University of Cape Town

Computational Psychiatry

Neuropsychological Bayesian Reinforcement Learning

Student:
Zach Wolpe
WLPZAC001

Supervisor:
A/Prof. Jonathan Shock
External supervisor:
A/Prof. Benjamin Cowley
Co-supervisor:
Mr. Allan Clark

Minor dissertation for the degree

M.Sc. Advanced Analytics

Department of Statistical Sciences

15 May 2022

Declaration of Authorship

Signed: Signed by candidate

Date: 15 May 2022

Abstract

Cognitive science draws inspiration from a myriad of disciplines, and has become increasingly reliant on computational methods. In particular, theories of learning, operant conditioning and decision making have shown a natural synergy with statistical learning algorithms. This offers a unique opportunity to derive novel insight into the conditioning process by leveraging computational ideas. Specifically, ideas from Bayesian Inference and Reinforcement Learning.

In this thesis, we examine the statistical properties of associative learning under uncertainty. We conducted a neuropsychological experiment on over 100 human subjects to measure a suite of executive functions. The primary experimental task (Card Sorting) gauges one's ability to learn, via inference, the structure of some latent pattern that drives the decision making process.

We were able to successfully predict the subjects' behaviour in this task by fitting a Bayesian Reinforcement Learning model, alluding to the mechanics of the latent biological decision generating process and executive functions. Primarily, we detail the relationship between working memory capacity and associative learning.

Keywords: Cognitive science, mathematical psychology, computational psychiatry, reinforcement learning, Bayesian inference, machine learning.

Acknowledgements

To my loving parents, who have always supported my decisions, even those that they did not understand, I am eternally grateful. Forever encouraging me to do what is bold, forge my own path and express my ideas. It is these principles that have shaped many of my personal philosophies.

I would also like to express my gratitude towards my humble but brilliant academic advisors, who's mentorship has allowed me to articulate my passion for science. In particular, their meticulous dedication to supporting my efforts, has not only acting as a catalyst to my critical thinking, but also significantly changed my perspective, approach, diligence and world view.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The world of medicine, neuroscience, psychology and science at large is becoming increasingly computational. Whilst, by definition, scientific endeavours rely on empiricism, the role of computational methods in both analysing and designing experiments is shifting. Traditional statistical methods form the basis of all evidence-based hypothesis testing, but beyond this we are now entering an era of research where computational methods inform the theoretical literature.

An elegant dance has begun between the worlds of machine learning and neuroscience that contains a number of sub-fields from medical machine learning, to computational psychiatry and mathematical psychology. Often conducted in interdisciplinary groups, researchers aim to both design algorithms based on our understanding of biological cognition, and, simultaneously, utilise our understanding of computational systems to postulate new psychological theory. Many works, including this work, rely on the premise that the mind arises from some fundamental computational process.

1.1 Medicine and psychology

Neuropsychology: Data-driven methods are also gaining popularity in psychological studies. Machine learning techniques are playing an increasingly more important role in decomposing human behaviour and providing novel insight. Neuropsychology in particular - which is concerned with the neurological executive function that motivates behaviour - is a central beneficiary of statistical learning theory as it offers a paradigm for conceptualising much of the function of the cerebral cortex.

Computational psychiatry: Moreover, psychiatric research is showing a growing interest in leveraging algorithmic methods that argues for modularised executive functionality (Parr, Rees, and Friston, 2018). Algorithms, that have historically attempted to infer modularity of the cerebral cortex by studying patients with severe injuries (from H.M. to Phineas Gage) (García-Molina, 2012), may allow for the non-invasive decomposition and isolation of executive function into biologically plausible constructs and allow one to reason about the isolated neurological activity.

It is through these developments that much of the future of medical and psychiatric knowledge may rest on the successful integration of computational methods.

1.2 Computational paradigms

When applied to neuroscience, psychology and psychiatry, the realm of statistical learning constitutes three broad schools (that are segmented by physiological and behavioural relevance and not purely by mathematical constructs) namely: Dynamic systems, Bayesian inference and reinforcement learning.

Bayesian inference: A growing community of computational neuroscientists have become inspired by the idea of the brain operating in a fundamentally Bayesian fashion. The Bayesian brain hypothesis (Knill and Pouget, 2004) is one succinct articulation of this idea: where cognition is posed as a Bayesian update after gathering experience. The mathematical flexibility, robust statistical theory, ability to combine expert knowledge (priors) with learning algorithms, coupled with natural theoretical interpretations, all bolster Bayesian inference as a budding candidate for contemporary neurological research (Knill and Pouget, 2004). Furthermore, its generality allows for an abstract framework that represents a vast array of both Bayesian and frequentist techniques, and, as such, is extremely appealing to mathematicians and statisticians who aim to build strong foundational theory.

Reinforcement learning: Optimal control and reinforcement learning - a set of algorithms tasked with learning from experience, directly premised on many ideas of psychological learning theory - naturally extends to model behaviour (agnostic of whether the behaviour pertains to some automaton or human). Though showing great success in optimisation algorithms, Reinforcement Learning is largely inspired by intuitively plausible operant learning techniques. By this logic, it is natural to frame Reinforcement Learning models as special-case graphical model that capture iterative stochastic transitions.

It is in the union of Bayesian inference and Reinforcement Learning that we take inspiration to model the data generating process of associative learning in individuals, more specifically, dynamic learning under uncertainty.

1.3 Preceding work

The cultivation and convergence of computational methods and traditional neuroscience have been accelerated by key works in which computational models have been shown to intuitively map to cognitive learning (Huys, 2011). This enables one to test the many hypothetical relationships between the mechanics of learning and executive functionality (Zhang and Gläscher, 2020). Extracting information about the latent learning process may, in future, allow one to understand the conditions or personal attributes that relate to one's learning ability. We turn our attention to one niche of learning, *operant conditioning*, whereby individuals are tasked with forming some association between stimuli and feedback.

1.4 Objective

In this thesis we conduct an experiment consisting of a task battery to assess a series of neuropsychological executive functions with the aim of examining the relationships between the latent processes that govern probabilistic learning and various elements of working memory. We leverage the Wisconsin Card Sorting Task (a working memory and set shifting game) to study the learning process, modelling a biologically plausible, abstract representation of associative learning. We then aim to relate the recovered set of parameters - describing the learning process - to other areas of working memory, attentiveness and broader executive functions. Arriving at our primary scientific enquiry:

Can we recover a model of the latent associative learning process that governs decision making under uncertainty, offering reason about the underlying cognition? Furthermore, we hope to illustrate how to relate the parameters that govern this process to auxiliary executive functionality, and in particular examine the relationship between the learning mechanics and various forms of working memory.

More specifically, this requires finding an optimal generative statistical model that adequately represents human learning and illustrating how the parameters extracted from the generative learning process may be used in analysing the observable discrepancies across individuals.

Not only does this examine the viability of using reinforcement learning as a framework for representing associative learning and operant conditioning - offering insight into the latent generative process - but it also investigates one's reliance on working memory when performing learning tasks. Lastly, the modelling procedure demonstrates how arbitrary complexity can be encoded when fitting Reinforcement Learning models to behavioural data-sets and still achieve generalisation, thus laying foundations for further scientific exploration.

Chapter 2

Literature Review

In addressing our research objective, we turn our attention to the space of most salient ideas. Given the inherent interdisciplinary nature of the project, we draw inspiration from many fields, most of which converge under the general umbrella of computational cognitive science.

An abundance of works have been explored in recent years, of which many have used psychological and biological data to characterise individual's cerebral processes.

This chapter first explains how neuropsychological metrics are assessed - primarily from a neuroscientific and psychological perspective - and thereafter introduce mathematical and statistical ideas that are used to model behavioural data. These two vantage points then converge to describe the computational approach taken to cognitive science.

2.1 Neuropsychological experiments

How do we best assess an individuals' cognitive and psychological attributes? Neuropsychological tasks are designed to rely on specific, and where possible, isolated sets of cognitive abilities offering a quantification of an individual's cognitive capacity.

In general, a neurological test is a gauge of cognitive performance across a - or multiple - dimension/s, and is often undertaken in psychological studies or for more detailed patient diagnosis (Baker, 2012). These methods allow professionals to gauge the severity of a deficit or impairment of certain cognitive functions, or capture the decay of said functions over multiple tests.

Conducted to measure psychological functionality, that is hypothesised to be linked to a particular neurological structure or pathway, neuropsychological tests are used for research into brain functionality or for diagnosis of clinical deficits (Boyle, Saklofske, and Matthews, 2012).

Often administered in idealised settings (where participants are free from distraction, focused on an isolated task), the assessments are considered to estimate peak cognitive performance (Baker, 2012).

Most neuropsychological tests utilise traditional psychometric theory - whereby an individual's scores have no absolute meaning but are rather interpreted with respect to their performance relative to the other sample subjects (Lezak, Howieson, Bigler, and Tranel, 2012).

Conducting neuropsychological experiments are regularly motivated by the desire to measure:

1. Cognitive performance: indicative of how well individuals perform across some cognitive task with respect to some sample.
2. Left-right comparisons: contrasting the left and right sides of the body/brain.
3. Pathognomonic signs: tests that diagnose or measure the severity of distinct disorders. Pathognomonics relate to the symptoms indicative of a particular disease, condition or pathology.
4. Differential patterns: signals that are symptomatic of particular diseases or cognitive impairments.

Tests are characterized by the cognitive function under examination (Lezak et al., 2012). Although the classification can be blurred, the broad categories constitute:

- **Intelligence**: IQ and related metric tests - when one is working in a clinical setting, mental deterioration/decay ought to be considered.
- **Memory**: although debated, a clinical consensus suggests there are five distinct types of memory (Lezak et al., 2012). *Working memory* describes short term memory and long term memory is divided into *declarative/explicit memory* - which includes *Semantic memory* and *episodic memory*- and *non-declarative/implicit memory* which is decomposed into *procedural memory* and *priming/perceptual learning*.
- **Language**: tests associated with speech, reading and writing.
- **Executive function**: which constitutes various cognitive processes and sub-processes, capturing abstractions like: problem solving, planning, organizational skills, selective attention, inhibitory control and some aspects of short term memory.
- **Visuospatial**: which is concerned with visual perception, construction and integration - often associated with the parietal lobe.
- **Dementia specific**: An attempt to quantify one's severity of dementia.
- **Batteries**: assessing multiple neuropsychological functions by combining a series of tests to provide an overview of cognitive skills or an investigation into the relationship between neuropsychological attributes.

Neuropsychological experiments are an attempt to isolate, or modularise, brain activity and attributes in order to better understand an individual's cognitive constructs. One illustrative set of these constructs are executive functions (EFs).

2.2 Executive function

Quintessential to the human experience, executive functions describe a set of higher order cognitive processes that allow individuals to plan, self-regulate, prioritise and sustain long-term efforts towards goals (Duncan, 2010a). Although further dissections are available, executive functions are often categorised into three primary groups (Hooker, 1960):

1. *Working memory*: the ability to temporarily hold and manipulate information.
2. *Cognitive flexibility*: (also known as *set-shifting*), the ability to shift between tasks that have distinct cognitive demands.
3. *Inhibitory control*: (or cognitive control) is the ability to self-regulate and postpone immediate gratification for potential future gain (Hooker, 1960).

Learning and adaptability are great examples of sophisticated cognition which we are particularly interested in is the role of working memory in associative learning. Associative learning is the process of forming an enduring connection between two elements in a system. This can be learning a stimulus and response - a mental representation of an event or elements in a neural network association (Duncan, 2010a).

Believed to largely be performed in the prefrontal cortex (Duncan, 2010a), executive functions (EFs) are undoubtedly pivotal to associative learning being heavily reliant on working memory and set-shifting (Hooker, 1960).

Quantifying the discrepancies between individuals will almost always rely on some statistical procedure. While traditional statistical tests can be used to test mean or variance differences across groups, more sophisticated paradigms are used to model these discrepancies - as detailed in the sections to follow.

EFs have been shown to rely on hierarchical mental models (Kopp, 2012). This may be naturally represented in a Bayesian manner, encoding bias beliefs as Bayesian priors (Knill and Pouget, 2004). Kopp, 2012 displayed the updating nature of executive functions, whereby the mechanism by which executive functions mature has been shown to be adequately represented by a trial and error process (Duncan, 2010b).

Intelligent behaviour is a consequence of assembling a series of subtasks, creating structured mental programs, to achieve some set of goals. In recent years, the relationship between executive functions (governing intelligent behaviour) and various genetic (temperament/disposition) and behavioural (personality) factors have been examined (Thomson and Jaque, 2017). These effects are, however, interactive in nature requiring careful consideration when interpreting any experimental data. For instance, Thomson and Jaque, 2017 were able to describe an individuals' ability to self-regulate as a function of their temperament, and early developmental experiences - illustrating the complexity of neuropsychological characteristics and executive functions, thus requiring nonlinear (or multilinear) models to represent this information. Similarly, Braver, 2012 demonstrated the importance of flexibility in cognitive control (inhibition), postulating self-regulation as analogous to an internal update equation in the evidence of stimuli.

These papers emphasise the complexity and nonlinear nature of trying to model executive functions, and further solicit both updating (learning) and Bayesian (hierarchical) mechanics to represent executive functions adequately.

2.2.1 Measuring EFs

How do we measure neuropsychological attributes driving executive functions? Researchers regularly measure psychological constructs by requiring participants to perform some task and subsequently monitoring behaviour and performance (Duncan, 2010a). The availability of more accurate tools (functional magnetic resonance imaging (fMRIs), electroencephalogram (EEG), electrocardiography (EKG), etc) and techniques - many of which are driven by machine learning (signal processing, computer vision, etc) - have extended our reach into connecting advanced mathematical models to biological phenomena (Braver, 2012). For example, Sherman et al., 2016 show the relationship between learning and cortical activity observable in EEG. In particular, their findings suggest that alpha-band neural oscillations (a particular frequency of power spectral bands measuring electrical activity in the brain captured by EEG readings) periodically transmit prior evidence to visual cortex. Increasing evidence suggests an association between expectations and early visual processing, alluding to the link between expectations and attention.

If EFs are so universal to the human condition, offering the basic constituents of higher-order cognition, can we observe meaningful differences in EFs in healthy individuals? Friedman and Miyake, 2017 show that individual attributes in EFs can be observed at both the genetic and behavioural levels, arguing that EFs:

1. are correlated but separable when measured by observable variables,
2. are distinct from general intelligence (often referred to as g),
3. are highly heritable and possibly highly polygenic (influenced by more than one gene), and finally,
4. activate both common, specific unique neural pathways.

A case for observing individual discrepancies in EFs through behaviour is made by (Braver, 2012), where cognitive control is the latent process that regulates behaviour. Furthermore, Miyake and Friedman, 2012 show that individual differences in EFs: (1) show unity and diversity, meaning that they are correlated yet separable; (2) reflect substantial genetic contributions; (3) have direct clinical and cultural relevance; and (4) show developmental stability.

Our primary interests are working memory and set-shifting - the measurable latent processes that drive associative learning under uncertainty (Capon, Handley, and Dennis, 2003). Working memory capacity (WMC) has often been linked to decision making, more specifically syllogistic spatial inference, which is a specific kind of premise, based logical reasoning. Playing a pivotal role in most EFs, verbal and spatial working memory capacity are able to predict performance in syllogistic (deductive reasoning (Duncan, 2010a)) and spatial reasoning assessments (the ability to process and manipulate objects in space).

Working memory capacity has also been shown a significant association with fluid intelligence - the ability to perform abstract reasoning and problem solving independent of prior knowledge (Duncan, 2010a). Unsworth and Engle, 2005 examined the link between working memory and fluid intelligence as an illustration of broader intelligent behaviour; detailing how learning and higher order functions necessitate the manipulation and temporary retention of information.

When investigating the constituents of top-down (hierarchical) processing, which can be intuitively modelled in a Bayesian framework, Deco and Rolls, 2005 were able to link attention to decision making, by examining the relationship between attentive states and motor outputs.

Panichello and Buschman, 2021 show the neurological links between attention and working memory when executing cognitive control. The paper illustrates how sensory processing, in the form of stimuli and feedback, is localised to the prefrontal cortex and how working memory and attention show significant neurological overlap.

Executive functions and other higher order cognitive abilities are frequently measured through studying subject behaviour while performing tasks. In the next section we describe how and why certain tasks may be used.

2.3 Neuropsychological task battery

This section provides a descriptive account of psychological and neurological tasks used to measure cognitive attributes and a theoretical breakdown of the procedure, protocol

and implications of the tests conducted. To capture the relationships between executive functions, we are particularly interested in the correlation/interaction across tests.

2.3.1 The Wisconsin Card Sorting Test (WCST)

The **WCST** is a working memory and set shifting (or task switching) test: testing one's ability to display flexibility under changing conditions, that is, one's ability to shift attention. A participant or subject in a study taking the WCST is required to perform associative feedback driven learning by mapping stimuli to reward. In doing so, they are implicitly tasked with associating a value with a stimuli and to dynamically update these value approximations (Barcelo, 2020).

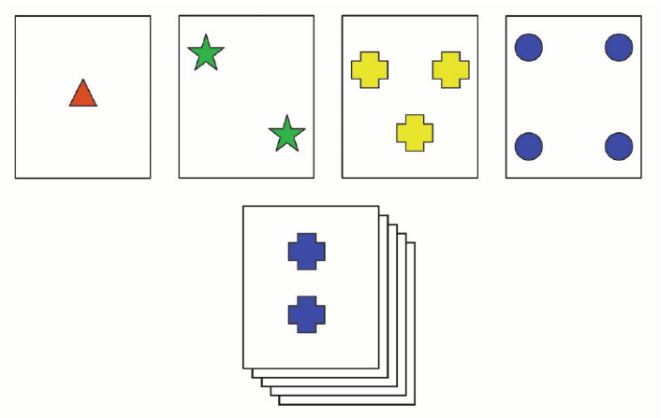


FIGURE 2.1: A WCST instance, the participant is tasked with matching the card below with one of the four cards above. The card could be matched on a number of dimensions (colour, shape, orientation or number of symbols). After an action is taken, a reward is received indicative of whether or not the correct matching rule was applied (Jonides and Nee, 2005).

Methodology

The participant is provided with a set of stimulus cards and told to match the card with one of the available options. Without being told the matching rule, the participant is simply given binary feedback indicating whether or not the chosen match is correct (Nyhus and Barceló, 2009). The experiment takes between 12 and 20 minutes and generates a number of psychometric scores pertaining to the relative numbers of categories achieved, trials, errors and perseverative errors.

Throughout a trial sequence, one stimulus (card matching rule) will return a positive reward, indicating that the matching rule is correct. All other options return a negative reward indicative of an error. The matching rule changes stochastically after a number of trials, requiring the user to update their beliefs. A perseverative error is one in which the rule has changed but the previous rule is employed incorrectly, showing an inability to adapt.

Clinical use

Often used to measure a patient's frontal lobe dysfunction, neuropsychologists and clinical psychologists utilise the WCST in patients with acquired brain injury, neurodegenerative disease, or mental illnesses such as schizophrenia (Lezak et al., 2012).

The frontal lobe has been shown to be active during feedback driven learning and strategic planning. It has also been linked to organised searching, often in the context of directing

behaviour towards achieving a goal. The same process requires self-regulation to modulate impulsive responses when conducting multi-step planning (Lezak et al., 2012). The WCST directly relies upon an array of cognitive functions including working memory, attention and visual processing. The task can measure a patient's competence in abstract reasoning and the ability to change problem-solving strategies when needed.

2.3.2 N-back task

The N-back task tests working memory and working memory capacity (WMC), by requiring participants to continuously recall stimuli delivered n time steps back (Gazzaniga, 2009).

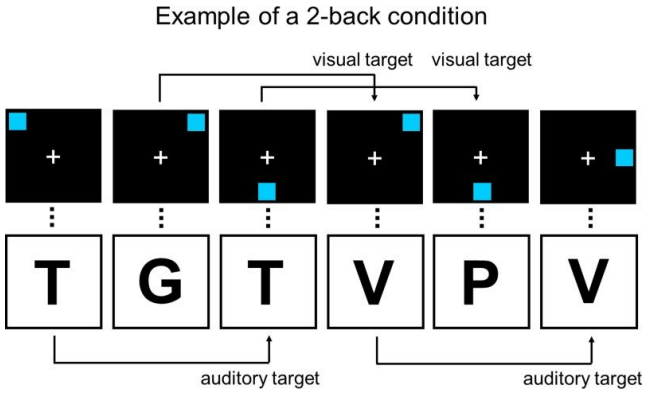


FIGURE 2.2: A depiction of a visual (above) and auditory (below) $\{n = 2\}$-back task. In our instance, a visual stimulus task is employed using letters as stimulus (Salminen and Schubert, 2012).

Methodology

The participant is presented with stimuli and required to indicate when the current stimulus matches that of n prior time steps (Gazzaniga, 2009). n is considered a *loading factor* - that scales task difficulty.

The task challenges the active part of working memory. The subject is required to both maintain and manipulate information in working memory. Not only does the candidate need to keep a representation of recently presented items in mind, but also continuously update the point of comparison (Gazzaniga, 2009).

Construct validity

Correlation analysis has brought into question the construct validity (that is the psychological reliability of the method to measure the desired quantity in clinical research) of the N-back task (Kane et al., 2007). Although it has achieved widespread adoption in both clinical and experimental settings, there are a few studies which find weak correlations between individuals' performance on the N-back task and performance on other widely accepted assessments of working memory (Jaeggi, Buschkuehl, Perrig, and Meier, 2010).

It has been hypothesised that this discrepancy is either due to the N-back task assessing access of different "sub-components" of working memory or, more seriously, the task depends greatly on familiarity and recognition-based discriminative processes that are heavily reliant on "active recall" (Kane, Conway, Miura, and Colflesh, 2007).

There have been a series of publications and media articles that explore the relationship between the N-back and improved fluid intelligence, and other working memory capacities

(Kane et al., 2007). The findings, however, are inconclusive and controversial - many of the original studies proving unreproducible.

Wider use

The popularisation of the N-back task has resulted in its adoption outside of experimental, clinical and medical settings (Owen, McMillan, Laird, and Bullmore, 2005). Many non-clinical applications utilise the task in attempt to improve fluid intelligence. It has also been applied to improve focus in individuals with ADHD, and to rehabilitate sufferers of traumatic brain injury. Many in the industry claim that the effects are not transient and generalist, or transferable to general cognitive processing (for example, fluid intelligence) - although in reality the same claims are controversial and often unfounded (Owen et al., 2005).

Neurobiology

Neuroimaging has allowed for detailed examination of the specific neurological activity during cognitive tasks. The lateral premotor cortex; dorsal cingulate and medial premotor cortex; dorsolateral and ventrolateral prefrontal cortex; frontal poles; and medial and lateral posterior parietal cortex have all be shown to activate during the N-back task (Owen et al., 2005).

2.3.3 Corsi Block Span task

Originating in the 1970's, this psychological task engages an individual's visuospatial short term working memory (Kessels, Van Zandvoort, Postma, Kappelle, and De Haan, 2000). A researcher administering the task lays out a sequence by tapping the blocks in a particular pattern, the candidate is then required to recall the pattern. Typically, the patterns start off simple (shorter length sequences) and grow in complexity or length until performance suffers, with the average person achieving a sequence length - or Corsi Span - of about 5-6.

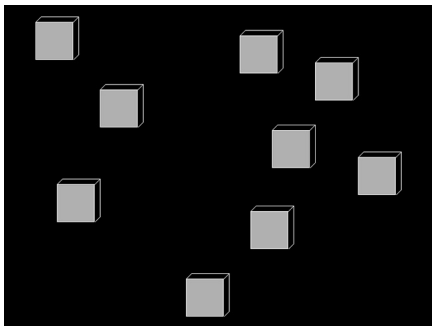


FIGURE 2.3: A typical layout of the Corsi block tapping task, prior to receiving a sequence (Kessels et al., 2008a).

Clinical use

The Corsi test is often used in experiments to measure the deficits in verbal and spatial memory spans in individuals with some neurological impairments. The task can be used to exhibit the reduction in working memory in individuals with Alzheimer's and multi-infarct dementia (Carlesimo, Fadda, Lorusso, and Caltagirone, 1994).

The Corsi block test is most commonly used to assess: memory loss; brain damage; spatial memory and nonverbal working memory (Gazzaniga, 2009).

Neurobiology

FMRI studies reveal that the ventrolateral prefrontal cortex is highly active when performing the task (Toepper et al., 2010). Relating the task to working memory models, one's visuospatial sketchpad is required during the task, however one's phonological loop is not (Vandierendonck, Kemps, Fastame, and Szmalec, 2004). Central executive resources are required as the sequence grows in length past 3 or 4 items. Another FMRI study indicates that brain activity does not appear to increase as the length of the sequences grow (Toepper et al., 2010). Furthermore, there are no significant differences in scores achieved by different genders, nor are there age related advantages past the age of 14 (Farrell Pagulayan, Busch, Medina, Bartok, and Krikorian, 2006).

2.3.4 Backward Corsi Block Span task

A slight alternative to its originator, the Backward Corsi Block span task requires a candidate to recall each sequence backwards. There appears to be no difference in difficulty between the forward and backward Corsi block tasks, as test scores do not differ significantly (Kessels et al., 2008b).

However, one notable study found that visuospatial learning disabled (VSLD) children performed far worse on the backward version of the task than they do on the forward version, whilst other children do not, thus indicating the backward task utilises specific spatial processes (Mammarella and Cornoldi, 2005).

Methodology

Cognitive impairments, deficiencies or hindrances may present themselves in the discrepancy between the forward and backward Corsi-Block span task. As such, a binary variable may be added to the experiment to capture whether or not an individual subject shows significant differences in the forward versus backward Corsi task. This may be indicative of an inability to generalise visuospatial learning to unique domains, or may offer insight into inertia effects that arise as a consequence of a subject becoming too invested in the strategy of the previous Corsi instance, as experiments are conducted sequentially.

2.3.5 Fitts' Law

Fitts' law is an equation that describes the speed of movement in human-computer interactions, which predicts how easily an individual is able to reach some target area, and as a function of the distance to and size of the target (Fitts, 1954a). Naturally, this may be inflated by one's ergonomic setting, whereby comfort and familiarity with one's system results in faster functionality.

Interestingly, the model has been adopted as one of the fundamental principles of artistic design as it allows illustrators to ensure their systems are user friendly (Fitts, 1954a). The model is described by the equation:

$$\text{Movement Time} = \log_2(2 \times \frac{\text{Distance}}{\text{Size}}).$$

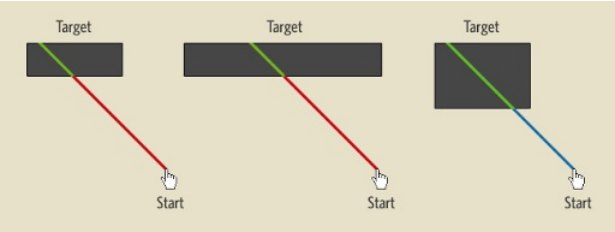


FIGURE 2.4: A instance of a Fitts' law experiment, quantifying movement as a function of distance and size (Fitts, 1954b)

This theoretical model has been studied in depth and achieves remarkable predictive power when contrasted with real experiments/data.

Methodology

The implementation of a Fitts' law task exploits its known properties to assess cognitive abilities. As a cognitive task, it requires that participants move a cursor from a given starting location to a target. The speed can be computed and compared to the expected behaviour given by the law formulation (Fitts, 1954a). The discrepancy may be indicative of cognitive differences in individuals.

2.3.6 Navon Task

A Navon figure is defined as a large recognisable shape - such as a letter - that is made up of a collection of smaller, again easily recognisable, objects (Navon, 1977a). It has been shown that individuals perceive global features before perceiving local features. Work done by Davidoff, Fonteneau, and Fagot, 2008 found that a culture isolated from western influence exhibited the opposite result: local features were identified before global features. Another paper found patients with simultanagnosia - a rare condition where individuals are unable to identify more than a single object at a time - are unable to detect the local structure, only identifying global features (McKone et al., 2010). Related work found that East Asians demonstrated significantly stronger global processing than Caucasians.

