10.2147/NSS.S270705>
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. American chemical society.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), 307. <https://doi.org/10.1186/1471-2105-9-307>
- Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>
- Studholme, B. (2023, 06/01/2023). *The esports salary market is headed toward a correction*. Digiday. Retrieved 20/06/2023 from <https://digiday.com/marketing/the-esports-salary-market-is-headed-toward-a-correction/#:~:text=Hal%20Biagas%2C%20then%20the%20executive,one%20league%20source%20told%20Digiday>
- Sturzenegger, C., & Bassetti, C. L. (2004). The clinical spectrum of narcolepsy with cataplexy: a reappraisal. *Journal of sleep research*, 13(4), 395-406. <https://doi.org/10.1111/j.1365-2869.2004.00422.x>
- Sturzenegger, C., Baumann, C. R., Lammers, G. J., Kallweit, U., van der Zande, W. L., & Bassetti, C. L. (2018). Swiss Narcolepsy Scale: A simple screening tool for hypocretin-deficient narcolepsy with cataplexy. *Clinical and Translational Neuroscience*, 2(2). <https://doi.org/10.1177/2514183x18794175>

- Sullan, M. J., Drummond, S. P. A., & Granholm, E. (2021). Sleep deprivation and compensatory cognitive effort on a visual information processing task. *Sleep*, 44(2), zsaa177. <https://doi.org/10.1093/sleep/zsaa177>
- Sullivan, S. S., Carskadon, M. A., Dement, W. C., & Jackson, C. L. (2021). Normal Human Sleep: An Overview. In M. H. Kryger, T. Roth, C. A. Goldstein, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine: Seventh Edition* (pp. 16-26). Elsevier.
- Swann, C., Moran, A., & Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise*, 16, 3-14. <https://doi.org/10.1016/j.psychsport.2014.07.004>
- Szelenberger, W., Piotrowski, T., & Dąbrowska, A. J. (2005). Increased prefrontal event-related current density after sleep deprivation. *Acta Neurobiologiae Experimentalis*, 65(1), 19-28.
- Taheri, M., & Irandoust, K. (2019). Morning exercise improves cognitive performance decrements induced by partial sleep deprivation in elite athletes. *Biological Rhythm Research*, 51(4). <https://doi.org/10.1080/09291016.2019.1576279>
- Taylor, J. L., O'Hara, R., Mumenthaler, M. S., Rosen, A. C., & Yesavage, J. A. (2005). Cognitive Ability, Expertise, and Age Differences in Following Air-Traffic Control Instructions. *Psychology and Aging*, 20(1), 117-133. <https://doi.org/10.1037/0882-7974.20.1.117>
- Taylor, N., & Stout, B. (2020). Gender and the two-tiered system of collegiate esports. *Critical Studies in Media Communication*, 37(5), 451-465. <https://doi.org/10.1080/15295036.2020.1813901>
- Thomann, J., Baumann, C. R., Landolt, H.-P., & Werth, E. (2014). Psychomotor Vigilance Task Demonstrates Impaired Vigilance in Disorders with Excessive Daytime Sleepiness. *Journal of Clinical Sleep Medicine*, 10(09), 1019-1024. <https://doi.org/10.5664/jcsm.4042>
- Thomas, C. J., Rothschild, J., Earnest, C. P., & Blaisdell, A. (2019). The Effects of Energy Drink Consumption on Cognitive and Physical Performance in Elite League of Legends Players. *Sports*, 7, 1-10. <https://doi.org/10.3390/sports7090196>
- Thun, E., Bjorvatn, B., Osland, T., Martin Steen, V., Sivertsen, B., Johansen, T., Halvor Lilleholt, T., Udnes, I., Hilde Nordhus, I., & Pallesen, S. (2012). An Actigraphic Validation Study of Seven Morningness-Eveningness Inventories. *European Psychologist*, 17(3), 222-230. <https://doi.org/10.1027/1016-9040/a000097>