FIGURE 2.5: An example of a Navon figure. It is probable that the observer noticed the macro-structure (the H) before noticing the micro-structure (the $x's$). Subjects are tasked with identifying specific letters, with the aim of measuring whether the global or local features are identified first (Wen and Kawabata, 2018).

The Navon effect

Participating in a Navon figure related tasks has revealed fascinating short term effects. Individuals tasked with reading the micro-structure of Navon figures for as little as five minutes have shown a deteriorating ability in facial recognition tasks (Macrae and Lewis,

2002). The transient effects also correlate with the properties of the image. The effect has been shown across vastly different tasks, from facial recognition to golf putting. Many recent papers rely in this effect to prime individuals, addressing whether or not soliciting a local or global attention bias correlates with spatial-visual task performance.

Navon, 1977a illustrated how the substantial discrepancy between global and local processing (of both visual and auditory stimuli) may predict downstream cognitive tasks (in their case a visual-spatial task). Their findings support the prevalence of global over local processing, observing that global factors may influence, bias, or inhibit local processes, but the reverse does not hold.

2.4 Neuropsychological relevance

Now that the tasks have been described, it is important to understand where they are utilised in the literature, first from a purely psychological and biological perspective, and subsequently how best the behaviour of participants may be modeled. It is important to understand the assumptions and theoretical expectations associated with a task battery, with the ultimate goal of addressing some research question.

The associative, feedback induced, learning ability is captured by the WCST in our experiment. For the last four decades the WCST has been the most distinctive test of prefrontal function (Nyhus and Barceló, 2009). Although advancements in brain imaging and clinical research have brought into question the task's ability to discriminate between frontal and non-frontal lesions, it still holds as a reliable mechanism for gauging working memory constructs.

It has been shown that the development of EFs, in particular driven by working memory and set-shifting, continues through adolescence (Huizinga, Dolan, and van der Molen, 2006). Working memory is believed to be the primary contributor to WCST performance. While set-shifting appears to stop developing through adolescence, it can be observed that working memory continues to develop into young-adulthood (Huizinga, Dolan, and van der Molen, 2006).

In statistical analysis we are often interested in latent constructs - that is information that is thought to generate the observed data but is separate from the observed data as it is either far less noisy or can be broken down into constituents (Hastie, 2001). Similarly, neuropsychological tasks are thought to examine a set of psychological latent constructs (Whyte, 2019). The WCST task is regularly coupled with psychological task batteries to assess latent psychological factors. Statistical techniques that decompose data in independent components, such as confirmatory factor analysis, has been used to show moderate correlations in set shifting, information updating and monitoring, and inhibition when using executive functions (Miyake et al., 2000). These multi-faceted psychological experiments repeatedly observe that WCST relates most prominently to set-shifting and working memory.

The examination of higher order executive functions can readily be broken down into latent constituents, suggesting that latent variable analysis is a fruitful approach to understanding the mechanics of executive functions, and that learning latent parameters that govern the learning process may yield non-trivial and theoretically aligned insights (Miyake, Friedman, Emerson, Witzki, Howerter, and Wager, 2000).

The reliance on working memory in associative learning is undeniable (Schultz, Dayan, and Montague, 1997), as such, measuring different working memory assessments may offer insight into the learning process.

An increase in 4-8 Hz power spectral density bands and a decrease in 8-25 Hz bands - measured via electroencephalography (EEG) - are observable during the N-back assessment: neurological activity often associated with WMC (Palomäki et al., 2012). Examination of the N-back's utility has been rigorous, it is considered consistent and reproducible, vouching for it is reliability as a working memory assessment (Ahonen, Huotilainen, and Brattico, 2016).

Exhibiting similar robustness, the Corsi block span test has been shown to offer consistent accuracy in both forward and backward instances (Brunetti, Del Gatto, and Delogu, 2014). Additional to the standard visuospatial working memory assessment (Smyth and Scholey, 1994), subject response times, when completing the assessment, may offer insight into the mechanisms underlying spatial sequence processing (Brunetti, Del Gatto, and Delogu, 2014). The task has been shown to directly correlate with associative learning (Miyake et al., 2000).

Biologically, neuroimaging has been used to link the Corsi task to right brain hemisphere processing (MILNER, 1971), and it has regularly been observed that subjects with either right hemisphere damage or visual field defects (VFDs) display significantly delayed reproduction of Corsi sequences (De Renzi, Faglioni, and Previdi, 1977).

The measurable dependence on working memory during associative learning may be obscure and hard to observe. For example, individuals with Parkinson's have been shown to exhibit deficits in the coordinating and integrating function of the central executive (a brain region associated with control and regulation of internal cognitive processes) that is governed by working memory (Dalrymple-Alford et al., 1994). The same individuals display poorer work fluency and report higher levels of depression. The decay in spatial working memory is equally present in individuals with schizophrenia (Chey et al., 2002). The same experiment, however, showed that these individuals exhibited no reduction in performance in the WCST (Dalrymple-Alford et al., 1994).

The Navon task allows us to quantify the dependency of associative learning on attention. Although it remains unclear whether or not global vs local attention relates directly to associative learning when performing probabilistic tasks, Tan, Lim, and Manalo, 2017 show links between risk taking and a tendency to prioritise global over local attention. Their findings are robust, controlling for participants' needs for cognitive stimulation (by measuring how much they enjoy effortful cognitive activities), alluding to possible dependency between relative attention mechanisms and exploration in learning (risk taking).

Finally, elements of motor control and computer literacy can be assessed by the Fitts task. Shown to be a reliable source of quantifying motor skills, measuring how Fitt's law deviates from the expected value may offer insight into the individuals' motor systems (or indirectly computer familiarity) (Fitts, 1954b). It is important to account for these motor skills so as not to erroneously conclude correlative relationships that may be better explained through these confounding effects.

We anticipate cognitive aging to show a negative correlation with motor and executive functions which may, consequently, manifest in one's learning ability (Chang, 2021). Chang, 2021 observed that the linear functions between computational load and performance time (captured by the Fitts task) differed significantly between motor and executive tasks in younger and older participants.

There is also evidence to suggest that cognitively impaired subjects' performances on the task decay at a substantially faster rate than healthy older subjects (Poletti et al., 2017).

Lower information processing speeds and deficits in executive function are correlated with the decline of sensorimotor performance in Fitts' task, and as such our Fitts measurements may offer insight into these mechanisms. Poletti et al., 2017 subsequently investigated the relationship between age, cognitive impairment and strategy execution and distribution (that is, those strategies used to solve sensorimotor tasks). They found that significant differences are observable in both the types and distributions of strategies employed to solve simple sensorimotor tasks, alluding to possible distinctions between subjects in our learning experiment.

We now turn our attention to the state-of-the-art computational methods used to model, investigate and simulate these neurological processes.

Before going into great detail of how these cognitive attributes are modelled, in the next section we introduce statistical ideas essential to formulating a mathematical approach to cognitive science.

2.5 Foundations of Reinforcement Learning

Reinforcement learning (RL) offers an intuitive framework to formalise any stochastic, sequential decision making process (Silver, 2015). As a consequence of its broad utility, RL has been studied from many directions. Engineers study optimal control, mathematicians and economists study operations research and bounded rationality, and neuroscientists and psychologists study reward systems and classical/operant conditioning. Each of these disciplines can be considered special cases of the broader RL/Markovian paradigm (Sutton and Barto, 2018).

There are four definitive characteristics that distinguish RL from other machine learning paradigms:

1. The models are unsupervised but rewards are provided to proxy performance.
2. Feedback is regularly delayed.
3. Data is sequential and initially violates i.i.d. assumptions (although techniques exist circumvent this).
4. Actions affect the subsequent data received (Silver, 2015).

RL is primarily concerned with describing a sequence of choices that an *agent* takes in an *environment*. The agent could describe a robot, car, person in an economy or a trading agent - anything that requires sequential decision-making (Silver, 2015).

Here we discuss one specific instance of RL to illustrate its practicalities, as well as the theory in the abstract. Consider a simple instance of a robot that wishes to navigate a two-dimensional maze.

Let a_t denote the action taken at time t. Given the maze, an agent placed in some location $State : s$ aims to move to the terminal state by taking some sequence of *actions* drawn from some policy π ($a_t \sim \pi$). More generally, some agent wishes to achieve a specified objective by taking some set of actions from the space of all possible actions A and thus traversing through the state space S. It is important to note the universality of this state space formalisation: though illustrated with a maze, it is applicable to any discrete sequential decision making process.

We then quantify the quality of the actions taken by an agent as the total *reward* attained over the agents lifetime, defined as some metric of success describing the agents objective (Szepesvari, 2010). Let $G_t = \sum_{t=0}^{\infty} r_{t+1}$, where r_t is the reward received at time t. We begin at time $t = 0$ and therefore only receive rewards after taking an action r_{t+1}. Reward is discounted to avoid infinite yields in the limit as well as logical consistency (representing the time value of returns, capturing uncertainty of future yield and maintaining consistency with empirical evidence of near term preference) (Silver, 2015) thus $G_t = \sum_{t=0}^{\infty} \gamma^t r_{t+1} < \infty$ where $\gamma \in [0, 1]$. Our agent's objective is to act (that is to select a policy π) such that we maximise the expected return:

$$\underset{\pi}{\operatorname{argmax}}\, G_t = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t r_{t+1}^{\pi} \right].$$

2.5.1 General RL notation

Stochastic RL models are a special case of Markov Processes (Gagniuc, 2017). Markov processes (or chains) are stochastic processes that model pseudo-random dynamic systems. A key tenet of Markov chains, known as the Markov property, is that future states depend only on the current state and not prior states (Silver, 2015). This is a pragmatic quality as it makes the otherwise intractable computable by leveraging conditional probabilities where marginal probabilities are unattainable, rendering the path probabilities computationally feasible in a reasonable number of operations. We define the following probabilities:

$$\begin{aligned}
s_t \in \mathcal{S} &: \text{an instance from the } \textit{state} \text{ of the system at time } t. \\
a_t \in \mathcal{A} &: \text{an instance from the } \textit{action} \text{ space.} \\
r_t \sim \mathcal{R}(s_{t+1}, a_t, s_t) &: \text{a sample from possible rewards.} \\
\gamma \in [0, 1] &: \text{ is the discount factor.} \\
\mathcal{P}^a_{ss'} = \mathbb{P}(S_{t+1} = s' | S_t = s, A_t = a) &: \text{state action transition probabilities.} \\
&\quad \text{The probability of traversing from state } s_{t-1} \text{ to state } s'_t \\
&\quad \text{when employing action } a. \\
\pi(a|s) = \mathbb{P}(A_t = a | S_t = s) &: \text{the policy } \pi \text{ is a distribution over actions} \\
&\quad \text{given states.} \\
&\quad \text{The } \textit{policy} \text{ - describes the probability distribution} \\
&\quad \text{over actions given the current state (Szepesvari, 2010).}
\end{aligned}$$

2.5.2 Deriving the value function

We aim to select a policy that maximises all future discounted rewards G. Let us denote the value v of a state s when following a particular policy π as $v^\pi(s)$. The value is the expected total reward earned starting from state s and employing policy π until reaching a terminal state. If the state values are known, our agent can simply act greedily with respect to the state values - opting for the policy that maximizes G (Sutton and Barto, 2018). The policy determines the expected reward in each state thus we define a state value function by:

$$\begin{aligned} v^{\pi}(s) &= \mathbb{E}_{\pi}\left[G_t | S_t = s\right] \\ &= \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r_{t+1} | s_t\right]. \end{aligned} \tag{2.1}$$

It is paramount to note that the value function can be compartmentalised into the immediate reward r_{t+1} and all future rewards that collectively form the discounted value of successor state $\gamma v^{\pi}(S_{t+1})$, shown here (Silver, 2015):

$$\begin{aligned} v^{\pi}(s) &= \mathbb{E}_{\pi}\left[G_t | S_t = s\right] \\ &= \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r_{t+1} | S_t = s\right] \\ &= \mathbb{E}_{\pi}\left[r_{t+1} + \gamma G_{t+1} | S_t = s\right] \\ &= \mathbb{E}_{\pi}\left[r_{t+1} + \gamma v^{\pi}(S_{t+1}) | S_t = s\right] \\ &= \mathbb{E}_{\pi}\left[r_{t+1} | S_t = s\right] + \mathbb{E}_{\pi}\left[\gamma v^{\pi}(S_{t+1}) | S_t = s\right]. \end{aligned} \tag{2.2}$$

Another important quantity, the action-value function $q^{\pi}(s, a)$ is the expect return following policy π starting in state s after taking action a:

$$q^{\pi}(s, a) = \mathbb{E}_{\pi}\left[G_t | S_t = s, A_t = a\right]. \tag{2.3}$$

The action value function can be decomposed in a similar way:

$$q^{\pi}(s, a) = \mathbb{E}_{\pi}\left[r_{t+1} + \gamma q^{\pi}(S_{t+1}, A_{t+1}) | S_t = s, A_t = a\right]. \tag{2.4}$$

For readability we let $\mathcal{R}_s^a = \mathbb{E}_{\pi}\left[r_{t+1} | S_t = s, A_t = a\right]$. It then follows that because $q^{\pi}(s, a)$ is one step (action) ahead of $v^{\pi}(s)$, we can formulate the relationship between the state and action value functions as:

$$\begin{aligned} v^{\pi}(s) &= \sum_{a \in \mathcal{A}} \pi(a|s) q^{\pi}(s, a) \\ q^{\pi}(s, a) &= \mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^{\pi}(s'). \end{aligned} \tag{2.5}$$

We can now substitute the above expression for $q^{\pi}(s, a)$ into the expression for $v^{\pi}(s)$, arriving at the quintessential *Bellman Equation* for $v^{\pi}(s)$:

$$v^{\pi}(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^{\pi}(s')\right). \tag{2.6}$$

The Bellman equation is an important quantity in RL. By formulating a state value function as a function of other states, one arrives at a recursive update equation. Similarly, we derive the Bellman equation for the action value function:

$$q^{\pi}(s,a) = \mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a \sum_{a' \in \mathcal{A}} \pi(a'|s') q^{\pi}(s',a). \tag{2.7}$$

If we have a Markov Decision Process (MDP) with n states, we can combine all the Bellman equations to get n linear equations for the n unknown value functions. We are then able to solve the linear equations for the value functions. As the state space grows, however, it quickly becomes intractable to solve the MDP in this way, thus necessitating faster approximation techniques.

2.5.3 Optimal value function and policy

The optimal state value function $v^*(s)$ and action value function $q^*(s,a)$ are defined as the maximum (respective) functions over all policies:

$$\begin{aligned} v^*(s) &= \max_{\pi} v^{\pi}(s) \\ q^*(s,a) &= \max_{\pi} q^{\pi}(s,a). \end{aligned} \tag{2.8}$$

An MDP is "solved" if we know the optimal value function - and therefore the best actions to take to maximise reward. Substituting these optimal polices $\pi = *$ into our Bellman equations yields the Bellman optimality equations (Szepesvari, 2010):

$$\begin{aligned} v^*(s) &= \max_a \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^*(s') \right) \\ q^*(s,a) &= \mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a \max_a \pi(a'|s') q^*(s',a). \end{aligned} \tag{2.9}$$

2.5.4 Exhaustive Search

It is important to intuit the data generating process utilised by Reinforcement Learning - a key dissimilarity from other machine learning techniques (Sutton and Barto, 2018). The data are the rewards generated by interacting with the environment.

Once our problem has been formalised, that is to say we have defined the state space S, action space A it is (in theory) possible to try every permissible action to compute the true values of each state. This method, known as exhaustive search (figure 2.6), is the process of computing all possible trajectories across the action space and simply choosing the action that maximises aggregated rewards. This is of course idealistic and impractical in most non-trivial problems, as the state space is generally too large to allow for this brute force method (Szepesvari, 2010).

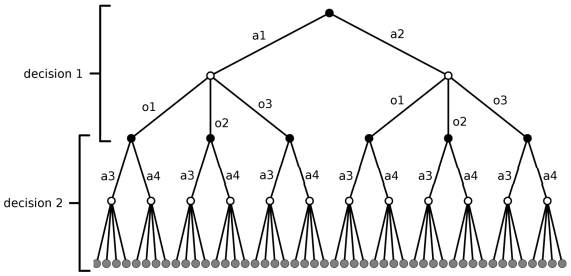


FIGURE 2.6: An illustration of the branching factor (number of possible actions at each node) of exhaustive search (Silver, 2015). a_i represents possible actions (action i) and o_i represent possible outcomes (rewards) that follow an action. It is clear to see that the space of possible trajectories compounds exponentially.

Although exhaustive search may be intractable, various dimensions of sampling the state/action space can result in adequately estimating state values, thus learning the optimal policy through some efficient search or optimisation algorithm (Silver, 2015). Next, we discuss some of the techniques used to efficiently search the value function space.

2.5.5 Dynamic Programming

Dynamic programming is the approach taken to solve MDPs. If complex problems can decomposed into simpler recursive sub-problems, that may be solved and reused, then dynamic programming can be used to solve the problem in its entirety.

The Bellman equations allow us to decompose a Reinforcement Learning problem such that dynamic programming becomes applicable. Here, we introduce three prominent approaches to dynamic programming used in RL: policy evaluation, policy iteration and value iteration.

Policy evaluation: Used to find the value function of a given policy, policy evaluation iteratively updates our value function estimate until it converges to the true value function. After initialising (possibly randomly) $v^1(s)$ we iteratively apply the Bellman equation until convergence. Synchronous updates are used too, meaning that each state's value estimates are updated in a single iteration. At each iteration $k+1$ for all states $s \in \mathcal{S}$ (with successor state s') we update our values estimates $v^{k+1}(s')$ with the Bellman equation:

$$v^{k+1}(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^k(s') \right). \tag{2.10}$$

The estimate is updated until convergence- that is until the maximum change in value estimates is below some very small threshold ϵ:

$$\max_{s \in S} \left| v^{k+1}(s) - v^k(s) \right| < \epsilon$$

Again, this assumes a policy π is known, however how may we proceed in the absence of having selected a policy?

Policy iteration: policy evaluation is used to find $v^\pi(s)$, however this search algorithm can be naturally extended to search for the optimal policy. Policy iteration is implemented as follows:

1. Initialise the policy π randomly.
2. Repeat until the policy converges to a stable estimate:
   (a) Evaluate v^π by policy evaluation.
   (b) For each state s use synchronous updates:

$$\pi(s) := \arg\max_{a \in A} q(s,a) = \arg\max_{a \in A} \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^\pi(s') \right). \tag{2.11}$$

The policy iteration can be shown to satisfy the Bellman optimality equation. Assuming a deterministic policy π that transitions to π' after an iteration, over any state s. By definition we improve the value function by:

$$q^\pi(s, \pi'(s)) = \max_{a \in \mathcal{A}} q^\pi(s,a) \geq q^\pi(s, \pi(s)) = v^\pi(s).$$

The value function is also either improved or remains constant $v^{\pi'}(s) \geq v^\pi(s)$, because:

$$\begin{aligned} v^\pi(s) &\leq q^\pi(s, \pi'(s)) = \mathbb{E}_{\pi'} \left[r_{t+1} + \gamma v^\pi(S_{t+1}) | S_t = s \right)] \\ &\leq q^\pi(s, \pi'(s)) = \mathbb{E}_{\pi'} \left[r_{t+1} + \gamma q^\pi(s, \pi'(s)) | S_t = s \right] \\ &\leq q^\pi(s, \pi'(s)) = \mathbb{E}_{\pi'} \left[r_{t+1} + \gamma r_{t+2} + \ldots | S_t = s \right] = v^{\pi'}(s). \end{aligned} \tag{2.12}$$

It follows that improvement stops if:

$$q^\pi(s, \pi'(s)) = \max_{a \in A} q^\pi(s,a) = q^\pi(s, \pi(s)) = v^\pi(s). \tag{2.13}$$

It then follows that the Bellman optimality equality has been satisfied:

$$v^\pi(s) = \max_{a \in A} q^\pi(s,a)$$

and thus $v^\pi(s) = v^*(s) \forall s \in S$ and that π is the optimal policy.

Value iteration: Consider a different method to iteratively search for the optimal policy, whereby the value function is updated directly. The value iteration procedure is executed as follows:

1. Initialise $v^\pi(s) = 0 \; \forall$ states s.
2. Repeat until convergence, that is until the maximum change in value estimates are smaller than some very small quantity $\max_{s \in S} |v^{k+1}(s) - v^k(s)| < \epsilon$
   (a) Synchronously update the value function for each state by the maximum reward for taking a further step from s to s' over all actions A, formally: $v^{k+1}(s) := \max_{a \in A} \left(\mathcal{R}_s^a + \gamma \sum_{s' \in S} P_{ss'}^a v^k(s') \right)$

It can be shown that with sufficient run time the value iteration algorithm will converge to the true, unknown, state values. In practise, however, this may require unrealistic computational time (Szepesvari, 2010).

While many other dynamic programming variants exist, the chief principles can be summarised as:

- Modularising the problem to avoid redundant computation (shared states value estimates) and then,
- Selectively sweeping over the state space to update value estimates (Silver, 2015).

Dynamic programming can thus be considered a breadth first search approach as we sweep over the action space, as apposed to rolling out a particular action sequence (Gagniuc, 2017).

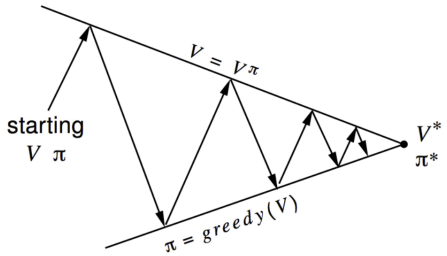


FIGURE 2.7: A graphical representation of dynamic programming - iteratively improving value estimates whilst following the current optimal policy greedily. Guaranteed to converge to the optimal state values and policy (v^* and π^*) in the limit (Sutton and Barto, 2018). Here v and π are initialised (labelled "starting") and thereafter estimates are updated iteratively until the optimal estimates are reached. "greedy" simply refers to a policy that selects the best s' deterministically.

The solutions discussed here require an accurate model of the environment - knowledge of transition dynamics T and reward distribution R. This is the domain of ***model-based RL***, an alternative ***model-free RL*** aims to estimate v and q directly from the data generated by interacting with the environment (Szepesvari, 2010).

2.5.6 Model-free RL

Dynamic programming sweeps over all states $s \in S$ to search for v^* and π^*, instead we may wish to take actions in the environment and update our estimates after observing the returned results (Silver, 2015). Once our state space, action space and objective are defined we need to search for value function estimates that will resultantly act in accordance with the objective - that is, to find the optimal policy.

In the absence of known transition dynamics and reward distributions the exact solution would require iterating over all possible sequences - which can often become intractable, as illustrated in figure 2.6. We may wish to employ approximation methods to iteratively search the solution space while updating our value function estimates (Szepesvari, 2010).

2.5.7 Monte Carlo learning

When determining when to update our state values, we may conduct purely depth-first search via Monte Carlo sampling (Sutton and Barto, 2018). Whereby we roll-out a single

policy until reaching the terminal state, and then approximate the current state value by the weighted average of the yielded returns. Formally, we may estimate a value function $v^{\pi}(s)$ by sampling T steps (following some policy π) N times and averaging the results to estimate an expected reward:

$$v^{\pi}(s) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \sum_{t=0}^{T} \gamma^t r_t^i | s_0 = s \right\}.$$

where $r_t^i \sim \mathcal{R}(s, a, s^{`})$ is a reward received by taking an action at time t on run i.

This approach follows a sequence of actions from some starting state s_0 to the terminal state s_T (assuming a finite RL problem the terminal state refers to the condition where no the game ceases), and averages the rewards received to estimate the true value function of the initial state $v(s)$. An alternative would be to update the estimated value before reaching the terminal state T.

2.5.8 Temporal Difference (TD) learning

In contrast with Monte Carlo methods, TD learning updates the estimates of $v(s)$ after taking a single action. This method - also referred to as bootstrapping - computes the weighted average over all possible actions at a given state s to update the estimate of $v^{\pi}(s)$ after a single action (Sutton and Barto, 2018).

Recall that the Bellman equation for the value function is given by:

$$v^{\pi}(s) = \sum_{a \in \mathcal{A}} \pi(a|s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^{\pi}(s') \right).$$

The marginal change in $v(s)$ is then defined as:

$$dv(s) = -v(s) + \sum_{a \in \mathcal{A}} \pi(a|s) \left(\mathcal{R}_s^a + \gamma \sum_{s' \in \mathcal{S}} \mathcal{P}_{ss'}^a v^{\pi}(s') \right).$$

which results in the update rule (Sutton and Barto, 2018):

$$v^{i+1}(s) = v^i(s) + \alpha \, dv(s).$$

where $\alpha \in [0, 1]$ is a learning rate that dictates how aggressively the estimate is updated. TD learning samples from the respective distributions:

$$\begin{aligned} a_t &\sim \pi(a|s_t) \\ s_{t+1} &\sim \mathcal{P}_{s_t, s_{t+1}}^{a_t} \\ r_t &\sim \mathcal{R}(s_{t+1}, a_t, s_t). \end{aligned}$$

Therefore a single action a_t, reward r_t and subsequent state s_{t+1} form $dv(s)$. Let us denote this sample δ_t:

$$\delta_t = v^i(s_t) + r_t + v^i(s_{t+1})$$
$$\implies v^{i+1}t(s) = v^i(s) + \alpha\delta_t.$$

One can see how this can be interpreted as reward prediction error (RPE), updating our belief proportionally to the discrepancy between our current belief and feedback from the environment (Sutton and Barto, 2018).

Bias-variance trade off: It is also intuitive to see the bias-variance trade off when sampling the solution space: TD learning updates value estimates with a single sample return r_t, whilst Monte Carlo learning runs a policy to completion utilising many returns $\sum_{t=0}^{T} \gamma^t r_t$. Whilst Monte Carlo is then less biased and reliant on a greater sample from the state space, it exhibits vastly greater variance as each individual r_t is a sample reward with associated stochastic fluctuations (Sutton and Barto, 2018).

Finally, TD-learning and Monte Carlo can be consolidated by means of TD(λ) update rules - where λ dictates the number of samples required to perform an update (Sutton and Barto, 2018). Figure 2.8 exhibits the state of possible search algorithms captured in the Markovian RL framework (Szepesvari, 2010).

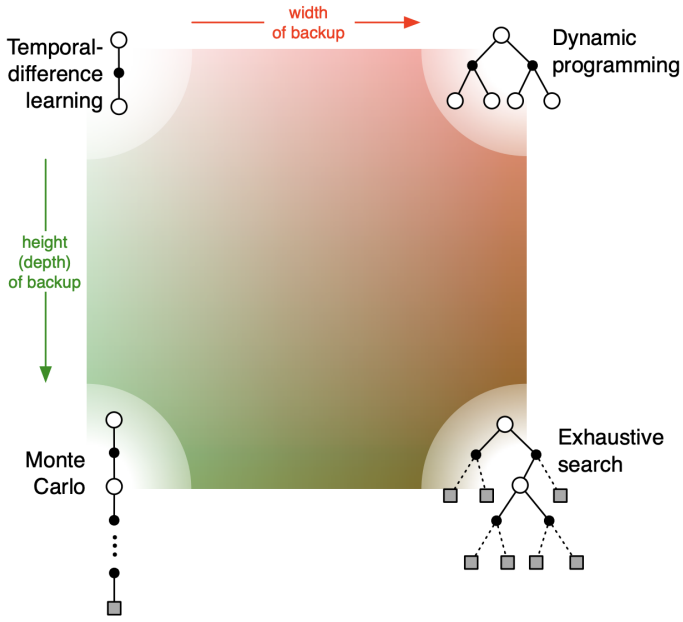


FIGURE 2.8: An illustration of the space of reinforcement learning methods (Sutton and Barto, 2018). Round nodes represent states (empty) and rewards (coloured) and the links between nodes represent policy trajectories. The square nodes represent the terminal states, ending a sequence. Exhaustive search traverses all possible branches in the search space, the other methods focus on updating estimated state values after only searching a section of the possible trajectories. The depth of the search shows how many actions are taken before an update to the value function is performed.

The manner in which we alternate between policies is governed by the infamous ***exploration-exploitation*** trade-off (Szepesvari, 2010). The exploration-exploitation trade off can be interpreted as a definitive characteristic that details the transitions between policies.

In the context of measuring executive functions, an exploration parameter is often of keen interest in computational psychiatry as it pertains to an individuals risk aversion by capturing the subject's tendency to try new policies.

There are copious extensions and variations to these methods, however they are beyond our requirements. The methods discussed here can be considered look up tables, meaning that we store (and iteratively update) state values over the entire state space. Many applications of interest instead attempt to fit some parametric function or distribution to the data (states, actions and additional attributes) to approximate state values (Silver, 2015).

It is worth noting that a distinct school of RL algorithms (*policy gradient methods*) focuses on learning the optimal policy π^* directly (Szepesvari, 2010). The methods we describe above are concerned with solving for state values to inform the optimal policy (Szepesvari, 2010). The latter techniques are of greater interest in our research because in neuropsychology we're concerned with understanding the driving forces behind actions - relying on RL as an approximation of cognition and thus deriving interpretable parameters.

Now that we have covered the essential elements of RL, we examine how computational methods are used in cognitive science.

2.6 Motivation behind computational psychiatry

Many, if not all, modern scientific endeavours have become increasingly data driven - capitalising on the exponentially decreasing cost to compute (Huys, 2011). Recent advancements in computational neuroscience, psychiatry and psychology, however, transcend a purely practical utilisation of large data and available compute and have begun to utilise computational methods in designing theoretical models. Consequently, we have now entered a new era of neuroscience and psychiatry that poses the brain as a computational system, taking theory from our knowledge of computational systems to infer characteristics of the mind (Adams, Huys, and Roiser, 2015).

One compelling example of this is the *Bayesian brain* hypothesis, which poses the brain as a Bayesian system that acts to collect evidence to support or oppose current beliefs (Knill and Pouget, 2004). We are constantly inundated with sensory information and are tasked with distilling and inferring reasonable responses to these signals. There is growing evidence to support the idea that neural computation is analogous to Bayesian optimisation in that we act in accordance with Bayesian optimisation principles (Parr, Rees, and Friston, 2018).

Information about uncertainty and prior beliefs are iteratively updated in light of new evidence, guiding our perception and sensorimotor control (Knill and Pouget, 2004). It then naturally follows that the brain may be representing sensory information probabilistically, assuming probability distributions over the physical world. In this way, brain function can be posed as an inferential process, combining prior beliefs with a generative (predictive) model to infer the causes of sensations (Parr, Rees, and Friston, 2018).

A fascinating sub-field has emerged where neuropsychological deficits, psychiatric disorders and pathological mental illnesses are posed as improperly updating model evidence to one's generative model by acting on aberrant priors - those with a poor fit to the real world (Parr, Rees, and Friston, 2018). That is to suggest these cognitive impairments as false inferences. Applicable to a vast array of systems - from visual neglect through hallucinations or autism - the utility of this theory of the mind surpasses simple intellectual gymnastics, offering plausible, theoretically rich, and implementable frameworks to generate (and test) hypotheses

and deriving novel solutions to exceedingly complex neuropsychological postulates (Knill and Pouget, 2004).

The key theoretical premise of this link between biology and computation is based on the idea that neurological activity is designed to compute estimates about the physical world, analogous to state-value updates in Reinforcement learning (Huys, 2013).

2.6.1 Biological Reinforcement Learning

The primary cognitive task for active organisms can be described as deciding how to act in a changing environment (Myin and Hutto, 2015). Referred to as an action-oriented view of cognitive systems, compelling arguments can be made to represent both motor and action control as RL algorithms (Rusanen, Lappi, Pekkanen, and Kuokkanen, 2021).

Given the complexity of reality, it is natural to assume that organisms act under their assumptions, or internal-model, of reality (Myin and Hutto, 2015). This idea, the premise of a representationalist view of neuroscience, suggests neurocognitive systems rely on (flexible) representations to problem solve (Rusanen et al., 2021). This framework is regularly used to describe adaptive, flexible and goal-directed behaviour.

This approach to cognition - in opposition to the view of enactivists who argue behaviour is better captured by instinctive reactivations and re-enactments - allows one to represent arbitrarily complex behaviour (Rusanen et al., 2021). Consider a human grasping a swirling object. Information across multiple sensory sources (vision, touch and kinesthesia) must be integrated. The human must adequately control multiple effectors (eyes, limbs, posture) to successfully complete the objective, in a purposeful goal-driven way.

This behaviour, requiring anticipation, preparation and planning, relies on sophisticated cognitive synchronisation, coordination and prediction (Stepp, Chemero, and Turvey, 2011).

Framing cognition in this light naturally lends itself to leveraging reinforcement learning to represent cognitive dynamics of action control (Rusanen et al., 2021). In a similar fashion, RL, posed with maximising long term rewards, utilises internal representational states to guide actions (Sutton and Barto, 2018). It is important to clarify that the representational states are not sensory-like "perceptions" but rather goal-directed abstractions, implying the need for updating personal intuitions about reality rather than simply processing sensory input.

RL allows one to quantify active control as a control system governing a cognitive agent who can learn, anticipate and adapt, thus forming a learning-as-decision making process (Sutton and Barto, 2018). This approach, showing recent success in computational cognitive sciences, artificial intelligence and robotics, offers a formal language for articulating various aspects of control processes (Rusanen et al., 2021).

RL offers a rich literature that allows for more sophisticated control theories than those from the 1960s (often simply proportional feedback models) (Chemero and Silberstein, 2008). A growing body of empirical and theoretical evidence illustrates its relevance in modeling action and motor control, memory, decision-making and learning (Hutto and Myin, 2020). Where many opposing theories of cognition are concerned with constructing descriptions of reality, RL is instead concerned with finding the best set of actions given limited information (Rusanen et al., 2021).

2.6.2 Predictive processing

Predictive processing (predictive coding or PP) refers to the theory of brain function in which the brain is constantly updating mental models of the environment - directly analogous to an agent updating state-value estimates in RL (Whyte, 2019), or posterior update in the Bayesian Brain context (Knill and Pouget, 2004).

EEG and magnetoencephalographic (MEG) have been used to illustrate the neurological relevance of utilising the WCST as a tool to represent predictive processing (Barcelo, 2020). By assessing neurological activity during 'model updating', there is clearly some link between the psychological assessment and corresponding theories of PP and active Bayesian inference (Barcelo, 2020). The WCST can be interpreted as predictive belief updating in light of negative and positive feedback.

PP has offered compelling evidence in a number of perceptual, cognitive and psychiatric phenomena. Euler, 2018 argue that hierarchical PP offers a robust framework to conceptualise the neuroscience of intelligence, and further illustrate how environmental contributions (that manifest in individual differences) play a large role in unique neural signatures that may account for differences in PP responses. This builds a significant case justifying PP (in our case represented as reinforcement learning) as the basis of a model to capture associative learning differences in individuals.

Closely aligned with predictive processing, psychological attributes such as risk aversion may be represented in the update equation in an RL model: exploratory tendencies dictating the speed at which state-value estimates are updated. The global vs local attention theory is best articulated by predictive and reactive control systems (PARCS) theory (Tan, Lim, and Manalo, 2017); as such, if we observe a tangible relationship between global/local attention and probabilistic learning, an extension into PARCS theory may offer a natural progression.

2.6.3 Applications of RL and Bayesian methods in modelling neuropsychological tasks

When analysing cognitive or behaviour data, computational models offer much deeper insight than the once ubiquitous approach of using standard aggregate statistics of simple heuristic metrics (D'Alessandro, Radev, Voss, and Lombardi, 2020). Bayesian models in particular have become an intuitive framework for modelling cognitive tasks, as the updating of priors naturally aligns with acting in light of new information (Knill and Pouget, 2004).

D'Alessandro et al., 2020 were able to effectively demonstrate the use of a Bayesian reinforcement learning model to represent the cognitive process governing the WCST, formalising the behavioural data into neurobiologically plausible, information-theoretic constructs. In doing so, they argue that model based neuroscience is an effective tool to extract meaningful information about the latent biological or anatomical processes.

A wide range of statistical techniques have been applied to model WCST, from behavioural models (that attempt to abstract high-level cognitive processes) to neural networks (that attempt to emulate biological neural networks) aiming to provide psychologically interpretable parameters or biologically inspired network structures (Steinke, Lange, and Kopp, 2020). Reinforcement learning is often used to disentangle psychological sub-processes to explain an individuals' behaviour (Silver, 2015). Bayesian models, building off the Bayesian brain hypothesis, offer the benefit of describing a subjects' performance in information theoretic quantities (D'Alessandro et al., 2020). For these reasons, it is in the combination of RL and Bayesian methods that best address our research enquiry.

A number of works have been conducted to link executive functions to different brain regions. Barcelo, 2020 in particular were able to illustrate (through examining experimental electroencephalographic (EEG) and magnetoencephalography (MEG) data of subjects performing the WCST) that the positive and negative feedback predictive cues (a system's reaction that may allude to some underlying process) - and thus abstract believe updating - are adequately modeled by an RL-like update that emulates the Bayesian brain hypothesis (Barcelo, 2020).

They posit that areas of the brain associated with uncertainty resolution (Silver, 2015) and memory consolidation are activated during the WCST. More specifically, they show that the two temporally distinct stages of cognition (1) inference about the hidden perceptual category; and (2) updating parameter estimates (learning); cause neurological activation across a distributed fronto-parietal network; linking the abstract RL theory to neurological response (Barcelo, 2020).

There is also evidence to suggest that neurological activity following stimuli during associative learning tasks can be explained as a function of novelty, task relevance, feedback (positive or negative) and information entropy with respect to novel information (Barcelo et al., 2006) - bolstering the argument for a predictive processing view of higher-order cognition.

Applying this computational RL approach to modeling learning has been shown to successfully extract meaningful (theoretically plausible) constructs that govern the underlying cognition (Van Slooten et al., 2018). Van Slooten et al., 2018 were able to fit and extract RL parameters that describe the latent cognitive process during a WCST and use these latent objects to explain the fluctuations in pupils of participants. This demonstrates how non-invasive metrics (in this case, pupillometry) may offer information about latent cognitive processes. In particular, in the future these non-invasive examinations may offer insight into individuals with psychiatric disorders (who often suffer from impediments in certain executive functions). More relevantly, RL was successfully chosen as a reliable tool to capture constituents of the latent learning process.

Similarly, Slooten, Jahfari, and Theeuwes, 2019 were able to map the relationship between latent parameters extracted from an RL model (in this case detailing the exploration exploitation trade-off measuring an individual's risk aversion and update responsiveness) to spontaneous eye blink rate.

These papers demonstrate the utility and applicability of using RL parameters to represent abstract cognitive processes during learning and decision making. Illustrating the ability to map these cognitive processes to physiological responses and (most applicable to our experimentation) generate reliable data for downstream analysis by quantifying the learning process.

A core interest in our experiments is assessing the reliance of working memory when performing associative learning/operant conditioning (captured by the WCST). Research suggests that the cognitive processes required for working memory capacity (WMC) greatly overlap with those governing instrumental learning, particularly in dynamic environments (Humann, Fischer, and Ullsperger, 2020).

Instrumental learning requires the ability to maintain items in WMC as well as update and shield items against interference or distortion (maintaining an accurate representation in memory). Humann, Fischer, and Ullsperger, 2020 were able to use reinforcement learning to illustrate that low working memory capacity individuals modulate learning rates less dynamically around value estimates; exhibiting slower updates to their internal models.

Their behaviour is also suggested to be more stochastic, less stable and highly susceptible to misleading probabilistic feedback.

2.6.4 Forms of neural computation

If we then rely on the assumption that neurological activity is fundamentally analogous to computation, what computational paradigms can be used to represent mental illness? One pragmatic dichotomisation is between errors arising in *inference* and those that arise during *learning* (Huys, 2013). Psychosis or cognitive distortions are indicative of poor inference, as the individual is incapable of accurately assessing reality (Ball and Goldstein-Piekarski, 2017). Additionally, neurological defects may arise through learning. The manner in which an individual's brain assimilates information is a function of the information to which it is exposed; which in turn can change how future information is processed. As such, exposure to certain information - such as trauma, ill-parenting or adverse life-events - may substantially distort the way in which an individual processes information (Huys, 2013).

2.6.5 Computational approaches to modeling neuroscientific data

Whilst nuances exist, the computational approaches taken to model neurological activity can be split into three categories (Huys, Guitart-Masip, Dolan, and Dayan, 2015):

- Dynamical systems
- Inferential models
- Associative learning

Dynamical systems

Often modelled by a series of differential equations, dynamical systems can be used to describe how *nodes* interact to produce certain behaviours (Ball and Goldstein-Piekarski, 2017). This may arise when modeling the anatomy of the brain - understanding the relationship between action potentials - but can be extended to any level of abstractions. Dynamical systems can capture how an individual's symptoms affect their condition (for example, how a chronically lonely individual may act in anti-social manner perpetuating their condition) (Huys, 2013).

Inferential models

Inference in this context is concerned with learning the generative process behind certain actions/behaviours (Huys, 2013). That is, given some sensory input, can we infer the dynamics of the underlying latent structure that produces some output.

Associative learning

Finally, learning mechanisms are concerned with how information changes future behaviour - how internal state, priors, assumptions and habits are updating in light of additional data (Huys, 2013). Consider a person who wishes to play chess. One likely approach taken would be to learn the rules of the game and attempt to simulate (consider different plays) at each step of a given game in order to predict an optimal outcome - acting in a manner to defeat their opponent. If the person could consider all possible game trajectories, it would be trivial to locate the optimal policy - comparable to exhaustive search in the Reinforcement Learning literature. If, instead, the individual relies on pattern matching, habitual behaviour and past experience to estimate the next best move, we can conclude

the individual is building a *model* of the *environment* and acting in accordance with the model's projections. The psychological literature refers to these learning mechanisms as *model-based* or *goal-driving* learning. The mental model is informed and updated in light of stimuli-response couplings, as the individual learns the association between stimuli and responses. Hence the term associative learning.

A seminal finding in contemporary neuroscience is the mechanism under which conditioning - a type of learning where a stimulus is associated with an outcome - pertains to dopamine neurons (Schultz, Dayan, and Montague, 1997). The formative finding shows that the release of dopamine (and thus the reward received) does not directly correspond to the actual reward received in reality, but rather the difference between the reward and the ***expected*** reward. As detailed in figure 2.9 extensive experimentation has been documented to show that the release of dopamine in the midbrain is dependent on expected (rather than absolute) reward. Examined by Schultz, Dayan, and Montague (1997) participants were conditioned on the conditional stimulus (CS) to expect rewards from beverages or sweets. Before learning the association, and when no stimulus is provided, the onset of the reward r results in a spike in dopamine (seen in the top panel of figure 2.9). Individuals are not expecting this rewards and thus are pleasantly surprised on its arrival. That is, there is a great positive discrepancy between the expected reward and actual reward. The middle and bottom panel of figure 2.9 show the dopamine response after being conditioned on the association. In both instances the stimulus alone is sufficient to trigger the dopamine release, which precedes the actual reward. After the stimulus, the participant's expectations have been shifted to expect a reward, so when the actual reward is received, dopamine remains relatively stable. Interestingly, if the reward is then omitted, a notable drop in dopamine occurs - indicative of the misalignment with expectations (Schultz, Dayan, and Montague, 1997). This expectation-reality discrepancy is known as ***RPE: reward prediction error***.

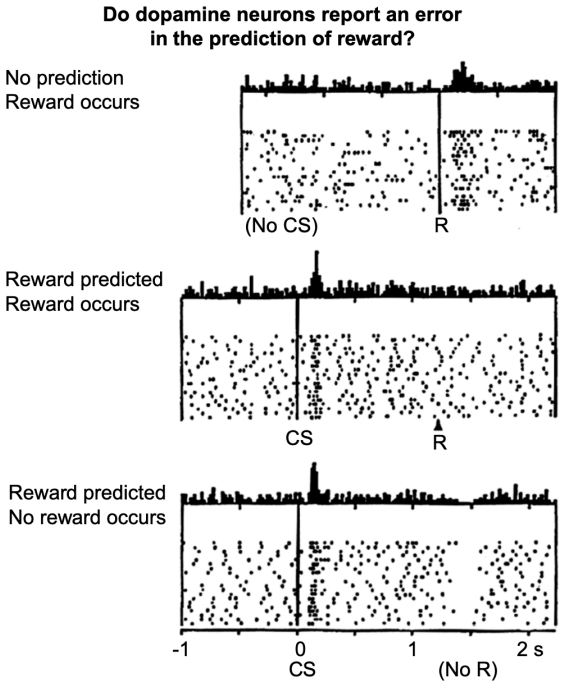


FIGURE 2.9: Seminal work detailing the relationship between releases of dopamine in the midbrain and reward prediction error (RPE). *CS* denotes conditional stimulus, R denotes reward and the x-axis denotes time. The figures show the spike of dopamine that follow the expectation of a reward (Schultz, Dayan, and Montague, 1997).

This important finding is the origin of leveraging reinforcement learning theory as a plausible framework for human associative learning - allowing the parameterisation of abstract cognitive phenomena in accordance with reward prediction errors (Adams, Huys, and Roiser, 2015).

2.6.6 Psychiatric analysis through computational models

This computational-theory driven approach to psychiatry is distinct from pure machine learning and statistical endeavours, not by the means under which the models are fit, but also by the apriori decisions made in selecting the model architecture (Adams, Huys, and Roiser, 2015). The data driven approach avoids theoretical assumptions, relying purely on statistical inference to reveal relationships in the (physiological and psychological) data; whilst the theoretical approach utilises the rich mathematical and statistical literature to define mathematically precise hypotheses about the data, by implementing the following procedure (Ball and Goldstein-Piekarski, 2017):

- ***Specification***: a statistical formulation about the (latent) data generating process.
- ***Estimation***: Fitting parameters to the observed data.
- ***Comparison***: Finding an optimal balance between model complexity and explainability to fit the best parsimonious model.
- ***Testing***: Assessing how well the optimal model recapitulates the observed data (Ball and Goldstein-Piekarski, 2017).

2.6.7 Rescorla-Wagner model

This theoretical model formulation can be illustrated by the RPE example provided in figure 2.9. A Reinforcement Learning state function estimation model called with Rescola-Wagner (RW) model formalises this cognitive process (Hazy, Frank, and O'Reilly, 2009). The RW model is a special case of Q-learning, where the value estimates of a state $Q_t(a)$ are updated at time t conditioned on stimuli a.

The Rescorla-Wagner model discretizes the RPE process, capturing the relationship between conditioned stimuli and learning (the associative learning process).

The model relates to Q-learning as state value estimates $Q_{t+1}(a)$ - associated with stimuli a - are updated in light of new information (positive or negative reward) r_t at time t.

$$Q_{t+1}(a) = Q_t(a) + \alpha \left[r_t - Q_t(a) \right].$$

The quantity $r_t - Q_t(a)$ measures the discrepancy between the individuals current expected value $Q_t(a)$ and the actual return r_t. Therefore the individual's value estimate is updated proportionally to their surprise (RPE):

$$Q_{t+1}(a) = Q_t(a) + \alpha RPE.$$

A more or less conservative update (that is the degree to which the individual weighs new information relative to old information) may be governed by the α.

The parameter α in this example is indicative of the individuals learning rate (Huys et al., 2015). Importantly, α has interpretable theoretical grounds, which may be used to

formulate hypotheses (such as whether or not different groups have different learning rates - the exact focus of our study) (Konstantakopoulos, 2019).

This Rescorla-Wagner Q-learning model formulation has successfully demonstrated statistical relevance when fitting models to cognitive data.

Many research questions examine the discrepancy between different population groups. For example Joue et al. (2021) were able to show that individuals of different genders (and hormonal groupings) exhibit different learning parameters in some tasks. The authors also demonstrate the reliability of these RL methods when contrasted with reliable medical devices (fMRI). This latter contribution was similarly demonstrated by Humann, Fischer, and Ullsperger (2020), who successfully mapped learning rate parameters to electroencephalogram (EEG) readings.

Notable contributions of the computational approach

Whilst of course these models are limited, we are effectively testing a mathematical theory of cognition, the advantages of this approach over purely theoretical interpretations cannot be understated. These models offer succinct methods to:

1. ***Traverse neurological levels of abstraction***. Computational models allow us to better move between different levels of explanation. In neuroscience and psychology we examine phenomena from an extremely varied range of perspectives - from the synaptic level to observed behaviour - these abstract model parameters may be interpreted at various levels; allowing for links between anatomical and social/behavioural phenomena (Adams, Huys, and Roiser, 2015).
2. ***Quantify the role of the environment***. These models allow us to explicitly capture environmental/social/external phenomena - better representing the state of reality. To be ignorant of our external environment is insufficient in many scientific enquiries.
3. ***Allow for better categorisation of the discrepancies between individuals***. We can design and implement very modular theories of cognition - thus learning distinct categories of pathology that may be indispensable given our theoretical knowledge of the world (Adams, Huys, and Roiser, 2015).

2.7 Theoretically plausible models

Returning to the observation that that firing of the midbrain dopamine neurons resembles a *reward prediction error* (Daw, 2011a), we aim to capitalise on the rich reinforcement learning literature to design a comparable computational framework that offers rigorous scientific enquiry. The salient contribution of utilising a reinforcement learning framework is to explicitly model the dynamics dictating the *trial by trial* response to feedback, in contrast with traditional statistical techniques that emphasis modelling average behaviour (O'Doherty, Hampton, and Kim, 2007). Indicative of the central issues of learning: how behaviour (or possibly neural activity) changes in response to feedback (Daw, 2011a).

Data will often consist of a series of experimental choices and outcomes. In theory any arbitrary relationship can be used to describe the effect of outcomes on later choices, the task of the modeller is to design a parsimonious model that plausibly represents neurological activity - attempting to capture some latent choice generating process (Sutton and Barto, 2018). This equates to modelling the expected state action values $Q(a, s)$. It's worth noting that the approach allows one to quantify neuropsychological parameters that would,

in a traditional model, be purely subjective - such as state value estimation, expectations, exploratory tendencies, risk aversion etc (Daw, 2011a).

2.7.1 Theoretical vs empirical parameters

These models are often complex systems, however, if we consider the model parameterisation one is able to search for a theoretically plausible, parsimonious formulation. A combination of frequentist and Bayesian, though largely Bayesian, inference techniques are drawn upon to fit behavioural data (Daw, 2011a).

A model is essentially a quantitative hypothesis about how the brain approaches a problem. Researchers regularly fit, compare and test an array of models with various levels of complexity and different covariates to arrive at a model that best articulates the underlying cognitive data generating process. Whilst models contain numerous free parameters, one salient distinction is the difference between:

1. ***Neuropsychological parameters***: inferred by the data, these parameters (typically the learning rate α and exploratory dynamics β) describe the abstract cognitive process.
2. ***Covariates***: as in standard statistical models, covariates offer the explanation of further variation in the data.

Both parameter types are estimated from the data but are distinct in that *neuropsychological* parameters are purely a choice of the statistician designing the model - inferred only indirectly by observed behaviour - whilst *covariates* are collected data that should be tested for explanatory value before inclusion - to adhere to the principle of parsimony. The principle, also known as Occam's razor, states that the simplest solution or relationship between things is often the correct explanation.

2.7.2 Learning vs observation models

A further imperative distinction is the separation of computational theory into the:

1. ***Learning model***: describing the dynamics of the model's internal variables - such as reward prediction error (RPE) (Daw, 2011a).
2. ***Observation model***: describes the relationship between the internal, learning, model and the observed data - for example, how RPE drives choices or how prediction errors produce a neurological spike.

The observed model regresses the internal variables onto the observed data (Daw, 2011a). The learning model is typically deterministic, the observation model, however, captures the stochasticity in the actual data, incorporating noise and assigning probabilities to the observations.

2.7.3 Parameter estimation

Suppose some model m, parameterised by a vector of free parameters θ_M (Gelman et al., 2004). The model is the composite of the learning and observation models that describe the learning (often RW) state value update and the link between the state values and choice behaviour respectively. The model M describes a probability distribution - or *likelihood* function $P(D|M, \theta_M)$ over possible data sets D. Bayes' rule allows us to derive the probability of the parameters having observed a data set D:

$$P(\theta_M|D, M) \propto P(D|M, \theta_M)P(\theta_M|M). \tag{2.14}$$

Formally, the *posterior* probability distribution over the free parameters, given the data, is proportional to the *likelihood* of the data, given the model and free parameters; and the *prior* probability distribution of the parameters (Gelman et al., 2004).

Bayes' theorem allows one to begin with a theory of some data generating process - that is a set of parameters that noisily produce data - and invert it into a problem by which data noisily reveal the parameters that generated it (Daw, 2011a). If we negate the influence of the prior, treating it as flat and uninformative, the most probable parameter estimates are those obtained by maximising the likelihood function $P(D|M, \theta_M)$ (that is the *maximum likelihood estimate* (Daw, 2011a). Let us denote this vector of estimates $\hat{\theta_M}$. Classical statistics is concerned with deriving these point estimates, often the maximum likelihood estimate, whilst this Bayesian paradigm is instead concerned with learning entire probability distributions over the parameter estimates (Gelman et al., 2004).

2.7.4 Maximum likelihood estimation for RL models

Consider a simple game in which an agent is tasked, at each discrete trial t, with making a choice c_t between two machines left L and right R. The agent stochastically receives a reward r_t after each action, where $r_t \in \{0, 1\}$. If we relate this dichotomous choice to Q-learning, the agent assigns an expected value to each machine $Q_t(L)$ and $Q_t(R)$ (where there is only a single state so this requires no index) (Daw, 2011a). These values are initialised neutrally, say at 0, and then updated on each trial, which forms the *learning model*:

$$Q_{t+1}(c_t) = Q_t(c_t) + \alpha\delta_t. \tag{2.15}$$

where $0 \leq \alpha \leq 1$ is a free *learning rate* parameter and $\delta_t = r_t - Q_t(c_t)$ is the RPE (Sutton and Barto, 2018). We then need to assume an *observation model* - to related the latent learning process to the observed behaviour. It is natural to assume that choices are made probabilistically according to a *softmax* distribution (Daw, 2011a):

$$P(c_t = L|Q_t(L), Q_t(R)) = \frac{\exp\{\beta Q_t(L)\}}{\sum_{i=R,L}\exp\{\beta Q_t(i)\}}. \tag{2.16}$$

where β is a free parameter called the *inverse temperature* that loosely describes the agents willingness to *explore* vs *exploit* knowledge - the famous trade-off in search algorithms. β gives a weight to each choice value, such that agents with equivalent state-value estimates $Q_t(L)$ and $Q_t(R)$ but vastly different exploratory coefficients β may make greatly different choices, as this weighting describes how the agent samples actions (Sutton and Barto, 2018). It is worth noting this simple observation model is equivalent to a *logistic regression link function* where c_t is the response variable; $Q_t(L) - Q_t(R)$ is the independent variable and β is the regression weight coefficient (Daw, 2011a). Note that a link function is a map from a non-linear relationship to a linear one, used in statistical analysis to exploit linear model (Hastie, 2001).

It can also be shown that the above model is a special case of a Kalman filter, a Bayesian smoothing technique that utilises the sample learning model but allows for a dynamic learning rate α (Daw, 2011a).

Although referred to as biological parameters throughout, which naturally fits both the literature and intuition, it is important to note that α and β are abstract conceptual cognitive processes and do not map to biological processes directly; but are widely accepted under the predictive processing paradigm to understanding learning (Barcelo, 2020), (Gershman, 2016), (Joue, Chakroun, Bayer, Gläscher, Zhang, Fuss, Hennies, and Sommer, 2021). Rather than well understood neurological processes these parameters are loose abstract metaphors, though closely related to striatal dopamine signalling (Joue et al., 2021).