- Tidy, J. (2022, 12/10/2022). *Commonwealth Games makes mystery esports U-turn*. BBC. Retrieved 20/06/2023 from <https://www.bbc.com/news/technology-63232333>
- Timmer, K., Calabria, M., Branzi, F. M., Baus, C., & Costa, A. (2018). On the Reliability of Switching Costs Across Time and Domains. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01032>
- Tkachenko, O., & Dinges, D. F. (2018). Interindividual variability in neurobehavioral response to sleep loss: A comprehensive review. *Neuroscience & Biobehavioral Reviews*, 89, 29-48. <https://doi.org/10.1016/j.neubiorev.2018.03.017>
- Todt, N., Pase, A. F., Scarton, A., Rolim, L. H., Berlitz, G. Z., & Baptista, L. V. (2020). The eSports and Olympic Games: Perspectives of an ongoing debate. *Journal of Human Sport and Exercise*, 15(1proc), S94-S110. <https://doi.org/10.14198/jhse.2020.15.Proc1.10>
- Toth, A. J., Conroy, E., & Campbell, M. J. (2021). Beyond action video games: Differences in gameplay and ability preferences among gaming genres. *Entertainment Computing*, 38, 100408. <https://doi.org/10.1016/j.entcom.2021.100408>
- Toth, A. J., Ramsbottom, N., Constantin, C., Milliet, A., & Campbell, M. J. (2021). The effect of expertise, training and neurostimulation on sensory-motor skill in esports. *Computers in Human Behavior*, 121, 106782. <https://doi.org/10.1016/j.chb.2021.106782>
- Toth, A. J., Ramsbottom, N., Kowal, M., & Campbell, M. J. (2020). Converging Evidence Supporting the Cognitive Link between Exercise and Esport Performance: A Dual Systematic Review. *Brain Sciences*, 10(11), 859. <https://doi.org/10.3390/brainsci10110859>
- Tregear, S., Reston, J., Schoelles, K., & Phillips, B. (2009). Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. *Journal of Clinical Sleep Medicine*, 5(6), 573-581. <https://doi.org/10.5664/jcsm.27662>
- Troester, M., Quan, S., Berry, R., Plante, D., Abreu, A., & Alzoubaidi, M. (2023). *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications* (Vol. 3). American Academy of Sleep Medicine.
- Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2). Reading, MA.
- Twooroger, S. S., Davis, S., Vitiello, M. V., Lentz, M. J., & McTiernan, A. (2005). Factors associated with objective (actigraphic) and subjective sleep quality in young adult

- women. *Journal of Psychosomatic Research*, 59(1), 11-19.  
<https://doi.org/10.1016/j.jpsychores.2005.03.008>
- Uddin, L. Q. (2021). Cognitive and behavioural flexibility: neural mechanisms and clinical considerations. *Nature Reviews Neuroscience*, 22(3), 167-179.  
<https://doi.org/10.1038/s41583-021-00428-w>
- Uehli, K., Mehta, A. J., Miedinger, D., Hug, K., Schindler, C., Holsboer-Trachsler, E., Leuppi, J. D., & Künzli, N. (2014). Sleep problems and work injuries: A systematic review and meta-analysis. *Sleep Medicine Reviews*, 18(1), 61-73.  
<https://doi.org/10.1016/j.smrv.2013.01.004>
- Ustinov, Y., & Lichstein, K. L. (2013). Actigraphy Reliability with Normal Sleepers. *Behavioral Sleep Medicine*, 11(5), 313-320.  
<https://doi.org/10.1080/15402002.2012.688779>
- Valdez, P., Reilly, T., & Waterhouse, J. (2008). Rhythms of Mental Performance. *Mind Brain and Education*, 2(1), 7-16. <https://doi.org/10.1111/j.1751-228X.2008.00023.x>
- Valls-Serrano, C., de Francisco, C., Caballero-López, E., & Caracuel, A. (2022). Cognitive Flexibility and Decision Making Predicts Expertise in the MOBA Esport, League of Legends. *SAGE Open*, 12(4), 21582440221142728.  
<https://doi.org/10.1177/21582440221142728>
- Van Dessel, P., Liefvooghe, B., & De Houwer, J. (2020). The Instructed Task-Switch Evaluation Effect: Is the Instruction to Switch Tasks Sufficient to Dislike Task Switch Cues? *Journal of cognition*, 3(1), 1. <https://doi.org/10.5334/joc.90>
- Van Dongen, H. P. A. (2017). Fatigue as a challenge in 24/7 operations - how to survive and thrive. In. University of South Australia, Adelaide, Australia.
- Van Dongen, H. P. A., Balkin, T. J., Hursh, S. R., & Dorrian, J. (2022). Performance Deficits During Sleep Loss and Their Operational Consequences. In M. H. Kryger, T. Roth, C. A. Goldstein, & W. C. Dement (Eds.), *Principles and Practice of Sleep Medicine: Seventh Edition* (pp. 742-749). Elsevier.
- Van Dongen, H. P. A., & Belenky, G. (2009). Individual differences in vulnerability to sleep loss in the work environment. *Industrial health*, 47(5), 518-526.  
<https://doi.org/10.2486/indhealth.47.518>
- Van Dongen, H. P. A., Belenky, G., & Krueger, J. M. (2011). A Local, Bottom-Up Perspective on Sleep Deprivation and Neurobehavioral Performance. *Current Topics in Medicinal Chemistry*, 11(19), 2414-2422.  
<https://doi.org/10.2174/156802611797470286>

- Van Dongen, H. P. A., Caldwell, J. A., & Caldwell, J. L. (2011). Individual differences in cognitive vulnerability to fatigue in the laboratory and in the workplace. In H. P. A. Van Dongen & G. A. Kerkhof (Eds.), *Progress in Brain Research* (Vol. 190, pp. 145-153). Elsevier. <https://doi.org/10.1016/b978-0-444-53817-8.00009-8>
- Van Dongen, H. P. A., Caldwell, J. A., Jr., & Caldwell, J. L. (2006). Investigating systematic individual differences in sleep-deprived performance on a high-fidelity flight simulator. *Behavior Research Methods*, 38(2), 333-343. <https://doi.org/10.3758/BF03192785>
- Van Dongen, H. P. A., & Dinges, D. F. (2000). Circadian rhythms in fatigue, alertness, and performance. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (3 ed., pp. 391-399). W. B. Saunders.
- Van Dongen, H. P. A., & Dinges, D. F. (2003). Investigating the interaction between the homeostatic and circadian processes of sleep–wake regulation for the prediction of waking neurobehavioural performance. *Journal of sleep research*, 12(3), 181-187. <https://doi.org/10.1046/j.1365-2869.2003.00357.x>
- Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: Dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26(2), 117-126. <https://doi.org/10.1093/sleep/26.2.117>
- Vaz, L., Rooyen, M. V., & Sampaio, J. (2010). Rugby Game-Related Statistics that Discriminate Between Winning and Losing Teams in Irb and Super Twelve Close Games. *Journal of sports science & medicine*, 9(1), 51-55. <https://pubmed.ncbi.nlm.nih.gov/24149385>
- Veasey, S., Rosen, R., Barzansky, B., Rosen, I., & Owens, J. (2002). Sleep Loss and Fatigue in Residency Training: A Reappraisal. *JAMA*, 288(9), 1116-1124. <https://doi.org/10.1001/jama.288.9.1116>
- Vermeulen, L., Braem, S., & Notebaert, W. (2019). The affective twitches of task switches: Task switch cues are evaluated as negative. *Cognition*, 183, 124-130. <https://doi.org/10.1016/j.cognition.2018.11.002>
- Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ*, 323(7321), 1123-1124. <https://doi.org/10.1136/bmj.323.7321.1123>