Likelihood function

Given the model described above, the data set D consists of an entire sequence of choices $c_{1...T}$ and the associated rewards $r_{1...T}$ - note that this describes the actions of a single person. The likelihood function is the probability of the whole observed data set D, computed as the product of their probabilities from equations 2.16 (Daw, 2011a):

$$\prod_t P(c_t = L|Q_t(L), Q_t(R)). \tag{2.17}$$

where Q_t estimates are determined by equation 2.15 given the observed rewards and choices. Equations 2.15 and 2.17 constitute the full likelihood, we can then estimate the free parameters $\theta_M = \langle \alpha, \beta \rangle$ by maximum likelihood (Daw, 2011a).

Confidence intervals

Performing statistical hypothesis testing naturally requires confidence intervals around the parameter estimate $\hat{\theta_M}$. Intuitively, the reliability of the parameter estimate should be assessed proportionally to how probable alternative parameter choices are, that is the steepness of the gradient of the surface of the likelihood function (Daw, 2011a). The second derivative of the likelihood function with respect to the parameters - the *Hessian* - quantifies the steepness of slope of slope in the likelihood surface. The Hessian is a square matrix with a row and column for each parameter, with larger values indicative of a steeper slope. If H is the Hessian of the negative log likelihood function at the maximum likelihood point $\hat{\theta}_M$, then it is inverse H^{-1} is the standard estimate for the covariance of the parameter estimates. The variance of each parameter is then the diagonal of H^{-1}, thus the square root provides the respective standard error. $\hat{\theta}_M + / - 1.96$ *standard errors* is used to compute the 95% confidence intervals around parameter estimates.

Covariance between parameters

The non-diagonal elements of the inverse Hessian H^{-1} provide the covariance between parameter estimates, where larger values are symptomatic of multicollinearity (a condition where multiple covariates are highly correlated) and therefore the model may be unable to reach a unique optimum (as covariate effects are confounded) (Daw, 2011a). Additional complexity arises when fitting Q-learning models because the reward r_t is multiplied by both α (when updating Q_t) and β (when computing the choice probability) before affecting the action a_t - making it very difficult to discern the effects of each parameter. Empirically α and β estimates tend to be negatively correlated - as they are inversely coupled. The parameters offset the effects of one another and therefore the same likelihood can be attained with different α, β combinations.

2.7.5 Pragmatic implications of model fitting

Specified in this way, choice probabilities are equivalent to a General Linear Model (GLM) - a statistical model that links non-linear relationships to a linear form for simpler computation (Hastie, 2001). GLMs are usually fit with some optimisation procedure that efficiently searches the parameter space to maximise the likelihood. Standard open source maximum likelihood optimisers, however, cannot be used because the Q_t values enter the model as a function of free parameters and thus are stochastic and do no enter the likelihood linearly. Is non-linearity cannot be optimised with a linear optimiser requiring a different approach. As a result parameters cannot be estimated by a general linear model (GLM) (Daw, 2011a).

Computing the likelihood given a parameter vector θ^i_M: given a data set - a sequence of *actions* a_t and *rewards* r_t - and a vector of parameters θ^i_M is it straightforward to iteratively cycle through the data, update Q_t and compute $P(c_t|Q_t, \theta^i_M)$; the product of which produces the likelihood function. In reality, however, it is exceedingly likely that some choice probabilities $P(c_t|Q_t, \theta^i_M)$ are so small that they exceed the floating point value of the computer performing the task (making it impossible to distinguish between different nearby likelihood estimates). It is thus favourable to instead compute the numerically stable *log likelihood* $\prod_t P(c_t|Q_t, \theta^i_M) = \sum_t log\left(P(c_t|Q_t, \theta^i_M)\right)$ - a monotonic transformation with an equivalent maximum/minimum value. In addition, the likelihood surface is invariant to any addition or subtraction of a constant. Thus under/overflow issue can further be mitigated by normalising the Q_t values, subtracting the mean from each value before computing the likelihood (Daw, 2011a).

Searching the parameter space θ_M: One naive approach would be to discretise the parameter space of possible parameters that constitute the θ vector and simply enumerate over the possible parameters, computing the log likelihood and selecting the parameter configuration that maximises the log likelihood. This is impractical for a multitude reasons: (1) as the number of free parameters grows, it becomes increasingly difficult - and often intractable - to compute all likelihood functions; (2) the granularity and boundaries of the search are predefined which may erroneously exclude certain regions or granularity of the search space, leading to poor results or at best inefficient search (Daw, 2011a). In the case of non-linear models the tight coupling between variables further exaggerates these shortcomings.

Nonlinear optimization: A prudent alternative to grid search is to utilise a nonlinear function minimiser (thus requiring the negative log likelihood) - readily available in many open source software packages. Not only do these packages intelligently search the parameter space - efficiently sampling based on variations of hill climbing strategies - but also search continuously (without discretising the space) which has the effect of increasing or decreasing granularity when required in order to better locate an optimum (Gelman et al., 2004). It is important to remember that the search can often find a local minimum, thus it is prudent to stochastically or systematically initialise the parameter search many times and simply use the best run (Daw, 2011a). It also warrants noting that the Hessian or gradient vectors are often computed routinely by these non-linear optimisers, but in the event that they are omitted one can apply the chain rule to compute the matrices (for the purpose of confidence intervals) after searching the parameter space.

Bounded search: These nonlinear search algorithms often allow the statistician to impose boundaries that, in theory, limit the space for efficiency. In the case of biologically plausible models, it is tempting to limit the search space to those that the theory allows (those that have semantic interpretations), in our case (Daw, 2011a):

- $0 \leq \alpha \leq 1$: learning rate. It is outside of the scope of this theoretical learning model to allow for large α values. Furthermore, a large learning rate can lead to unstable estimates that grow rapidly and diverge.
- $0 \leq \beta \leq l$: where negative values have no interpretation and large values l will likely lead to arithmetic overflow.

Boundaries should, however, be used sparingly and with caution for a number of reasons: (1) in the case of a highly non-linear system the parameters are intertwined, thus constraints on one will effect the other; (2) an optimum is found outside of the theoretically plausible range may be indicative of a poor fit, noisy data or possibly the data revealing some truth outside of the current scientific consensus; and most saliently (3) most confidence interval estimates as well as model comparison techniques rely heavily on the inverse Hessian to compute the gradient of the likelihood surface - imposing boundaries will severely truncate or limit the likelihood surface and thus may impede the usability of this assessment criterion (Gelman et al., 2004).

Bayesian regularization: It is of course, possible that the MLE reaches poor or uninterpretable estimates due to noise in the data, one flexible paradigm to regularize the parameters is to specify the model as a Bayesian system and utilise a prior $P(\theta_M|M)$ to constrain the possible parameter estimates (Gelman et al., 2004). This is, in fact, a generalisation as specific boundaries can be show to be equivalent to some Bayesian prior - for example a hard $0 \leq \alpha \leq 1$ boundary is equivalent to imposing a uniform prior over the same domain. If the Bayesian approach is adopted, one simply substitutes the MLE objective function (log likelihood) for the posterior from equation 2.14. Parameter estimates for a given participant may also be regularized by the broader sample of participants or groups in which they operate - a pooling variance technique known as Hierarchical mixture models, that is the focus of the following section.

2.7.6 Hierarchical models

The binary choice Q-learning model discussed thus far models the choice behaviour of an individual participant. How then do we extend the model to account for multiple participants? One naive solution would be to average behaviour over all subjects. Although intuitive, this approach negates the importance of individuals' differences, failing to address most interesting research questions. This approach treats all parameters as *fixed effects* that do not allow variation within subjects (Daw, 2011a) - illustrated in figure 2.10. Instead, we aim to capture some variation *between* subjects (Gelman et al., 2004). Distinguishing between *within*-and-*between*-subject variability is of utmost relevance in answering interesting questions; further, the failure to do so can result in overstated result significance.

Independent models: Another common approach would be to fit individual models to each subject and then use some summary statistics (such as average performance or parameter estimates) per model to test whether or not significant differences exist between subjects (or possibly averages over groups) by utilising ANOVA, or one-or-two sided t-tests (Daw, 2011a). Treating each parameter estimate as a random variable equates to *random effects*. Whilst largely used and justifiable for many research questions, this approach fails to adequately capture the possible dependency between participants in a population, ignoring within-subject error bars (failing to smooth for irregular behaviour, noise and/or anomalies).

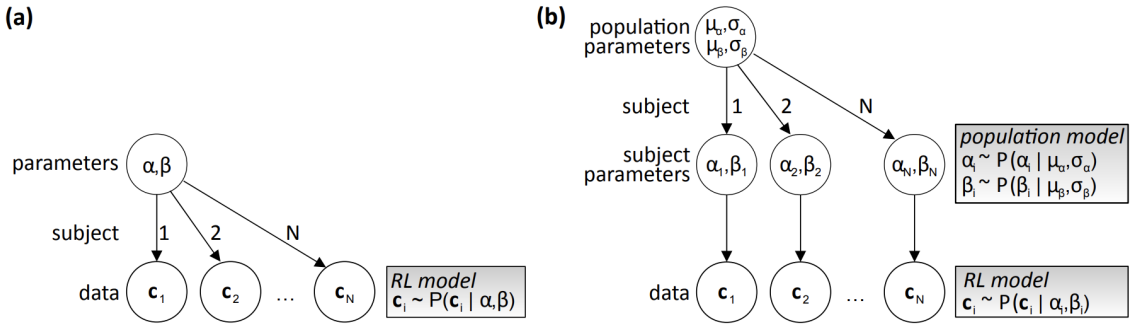


FIGURE 2.10: Capturing the inherent hierarchical structure of the data by imposing fixed vs random effects. (a) *Fixed effects*: parameter estimates are shared across subjects. (b) *Random effects* each subject's parameter estimates are drawn from a common population distribution - that becomes the regularising prior (Daw, 2011a)

.

Hierarchical structure over the population distribution: It is possible to instead explicitly model how individual parameters vary across the common distribution by imposing some distributional assumptions about the data generating process (Hastie, 2001).

We refers to a **population distribution** as a pooled common distribution that captures the aggregated behaviour of the sample. This terminology is adapted as it is frequently used in mixed/hierarchical Bayesian models, but should not be conflated with the classical idea of extrapolating one's (sample based) findings to a population at large.

We can assume that, just as an individual is sampled from the population, an individual's parameters $\theta : \{\alpha, \beta\}$ are drawn from some population distribution over possible parameters (Hastie, 2001). For example, we may assume a given subjects parameters $\theta : \{\alpha, \beta\}$ are sampled from distinct Gaussian priors with some mean values μ_α, μ_β and variance values $\sigma_\alpha, \sigma_\beta$ - denoted as $P(\alpha \mid \mu_\alpha, \sigma_\alpha)$ and $P(\beta \mid \mu_\beta, \sigma_\beta)$ (Daw, 2011a). Furthermore, parameter boundaries such as the previously discussed $0 \leq \alpha \leq 1$ can be imposed by these probability distributions by limiting the supported range - in this case utilising a β distribution or normal distribution transformed through a logistic function. These population distributions, of course, form regularising priors over the space of possible parameters.

The resulting two-level Hierarchical model now assumes a data generating process whereby an individual's parameters α, β are sampled from a population and thereafter used to generate the data c_t by interacting with the environment (receiving stimuli and rewards r_t) (Daw, 2011a). The parameters of interest are usually the population level parameters $\mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta$ - as we are pooling variance to better estimate the pooled behaviour - allowing us to answer questions about the difference between population groups.

The probability of the observed choice behaviour of participant $i \in \{1 \dots N\}$ $\mathbf{c_i}$ (where $\mathbf{c_i}$ is a vector of choices over time for participant i) is then the probability given to them by the RL model $P(\mathbf{c_i}|\alpha_i, \beta_i)$ averaged over all possible hyper-parameter settings of the individual subject's parameters according to their population distribution (Daw, 2011a):

$$P(\mathbf{c_i}|\mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta) = \int P(\alpha_i|\mu_\alpha, \sigma_\alpha)P(\beta_i|\mu_\beta, \sigma_\beta)P(\mathbf{c_i}|\alpha_i, \beta_i)d\alpha_i d\beta_i. \tag{2.18}$$

Intuitively, equation 2.18 emphasises that from the perspective of performing inference on the population parameters, an observed individual α_i or β_i are auxiliary variables to be

averaged out (Daw, 2011a). The product over these individual distributions gives us the probability of the entire observed data set (all N participants) (Gelman et al., 2004):

$$P(\mathbf{c_1} \dots \mathbf{c_N} | \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta) = \prod_i P(\mathbf{c_i} | \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta). \tag{2.19}$$

Bayes' rule is used to recover the the population parameters given the entire dataset:

$$P(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta | \mathbf{c_1} \dots \mathbf{c_N}) \propto P(\mathbf{c_1} \dots \mathbf{c_N} | \mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta) P(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta). \tag{2.20}$$

Estimating population parameters in a hierarchical model: Utilising equation 2.20 we are now - in theory - able to estimate population parameters $\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta$ using maximum likelihood (MLE) or maximum a priori (MAP) and estimate confidence intervals with the inverse Hessian - allowing one to compare between-group population estimates (Huys, 2013). This pseudo algorithm is implementable as follows:

1. Write a function that returns the log probability of choices over a population of participants given the population parameters $\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta$: equation 2.19.
2. Which in turn requires computing equation 2.18 - providing the average over parameter values for a given subject.
3. This nonlinear system may be optimised through some standard off-the-shelf nonlinear optimisation algorithm.

As in almost all non trivial Bayesian applications, however, equations 2.18 is ***intractable*** and requires some approximate method (Gelman et al., 2004). A common remedy is to use sample techniques to estimate choice probabilities by the posterior mean. One can draw k samples from distributions $P(\alpha_i|\mu_\alpha\sigma_\alpha)$ and $P(\beta_i|\mu_\beta\sigma_\beta)$; and use these samples to approximate the integral by averaging $P(\mathbf{c_i}|\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta) \approx \frac{1}{k}\sum_{j=1}^{k} P(\mathbf{c_i}|\alpha_j, \beta_j)$. One pragmatic caveat is that off-the-shelf non-linear optimisers often require smooth parameter updating, which can become problematic with random sampling, thus it is recommended that that the same random seed is used for each subject iteration (Daw, 2011a).

Estimating population parameters via summary statistics: An alternative approach would be to simply estimate all the individual parameter $\theta_M^i : \{\alpha_i, \beta_i\}$ and thereafter make some distribution assumption about the data generating process - say assuming that individual parameters $P(\alpha_i|\mu_{alpha}, \sigma_\alpha)$ are drawn from Gaussian distributions $\alpha_i \sim \mathcal{N}(\mu_{alpha}, \sigma_\alpha)$ (and similarly for β_i) (Gelman et al., 2004). Thereafter it is straightforward to treat the individual parameter estimates as samples from the population distribution and fit Gaussian distributions to estimate the population parameters $\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta$ (Daw, 2011a). Importantly, making these distributional assumptions allows one to use standard t-test and confidence intervals to test the significance of both between and within group variation - avoiding the need of computing the Hessian (second derivative of the likelihood function) for confidence intervals.

It is possible that great noise in the data generating process can inflate the estimated population variance - as illustrated in figure 2.11 which warrants concern as this will greatly impact the utility of hypothesis testing (Daw, 2011a).

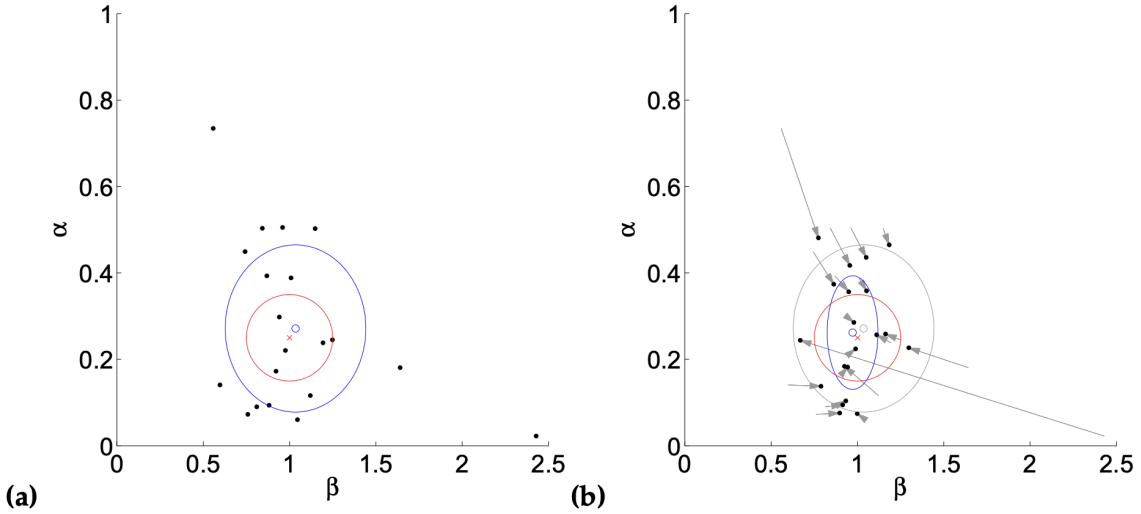


FIGURE 2.11: Simulated experiments detailing the benefits of a well specified prior distribution, where data are sampled from a bi-variate Gaussian and the true mean and standard deviation are depicted as the red dot and red circles respectively (Daw, 2011b). **(a)** Utilises the individual/summary statistic approach whereby individual parameters are fit to each subject and thereafter bi-variate Gaussian is fit to the population parameters by interpreting the individual subjects as samples; estimates are shown in blue (Daw, 2011a). Whilst the mean is well estimated and unbiased, it appears to exhibit inflated variance. **(b)** The individual estimates here were fit using MAP whereby the gray ellipse serves as the prior distribution, forcing the sample estimates towards the true mean, compressing the variance. Imposing the prior is equivalent to fitting the hierarchical model whereby the prior regulates population assumptions.

Estimating individual parameter estimates in a hierarchical model: Although in the above description the focus is on estimating population parameter values, and individual parameter estimates are treated as auxiliary variables, some research questions requires individual estimates. Assuming population level parameters, it is possible to recover individual estimates (Daw, 2011a):

$$P(\alpha_i, \beta_i | \mathbf{c_i}, \mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta) \propto P(\mathbf{c_i} | \alpha_i, \beta_i) P(\alpha_i, \beta_i | \mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta) \tag{2.21}$$

$P(\mathbf{c_i}|\alpha_i, \beta_i)$ is the likelihood of the individuals choice behaviour, and $P(\alpha_i, \beta_i|\mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta)$ serves as a prior, regularising the estimate towards characteristics of the population (Gelman et al., 2004). The estimates are regularised towards that of the the group mean, thus the Bayesian equation details the balance between group level importance and individuality.

An illustration of hierarchical Bayesian RL: It is important to thoroughly understand how these hierarchical models are constructed in the context of neuropsychology and physiology. Van Slooten et al., 2017 were able to map the pupillometric responses to an underlying cognitive learning process by fitting a hierarchical learning model (Van Slooten et al., 2017). Whilst specific to pupillometry, this speaks to a broader class of linking non-invasive observables to some generative learning model. A theoretically sound experiment, pupils have been shown to dilate during uncertainty and after the occurrence of unexpected events (Van Slooten et al., 2017). It has been observed to map directly to surprise during random positive reward tasks (gambling). A replication of this effect in the cognitive

learning model was achieved, showing concordance with the existing literature, providing confidence in the RL approach. Furthermore, in describing biological relevance, the authors were able to map:

- One's tendency to explore vs exploit information.
- One's ability to learn capturing the rate at which value beliefs are updated.

Particularly focused on *value-based decision making* and subsequent **decision evaluation**, the learning parameters extracted from the hierarchical model were used as features in a subsequent ridge regression model to map one's exploratory nature to pupillometric fluctuations (allowing for inverse reasoning). A graphical representation of the hierarchical RL model used by Van Slooten et al., 2017 is depicted in figure 2.12.

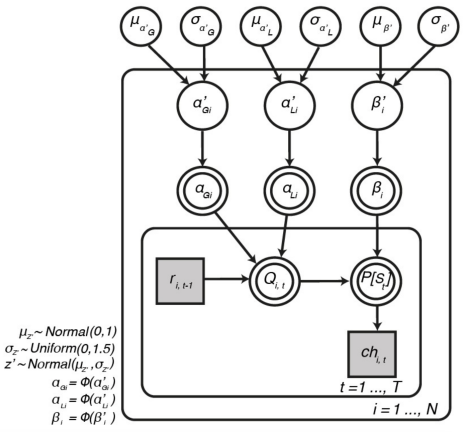


FIGURE 2.12: Graphical modeling: Bayesian Hierarchical Reinforcement Learning utilised in the pupillometric study by Van Slooten et al., 2017.

After extracting learning parameter β and fitting a ridge regression to predict pupil-size, the paper reported:

- Pupil dilation predicts an individual's tendency to exploit (vs explore) high value options. That is, the frequency at which they sample the action space (Van Slooten, Jahfari, Knapen, and Theeuwes, 2017).
- After feedback, biphasic pupil responses were observed - the amplitude of which correlated with participants' learning rates α.
- Feedback related dilation was mapped to value uncertainty.
- Later pupil constriction scaled with reward prediction errors (RPE).

These findings show that pupil size fluctuations can provide detailed information about the computations underlying valued-based decisions and the subsequent updating of value beliefs (Van Slooten et al., 2017).

More broadly, the paper serves as a clear illustration of how statistical inference can be used to approximate unknown theoretical quantities. More specifically, using RL we are able to reason about a latent cognitive process by designing a model that abstractly represents the underlying biological data generating process. Put another way, by simulating the learning process the authors could investigate how pupil size depends on trial-to-trial fluctuations in underlying computational variables such as: *value beliefs, uncertainty and RPE (Reward Prediction Error)* (Van Slooten et al., 2017).

2.8 Incorporating additional data

Suppose, as in most experiments, we are able to capture some additional data about the subjects such as demographic, behavioural or psychological attributes. This data may either:

1. Offer substantial explanatory power reducing variation in parameter estimates, or;
2. Allow additional hypothesis testing to measure the congruence with neuropsychological theory, asking whether or not observed covariates have an empirical relationship with choice behaviour $H_0 : \beta_j = 0$.

2.8.1 Explanatory power in the observation model

How then do we append these covariates to our model architecture? One would simply update the observational model to capture these parameter, defining the relationship between the data and the choice (response) (Daw, 2011a). The idea is that a common learning model may be observed by particular (observable) measurements given insight into the true underlying data generating process. One pragmatic illustration of this is to simply add Gaussian noise to the observed model, to better account for random fluctuations, however the idea is to add covariates to improve the model. Another common addition is reaction times (RT), for example state s at time t (s_t) can be a function of the RPE update δ_t, task reaction time RT and sample from a Gaussian distribution to capture random noise $\epsilon \sim \mathcal{N}(0, \sigma)$:

$$s_t = \beta_0 + \beta_1 \delta_T + \beta_2 RT + \epsilon \tag{2.22}$$

If the primary focus is to test the relevance of additional covariates on learning or choice behaviour, one may wish to utilise a single global α (and other learning model parameters) to yield more robust, stable results (Daw, 2011a) - which may be achieved by fitting a global fixed effects model to estimate some aggregate baseline learning rate parameter. This may greatly deflate variability in the individual parameter estimates (providing more accurate approximations) leading to more robust statistical diagnostics.

2.8.2 Explanatory power of the learning model

Just as the above example demonstrates how one might update the *observation model*, one can just as readily use an alternative parameterisation of the learning model to incorporate additional information or covariates.

It is salient to consider where covariates ought to be amended, to achieve the optimal theoretical relevance; for instance, adding an IQ parameter to the learning model may be appropriate as it might influence the underlying learning process (Daw, 2011a). As an illustration, one might assume that individual α_i values are sampled from a Gaussian where IQ effects the generation of the learning rate:

$$P(\alpha_i | \mu_\alpha, \sigma_\alpha, K_{IQ}, IQ_i) \sim \mathcal{N}(\mu_\alpha + K_{IQ}, \sigma_\alpha)$$

Different models may offer better or worse explanatory power, and therefore must be tested in accordance to how well they fit the empirical data.

2.8.3 Alternative population models

The standard implementation described here assumes parameters are sampled from unimodal Gaussian distributions $P(\alpha|\mu_\alpha, \sigma_\alpha) \sim N(\mu_\alpha, \sigma_\alpha)$. However, this too permits many variants and extensions (Gelman et al., 2004). One might imagine a situation where subjects cluster into different groupings, this can be modelled by using a multimodal mixture model of the parameters (Daw, 2011a). One would specify a bi-modal mixture population model as follows:

$$\pi_1 N(\mu_{\alpha 1}, \sigma_{\alpha 1}) N(\mu_{\beta 1}, \sigma_{\beta 1}) + (1 - \pi_1) N(\mu_{\alpha 2}, \sigma_{\alpha 2}) N(\mu_{\beta 2}, \sigma_{\beta 2}).$$

In this example μ's dictate the modal values and π the predominance of cluster one.

2.8.4 Parametric nonstationarity

The models described thus far assume stationary (constant) model parameters throughout the experiment, which may be an unrealistic assumption in many experiments. A high learning rate - whereby subject's update value estimates aggressively in an exploitatory fashion - promotes rapid acquisition but subsequent instability as estimates are continuously aggressively updated (Daw, 2011a). Similarly, a learning rate that adequately captures a subjects asymptotic insensitivity to feedback would predict unrealistically slow value estimation. It is natural to assume that rapid acquisition followed by asymptotic stability is desirable, ideally this would be fully captured by the converges of expected state values Q_t to approximations of the true (unknown) state values. This may be represented by an increasing softmax temperature, or decaying learning rate - with the learning rate decay capturing some value inertia. It should be noted that this, again, alludes to the difficulty when modelling highly correlated covariates, an additional level of complexity whereby (in this case) the relationship between the entangled parameters changes temporally (Gelman et al., 2004).

One approach to handling such nonstationarity is to model the dynamics of the non-stationary parameters by adding free parameters - expressing the system as a function of more elementary parameters (Daw, 2011a). Effectively adding granularity to the data generating process, allowing for variability in the learning rate. These, however, often add superfluous complexity and further confound the parameter estimation.

If instead, we do not wish to specify a meta-level data generating process, another approach would be to allow for multiple learning/exploratory (α/β) parameters (Daw, 2011a). On the extreme, if every trial had independent parameters the specification would saturate the model (n parameters exceeding the number of samples - prohibiting unique parameter estimation) (Gelman et al., 2004). One option would be to assume the β_t parameters are variable but change linearly and deterministically, for instance:

$$\beta_t = \beta_{start} + \frac{t}{T}(\beta_{end} - \beta_{start}).$$

Which would require learning a single additional parameter ($\beta_{end}, \beta_{start}$ as apposed to a single β) (and is a special case of the aforementioned strategy) (Sutton and Barto, 2018). One might wish to allow for some stochasticity, in which case the β_t parameters may be represented as some random process, for example: assuming β_t is captured by a Gaussian random walk (Daw, 2011a):

$$\beta_{t=1} = \beta_{start};\ \beta_{t+1} = \beta_t + \epsilon_t;\ \epsilon \sim \mathcal{N}(0, \sigma_\epsilon)$$

Also containing 2 free parameters $\beta_{start}, \sigma_\epsilon$. One caveat to adding this stochasticity is the added complexity to model fitting: because the transition dynamics are probabilistic, behaviour in the expectation is estimated by averaging over many random trajectories to converge to a suitable variance estimate - analogous to the aforementioned technique used over different subject-specific parameters (Daw, 2011a).

Finally, a more deliberate approach is often taken, whereby the experimental design is chosen in an attempt to minimize nonstationarity. For example, in a multi-armed bandit problem, one may specify some stochastic update rule over reward probabilities - requiring the subjects to continue to learn the reward dynamics (Daw, 2011a). The approach taken in our experimental design - described in the methodology - employs this strategy by randomly changing the underlying learning rule, requiring frequent attention. Ideally, estimated learning rates should be asymptotically stable.

2.9 Biological and neurological complexity

To more realistically describe the cognitive process governing some learning task, the baseline Rescorla–Wagner model can be extended to account for any abstract biological concepts (Gershman, 2016).

The first and most common extension, in light of strong theoretical evidence, is to allow for different learning rates α conditional on whether or not the corresponding reward is positive α_g or negative α_l.

Other common extensions are to account for inertia, stickiness or bias (Daw, 2011a), that is a candidates tendency to either overweight their current estimates or bias certain states. This can be encoded by the addition of a single - usually temporally independent, state dependent - term to the state-update equation (Gershman, 2016).

Any variant of complexity or biologically theoretic encoding can be amended, the model comparison techniques are sought to penalise complexity and thus expose unfounded model specifications.

2.9.1 Dynamic (meta) learning

As an example of the the above fitting produced, Humann, Fischer, and Ullsperger, 2020 were able to model the relationship between working memory and dynamic learning. The authors adapted the Rescorla-Wagner model to allow for meta-learning parameters η and k: that dynamically govern the learning process.

Let reward prediction error for stimulus X_t be denoted $\delta_t = R_t - V_t(X_t)$ where $V_t(X_t)$ is the value estimate of stimulus X at time t. The authors then allowed for a dynamically updating learning rate $\alpha_{t+1}(X_t) = \eta|\delta_t| + (1-\eta)a_t(X_t)$ (Humann, Fischer, and Ullsperger, 2020).

Additional intercept flexibility (allowing for variability by stimuli type in the estimates) in the model was specified with parameter k, such that the state-value approximations were updated according to $V_{t+1}(X_t) + a_t(X_t)k\delta_t$ (Humann, Fischer, and Ullsperger, 2020).

A logistic link function was used to map the value estimates to (binary) choice probabilities. The best model was chosen using information theoretic criterion, iBIC. Not only did the

authors report fitting a more generalisable model than the Rescorla-Wagner baseline, they also showed the relationship these meta-learning parameters η, k have with working memory (task performance).

It was concluded that working memory capacity (WMC) is relied upon when performing instrumental learning tasks. Finally, the paper illustrates one approach to linking learning models to EEG (prominent physiological data): mapping the relationship between working memory, observable RPE and electrical activity in the brain measured by EEG channels (Humann, Fischer, and Ullsperger, 2020).

2.9.2 Neuroscientific Bayesian interpretations

It is often natural to pose the RL problems in a Bayesian fashion (Daw, 2011a). Gershman, 2016 detail the benefits of well specified priors to not only speed convergence and set theoretical bounds but beyond this to reach more reliable parameter estimates.

In D'Alessandro et al., 2020, a Bayesian brain model is used to pose dynamic learning as a Bayesian update equation, supporting the Bayesian brain hypothesis (Knill and Pouget, 2004). The authors model the WCST as a sequential process where subjects recursively compute value estimates of state s_t by Bayesian belief updating:

$$p(s_t|x_{0:t}) = \frac{p(x_t|s_t, x_{0:t-1})p(s_t|x_{0:t-1})}{p(x_t|x_{0:t-1})}.$$

where $x_t = (a_t, f_t)$ is the observation vector consisting of action ($a_t \in \{1, 2, 3, 4\}$) and feedback $f_t \in \{0, 1\}$ pairs.

The reward-action pairings are then weighted to inform the likelihood:

$$p(x_t|s_t, x_{0:t-1}) = \frac{f_t p(a_t|s_t = i) + (1 - f_t)(1 - p(a_t|s_t = i))}{f_t \sum_j p(a_t|s_t = j) + (1 - f_t) \sum_j (1 - p(a_t|s_t = j))}.$$

Leveraging the Markov property, the likelihood assumes the current observation to be independent of the previous observations without loss of generality (D'Alessandro et al., 2020):

$$p(x_t|s_t, x_{0:t-1}) = p(x_t|s_t).$$

Prior beliefs at time t are computed from the posterior of the previous trial $p(s_{t-1}|x_{0:t-1})$ and the person's belief about transition dynamics between hidden states $p(s_t|s_{t-1})$ (D'Alessandro et al., 2020). Intuitively, the prior can be thought of as the predictive probability over the hidden states, computed according to the Chapman-Kolmogorov equation:

$$p(s_{t+1} = k|x_{0:t}) = \sum_{i=1}^{3} p(s_{t+1} = k, s_t = i, \Gamma(t))p(s_t = i, x_{0:t}).$$

where $\Gamma(t)$ represents a state-transition matrix computed as a square and asymmetric matrix related to the hidden state activation levels (D'Alessandro et al., 2020).

The paper uses this model to compute a number of information theoretic quantities to describe an individual's cognitive processes. *Bayesian surprise* can be measured by the

divergence between the current and previous state probability estimates, that is, the amount by which one changes their estimates in light of new information (D'Alessandro et al., 2020):

$$\begin{aligned}\mathcal{B}_t &= \mathbb{KL}\left[p(s_{t+1}|x_{0:t})||p(s_t|x0:t-1)\right] \\ &= \sum_{i=1}^{3}\left[p(s_{t+1}=i|x_{0:t})log\left(\frac{p(s_{t+1}=i|x_{0:t})}{p(s_t=i|x_{0:t-1})}\right)\right].\end{aligned}$$

were there are 3 states. The *Shannon surprise*, of a current observation contingent on the previous is denoted as the conditional information gained (given that the WCST has 3 states) (D'Alessandro et al., 2020):

$$\begin{aligned}\mathcal{I}_t &= -\log p(x_t|x_{0:t-1}) \\ &= -\log\sum_{i=1}^{3}\left[p(x_t|s_t=i)p(s_t=i|x_{0:t-1})\right].\end{aligned}$$

Entropy, accounting for the uncertainty in the agents' internal model, is computed over the predictive distribution:

$$\begin{aligned}\mathcal{H}_t &= \mathbb{E}\left[-log\ p(s_t|x_{0:t-1})\right] \\ &= -\sum_{i=1}^{3}p(s_t=i|x_{0:t-1})log\ p(s_t=i|x_{0:t-1}).\end{aligned}$$

The above Bayesian cognitive model was fitted to clinical card sorting (WCST) data, with the objective of evaluating the ability of the framework to account for dysfunctional cognitive dynamics of information processing in substance dependent individuals (SDI) when compared to healthy controls (D'Alessandro et al., 2020).

SDIs were shown to exhibit inefficient conceptualisation of the task and dysfunctional error-prone response strategies, when compared with healthy controls. This may be attributed defective error monitoring and behaviour modulation systems - functionally dependent on cingulate and frontal brain regions (D'Alessandro et al., 2020). The WCST should be relatively straightforward for healthy subjects. Their model captures the discrepancies in SDIs and their healthy counterparts.

2.10 Model comparisons

Thus far model fitting has been discussed, however how should one select a candidate model in a plethora of possible variants, nestings and extensions? Furthermore, to what extent does the data support different candidate models (Daw, 2011a)? Many empirical scientific endeavours are inherently an abstract model selection process - as in the case of many neuroscience studies - for the simple reason that the model defines the theory of interest. In our case, the optimal model corresponds to the best theory of the mechanics behind human cognition, presupposing the question: *"Can we specify a data generating process that loosely*

maps to the true underlying cognitive process?", thus the model is tested by how well our theory fits the data and how well we expect the model to generalise to out-of-sample data.

When applied to Reinforcement learning, the most salient dichotomy is the distinction between model-free learning (as performed in our standard Q learning model where candidates update state values directly); or model-based learning (whereby an agent evaluates actions indirectly by learning some more fundamental latent process and reasoning about said process (Daw, 2011a).

Further, in general more free parameters will improve the fit of a model: a consequence of the classical curse of dimensionality that has been plaguing statisticians since the dawn of empiricism (Hastie, 2001). Intuitively, added flexibility permits greater overfitting, saturating the model.

Similar techniques from the aforementioned model fitting section are utilised to discern the optimal fit; a corollary of the fact that similar reasoning is used to derive the solutions; that is we aim to discover the best fit (Daw, 2011a).

2.10.1 RL illustration

Policy and value models: When conducting model evaluation we simply need to, again, compute data likelihoods under a model, optimise parameters and estimate Hessians (Daw, 2011a). An architectural choice that needs to be made when designing reinforcement learning systems is the discrepancy between *policy and value models*. Formally known as the *representation* question: What is actually learned that guides behaviour? *Value based* models, such as Q learning, learn the value of actions; whilst *policy-based* algorithms learn the best course of actions directly to estimate the optimal choice strategy (Sutton and Barto, 2018). A simple instance of such a model, that updates the optimal policy directly, is illustrated here. This model tries to learn a sequence of actions that best fits the data. Note that this is independent of the estimated state values, but is rather only concerned with the sequence of actions. $\phi(c_t)$ denotes the action weighting at time t.

$$\phi_{t+1}(c_t) = \phi_t(c_t) + (r_t - \bar{r}). \tag{2.23}$$

Where $\bar{r}$ is a comparison constant, often taken as the mean overall reward and the model tracks a new free parameter π_t (Daw, 2011a).

As before, the choice behaviour is subsequently captured by an observational model:

$$P(c_t = L|\phi_t(L), \phi_t(R)) = \frac{exp\,(\beta\phi_t(L))}{exp\,(\beta\phi_t(R)) + exp\,(\beta\phi_t(L))}. \tag{2.24}$$

Equation 2.23 hypothesises a different model; where the previous model estimated expected average reward Q for each choice and makes decisions based on relative value differences; π offers general "knobs" that control the actions taken (Daw, 2011a). This alternative model - defined by a unique set of constraints on the relationship between the feedback and subsequent choices - offers an alternative hypothesis of the data generating process (Sutton and Barto, 2018).

It has been observed that choice values ϕ of better than average actions tends towards infinity, $\phi \to \infty$, resulting in a situation where the model exclusively selects the option that appears best in the initial sequence of actions failing to adjust and explore different choices later in the sequence (Daw, 2011a). Q-learning, however, tends towards the true unknown

expected reward value; allowing for less than complete preference for one option over others and thus implicitly capturing some quantity of uncertainty (Sutton and Barto, 2018).

Choice autocorrelation:

The policy based model has only a single parameter β in contrast to the two Q-learning model's two parameters $\theta : \{\alpha, \beta\}$ which adds complexity that may cause overfitting (Daw, 2011a). To illustrate the danger of saturating a model, consider the follow amendment to our standard Q-learning model:

$$P(c_t = L|Q_t(L), Q_t(R), L_{t-1}, R_{t-1}) = \frac{exp\left(\beta Q_t(L) + kL_{t-1}\right)}{exp\left(\beta Q_t(R) + kR_{t-1}\right) + exp\left(\beta Q_t(L) + kL_{t-1}\right)}. \tag{2.25}$$

Equation 2.25 describes a model with binary indicator variables L_{t-1} and R_{t-1} that take values $L_{t-1} \in \{0, 1\}$ according to whether the previous trial's $t-1$ response was L or R - capturing an inertia term (Daw, 2011a). The motivation for this model is choice autocorrelation: whereby candidates have a tendency to either persevere with existing preferences or switch readily. Positive k values promote sticking, whilst negative values support alternating.

2.10.2 Classical techniques

How do we assess how well some set of model parameters fits the data? Let M_i denote a vector of parameters of model i, In some sense maximising the likelihood function achieves this, finding the parameter estimates M_1 that maximise the likelihood of the observed data $P(D|M_1, \hat{\theta}_{M_1})$ Daw, 2011b. Although simple to compute, this approach inflates the measure of how well the model predicts the dataset because the same data is used to both fit and validate the model (Hastie, 2001).

Nested models: Returning to 2.25, it is straightforward to see that this model is an extension to the previous discussed model. Thus dropping the binary indicator variables L_{t-1} and R_{t-1} and associated free parameter k yields the original model. Known as *nested* models, the former model M_1 is a special case of the later M_2 where $k = 0$, thus all parameter specifications available to M_1 is available to M_2 and M_2 is by necessity at least as well fitted to the data as M_1 (Daw, 2011a). Even if the data is generated by M_1, it is likely that noise is sampling the observations exhibit some bias towards a non-null k value (Hastie, 2001). In general, more complex models will *overfit* the noise in the data generating processes. In the extreme case, the number of parameters equates or exceeds the number of data points, allowing for perfect (and completely non-general) interpolation of the data.

Cross-validation: One common approach to address this concern is fit a model to a dataset ("training" set) and thereafter use another dataset (the "holdout", "testing or "validation" set) to compute the likelihood of the testing set given the original (training) set parameters. If the model was greatly influenced by noise, it will fit the second data set poorly. That is, it will not predict the second dataset well. Conversely, if the model adequately captures the true latent data generating process, it will predict the second data set sufficiently (Hastie, 2001). Since this approach is primarily concerned with predicting the held out data set, it allows one to compare models with different numbers of parameters - the holdout data set likelihood score is not inflated by the number of parameters (in fact, may be hindered by fitting noise).

Whilst the prominent model section method of choice in many areas of neuroscience, it is not recommended that this approach is used in trial-by-trial analysis (Daw, 2011a). Given the temporal nature of the data, it is difficult to define a second, testing, dataset that is truly independent of the first (Hastie, 2001). Further, splitting the data temporally (training on earlier trials and testing on later trials) may lack the core assumption of being *identically distributed* - a consequence of non-stationary parameters in the data generating process (Daw, 2011a).

Likelihood ratio test:

Consider, again, the case when a single data set is used to fit the maximum likelihood estimate. Although the estimate is inflated - often fitting noise from the given sample - statistical theory allows us to quantify the probability of this likelihood inflation (Daw, 2011a). We need to discern whether adding additional parameters to the model truly improves the fit, or if the improvement is a result of fitting noise when granted the flexibility of superfluous parameters (Hastie, 2001). If, and only if, we are comparing nested models a likelihood ratio test can be used to address this question. Under the null hypothesis that the data is generated by a simpler model M_1, rejecting the null hypothesis (in light of a low p-value) is interpreted as rejecting the simpler model with confidence. The likelihood ratio test statistic is calculated by fitting a complex model M_2 and simpler nested model M_1 to the same data set and thereafter computing:

$$d = 2 \cdot \left[logP(D|M_2, \hat{\theta}_{M_2}) - logP(D|M_1, \hat{\theta}_{M_1}) \right].$$

Since M_2 nests M_1, $d \geq 0$ by necessity (Hastie, 2001). The probability of a difference d arising from M_1 follows a *chi-square* distribution with degrees of freedom n (where n is the number of additional parameters in M_2):

$$d \sim \chi^2(n).$$

Therefore, the probability of the test: *the probability of a distance d or larger arising due to chance* is $1 - \chi^2(d, n)$ (Daw, 2011a). Tested at some chosen level of significance, one can draw conclusions about the relevance of additional variables. While powerful, and frequently used in regression analysis, the likelihood ratio test is both limited to nested models and primarily a frequentist methodology - both issues that may be addressed by Bayesian methods (Hastie, 2001).

2.10.3 Theoretical Bayesian model comparison

Model evidence: When performing Bayesian inference, we are primarily concerned with computing the posterior distribution:

$$P(M|D) \propto P(D|M)P(M)$$

The probability of the data under the model $P(D|M)$ is known as model evidence (Daw, 2011a). It is salient to note that model evidence does not make any reference to any particular model parameters $\hat{\theta}_M$. It is for this reason that the score computed by a likelihood function $P(D|M, \hat{\theta}_M)$ is inflated by the number of free parameters: it takes as given parameters that fit the observed data (Hastie, 2001). Put succinctly, when asking how well a model predicts a dataset, it is a fallacy to retrospectively choose the parameters that would have

best fit the data, after observing the data (Daw, 2011a). Overstating the models predictive capacity. When making model comparisons using model evidence $P(D|M)$ - agnostic of optimal parameters - negates overfitting. This quantity is computed by the (weighted) average of all possible parameter configurations for a given model, *prior to examining the data* $P(\theta_M|M)$. Formally:

$$P(D|M) = \int P(D|M, \theta_M)P(\theta_M|M)\ d\theta_M.$$

Automatic Occam's razor: As a mathematical convenience, the posterior distribution favours simpler models. One might encode some regularising prior in $P(M)$, as if often done in Bayesian analysis (Gelman et al., 2004). External to this, inherent in it is formulation, the posterior computation has a preference towards simpler models by normalising the model evidence $P(D|M)$ (Daw, 2011a). $P(D|M)$ is a probability distribution, thus must sum to $1 \int P(D|M)dD = 1$ (Hastie, 2001). This means that more flexible models (with more free parameters) assign lower probabilities to each $P(D|M)$ value (as they must sum to 1) and similarly simpler models assign greater probabilities to each $P(D|M)$ - effectively imposing a penalty on complexity (Daw, 2011a).

Bayes factors: When comparing Bayesian models, the standard statistical quantity to quantify two models' relative fit is the ratio of their posterior probabilities (known as the Bayes factor):

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(D|M_1)P(M_1)}{P(D|M_2)P(M_2)}.$$

Conveniently, the Bayes rule denominator cancels out. The log of the Bayes factor is symmetric, positive and negative values favouring M_1 and M_2 respectively (Daw, 2011a). Although not identical to $p - values$, Bayes factors are often interpreted similarly. As a guiding heuristic: a Bayes factor of 20 (or log Bayes factor of ≈ 3) corresponds to a $20:1$ evidence in favour of M_1 which is analogous to a $p = 0.05$. A rich literature is available to convert Bayes factors to their familiar frequentist counterpart $p - values$ for ease of interpretation.

2.10.4 Practical Bayesian model comparison

Bayesian model comparisons circumvent the disadvantages of utilising purely likelihood driven techniques, mitigating the risk of drawing ill-founded conclusion on inflated estimates (Gelman et al., 2004). These techniques, however, are accompanied with their own set of complications, namely:

1. ***Integrating the model evidence:*** As aforementioned, the model evidence is almost always intractable - requiring approximate estimates.
2. ***Prior specification:*** The assumed prior distribution plays a pivotal role is many Bayesian methods, and although flat/uninformative priors can by used in the absence of a defensible parameters, these neutral decision, too, bare consequences (Gelman et al., 2004).

Priors: Prior distributions $P(\theta_M|M)$ act as a weighting kernel over the model evidence (Daw, 2011a). The prior, in the way, dictates the admissible range of the parameter space as well as the relative likelihood of parameter configurations before seeing the data (Gelman et al., 2004). Put another way, the prior regulates the parameter space by imposing

(soft) constraints over the flexibility of θ_M. Moreover, because we are in the realm of probability density functions (requiring that values are normalised such that they sum to 1) ignoring the prior (as is done in the aforementioned techniques) is often both incorrect and mathematically unstable. As shown below, BIC technique negates the need to specify a prior, however if one is able and willing to specify some admissible prior distribution more favourable results can be achieved (Daw, 2011a).

Sampling: As is often utilised in Bayesian statistics, the simplest method to approximate the model evidence is to average a sample (Daw, 2011a). This can be achieved by drawing candidate parameter estimates from the prior $\theta_M^i \sim P(\theta_M|M)$ and the compute the data likelihood $P(D|\theta_M, M)$ and averaging the results. This avoids traditional *optimisations*, and simple requires *evaluating* the likelihood at the sample points in the parameter space (Gelman et al., 2004). Although conceptually simple and easy to implement, this naive sampling technique is inadvisable for complex models: it is straightforward to see that as the number of free parameters grows, the curse of dimensionality ensures that the likelihood of sampling a suitable region of the parameter space diminishes exponentially (Daw, 2011a).

Laplace approximation: One powerful technique is to approximate the function being integrated with a Gaussian - for which the integral can be computed analytically (Gelman et al., 2004). In this instance, we can approximate the likelihood surface as a Gaussian centred around the maximum a prior estimates $\hat{\theta}_M$ - notably, this is the same approximation utilised to motivate the reliance of the inverse Hessian H^{-1} to approximate error bars (Daw, 2011a). The Leplacian approximation applied to the model evidence is derived as:

$$log\left(P(D|M)\right) \approx log\left(P(D|M,\hat{\theta}_M)\right) + log\left(P(\hat{\theta}_M|M)\right) + \frac{n}{2}log\left(2\pi\right) - \frac{1}{2}log\left|H\right|. \quad (2.26)$$

Where n is the number of free parameters and $|H|$ is the determinant of the Hessian - describing the covariance of the assumed Gaussian (Gelman et al., 2004). These quantities are straightforward to compute but do, however, require the specification of some prior over the parameter space (as computation is done with respect to the MAP estimate and not simply the MLE). Similarly, note that the Hessian is the Hessian of the posterior, and not merely the likelihood function.

The equation 2.26 is the log posterior

$$log\left(P(D|M)\right) \approx log\left(P(D|M,\hat{\theta}_M)\right) + log\left(P(\hat{\theta}_M|M)\right).$$

that is penalised by the last two factors

$$\frac{n}{2}log\left(2\pi\right) - \frac{1}{2}log\left|H\right|.$$

to account for the overparameterization/inflated likelihood estimates (Daw, 2011a).

2.10.5 WAIC: Watanabe–Akaike information criterion

Now that we have defined the staging and sequence of possible models, we need to determine some concise metric for model comparison. The widely applicable information criterion (WAIC) (Watanabe, 2010) - also known as the Watanabe–Akaike information criterion - is the generalized version of the Akaike information criterion (AIC) that has been

shown to sufficiently approximate leave-out-one cross validation (Watanabe, 2012). WAIC is frequently used to compare this class of Bayesian hierarchical models (Watanabe, 2010). Offering a simple, readily implementable and reliable basis for comparison.

BIC and related techniques: Bayesian Information Criterion (BIC) offers a simpler alternative, derived (Daw, 2011a):

$$logP(D|M) \approx logP(D|M, \hat{\theta}_M) - \frac{n}{2} log\ m. \tag{2.27}$$

Where n is the number of free parameters and m is the number of data point (loosely indicative of the confidence in the sample) (Daw, 2011a). Similar in that it takes a known computable quantity and adds a penalty for more complex models, it should be noted that BIC relies only on the likelihood $log\left(P(D|M, \hat{\theta}_M)\right)$, not the full posterior, and thus is prior agnostic. Whilst convenient, neglecting the prior can result in far poorer results (given the critical importance of the prior as described above).

It has been extensively shown that the Laplace approximation more adequately estimates the model evidence and should be used if one is willing to declare and defend some prior over the parameter space (Daw, 2011b). These findings hold true in the space case of uniform, uninformative, priors over a large range (Daw, 2011a). Parameters should only be penalised to the extend that they add explanatory power to the model: BIC blindly uses raw n and m values, irrespective of their actual fit, whilst the Laplacian approximation accounts for parameter uncertainty by the addition of the final term $-\frac{n}{2} log\,|H|$ (Daw, 2011b).

One should also note that other penalised scores for model comparisons exist, most famous the Aikaike Information Criterion (AIC) $logP(D|M, \hat{\theta}_M) - n$ (Daw, 2011a). A frequently used alternative is the WAIC statistics, that has been shown to estimate out of sample performance with confidence (Hastie, 2001). Consistently used in the literature, as shown by (Daw, 2011a).

Model comparison summary

It is perfectly plausible to compare likelihood functions directly (an intuitive solution as the likelihood directly measures the probability of the data given the parameter set) however, one would need to account for overfitting (fitting noise) when adding superfluous model complexity (free parameters) (Daw, 2011a). If models are *nested* the likelihood ratio test is a great method for comparing models that offers known statistical properties and thus an associated p-value test that is frequently used in the literature.

If models are not nested, approximate Bayes factors can be used as a basis of comparison. Given it's simplicity, BIC is widely used throughout the literature, however there is increasing evidence that - if one is able to specify some prior - Laplacian approximations may offer more reliable results. Instead of relying on simple parameter/data counting (as done in BIC), the Laplacian approximation sufficiently accounts for variability of the parameter estimates (captured by the Hessian).

If we are able to adequately estimate out-of-sample performance, this should be used as a basis of comparison Daw, 2011a. The WAIC metric circumvents the need to split data, impractical in sequential decision making tasks, proving to be a sound metric for model comparison.

2.10.6 Comparing population model

How then do we extend these model comparison techniques to the aforementioned hierarchical structures so often used in Bayesian data analysis?

We need to begin by determining whether the model itself is a fixed or random effect (Daw, 2011a). If, as is often the case, we wish to make categorical claims at the mechanisms of the brain, it may be natural to assume no variability across subjects in the model *identity* (as opposed to in it's parameters). Following this logic, the model identity is then taken as fixed effect across subjects. Naturally, although the underlying data generating process may be fixed across subjects, noise may enter the system at any stage resulting in unique fluctuations. Bayes' theorem then tells us:

$$P(M|c_1 \dots c_N) \propto P(c_1 \dots c_N|M)P(M). \tag{2.28}$$

Neglecting the hierarchical structure: One simple (and extreme) approach would then be to completely neglect any hierarchical prior and assume all individual subjects parameter values are sampled independently from some (known or unknown) prior distribution (Daw, 2011a). Assuming this independence, one can then decompose equation 2.28 across subject and perform inference separately (analogous to the summary statistics approach to parameter estimation):

$$log\left[P(c_1 \dots c_N|M)P(M)\right] = \sum_i logP(c_i|M) + logP(M). \tag{2.29}$$

This, naive but useful, approach computes the model evidence for the full dataset by aggregating the probability of the data given the model over each subject's fit (where this individual subjects probability of the data given the model is captured by either the WAIC, BIC or Laplacian approximation of the model evidence). Comparisons can then be drawn between models by using these aggregate values to compute Bayes factors (Daw, 2011a). Notably, if taking this approach one would likely report the number of independent subjects, and the proportion of individual subjects' models that are in agreement with the findings drawn by the population - to test the assumption of subject independence.

If the research question is primarily concerned with the population model, a more laborious - but perhaps robust - approach would be to integrate out the population level parameters $\theta_{pop} =< \mu_\alpha, \mu_\beta, \sigma_\alpha, \sigma_\beta >$ (Daw, 2011a). Effectively computing the model evidence over all possible population level models (Gelman et al., 2004), as detailed here:

$$P(c_1 \dots c_N|M) = \int P(c_1 \dots c_N|M, \theta_{pop})P(\theta_{pop}|M) \;\; d\theta_{pop}.$$

This quantity, of course, needs to be approximated by the aforementioned techniques (BIC, WAIC, Laplacian Approximation, etc) (Daw, 2011a). Notably, the inner function $P(c_1 \dots c_N|M, \theta_{pop})$ again requires an integral approximation.

Lastly, one could consider some probability distribution over the model *identity*. That is, to assume *random effects* over the individual subjects' model identity (Daw, 2011a). Whilst requiring an additional hierarchical structure to capture the variability across model identities, the implementation is largely unchanged.

It should be noted that a summary statistics approach to model selection - whereby researchers performing model section by averaging the Bayes Factors across individual subject

models - is unfounded. This is because of the nature of interpretation, this type of analysis would assume variability in model identity being sufficiently captured by that dispersion across subjects, an unjustified theoretically claim.

2.10.7 Caveats and notes

Here we leave some final remarks that may be important in both practical applications and in intuiting the aforementioned.

Why not assess models by counting the predictive accuracy? A natural approach may be to simply select the model that is able to best predict the (testing) data. This, however, is unjustified and is strongly advised against (Daw, 2011a). Firstly, this approach neglects the specification of a probabilistic observation model. Removing the reliance on statistical estimation forgoes the opportunity to make actual statistical claims (that is, inferring the results generalise to the population with sampling variability). Secondly, in our neuropsychological domain, we are frequently interested in model specification as a proxy for cognitive function. This too is abandoned if a pure predictive approach is taken. Finally when employing this approach magnitude is completely negated. Computing likelihood (or posterior) distributions allow one to quantify the degree to which the prediction deviates from the true response - as apposed to a binary indication - meaningful information about the variability of the system (Daw, 2011a).

Is it possible to discern between-group parameter differences if parameters are correlated? As previously emphasised: although the learning rate parameter α and exploratory temperature parameter β may be independent in the underlying data generating process; they may be correlated when assessed empirically because they have similar expected effects on the observed data. This may cause difficulty in interpreting the parameter estimates and testing quantities across populations. One may wish to investigate what subset of parameters vary significantly across populations. This can be achieved by posing the question as a model selection hypothesis whereby various models are tested that share certain parameters (shared α vs shared β vs shared α and β) (Daw, 2011a).