- Virge. (2020). *The Best Mechanics to Learn at Your Rank* [Video File]. <https://www.youtube.com/watch?v=4J7F5s-Mu9E>

- Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes 'expert' in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied Cognitive Psychology*, 24(6), 812-826. <https://doi.org/10.1002/acp.1588>
- 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. <https://doi.org/10.1016/j.neuroimage.2011.03.052>
- Vrijkotte, S., Roelands, B., Meeusen, R., & Pattyn, N. (2016). Sustained military operations and cognitive performance. *Aerospace medicine and human performance*, 87(8), 718-727. <https://doi.org/10.3357/AMHP.4468.2016>
- Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., & Tononi, G. (2011). Local sleep in awake rats. *Nature*, 472(7344), 443-447. <https://doi.org/10.1038/nature10009>
- Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. *Transactions of the American Mathematical Society*, 54(3), 426-482. <https://doi.org/10.2307/1990256>
- Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with Sleep Makes Perfect: Sleep-Dependent Motor Skill Learning. *Neuron*, 35(1), 205-211. [https://doi.org/10.1016/S0896-6273\(02\)00746-8](https://doi.org/10.1016/S0896-6273(02)00746-8)
- Walker, M. P., & Stickgold, R. (2004). Sleep-Dependent Learning and Memory Consolidation. *Neuron*, 44(1), 121-133. <https://doi.org/10.1016/j.neuron.2004.08.031>
- Wallace, A., & Bucks, R. S. (2013). Memory and Obstructive Sleep Apnea: A Meta-Analysis. *Sleep*, 36(2), 203-220. <https://doi.org/10.5665/sleep.2374>
- Walsh, N. P., Halson, S. L., Sargent, C., Roach, G. D., Nédélec, M., Gupta, L., Leeder, J., Fullagar, H. H. K., Coutts, A. J., Edwards, B. J., Pullinger, S. A., Robertson, C. M., Burniston, J. G., Lastella, M., Le Meur, Y., Hausswirth, C., Bender, A. M., Grandner Michael, A., & Samuels Charles, H. (2021). Sleep and the athlete: narrative review and 2021 expert consensus recommendations. *British Journal of Sports Medicine*, 55, 356-368. <https://doi.org/10.1136/bjsports-2020-102025>
- Wang, Y., Mei, H., Jiang, Y.-R., Sun, W.-Q., Song, Y.-J., Liu, S.-J., & Jiang, F. (2015). Relationship between Duration of Sleep and Hypertension in Adults: A Meta-

- Analysis. *Journal of Clinical Sleep Medicine*, 11(09), 1047-1056.  
<https://doi.org/10.5664/jcsm.5024>
- Ward, M. R., & Harmon, A. D. (2019). ESport Superstars. *Journal of Sports Economics*, 20(8), 987-1013. <https://doi.org/10.1177/1527002519859417>
- Warren, N., & Clark, B. (1937). Blocking in mental and motor tasks during a 65-hour vigil. *Journal of Experimental Psychology*, 21, 97-105.  
<https://doi.org/10.1037/h0057288>
- Waterhouse, J., Reilly, T., Atkinson, G., & Edwards, B. (2007). Jet lag: trends and coping strategies. *Lancet*, 369(9567), 1117-1129. [https://doi.org/10.1016/s0140-6736\(07\)60529-7](https://doi.org/10.1016/s0140-6736(07)60529-7)
- Watson, F. L., Pasteur, M.-A. L., Healy, D. T., & Hughes, E. A. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology: Clinical and Experimental*, 9(1), 51-61. <https://doi.org/10.1002/hup.470090107>
- Weissberger, G. H., Wierenga, C. E., Bondi, M. W., & Gollan, T. H. (2012). Partially overlapping mechanisms of language and task control in young and older bilinguals. *Psychology and Aging*, 27(4), 959-974.  