Assessing the model fit: Bayesian hierarchical inference lacks ubiquitous intuitive metrics to monitor model fit, however, a number of model diagnostics are frequently reported.

Data likelihoods (often BIC-corrected) are regularly reported. The data log likelihood under pure chance can be easily computed (Gelman et al., 2004). These quantities allow us to compute the *pseudo-*r^2. A (theoretical) perfect prediction model producing a unit likelihood $P(D|\theta_M, M) = 1$, it follows that the *pseudo-*r^2 is computed as the fraction reduction in this log-likelihood of the model from the log-likelihood of the chance null hypothesis.

Let R be the log-likelihood under chance (such as in 100 trial binary choice task $100{\cdot}log(0.5)$) and L be the log-likelihood under the fit model, then:

$$pseudo\text{-}r^2 = 1 - \frac{L}{R}.$$

It can be more interpretable to examine the average log-likelihood per trial (i.e. $\frac{L}{T}$ for T trials). When dealing with choice data, exponentiation this average log-likelihood produces probability distributions that are readily interpreted relative to the random null model (Gelman et al., 2004).

It is also straightforward to assess whether any model fits better than chance. Every model nests a 0 parameter empty model (which assumes all data is due to chance). A likelihood

ratio test can be used to assess whether or not a model bests it is 0-parameter nesting (Daw, 2011a). A more rigorous, frequently used, test is test the full model against a nested alternative that contains only parameters modeling mean response tendencies or biases - as is commonly done in regression analysis (Gelman et al., 2004).

2.11 Optimisation procedure

We have described, at depth, what models we wish to examine, however we have not commented on the (Bayesian) optimisation procedure to fit the parameters. The models are fit with the NUTS (No-U-Turn-Sampling) algorithm, detailed in figure 2.13, the NUTS algorithm implements MCMC sampling with efficient sampling bounds (Homan and Gelman, 2014).

Algorithm 1 Hamiltonian Monte Carlo

Given θ^0, ϵ, L, $\mathcal{L}$, M:
for $m = 1$ to M **do**
  Sample $r^0 \sim \mathcal{N}(0, I)$.
  Set $\theta^m \leftarrow \theta^{m-1}, \tilde{\theta} \leftarrow \theta^{m-1}, \tilde{r} \leftarrow r^0$.
  for $i = 1$ to L **do**
    Set $\tilde{\theta}, \tilde{r} \leftarrow \text{Leapfrog}(\tilde{\theta}, \tilde{r}, \epsilon)$.
  end for
  With probability $\alpha = \min\left\{1, \frac{\exp\{\mathcal{L}(\tilde{\theta}) - \frac{1}{2}\tilde{r}\cdot\tilde{r}\}}{\exp\{\mathcal{L}(\theta^{m-1}) - \frac{1}{2}r^0\cdot r^0\}}\right\}$, set $\theta^m \leftarrow \tilde{\theta}$, $r^m \leftarrow -\tilde{r}$.
end for

function Leapfrog(θ, r, ϵ)
Set $\tilde{r} \leftarrow r + (\epsilon/2)\nabla_\theta \mathcal{L}(\theta)$.
Set $\tilde{\theta} \leftarrow \theta + \epsilon\tilde{r}$.
Set $\tilde{r} \leftarrow \tilde{r} + (\epsilon/2)\nabla_\theta \mathcal{L}(\tilde{\theta})$.
return $\tilde{\theta}, \tilde{r}$.

FIGURE 2.13: Monte Carlo sampling procedure, the basis of NUTS (Homan and Gelman, 2014)

2.12 Model-free correlation analysis

When conducting statistical enquiry, it is pragmatic to precede model fitting with a model-free analysis in hopes of gaining better intuition into the data, reduce dimensionality and suppress the space of possible models by examining and fine-tuning theoretical ideas.

A thorough model-free analysis scrutinises linear, non-linear and possibly interactive relationships in the data; removing superfluous information.

F-test: A simple linear regression F-test fits a linear model to the data (in our case, the WCST aggregate performance as a linear function of the selected covariate) and then measures the relative variation explained producing an F-statistic:

$$F = \frac{explained\ variance}{unexplained\ variance}$$

This statistic follows an F-distribution and can test statistical significance.

In the discrete case, the calculation is performed:

$$F = \frac{\sum_{i=1}^{K} n_i \left(\bar{Y}_i - \bar{Y}\right)^2 / (K-1)}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_i\right)^2 / (N-K)}.$$

Where K is the number of groups in the discrete class; N is the number of data points; $\bar{Y}$ is mean response in the dataset & $\bar{Y}_i$ is a mean response of group i. In the continuous case the cross correlation statistic is utilised.

Whilst theoretically robust (offering statistical founded interpretations) the F-test is severely limited only capturing independent linear dependence.

2.12.1 Mutual Information: nonlinear variable ranking

Originating in information theory, mutual information (MI) captures the mutual dependence between two random variables (Cover and Thomas, 2006). More technically, it quantifies the amount of information (in Shannons bits, nats or hartleys) obtained about one random variable by observing the other. MI is intrinsically linked to entropy of a random variable: which itself quantifies the expected "amount of information" held in a random variable (analogous to variance of the random variable in the expectation).

Statistical intuition: In statistics and mathematical data analysis, mutual information can quantify non-linear relationships in random variables (Baudot et al., 2019). More specifically, MI determines how different the joint distribution of a pair of random variables $p(X, Y)$ is from the product of the marginal distributions $p(X) \otimes p(Y)$; that is, the expected value of the pointwise mutual information (PMI).

MI variable ranking: As illustrated in figure 2.14, MI is capable of capturing nonlinearities in the data. In our case, the relative scores of MI are of interest, provide a method for ranking covariates by the amount of mutual information with the response.

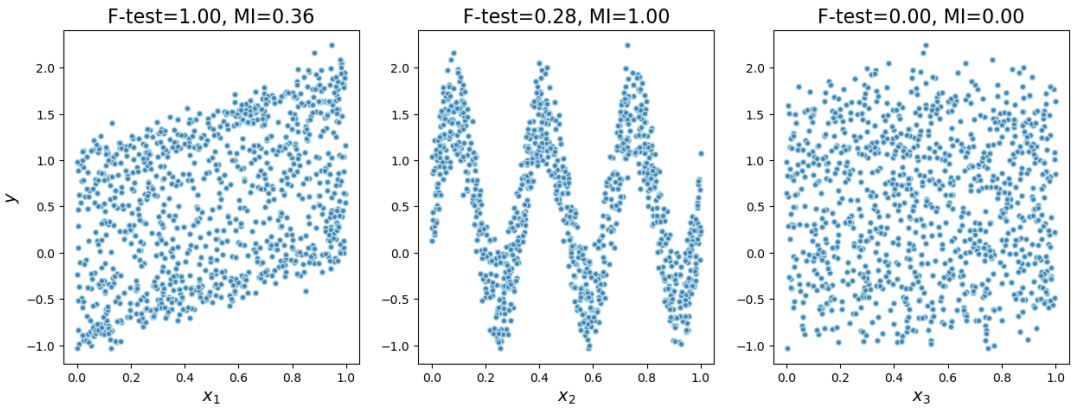


FIGURE 2.14: An illustration of non-linear dependence. The figures plot the relationship between 3 sets of random variables: the first with a high-variance linear relationship; the second with a low-variance non-linear relationship; and the third without any variable dependence. Both the (F-statistic) p-values and (normalized) mutual information are reported. It is clearly illustrated that MI is capable of capturing non-linear dependencies in the data by quantifying their joint probability density function relative to the tensor product of their individual density function.

.

Computation: Formally, the MI of two jointly continuous random variables is defined:

$$I(X,Y) = \int_{\mathrm{x}} \int_{y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} dxdy.$$

Where the integrals are replaced with summations for discrete instances.

Kullback–Leibler divergence: Intuitively is it straightforward to show the relationship between MI and another salient information-theory derived quantity the Kullback–Leibler (KL) divergence. The KL divergence $D_{KL}(P||Q)$ - or relative entropy - is an asymmetric statistical difference that measures how one probability distribution Q differs from another P (Baudot et al., 2019).

The divergence of P from Q can be thought of as the expected excess "surprise" from using Q as a model with the actual distribution over the space is P (Cover and Thomas, 2006).

For distributions P and Q of a continuous random variable, the relative entropy is defined as:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) log \frac{p(x)}{q(x)} dx.$$

MI and D_{KL}: MI can then be shown to be the D_{KL} from the product of the marginal distributions $p(x) \otimes p(y)$ of the joint $p(x,y)$:

$$I(X,Y) = \mathbb{E}_Y \left[D_{KL} \left(p_{X|Y} \parallel p_X \right) \right].$$

Quantifying the conditional dependency (after accounting for variation in the reference covariate) (Cover and Thomas, 2006).

Although F-test and MI offer sound approximations of shared distributions, we may wish to leverage ensemble methods: in hopes of providing a fuller analysis. In particular, we may wish to use ensemble methods for variable section.

2.12.2 Ensemble methods

mRMR: Maximum Relevance Minimum Redundancy

Intuition: Maximum Relevance Minimum Redundancy (mRMR) is a prominent and highly successful feature selection algorithm that offers a succinct and intuitive framework to curate a subset of covariate that are not only most predictive, but also the most uncorrelated and distinct (Zhao, Anand, and Wang, 2019). Many modern machine learning applications mine a vast collection sample covariates that require scientifically robust scrutiny and pruning in order to build generalisable and scalable models. In order to arrive at interpretable/actionable solutions feature redundancy ought to be carefully considered in addition to purely (linear or nonlinear) correlation with the response variable (Gelman et al., 2004). In essence, mRMR balances feature importance and mitigates feature redundancy by ranking covariates according to some score that captures both metrics:

$$f(X_i) = \frac{\phi(Y, X_i)}{\delta(X_i, X_i')}.$$

Where $\phi(.)$ is some function that captures the relevance of feature X_i when predicting response Y and $\delta(.)$ is some function that quantifies the redundancy between feature X_i and the set of compliment features X_i'. Any plausible function set may be used to quantify these metrics, in this way the original algorithm can readily be extended to any nonlinear and non-parametric functions (Zhao, Anand, and Wang, 2019).

Formalisation: Whilst many variants have been thoroughly tested on both real-world and simulated datasets - including nonlinear kernel transformations and Shannon negative entropy - two variants prove consistently superior and thus are used in our analysis (Zhao, Anand, and Wang, 2019).

1. FCQ (F-test correlation quotient):

$$f^{FCQ}(X_i) = \frac{F(Y, X_i)}{\frac{1}{|S|}\sum_{X_s \in S} \rho(X_s, X_i)}$$

where $F(.)$ is the F-statistic score for relevance; S is the complement set of covariates (excluding feature i) and $\rho(.)$ is the Pearson correlation. Note that the Q is used to denote the use of a quotient as apposed to a simple subtraction between relevance/redundancy scores (normalisaing the data).

2. RFCQ (Random-Forest correlation quotient):

Random forest: Decision trees are a simple discriminative algorithm that attempt to model dichotomous response variable by sequentially separating input data by their covariate values (Hastie, 2001). In the case of discrete covariates, the tree separates the data at each possible permutation, and selects the optimal configuration by computing the accuracy of the prediction and utilising the split that best fits the data. In the case of continuous predictor variables, each possible splitting value is computed and thereafter the same model score (denoted as the Gini importance or mean decrease impurity) is used to locate the optimal split (Hastie, 2001). The algorithm is flexible and naturally extends to both (probabilistic) multinomial data and regression settings.

Though simple and easy to implement, classical decision trees are greatly flexible and thus regularly overfit any particular data instance. The widespread success of the algorithm is a consequence of leveraging statistically reliable techniques to bootstrap aggregate performance (Hastie, 2001). By randomly sampling bootstrap-samples of the dataset and aggregating the predictions of many independently fit decision trees, statistically robust (in the limit) results can be achieved. This algorithm is known as a Random Forest.

Random forest feature selection: Each node of each decision tree represents a dichotomous split in the data over a single covariate. As such, a natural extension is to use decision trees - and thus random forests - for feature selection. The (nonlinear) predictive power of any particular variable can be assessed by quantifying the deterioration of the algorithms performance in its absence (Hastie, 2001). Random forest offers a nonlinear random sampling course of action to feature importance ranking and feature selection.

It follows, that RFCQ is given by:

$$f^{RFCQ}(X_i) = \frac{I_{RF}(Y, X_i)}{\sum_{X_s \in S} \rho(X_s, X_i)}.$$

where $I_{RF}(Y, X_i)$ is an importance score computed by random forest feature selection (Zhao, Anand, and Wang, 2019).

2.13 GAMs: General Additive Models

In aid of quantifying the relationships between neuropsychological quantities, it is desirable to fit some model to measure the relationships between psychological covariates and the recovered biological parameters. We aim to uncover potential neuropsychological or demographic relevance in the learning process. Although in highly complex learning tasks this may be encoded into the learning model directly. It is more prudent and interpretable to fit additional models on the recovered parameters - as done in the relevant literature (Ball and Goldstein-Piekarski, 2017). For our purposes, GAMs are the most suitable option because of three attributes they possess, GAMs:

1. Allow for nonlinearities in the data.
2. Decouple the covariates producing interpretable results. If not by quantity, at least by covariate significance and direction.
3. Regularise the fit by naturally, allowing for further variable selection and generalisation.

GAMs fit a linear model after performing some transformation to the input space of each covariate.

$$E(Y|X_1, X_2, ..., X_p) = \alpha + f_1(X_1) + f_2(X_2) + ... + f_p(X_p).$$

Whilst extendable to any set of link functions, $f(.)$ usually constitute splines and linear terms (Hastie, 2001).

Fitting GAMs is conducted by minimising some penalised least squares (and often tuning the hyperparameters λ_i):

$$PRSS(\alpha, f_1, f_2, ...f_p) = \sum_{i=1}^{N} \left(y_i - \alpha - \sum_{j=1}^{p} f_j(x_{ij}) \right)^2 + \sum_{j=1}^{p} \lambda_j \int f_j''(t_j)^2 dt_j.$$

Apart from it is pragmatism, GAMs are very intuitive models. The penalisation term computes the second derivative of the basis transformation. That is, the rate of change - favouring stable models.

Chapter 3

Methodology

Returning to our scientific enquiry, we are fundamentally interested in mapping learning rates to psychological attributes.

We require a task to measure probabilistic learning - in our case, dynamic learning under uncertainty - of subjects, as well as the data to categorise and describe subjects into psychologically relevant sub-populations. Demographic data is readily accessible by self-reporting, and we are primarily interested in how variations across different biological and psychological neurocorrelaties influence the learning process. These quantities are not directly observable and therefore require approximations - via additional auxiliary neuropsychological experiments.

This chapter aims to motivate and detail the choices made in our experimental design. It can be grouped into three categories: subsections 3.1 and 3.2 explain our experiments; subsections 3.3 and 3.5 describe how we reduce complexities in some covariates; and subsections 3.6, 3.7 and 3.8 detail the RL modeling process.

3.1 Experimental design

In conducting our study, a suite of psychological experiments were selected to measure the participants' (probabilistic) associative learning ability, as well as a series of auxiliary tasks to gauge executive functionality.

The Wisconsin Card Sorting Task (WCST) is our study's primary assessment, detailed in subsection 2.3.1, requiring subjects to perform associative learning by mapping stimuli to rewards.

As well as collecting demographic data, the auxiliary tasks include the N-back, Corsi, Fitts and Navon assessments; as explained in section 2.1. An $N = 2$ N-back is used to assess working memory and (mildly) examine fluid intelligence (Unsworth and Engle, 2005), (Palomäki et al., 2012). The Corsi Block Span task (Corsi) sheds light onto visospatial working memory (Brunetti, Del Gatto, and Delogu, 2014) - more specifically the activity in the ventrolateral prefrontal cortex. As the theoretical expected values are known, the Fitts task is used to proxy attentiveness, computer literacy and hand-eye co-ordinative motor skills by assessing the deviation from expected behaviour (Fitts, 1954a). Finally, the Navon task assesses the discrepancy between speed and accuracy of identifying patterns in both local and global structures, allowing us to measure global versus local stimulus processing, as well as attentiveness (Davidoff, Fonteneau, and Fagot, 2008). The experiments were implemented sequentially as follows:

1. Wisconsin card sorting task (WCST)

2. N-back task
3. Corsi block span task
4. Navon task

The experiments were conducted online utilising Psytoolkit - a research first state-of-the-art free Linux distribution for conducting Psychological research (Stoet, 2010). Participants were sampled from to represent a diverse population of individuals and demographics. The experiments were conducted online, using Amazon's Mechanical Turk (Crowston, 2012).

All individuals over 18 years of age with a desktop computer were eligible to participate. MTurk allows researchers to post "jobs" on the site, detailing the task specification and participation eligibility requirements, and thereafter eligible workers can opt to perform the task if they so desire. Subjects were paid $2.70 to participate in the experiment, with an expected time of 25 minutes (a reasonable estimate based on the task duration and complexity), and a maximum time of 40 minutes over which the experiment would automatically terminate.

MTurk offers easy access to a large populations of workers, and as such has been widely adopted in both research and industry (Lu et al., 2021). MTurk's ubiquity has proven it to be a reliable, generalisable, tool to generate samples of the eligible population (Lu et al., 2021) and for this reason we consider it a reliable source of data. The attentiveness of some workers, however, have been raised doubts (Lu et al., 2021). It is also intuitive to imagine a sub-set of the population of workers that attempt to work through available tasks as quickly as possible (without regard for the task) to earn money. As our study is not necessarily concerned with individuals with psychiatric illnesses or mental illnesses/impairments, we can safely assume that healthy individuals should be able to complete the chosen neuropsychological tasks with relative ease (supported by the literature, discussed in section 2.1). It is, therefore, reasonable to dismiss a portion of subjects who perform very poorly on the tasks, although there is limited indication as to what a reasonable threshold will be as it is very specific to the task in question.

MTurk returns the experimental results as text files. We then wrote a number of Python scripts to extract, tabulate and store the data for analysis. All experimentation, modelling, dimension reduction, and data analysis were then conducted in Python, available here. The dimension reduction, mutual information and statistical estimates utilised Scikit-Learn (Pedregosa et al., 2011) and the Bayesian RL modeling was conducted in pyStan - an open source probabilistic programming interface frequently used in cognitive science RL modeling (Konstantakopoulos, 2019).

3.2 Data encoding

The high-level objective of this analysis is to model differences in individuals' WCST performance, measuring their ability to perform associative learning under uncertainty (Miyake et al., 2000), (Huizinga, Dolan, and van der Molen, 2006) and, in doing so, quantifying meaningful differences in individual task performance.

The WCST model - biologically inspired - is designed to reveal abstract cognitive properties describing the learning parameters and exploratory nature of the individual. The model parameters were then analysed with respect to neuropsychological faculties, executive functions (measured by the auxiliary tasks) and demographic data, attempting to reveal possible relationships between associative learning and cognitive faculties/executive functions.

Throughout the thesis the subscript s refers to a *subject* and t refers to *time trial, indicating the trial number in a sequence.*

3.2.1 Reward encoding

In the context of cognitive science reinforcement learning models, discussed thoroughly in section 2.6, the response is simply encoded through the binary reward r_t^p that measures whether or not a subject was able to match the stimulus to the correct underlying rule. The WCST has three possible matching rules: colour, shape and number.

$$r_t^s = \begin{cases} 1 \text{ if the correct rule is chosen} \\ 0 \text{ otherwise.} \end{cases}$$

The matching rule changes stochastically. Individuals are tasked with learning (through association of stimulus to reward) the correct rule and updating their value estimates associated with a stimulus after stochastic changes in the underlying process (Slooten, Jahfari, and Theeuwes, 2019), (Van Slooten et al., 2018), (Barceló, 2021).

Although the Q-learning models used by Slooten, Jahfari, and Theeuwes, 2019, Van Slooten et al., 2018 and Barceló, 2021 represent the data directly as it is observed without computed summary statistics - as it exactly corresponds to how the individual receives feedback and is the approach taken in cognitive science RL research - a simplified encoding is required for exploratory data analysis. In order to approximate WCST performance, we want to generate a single figure that captures an individual's associative learning ability. We simply compute the proportion of correct actions taken:

$$y^s = \frac{1}{T} \sum_{t=1}^{T} r_t^s. \tag{3.1}$$

where T is the number of trials. This summary statistic is used in the initial exploratory data analysis (EDA); however, the full sequence of choices will be used in the subsequent RL model. This average performance y^s will serve as a metric to proxy the relationship between associative learning and demographics (for variable reduction). If the research in question was not concerned with extracting learning parameters, the proportion scores could quite easily be modeled directly with some flexible regression analysis (such as Gaussian processes (Hastie, 2001)). If the researchers are instead interested in the biologically inspired parameters, as in our research, a full RL specification is prefered as we may be able to capture elements of the underlying cognitive process.

3.2.2 Independent variables

Our study consists of multiple distinct experiments offering rich, detailed data sources that require variable pruning (selection and dimensionality reduction) in order to fit a parsimonious model (Hastie, 2001).

After ranking covariates for potential inclusion in the RL model, we fit a number of Bayesian RL models to the data and select the model with the best out-of-sample performance.

Theoretical biological parameters

Analogous to the work done by Barceló, 2021, Humann, Fischer, and Ullsperger, 2020 and Slooten, Jahfari, and Theeuwes, 2019 the most relevant covariates in our analysis are the RL parameters that loosely map to biological processes: describing an individual's learning characteristics. The two quintessential RL cognitive science parameters capture the learning rate α and exploratory temperature parameter β (Daw, 2011a).

These are the most significant parameters in the analysis: constituting the learning model and offering direct neuropsychological interpretations.

Demographics data

The following participant demographic information was self-reported, apart from the reaction times, by subjects participating in the MTurk study:

$$
\begin{aligned}
X_a^s &: \textit{age} && \textit{numeric} \\
X_g^s &: \textit{gender} : \{\text{male, female, other}\} && \textit{nominal} \\
X_h^s &: \textit{handedness} : \{\text{left, right, ambidextrous}\} && \textit{nominal} \\
X_i^s &: \textit{income} && \textit{ordinal} \\
X_e^s &: \textit{education level} && \textit{ordinal} \\
X_c^s &: \textit{computer hours per week} && \textit{ordinal} \\
X_{d-rt}^s &: \text{response time (ms) to complete the demographics questions} && \textit{ordinal}
\end{aligned}
$$

Computer hours per week is a numeric real number between 0 and 100. *Education level* is coded as a factor variable with four levels: {none, high-school, bachelors degree, graduate degree}.

Sample Demographics					
Gender		Handedness		Education	
Male	118	Right	195	High School	44
Female	94	Left	16	University	117
Other	2	Ambidextrous	3	Graduate School	53
Total	214		214		214

TABLE 3.1: Summary demographic information about the experiment's subjects. The sample is well distributed between men and women. Additionally, the average subject age was 37.6 years old with a large standard deviation of 12.3 years.

Neuropsychological data

Auxiliary neuropsychology experiments (detailed in sections 2.3.2 to 2.3.6) were conducted to gauge the different executive functionality of each participant. The following data was extracted as subjects performed the MTurk task battery, for each subject s:

$$
\begin{aligned}
X^s_{nback} &: N-Back\ task\ accuracy && numeric \\
X^s_{nback-rt} &: N-Back\ reaction\ time && numeric \\
X^s_{fitts} &: Fitts\ task\ accuracy && numeric \\
X^s_{corsi} &: Corsi\ span\ achieved && nominal \\
X^s_{navon} &: Navon\ task\ accuracy && numeric \\
X^s_{navon-rt} &: Navon\ reaction\ time && numeric \\
X^s_{wcst-rt} &: WCST\ reaction\ time && numeric
\end{aligned}
$$

3.3 Outlier removal and covariate compression

Before performing the four steps of the analysis, it is important to remove outliers and (if possible) reduce dimensionality by compressing covariates if empirically supported.

3.3.1 WCST

In preprocessing the data, the average WCST performance forms the basis of outlier removal. Our study does not consider individuals with severe mental illnesses, impairments or psychopathologies, nor do we have any information as to whether or not the MTurk subjects suffer from such conditions. It is exceedingly unlikely that healthy individuals score a very low score in the WCST - given its simplicity - and thus low scores are more likely a consequence of negligence or completing the task as quickly as possible to receive payment. This interpretation is supported in meta-studies that examine the validity of MTurk as a scientific platform (Lu et al., 2021). Furthermore, this is in accordance with the WCST literature whereby healthy participants are expected to have little trouble with the task (Baker, 2012), (Miyake et al., 2000). For this reason, in addition to incomplete data being removed, subjects with significantly low scores were excluded.

We assessed the aggregate WCST performance scores $\mathbf{Y}$ and set a threshold λ to exclude subjects with exceedingly poor scores, removing their data from the analysis.

Decision threshold: The decision was taken to remove all participants who scored under 0.4 for the WCST task. This is a plausible threshold given the complexity of the task, and removes about 10% of the participants. It is also only marginally above the expected result of a candidate who guesses at random (0.333). There is no literature, to our knowledge, that performed the WCST on MTurk, and thus no basis for this decision apart from an intuition developed in the task.

$$\lambda_{wcst} = 0.4.$$

3.3.2 Navon task

We examined the data generated by the Navon task in order to further simplify the covariates summary statistics. The question arises: are there significant differences in global and local performance? If not, values may be aggregated to reduce dimensionality.

As a component of the analysis, we are concerned with two distinct ideas regarding the relationship between Navon task performance and WCST performance:

1. Does the Navon task - a proxy for attentiveness - show correlation with the WCST performance and its underlying learning parameters? Secondly, in aid of this,
2. Can any insight be drawn between the relative performance of global or local Navon performance?

To assess these relationships statistically, we summarise the Navon task data by the following three covariates:

$$\begin{aligned} X^s_{nv-GL} &: \textit{Navon task global} - \textit{local performance} && \textit{numeric} \\ X^s_{nv-agg} &: \textit{Navon task aggregate score} && \textit{numeric} \\ X^s_{nv-rt} &: \textit{Navon reaction time } (ms) && \textit{numeric.} \end{aligned}$$

This encoding allows us to capture potential variations between localised attention mechanisms across participants. If there is no empirical evidence to support the separating of global and local attention differences, following the principle of parsimony, it is advantageous to favour simpler parameterisations (Hastie, 2001), and therefore, if compressing these variables to aggregate figures is possible without losing information about the task's correlation with associative learning, it is favourable.

3.4 Modeling pipeline

The covariates were combined in a design matrix. Similarly, the aggregate summary responses were vectorised.

$$\begin{aligned} \mathbf{X}^s &: [X^s_a, X^s_g, \ldots, X^s_{wcst-rt}] \\ \mathbf{Y} &: [y^1, \ldots, y^s]. \end{aligned} \tag{3.2}$$

where $\mathbf{X}^s$ is a vector that forms of the rows of the matrix $\mathbf{X}$.

Our modeling pipeline is thus a four fold sequential process:

1. **Covariate prioritisation**: In order to rank potential covariates (essentially preprocessing to order the covariates for inclusion in the subsequent RL model) we performed a correlation analysis between each covariate in $\mathbf{X}$ (all demographic and neuropsychological data) and the summary WCST average performance $\mathbf{Y}$. Variables are selected for the RL model on a purely empirical basis. This does not, however, exclude any covariates from the more important subsequent analysis whereby we examine the relationships between covariates in $\mathbf{X}$ and parameters extracted from the RL model (quantifying one's learning ability).
2. **Simulating RL models**: In the next section of our analysis, we perform simulations to provide confidence in the chosen model paradigm. First, we assess if, for a single subject, RL can adequately capture the decision making process during an associative learning task.

   Furthermore, what is the best method to scale RL models over a population? Although it is well understood that the summary statistics approach to Bayesian hierarchical models (described in section 2.7.6) insufficiently captures population variance (Gelman et al., 2004) we run these simulations because it is worth investigating

this assumption in the context of Bayesian reinforcement learning models. Although known to be statistically flawed (insufficiently accounting for variance across subjects) (Daw, 2011b), (Parr, Rees, and Friston, 2018), some advocate the summary statistics approach for its simplicity. To address this, we simulate a sequence of actions that are generated by an RL model - simulating the assumed data generating process - and assess the risk of leveraging the summary statistics approach over a full Bayesian hierarchical model.

The simulated dataset allows us to measure the discrepancy between the model fit and true (known) parameterisation, examining the validity of the approach. If a case for the summary statistics approach could be made, it allows one to ask a multitude of research questions with very little effort as individual models can be fit and simply added together to measure the average performance of different sub-populations of interest. If the variance is truly insufficiently captured, a full hierarchical model is required.

3. **Cognitive science RL models**: After ordering the covariates and assessing the necessity of a full Bayesian RL model, a series of RL models were fit to assess which parameterisation best describes the associative learning process (and thus latent executive functions) governing the WCST. Model comparison is conducted to select the optimal model that best articulates the data (explained in section 2.10). After assessing the model's ability to regenerate the data - indicative of capturing the data generating process - the best fitting model learning parameters (based on generalised performance) are extracted. These parameters quantify the learning process of the subject, and are the primary quantities of interest.

4. **Analysis of learning parameters**: Finally, the learning parameters are examined with respect to the covariates in X to address our research objective of assessing which executive functions, cognitive faculties and (possibly) demographic information describes one's learning process. A correlation analysis is used to describe the learning parameters. In the presence of sufficient correlation, a General Additive Model would have been used to model the relationships; however, the correlation analysis did not justify a full model.

3.5 Covariate prioritisation

Although we know we want to fit an RL model to explain the decision generating process that guides one's ability to perform the WCST, how do we reduce the infinite possible model configurations? The literature advocates a nested modeling approach (Daw, 2011b) which also requires variable ranking. Given the simplicity of the WCST, it is unlikely that adding many covariates to a baseline null model will improve the fit, and therefore it is important that we select the covariates with the best chance of improving the fit (Hastie, 2001), (Gelman and Hill, 2006).

The first step in our analysis - assessing the correlations between covariates and the average WCST performance scores in $\mathbf{Y}$ - is used to empirically rank variables by their (both linear and nonlinear) correlation with WCST performance y^s.

We thus need to establish an order of variable importance before modeling the data. Three schools of techniques are used to rank the covariates:

1. Linear relationships
2. Non-linear relationships

3. Ensemble relationships

In determining the order of variable relevance, each covariate is examined with respect to average WCST scores by the techniques discussed in section 2.12. The first, and simplest, comparison is to conduct a linear regression F-test to capture linear dependence. Nonlinearity is examined by computing the Mutual information between each covariate and the aggregated WCST scores (section 2.12.1). Finally, to paint a more complete picture of how variable redundancy and autocorrelation may render additional covariates superfluous, both the RFCQ and FCQ variants of mRMR are computed (as elucidated in section 2.12.2).

3.6 Simulating RL models

When performing statistical learning it is generally advisable to test your chosen school of algorithms by simulating datasets (Hastie, 2001). Simulated data, particularly in the realm of generative models, allows one to assume the data generating process and then subsequently assess the chosen algorithms' ability to recover the process. In our case, we would like to test two assumptions: (1) Can Reinforcement Learning adequately model the decision making process of an individual during an associative learning task (such as WCST); and (2) When modelling the entire population, is the summary statistics approach (discussed in section 2.7.6) justified, or is a full hierarchical Bayesian model required to capture the variation across subjects (Hastie, 2001), (Gelman et al., 2004)?

3.6.1 Bayesian reinforcement learning: single subject

We begin by generating data of a single subject performing an associative learning task. We simulated the subject's choice behaviour assuming they follow an underlying RL update. As an additional illustration, we assume the model to be Bayesian to illustrate how prior distributions can be placed over RL models (Gelman et al., 2004). In the frequentist case, the priors can simply be assumed to be uniform over the permissible range (Hastie, 2001).

Data generating process: Instantiating the WCST, a subject is tasked with learning the values associated with a set of stimuli through experience.

Each stimulus has an (unknown) true expected value associated with it - that is how likely the stimulus is to return positive feedback. The objective to maximise reward, and thus if the true expected values of each stimulus were known, the subject could achieve the maximum reward by always selecting the stimulis with the highest expected value.

In reality, however, the subject does not know the true expected value of each stimulus and needs to estimate these values through experience by undertaking different actions (Daw, 2011b), (Sutton and Barto, 2018).

Formally, the action space is defined as *action space*:

$$a \in \{c, s, n\}$$

where $\{c, s, n\}$ stand for $\{colour, shape \text{ and } number\}$.

Each action (selecting a stimulus) has some unknown probability of returning positive feedback. Furthermore, it is possible that the action values are non-stationary, and thus we define a *state* as an action value at a particular time t in the trail sequence. To simulate the data we arbitrarily set the true (unknown) probability of receiving a reward associated with each stimulus as:

$$Q^{true}(a) = \begin{cases} 0.8 \text{ if } a = c \\ 0.6 \text{ if } a = s \\ 0.7 \text{ if } a = n \end{cases}$$

The subject, with no prior knowledge of these values is assumed to initialise all state-values uniformly (completely unbiased) $Q_0(a) = \frac{1}{3}$. We then assume the subject uses a reinforcement learning update to amend their internal value estimates $Q_t(a)$ of each action a at time t:

$$Q_t(a) = Q_{t-1}(a) + \alpha \left[r_t(a_t) - Q_{t-1}(a) \right]. \tag{3.3}$$

where r_t is the reward received at time t. $r_t(a_t)$ is the feedback used to update the subject's internal model.

$$r_t(a_t) = \begin{cases} 1, \text{ with probability } Q^{true}(a) \\ 0, \text{ with probability } 1 - Q^{true}(a). \end{cases}$$

Simulated parameters: α, the learning rate, is the first parameter we wish to recover. The simulation aims to show that RL model parameters α and β can be recovered using Bayesian RL. Artificially, we simulate a sequence of actions with fixed learning parameters:

$$\alpha = 0.4, \beta = 10.$$

Given these parameters we can simulate choice behaviour.

The state-value approximations are used to sample actions probabilistically from a multinomial Boltzmann distribution (Hastie, 2001):

$$\pi_t(a) = \frac{exp\left\{-\beta Q_t(a)\right\}}{\sum_a exp\left\{-\beta\ Q_t(a)\right\}}.$$

Single subject sensitivity analysis

After showing the model's ability to recover the true parameter values α and β (by demonstrating that the fit parameters approximate the unknown true values) we repeat the experiment over a range of permissible α and β values to assess the stability of the approach. To ensure statistical robustness and generalisability, it is important test the system over a range of plausible parameter values, as well as incorporating extreme values in order to assess the reliability of the model at the ends of the permissible range.

To assess the approaches convergence properties, we repeated the above experiment over all combinations in the parameter space:

$$\alpha \in \{0.05, 0.1, 0.25, 0.5, 0.75, 1\},\ \beta \in [1:11]$$

.

the β values, spread in intervals of 1, appear to be spread in the roughly across this range in similar papers (Slooten, Jahfari, and Theeuwes, 2019), (Barcelo, 2020).

We then fit the same (Bayesian) RL model as described above to recover the parameters for each model.

α and β are known to be dependent - a consequence of the nonlinear update discussed in section 2.5, necessitating testing over a range of values.

3.6.2 Bayesian hierarchical models: many subjects

Premise: The premise of hierarchical models is to capture some shared variance across subjects (Gelman et al., 2004). One can, using the above approach, readily fit a model for each subject, and then simply use the individual model parameter estimates (in our case α^s and β^s associated with subject s) to fit a population distribution. Known as the *summary statistics* approach, detailed in section 2.7.6, there is merit in this method despite it being known to improperly represent variance across the population (Hastie, 2001). It is likely that members of a population share some information. Hierarchical Bayesian models allow one to regularise the parameters of any given individual model to better fit the entire population. They assume the individual parameters α^s and β^s are drawn from the distribution of the population level parameters. As discussed in section 2.7.6, this acts as a regularising function to modulate the parameters of an individual. By simulating this process we are able to show the potential dangers of an ill-specified population model.

Benefits of the summary statistics approach: despite its known inferiority, many practitioners deliberately use this approach as it does offer a number of very transparent advantages (see section 2.7.6 for furthermore details):

1. It is simpler to implement.
2. Large population hierarchical models are increasingly computationally expensive as the population grows.
3. Individual models allow one to easily examine different research questions. For example, if a set of individual subject parameter estimates α^s are used to fit a normal distribution to estimate the population parameter $\mu_\alpha, \sigma_\alpha$ by assuming $\alpha^s \sim N(\mu_\alpha, \sigma_\alpha)$ we can easily test any combination of sub-populations to compare different groups.

Simulation: To quantify the discrepancy between the summary statistics approach and the full hierarchical Bayesian model, we simulate a dataset that follows the assumptions under the hierarchical Bayesian framework (as detailed in section 2.7.6) and assess how well each method is able to recover the true parameter estimates.

Hierarchical data generating process

Depending on our research question, we may either wish to recover individual level parameters α^s, β^s (and simply use the hierarchical structure as a statistical method to regularise parameters over the population), or the population level parameters $a_\alpha, b_\alpha, \mu_\beta, \sigma_\beta$. Regardless, the assumed data generating process is identical.

It is assumed that individual subjects sample their unique data generating parameters from these global priors (Gelman et al., 2004). This random sampling functionality is a mathematical convenience that accounts for random fluctuations between individuals. Intuitively, this also holds, as we can assume some similarity across members of the same population.

Simulated dataset: To illustrate the discrepancy between the two methods, we must simulate a population of subjects performing the WCST. The data is generated the following sequence:

1. ***Individual parameters***: Individuals sample their parameters from some population level distributions.
2. ***Generate action sequences***: individuals generate their action sequences using their individual parameters (identical to the above single subject simulation).

For our demonstration, we assume that the α and β population distributions to be $Beta$ and $Normal$ distributions respectively. Any distribution may have been chosen, so long as it satisfies the permissible range. The α learning rate parameters are bound between $0 \leq \alpha^s \leq 1$ and are likely to be positively skewed towards higher values - both conditions that can be easily enforced with beta distribution. The β population parameters are positively bound (Barcelo, 2020), so the parameter are specified such that this normal distribution does not include negative values.

Therefore, we aim to recover the population level parameters:

$$\theta_{pop} : \{a_\alpha, b_\alpha, \mu_\beta, \sigma_\beta\}$$

where the individual subject parameters that generate the data are drawn from:

$$\alpha \sim Beta(a_\alpha, b_\alpha), \quad \beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta).$$

Simulated example: In performing our simulation, we assume the following population distributions:

$$\alpha \sim Beta(3, 9), \quad \beta \sim \mathcal{N}(3, 1)$$

After randomly sampling the individual parameters from these population distributions, we generate a dataset of a series of individual WCST instances. We then fit two models to recover the population distributions:

Summary statistics model: An unique model is fit for each subject. The same RL modeling procedure as that described in 3.6.1 is used to fit individual parameters to each subject's data. The individual parameters are then used as samples to fit the population level distributions.

Bayesian hierarchical model: Using the same generated dataset, a Bayesian hierarchical model is fit to recover the population parameters. The Bayesian hierarchical model requires more explicit assumptions about the parameters. For simplicity, it is assumed that normal distributions generate the individual parameters:

$$\alpha^s \sim \mathcal{N}(\mu_{alpha}, \sigma_\alpha); \quad \beta^s \sim \mathcal{N}(\mu_{beta}, \sigma_\beta).$$

Where the summary statistics approach infers variances by the distribution of the samples (Gelman and Hill, 2006), the variance parameters require explicit priors when fitting a Bayesian hierarchical model. As often done in Bayesian analysis, it is assumed that the variance parameters follow $half - cauchy$ distributions because variance is 0 bound and likely to follow some decreasing function (unlikely to be uniform) (Gelman et al., 2004). In the context of cognitive science RL models, we know that the variance of α is small (since $alpha$ is $[0, 1]$ bound), and, consequently, we specify a smaller prior relative to the variance over β. The following priors are chosen:

$$\sigma_{\alpha}^{s} \sim cauchy(0,1); \;\; \sigma_{\beta}^{s} \sim cauchy(0,3)$$

.

The hierarchical model was fit using pySTAN (Riddell, Hartikainen, and Carter, 2021) to recover the parameter estimates. pySTAN employs a Monte Carlo sampling procedure, as discussed in section 2.11, to learn parameter estimates.

3.7 Cognitive science RL models

The first component of the analysis (section 3.5) determines a covariate ranking by investigating the statistical relationships between the explanatory covariates and (aggregate) WCST performance. Thereafter, section 3.6 examines the necessity of using a full Bayesian hierarchical model in place of a simpler summary statistics approach to modelling the data. Now we turn our attention to modeling the WCST with Reinforcement learning.

The findings of the aforementioned are used to inform the cognitive science RL models specified in this section. The subsequent sections then ranks the covariates for possible inclusion in the RL model and details the necessity of a full Bayesian model.

The purpose of establishing covariate rank is to design a range of nested hypothetical models and thereafter search this function space for the model that best describes the data. We pose a *null hypothesis* as a baseline hierarchical model capturing the simplest biological mapping from the *action space* to the *reward space*. Afterwards, we add complexity by encoding parameter covariates. Allowing for model comparison by utilising the WAIC statistic (discussed in section 2.10) to approximating leave-one-out cross validation - indicative of the optimal parsimonious structure.

After selecting a model that best fits the data, the final stage of the analysis examines the correlation between extracted (learning) model parameters and executive functions/demographic data. On recovering the learning and exploratory parameters, we model the underlying biological process as a function of our neuropsychological covariates, thus capturing the relationship between executive function and learning dynamics.

3.7.1 RL model architectures

Building off the literature discussed in section 3.7 we designed a suite of RL architectures to model the associative learning process of subjects completing the Wisconsin Card Sorting Task.

RL model structure: There are two quintessential elements in specifying an RL model (Slooten, Jahfari, and Theeuwes, 2019). The state-value estimates are $Q_t^s(a)$ (for subject s and action a at time t), and subsequent sampling probability distribution $\pi_t^s(a) = f(Q_t^s(a)) \; \forall a$. The state-value estimates quantify the value of a given state, which in our case is the value of taking a particular action at a certain time in the sequence, and the sampling probabilities use these values to determine the distribution from which new actions are samples, and in turn update the state-value estimates. This says nothing, however, of the functional form that each component takes. Our goal is to search the function space to select a function form that best recapitulates the data.

Model categorisation: Models can be segmented into two broad classes: $\{$*biological covariate models and psychological covariate models*$\}$, more precisely:

1. *Biological models*: are governed by parameters that detail one's (abstract) physiological processes.
2. *Psychological models*: layer additional information on biological models by capturing neuropsychological metrics as a part of the data generating process.

We first determine a biological model and, thereafter, examine the value of encoding neuropsychological covariates directly into this data generating process. It is unlikely that the additional covariates add substantial explanatory power in the learning process; however, it is worth inclusion for illustration - in the face of more complex cognitive tasks it may be possible that RL models may insufficiently capture variance in the learning process. Most similar works from the surrounding literature, as described in section 2.6.5, work directly in the space of *biological models*, through it may be advantageous to encode additional covariates directly into the learning model (Schönberg et al., 2007).

WCST RL model: Directly comparable to the models discussed in section 2.6.6, to model the WCST we estimate the values of three possible actions: matching the cards on *colour, shape or number*. This defines our action space:

$$A \in (c, s, n).$$

State values are denoted $Q_t^s(a)$ representing subject s' estimate of the value of state/action a at time t value. These action value estimates are used to determine the frequency at which the subject samples actions, governed by probabilities $\pi_t^s(A = a)$.

Nested models: In designing our series of RL models, we begin with the simplest parsimonious structure and subsequently add complexity. First, we test a number of biologically inspired architectures, drawing from the broader cognitive science literature to specify potential parameter spaces that are loosely based on neuropsychological ideas. This is the approach taken by most relevant works, as done by Slooten, Jahfari, and Theeuwes, 2019 and Barcelo, 2020. For this reason we expect one of these models to best fit the data. We name this group of models ***bio-models***.

Adding covariates: Our objective is to extract these neuropsychological learning parameters (that quantify each subjects' associative learning ability), and thereafter analyse these parameters with respect to individual demographic and cognitive characteristics. Nonetheless, including additional covariates may better explain the discrepancies in $Q_t^s(a)$ and $\pi_t^s(a)$ updates between subjects, and as such we should explore the potential of including covariates directly in the learning or observation models. This approach is illustrated by Daw, 2011b and discussed in section 2.8. After selecting the *bio-model* that best describes the data, we explore adding covariates in two stages.

Separating covariates: Covariates are selected by their (linear and nonlinear) correlation with the subjects' aggregate WCST score (section 3.5). Further, we separate covariates by two classes: demographic data and neuropsychological data. This is to prioritise neuropsychological inference in alignment with our research question. To this end, we first investigate adding neuropsychological covariates to the RL model (naming these models *psychological models* or *psy-models* for short) and select the psychological models that best describes the data. Thereafter, if there is evidence for the data requiring a richer model specification we could repeat the process for demographic covariates (denoted *demographic models* or *demo-models*). If the models with the neuropsychological covariates produce the best out-of-sample WAIC scores, that could suggest that adding additional covariates

is warranted. If, however, a biological model is chosen we need not investigate adding furthermore parameters.

The purpose of testing a number of different functional forms of both $Q_t^s(a)$ and $\pi_t^s(A = a)$ (as described below) is to find a representation that best aligns with the data.

Now we describe the specific model tested:

Biological models

Null model: A direct instantiation of our simulated model, the *null* model contains the minimal parameterisation that captures the data generating process. We aim to recover the population level parameters:

$$\theta_{pop} :\{\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta\}. \tag{3.4}$$

We fit individual model parameters by assuming individual parameters for subject s are instantiations of these hierarchical (population) priors:

$$\alpha^s \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha); \;\; \beta^s \sim \mathcal{N}(\mu_\beta, \sigma_\beta). \tag{3.5}$$

These parameters are assumed to generate the data by determining the sequence of *actions* taken by the subject s through governing their individual update state-value estimates.

The state space estimates $Q_t^s(A)$ are initialised uniformly $Q_0^s(A) = \frac{1}{3}$ where there are 3 potential actions $a \in \{c, s, n\}$. These estimates are sequentially updated in light of environmental feedback by the **learning model**:

$$Q_{t+1}^s(A) = Q_t^s(A) - \alpha^s \left[r_t^s - Q_t^s(A)\right].$$

Subsequently, the state-value estimates are transformed to compute the probability of sampling each action, described by the **observation model**:

$$\pi_t^s(A = a) = P_t^s(A = a|\alpha^s, \beta^s, Q_t^s(A)) = \frac{exp\beta^s Q_t^s(a)}{\sum_{a\in A} exp\beta^s Q_t^s(a)}$$

where $A = \{s, n, c\}$. For brevity, moving forward, the Boltzmann distribution is represented by the following notation:

$$\zeta(\beta^s Q_t^s(A)) = \frac{exp\beta^s Q_t^s(a)}{\sum_{a\in A} exp\beta^s Q_t^s(a)}$$

Bio-model 1: Dynamic learning parameters: It is natural (and theoretically justified (Parr, Rees, and Friston, 2018)) to assume that subjects exhibit different learning rates when parsing *positive* and *negative* feedback. The first (additional) model offers an encoding of this phenomenon by replicating the *null* model, but allowing for two learning rates per subject by adjusting the **learning model**:

$$Q^s_{t+1}(A) = Q^s_t(A) - \begin{cases} \alpha^s_G \left[r_t - Q^s_t(A)\right], & r_t = 1 \\ \alpha^s_L \left[r_t - Q^s_t(A)\right], & r_t = 0. \end{cases}$$

and is therefore parameterised by:

$$\theta_{pop} : \{\mu_{\alpha_G}, \sigma_{\alpha_G}, \mu_{\alpha_L}, \sigma_{\alpha_L}, \mu_\beta, \sigma_\beta\}.$$

where G denotes a gain and L denotes a loss. Theoretically mapping to different update rules in light of positive and negative feedback, a phenomenon well-studied in the psychological literature (Parr, Rees, and Friston, 2018).

As with all furthermore additions, it is assumed the additional parameters are sampled from *Normal* population distributions, resulting in the individual subject parameters:

$$\begin{aligned} \alpha^s_G &\sim \mathcal{N}(\mu_{\alpha_G}, \sigma_{\alpha_G}) \\ \alpha^s_L &\sim \mathcal{N}(\mu_{\alpha_L}, \sigma_{\alpha_L}) \\ \beta^s &\sim \mathcal{N}(\mu_\beta, \sigma_\beta). \end{aligned} \tag{3.6}$$

This exactly matches the model used by Slooten, Jahfari, and Theeuwes, 2019 in their probabilistic choice task.

Linear component: Additional to the *learning* and *observation* models, we introduce an additional component in formalising the model architectures, the *Linear model* $\psi(x)$. Let the linear model be the link between additional covariates and observation model. Thus far, it has been superfluous to formalise this link function as it has essentially been an identity function such that:

$$\psi = Q^s_t(A); \;\; \pi^s_t(A = a) = \zeta\left(\beta^s \psi\right).$$

This general framework allows one to manipulate the linear model directly to capture additional covariates before multiplying by the β^s coefficient.

Bio-model 2: Action bias: Some research shows that subjects may exhibit state inertia or bias when performing probabilistic tasks (Huys, 2013). Inertia is highly unlikely given our experimental design, as it is antithetical to the task rules; however, it is plausible to imagine that subjects may exhibit some state-space bias. If a subject understands the rules of the WCST, they should exhibited no intentional inertia.

It is easy to imagine a situation where inertia is applicable: consider the exact same WCST as ours, but where the feedback is probabilistic. In such a setting, the subject would not have full confidence in negative feedback (as negative reward may simply be a consequence of probability) and thus subjects may exhibit some inertia (sticking to current choice probabilities). Our WCST's feedback is not probabilistic, and as such subjects should know that negative feedback can only result from a matching rule change - thus, no intentional inertia is expected.

A bias, on the other hand, could adequately represent a situation where subjects tend to try some actions over others (in absence of additional information). For example, when a

subject receives negative feedback, they can be certain that their current choice is incorrect but have no information about which action should be sampled next. A bias could represent the tendency to sample some options over others. A bias can be captured as an additional stationary parameter in the linear model (Huys, 2013).

$$\psi = Q_t^s(A) + \upsilon^s(A).$$

Adding a bias per subject requires the additional population parameters: $\{\upsilon^s(A) \sim \mathcal{N}(\mu_\upsilon, \sigma_\upsilon)\}$. The likelihood is now defined:

$$\pi_t^s = \zeta\left(\beta^s \psi\right)$$

The above parameterisations constitute the *biological models*. After determining the parsimonious biological model we: **1.** Illustrate the potential utility of encoding neuropsychological covariates directly into the learning process (examining two covariates selected by their model free ranking); and, more critically **2.** Model the learning parameters as a function of neuropsychological and demographic covariates.

3.7.2 Psychological models

Neither the learning nor observation models require updating after selecting the biological model. Here, the linear model adaptations are described - it is assumed that both the linear and observation models are inherited from the chosen biological model.

We examine the utility of adding covariates to the learning process by assessing potential model improvements. All of the psychological covariates are numeric in nature, allowing one to directly transfer the hierarchical techniques used thus far.

Expanding the parameter space: Each additional covariate is assumed to be sampled from a normal distribution $\sim \mathcal{N}$ population prior. As a consequence each additional covariate cv results in $s+2$ additional parameters, where s is the number of subjects. s individual parameters and 2 additional population parameters.

$$cv^s \sim \mathcal{N}(\mu_{cv}, \sigma_{cv}).$$

The covariates in both the psychological and demographic models were chosen retrospectively, after performing the model-free analysis (section 3.5).

Letting ξ denote the chosen biological linear model, we select the WCST reaction time (RT) because it is the most informative psychological covariate - as ranked by our mutual information and F-statistic model free analysis - and build an additional model:

1. **Psy-model 1: WCST response time (RT)**: $\psi = \xi + rt_{wcst}^s$

The covariate is stationary and choice agnostic (independent of a) by nature as rt_{wcst}^s is the reaction time over the entire experiment. As before we determine whether or not this addition is warranted by computing the WAIC score of this addition and contrasting the result with the optimal biological model.

3.7.3 Model selection

We utilised the WAIC (see selection 2.10 for details) to select the optimal parsimonious model. WAIC approximates leave-one-out cross validation, as discussed in section 2.10.5,

making it a frequently used choice when comparing Bayesian hierarchical models (Zhang and Gläscher, 2020).

3.7.4 Computational limitations

When performing Bayesian analysis, some quintessential pragmatics arise that warrant mention.

Reparameterisation: the data generating process may constitute a plethora of constrained, abnormal, distributions - often imposing bounds with improper priors (Gelman et al., 2004). As such, model convergence can be problematic in an abnormal setting. Monte Carlo techniques rely on sampling from the posterior and perform far better under the assumption of simple (often unit normal) distributions (Silver, 2015). It is, therefore, often necessary to re-parameterise the model parameters to conform to simple distributions and, subsequently, transform the model to the desired parameter-space (Gelman and Hill, 2006). In practice, this requires a simple transformation of the data. Assuming $\tilde{\theta} \sim \mathcal{N}(0, 1)$, and given the desired permissible range, the parameters are transformed accordingly:

$$
\begin{aligned}
\theta \in (-\infty, +\infty) &: \theta = \mu_\theta + \sigma_\theta \tilde{\theta} \\
\theta \in [0, N] &: \theta = Probit^{-1}(\mu_\theta + \sigma_\theta \tilde{\theta}) \times N \\
\theta \in [M, N] &: \theta = Probit^{-1}(\mu_\theta + \sigma_\theta \tilde{\theta}) \times (N - M) + M \\
\theta \in (0, +\infty) &: \theta = exp(\mu_\theta + \sigma_\theta \tilde{\theta}).
\end{aligned}
$$

In fitting the hierarchical models, we employed this re-parameterisation technique to achieve model convergence.

Monte Carlo sampling limitations: sub-sampling: The NUTS (No-U-Turn-Sampling) variant of MCMC is used fit the Bayesian RL models (Homan and Gelman, 2014).

When contrasted with that of the literature, our sample is far larger than the norm. With over 270 subjects, our sample exceeds 10× the sample size of many comparable studies (Slooten, Jahfari, and Theeuwes, 2019). Although allowing for extremely generalisable results, this scale is not without problems. These deeply hierarchical Bayesian methods rely on statistical properties that are antithetical to scale. In particular the Monte Carlo optimisation procedures used to estimate parameters fail to scale to large dataset (an active area of research in Bayesian inference) (Gelman et al., 2004). Bearing in mind that each subject executes 100 trials, a dataset of 270 subjects translates to $270 \times 100 = 27000$ actions-rewards pairs. This scale causes divergence (failing to converge to stable estimates as the chain continues to explore the search space) of MCMC methods in even simple model specifications, which is only exacerbated in the face of multiple covariates.

We took the decision to reduce the sample size in order to continue with the analysis. Two sub-samples are generated: the first takes the top performing 100 subjects; the second samples 100 subjects at random. This decision was taken to: assess the learning properties of the most attentive subjects, most likely indicative of clean data, and in parallel assess the learning properties of a random set to mitigate cherry picking.

These sample sets are still much larger than that of the standard. We intend to scale the analysis in future work, which will most likely translate to specifying a model that optimises latent space variables with Variational Inference (the approach taken to scale text based Bayesian methods in recent literature) (Sutton and Barto, 2018).

3.8 Analysis of learning parameters

The final component of our analysis is concerned with predicting the individual learning characteristics. The hierarchical Bayesian RL model describes a subject's ability to perform the associative learning task. The parameterisation of the chosen model maps to abstract cognitive processes (like learning rates α and exploratory tendency β). Similar to the approach taken by Slooten, Jahfari, and Theeuwes, 2019 and Barcelo, 2020, we are interested in mapping these cognitive quantities to executive functions and/or demographics.