<https://doi.org/10.1037/a0028281>
- Whitehead, S., Till, K., Jones, B., Beggs, C., Dalton-Barron, N., & Weaving, D. (2020). The use of technical-tactical and physical performance indicators to classify between levels of match-play in elite rugby league. *Science and Medicine in Football*, 1-7. <https://doi.org/10.1080/24733938.2020.1814492>
- Whitmore, J., & Fisher, S. (1996). Speech during sustained operations. *Speech Communication*, 20(1), 55-70. [https://doi.org/10.1016/S0167-6393\(96\)00044-1](https://doi.org/10.1016/S0167-6393(96)00044-1)
- Whitney, P., Hinson, J. M., Jackson, M. L., & Van Dongen, H. P. A. (2015). Feedback blunting: Total sleep deprivation impairs decision making that requires updating based on feedback. *Sleep*, 38(5), 745-754. <https://doi.org/10.5665/sleep.4668>
- Whitney, P., Hinson, J. M., & Nusbaum, A. T. (2019). A dynamic attentional control framework for understanding sleep deprivation effects on cognition. In H. P. A. Van Dongen, P. Whitney, J. M. Hinson, K. A. Honn, & M. W. L. Chee (Eds.), *Progress in Brain Research* (Vol. 246, pp. 111-126). Elsevier.  
<https://doi.org/10.1016/bs.pbr.2019.03.015>
- Whitney, P., Hinson, J. M., Satterfield, B. C., Grant, D. A., Honn, K. A., & Van Dongen, H. P. A. (2017). Sleep Deprivation Diminishes Attentional Control Effectiveness

- and Impairs Flexible Adaptation to Changing Conditions. *Scientific Reports*, 7(1), 16020. <https://doi.org/10.1038/s41598-017-16165-z>
- Wickens, C. D., Hutchins, S. D., Laux, L., & Sebok, A. (2015). The Impact of Sleep Disruption on Complex Cognitive Tasks: A Meta-Analysis. *Human factors*, 57(6), 930-946. <https://doi.org/10.1177/0018720815571935>
- Williams, A. M., Ford, P. R., Eccles, D. W., & Ward, P. (2011). Perceptual-cognitive expertise in sport and its acquisition: Implications for applied cognitive psychology. *Applied Cognitive Psychology*, 25(3), 432-442. <https://doi.org/10.1002/acp.1710>
- Williams, H. L., Lubin, A., & Goodnow, J. J. (1959). Impaired performance with acute sleep loss. *Psychological Monographs: General and Applied*, 73, 1-26. <https://doi.org/10.1037/h0093749>
- Windred, D. P., Burns, A. C., Lane, J. M., Saxena, R., Rutter, M. K., Cain, S. W., & Phillips, A. J. K. (2023). Sleep regularity is a stronger predictor of mortality risk than sleep duration: A prospective cohort study. *Sleep*, zsad253. 10.1093/sleep/zsad253
- Wohlin, C. (2014). *Guidelines for snowballing in systematic literature studies and a replication in software engineering* Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, England, United Kingdom.
- Womack, S. D., Hook, J. N., Reyna, S. H., & Ramos, M. (2013). Sleep Loss and Risk-Taking Behavior: A Review of the Literature. *Behavioral Sleep Medicine*, 11(5), 343-359. <https://doi.org/10.1080/15402002.2012.703628>
- Woods, C. T., Sinclair, W., & Robertson, S. (2017). Explaining match outcome and ladder position in the National Rugby League using team performance indicators. *Journal of Science and Medicine in Sport*, 20(12), 1107-1111. <https://doi.org/10.1016/j.jsams.2017.04.005>
- Xia, B., Wang, H., & Zhou, R. (2017). What Contributes to Success in MOBA Games? An Empirical Study of Defense of the Ancients 2. *Games and Culture*, 14(5), 498-522. <https://doi.org/10.1177/1555412017710599>
- Yildiz, A., Quetscher, C., Dharmadhikari, S., Chmielewski, W., Glaubitz, B., Schmidt-Wilcke, T., Edden, R., Dydak, U., & Beste, C. (2014). Feeling safe in the plane: Neural mechanisms underlying superior action control in airplane pilot trainees—A combined EEG/MRS study. *Human Brain Mapping*, 35(10), 5040-5051. <https://doi.org/10.1002/hbm.22530>

## **Chapter 11. Appendices (separate digital file)**