Parameters of interest: We are interested in the biological parameters and their associated variances. In the event that the chosen model includes additional psychological or demographic covariates, they are ignored. This approach was taken by Slooten, Jahfari, and Theeuwes, 2019 to map non-invasive metrics (in their case pupilometry) to a cognitive process (exploratory-exploitative tendencies captured by the β parameter). Here we examine the possibility of making a similar mapping by examining the correlations associated across learning parameters and executive functions.

Correlation analysis: The same techniques used to understand the correlation between average WCST performance y^s in section 3.5 are directly applicable here. We examine all parameters with respect to linear correlation, mutual information and ensemble covariate ranking methods detailed in section 2.12.

GAMs: In light of robust correlative relationships, one may wish to quantify the predictive power of some executive function on the learning parameters by building a General Additive Model (Hastie, 2001). As discussed in section 2.13, GAMs offers a flexible, non-linear but still interpretable modelling framework, making them ideal for this application. Unfortunately the GAM was not supported, given insufficient relationships in the data.

In the next chapter we elucidate the findings of our experiments, working through the above four sections of analysis (3.5, 3.6, 3.7 and 3.8) to uncover the associative learning process.

Chapter 4

Results

In this chapter we showcase our analysis. After detailing the sample demographic information in section 4.1, we remove outliers from the WCST dataset (section 4.2) and assess the viability of compressing the Navon task covariates into a single predictor (section 4.3).

Thereafter, the same structure used in the methodology is employed: section 3.6 uses simulated data to quantify the risk of improperly specifying hierarchical models; section 3.7 fits a set of hierarchical RL models to behaviour (WCST) data; and section 3.8 elucidates the potential links between psychological covariates and learning characteristics.

4.1 Sample demographics

The two sub-samples contain 100 participants each, containing the best performing subjects and a random set (referred to as the *top* and *random* sample respectively). Although sampled differently, the distributions of demographic information appear similar in both sub samples, as shown in figure 4.1. *Age*, *demographic reaction times (RT)*, and *computer hours* are all positively skewed.

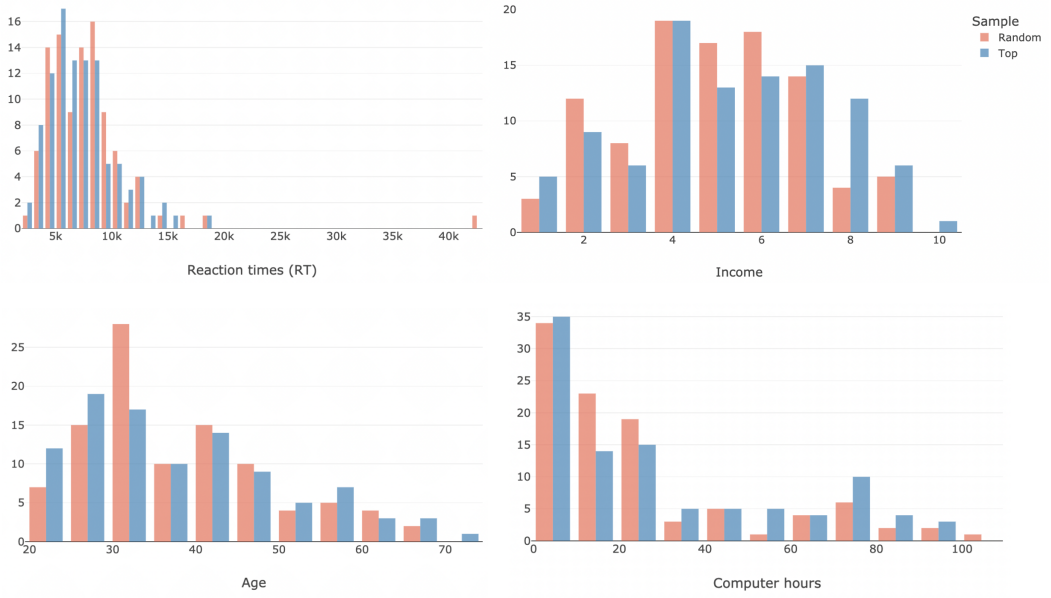


Figure 4.1: Demographic distributions of the two samples.

4.2 WCST outlier removal

As discussed in 3.3, we classify subjects who score under an average WCST accuracy of $\lambda_{wcst} = 0.4\%$ as outliers. These subjects are removed from the dataset. The distribution of subjects' average performances is illustrated in figures 4.2.

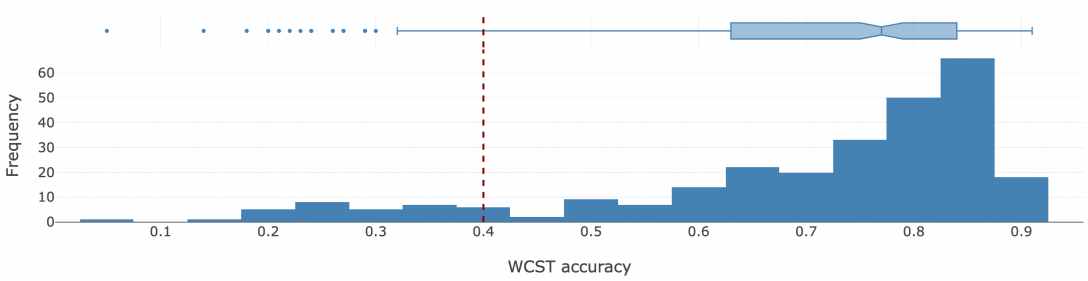


FIGURE 4.2: Distribution of WCST aggregate performance scores used to identify outliers. A subject acting randomly would (on average) achieve a score of 0.3. Marked with a red dotted line, 0.4, is the threshold we employed to remove outliers. The boxplot above the distribution highlights outliers (single points) as well as the $25th$, $50th$ and $75th$ percentiles corresponding to the start, middle line, and end of the box respectively.

.

These subjects' data were removed from the analysis and not considered moving forward.

4.2.1 Navon task data compression

The next step of the analysis was to determine the viability of compressing the Navon task data. Reducing the number of covariates is always favourable, given the principle of parsimony. If there is no meaningful statistical difference between the different Navon task subgroups, the data can be compressed to a single covariate to represent attentiveness.

Recall from section 3.3.2 that the Navon task measures one's ability to accurately identify local and global patterns in visual stimuli. With respect to the Navon task, our study is interested in the relationship between one's global and local attention (measured by the Navon), and one's ability to perform the associative learning task (WCST). If there is no statistical difference between global and local attention performance, compressing these into a single covariate that represents attentiveness is justified as no information is lost.

The Navon performance scores are computed as the percentage of correct actions taken in each respective subclass. More specifically, the *global* Navon performance score is the percentage of times the subject correctly identified a global signal (and similarly for *local* scores). The Navon scores labelled *none* represent the actions taken in absence of both global and local patterns - both were included for completeness and as a control.

As seen in figures 4.3 and 4.4, visually, the distributions of Navon performance scores are very similar. Figure 4.3 plots the WCST performance (y-axis) against the Navon performance scores (x-axis) - revealing no obvious pattern. The Navon performance distributions (over all subjects) are plotted in figure 4.4. While the *none* category is slightly more positively skewed, visually the *global* and *local* do not appear substantially different.

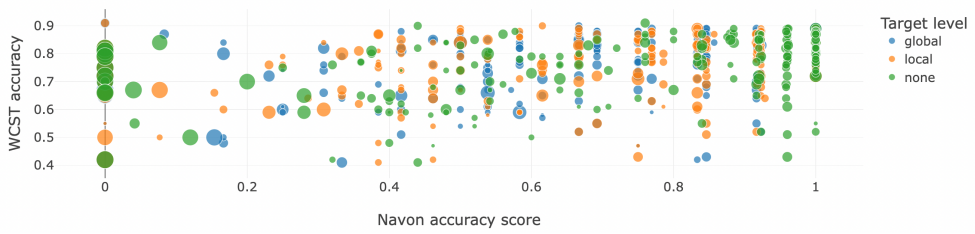


FIGURE 4.3: WCST performance as a function of different Navon class scores. The task requires the participant to identify patterns in both global and local settings, as well as an additional null state labelled "none". The graph plots a point per subject, placing their Navon score on the x-axis (coloured by Navon class), and average WCST performance on the y-axis. The size of the point represents how long the candidate took to complete the Navon task (the Navon task reaction time). No visual pattern emerges.

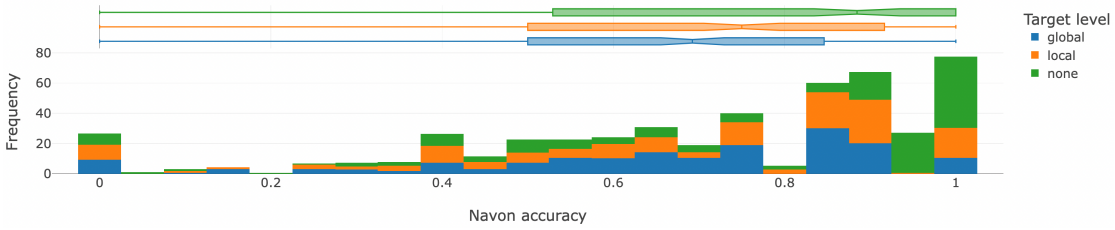


FIGURE 4.4: Distribution of Navon scores over the population of subjects. Although the "none" category is positively skewed, it is the task default behaviour and is merely included for completeness. There does not appear to be a meaningful (visual) difference between the "local" and "global" groupings.

The fact that there is no visual discrepancy between the Navon performance groups suggests that combining the covariates is justified. Before doing so, however, one should perform some statistical or correlation analysis to test these assumptions. Table 4.1 supports these findings statistically by measuring the relationship between the three Navon task groupings across a number of statistical metrics: Pearson correlation, T-test p-value and mutual information.

The T-test is used to test whether two groups have significantly different means. It is worth noting that we have not taken measures to ensure the t-test assumptions (normality and equal population variance) are satisfied, and, consequently, these t-test results should be considered loose approximations that may solicit further investigation and final results. This serves the purpose of our preemptive analysis.

All three groups are highly correlated with the *local* and *global* groups exhibiting the highest correlation of 0.73. The same pairing's means do not differ significantly, with a t-test p-value of 0.41. The Mutual information (distributional overlap) is relatively high for all pairings. In light of these figures, it is justified to combine the covariates into a single covariate.

Navon Task Accuracy						
	Pearson Correlation		T-test		Mutual Information	
	Local	Global	Local	Global	Local	Global
Global	0.73		0.41		0.44	
None	0.60	0.69	0.00	0.01	0.39	0.44

TABLE 4.1: Statistical analysis to measure the relationships between different Navon groupings. Only one set of off-diagonal figures are reported as values on the diagonal offer no meaning, and values are mirrored on the diagonal. The table can be broken down into three sections reporting Pearson correlation, T-test p-values, and Mutual Information over the three Navon task categories. *local* and *global* groupings exhibit very high correlation (0.73), no significant difference between means (t-test p-value= 0.41) and high Mutual Information (0.44). These figures support the idea of combining the covariates into a single feature.

Since we have determined that combining the covariates is warranted, we compress the individuals' Navon scores by simply computing their respective Navon accuracy averaged over all groupings (local, global and none).

The Navon task, however, emphasises the importance of capturing global vs local attention and this information should not be dismissed without further examination. To this end, an additional covariate (labeled *Nv global-local* in table 4.2) is created that is computed by the difference between a subjects average performance on the global and local groupings of the Navon task. Now that our covariates have been finalised we perform a preemptive statistical analysis between the covariates and the WCST average performance.

4.3 Covariate prioritisation

Now that outliers have been removed and the Navon task data has been compressed, we examine the statistical relationships between the covariates and WCST average performance. As detailed in section 3.5, we examine the relations with a number of metrics to adequately represent both linear and nonlinear relationships. This preemptive analysis offers insight into both approximate information, and serves as a basis for adding covariates to the Reinforcement Learning model in the event that additional covariates are warranted.

F-tests, Mutual Information, RFCQ, and FCQ were employed to measure covariate relevance - all of which are detailed in section 3.5. Each of the above methods are applied to our summary datasets. The response vector is the average *wcst* score earned by a participant over the entire trial. The results, available in table 4.2, are used to select and rank features. The **F-test** and **MI** are the primary variables in our consideration. mRMR provides supportive information - capturing redundancy, and is included to indicate potential overlap between covariates.

WCST Score Correlation Analysis							
Covariate		data type	*F*-statistic	*p*-value	MI	mRMR (rfcq)	mRMR (fcq)
Psy	wcst RT	float	111.38	< 0.01	0.19	4	0
	Fitts	float	31.37	< 0.01	0.09	8	4
	NBack	float	60.95	< 0.01	0.07	7	2
	NB RT	float	5.31	0.02	0.04	5	1
	Navon	float	25.09	< 0.01	0.01	9	6
	Nv global-local	float	0.99	0.32	0	12	14
	Nv RT	float	5.05	0.03	0	13	12
	Corsi	float	20.72	< 0.01	0	15	3
Dem	Dem RT	float	5.95	0.02	0.16	1	10
	Age	float	2.02	0.16	0	10	13
	Comp hours	float	5.24	0.02	0	14	9
	Handedness	object	1.52	0.23	0	0	5
	Education	object	3.02	0.05	0	2	11
	Gender	object	1.64	0.2	0	3	8
	Age group	object	0.48	0.79	0	6	15
	Income	float	9.47	< 0.01	0	11	7

TABLE 4.2: Covariate feature selection metrics. Covariates are separated into *Psy* and *Dem* data corresponding to *neuropsychological* and *demographic* respectively. The p-value column (p-value associated with F-test) is marked green if significant at a 5% level. Mutual Information (MI, a purely relative measure) is marked green if it exceeds 0.10. The two columns providing the results of the *mRMR* feature ranking mark the features 0-to-5 as green and 6-to-10 as blue. The *data type* column refers to the data structure used to encode the information.

Demographic covariate analysis: Although less correlated with WCST than the psychological covariates, some statistical relationships are observable in the demographic covariates (as detailed in table 4.2).

Handedness, *education level*, *gender*, and *age* are not significant in the F-test nor produce meaningful MI results. *Computer hours*, *Demographic reaction time*, and *Income* appear to have a significant linear relationship with the WCST.

Notably *Computer hours* and *Demographic reaction time* likely capture the same underlying generative process: the former being a self-described account of how frequently one uses computers, but the latter capturing the time taken to complete the demographic information which is likely to constitute attentiveness, alertness, and how comfortable one is with the machine on which they took the task.

Demographic reaction time boasts a lower p-value, offers a substantially larger MI, and ranks highly on the RFCQ variant of mRMR feature ranking. For this reason, reaction times are included (assumed to incorporate computer hours) as an additional covariate in the hierarchical RL model.

Three of the demographic variables, that show no significant relationship with WCST, rank highly on the RFCQ variant of mRMR. One may falsely assume spurious correlation if one fails to consider how random forest decision trees are fit. Instead, it is highly likely that the discrete nature of these covariates result in inflated importance in the compilation of

decision trees, as discrete variables have a much smaller permutation space when compared with the continuous variables. We do not feel that these findings circumvent the negligible F-test and Mutual Information scores, and, consequently, these variables are not considered highly correlated with the WCST average performance.

Neuropsychological covariates: As exhibited in table 4.2, most of the neuropsychological covariates are linearly related to the WCST. This may serve to illustrate the dependence between one's neuropsychological attributes and operant/associative learning abilities.

The *Navon global vs local* variable yields no relationship with the WCST response, thus supporting our prior analysis. While producing significant F-statistic p-values, both *Nback* and *Navon RT* (reaction time) score very low mutual information. *Navon RT* ranks low on both mRMR variants. *Nback RT* can be assumed subsumed by the *WCST RT*, as it scores lower on all metrics - it may be excluded as multiple *RT* (reaction time) variables are superfluous because they represent much of the same information. All remaining variables (*Fitts, Nback score, WCST RT and Corsi*) show significant linear relationships with the WCST - a promising finding detailing the potential relationships between these neuropsychological phenomena and associative learning.

In the next subsection we move away from the neuropsychological data and examine the importance of properly specifying hierarchical models to represent variation in a population.

4.4 Simulating RL models

In this section we simulate a number of Reinforcement Learning models to examine the properties of our chosen paradigm. Simulating data allows one to assess the expected behaviour of a system. If the true underlying process is known, one can gauge the model's ability to recover the data generating process. It also makes the assumptions of the system more explicit, which is important before drawing inclusion on models fit to real data.

First, we simulated an RL model that captures a sequence of choice behaviour of a single subject and, thereafter, we simulated an entire population of subjects. The simulation contains 100 actions, as this corresponds to the WCST. One would expect, however, a longer sequence to have exhibited better results as more information is supplied.

4.4.1 Single subject RL simulation

Data generating process: As detailed in 3.6.1, we simulated an action sequence of a single agent that updates their value estimates with the standard RL formulation:

$$Q_t(a) = Q_{t-1}(a) + \alpha \left[r_t - Q_{t-1}(a) \right]$$

where $r_t \in [0, 1]$. The true (simulated) parameter estimates are chosen to be:

$$\alpha = 0.4, \beta = 10.$$

These parameters are used to generate the data (the action choice sequence and update rule). This sequence is visualised in figure 4.5.

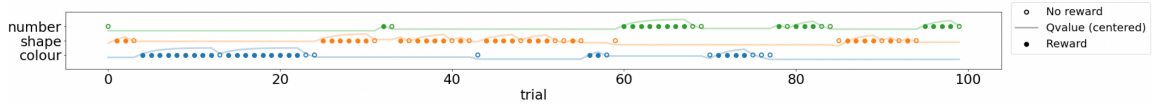


FIGURE 4.5: The data generating process can be visualised as a series of actions and corresponding feedback received as a function of time. Here, we visualize the simulated data of the three choices (y-axis) and represent the actions taken as nodes over time (x-axis). A node is coloured if the (probabilistic) rewards were positive, and transparent if otherwise. The three lines tracing horizontally through each choice represents the subjects' current $Q_t(a)$ state value approximation.

.

RL simulation results: After assessing the convergent properties of the Monte Carlo Markov Chain, we are confident in the procedure's ability to recover the true parameter estimates.

The posterior mean of both parameters are near their true counterparts, with the estimated values taken as the posterior mean: $\hat{\theta} : \{\alpha = 0.36, \beta = 11.22\}$; which statistically approximate the true values $\theta : \{\alpha = 0.40, \beta = 10.0\}$.

The model was fit via MCMC (Hastie, 2001). The approximated posterior distributions over the parameter space can be observed in figure 4.6.

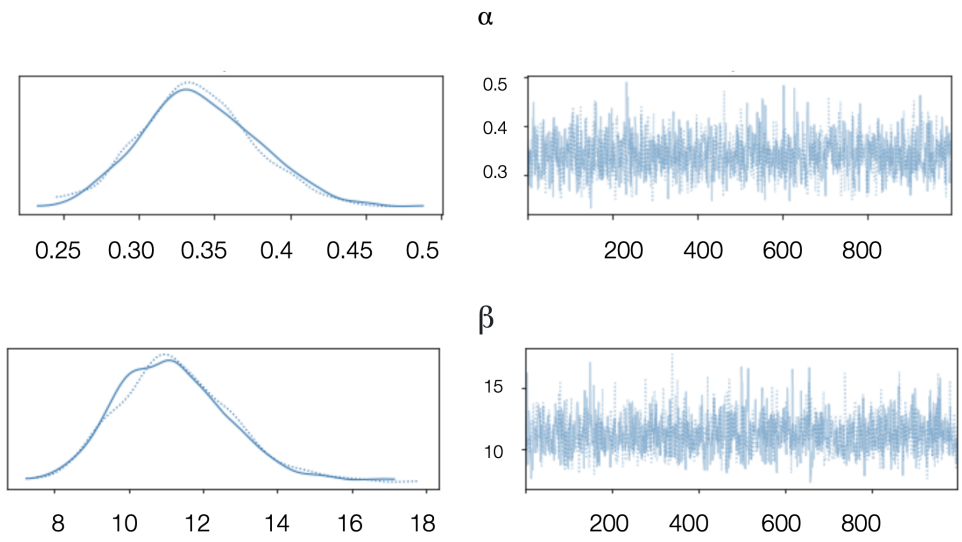


FIGURE 4.6: Parameter posterior distributions (left) and trace plots (right) of the learning model parameters α and β shown in the top and bottom plots respectively. The trace plots show the Monte Carlo posterior sampling procedure over 1000 samples. The model has converged near the true parameter values. The lighter blue lines on the posterior plots (left) show the posteriors reached during the burn-in period.

Assessing the convergence properties: To ensure statistical robustness and generalisability, it is important to test the system over a range of plausible models as well as incorporating extreme values. We repeated the above experiment over a set of combinations in the parameter space: $\alpha : \{0.05, 0.1, 0.25, 0.5, 0.75, 1\}$ and $\beta : [1 : 11]$, testing each parameter combination.

Figure 4.7 and 4.8 summarise the results of these experiments by plotting the true (data generating) parameter values on the $x - axis$ and corresponding the estimated values on the $y - axis$.

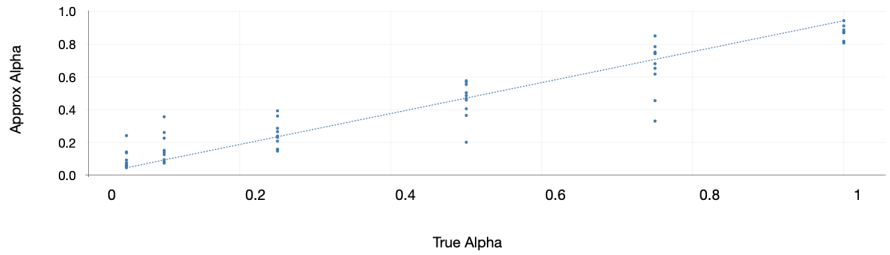


FIGURE 4.7: Parameter sensitivity analysis: In order to assess the robustness of this Bayesian fitting procedure, we simulate a wide range of both α and β values. The graphs plot the true data generating α on the x-axis and the estimated value on the y-axis. The $y = x$ line represents a perfect model estimate. We observe that the model parameters appear somewhat clustered around the true value, offering sufficient confidence in the technique; however, they are clearly skewed towards or tempered central values. The same experiment was conducted in the β parameter, as shown in figure 4.8. The skewness of estimates around extreme values is likely a consequence of sampling within the permissible range, as enforced by the prior. For this reason the model overestimates low values and underestimates for high values- this is observed in both the α and β (figure 4.8) parameters. This skewed behaviour is expected as the prior biases the results. Each MCMC ran for 1000 iteration after the burn-in period.

.

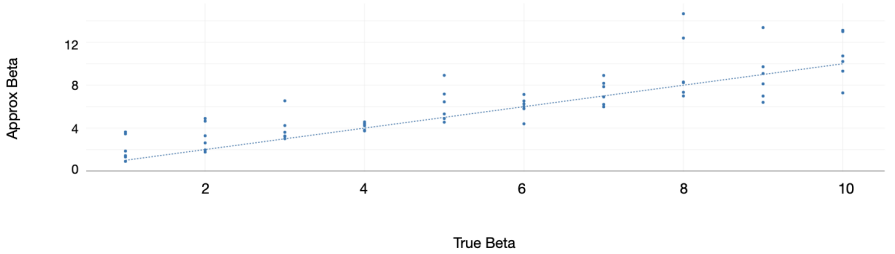


FIGURE 4.8: Plotting the estimated β values (y-axis) against the true data generating β parameters. As observed with the α estimates, the model tends to over estimate very low values but under estimate high values - a consequence of the regulating priors.

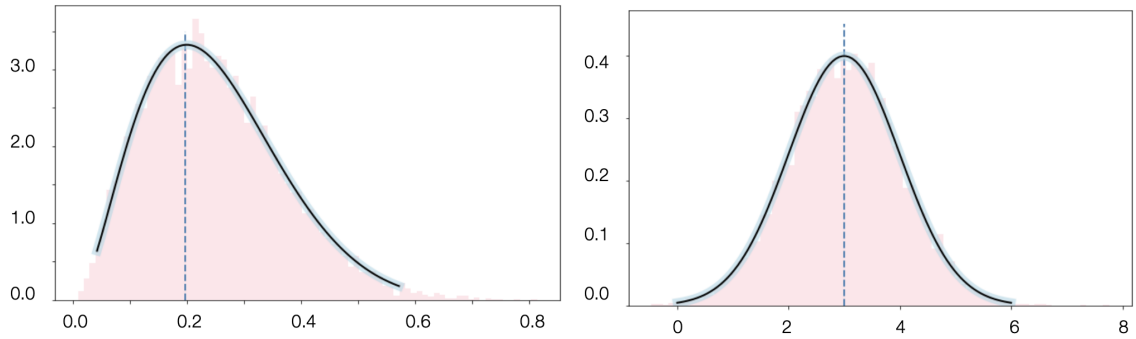
The estimates appear biased by prior, however, they cluster around the true data generating parameters and exhibit marginal variance over their true counterparts- we can infer that the MCMC fitting procedure is robust and applicable over a wide range of plausible parameter values, although it may be problematic over the edges of the permissible/theoretical range.

Generalising these results: These simulated results suggest this fitting technique may be extrapolated to a multitude of model variants; adding explanatory covariates, transforming the decision space non-linearly by a set of basis functions, utilising different prior (regularisation) schemes, engineering a better learning equation, et cetera. The fundamental approach remains unchanged and, thus, serves as proof of concept of many plausible variants. One salient statistical idea, however, is not captured by this process - that is

hierarchical variance pooling, requiring an additional simulation to assess hierarchical convergence properties.

4.4.2 Hierarchical population RL models

Here, we repeat the above experiment but for an entire population of subjects. As described in section 3.6.2, we simulate a population of subjects performing the same task, drawing their true generating parameters from the population distribution (visualised in figure 4.9). We then aim to recover the population distributions from the observable choice data.



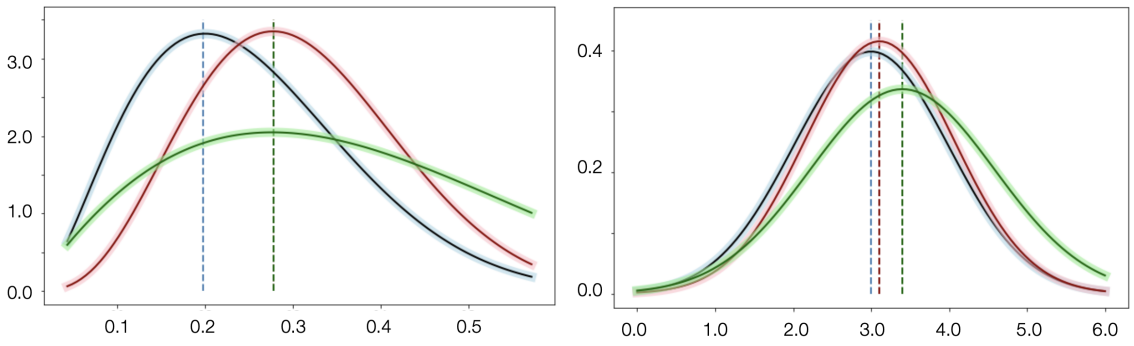
(A) $\alpha \sim Beta(3, 9)$ (B) $\beta \sim \mathcal{N}(3, 1)$

FIGURE 4.9: Hyperpriors governing the data generating process of the hierarchical model. Each individual subject's learning parameters are stochastically sampled from these population distributions.

Summary statistics approach

Before constructing the complete hierarchical Bayesian model, we implement - for the purpose of comparison - the naive summary statistics approach to estimate population parameters.

This approach fits n individual models - without considering mutual information or variance - and thereafter, assumes the individual sample estimates can be used to fit the population distributions. Figure 4.10 exhibits the results of this method. As expected, detailed in the literature (Daw, 2011a), the estimates are unbiased (approximating their optimal values when accounting for random fluctuations) but may exhibit high variance (particularly in the case of smaller samples).



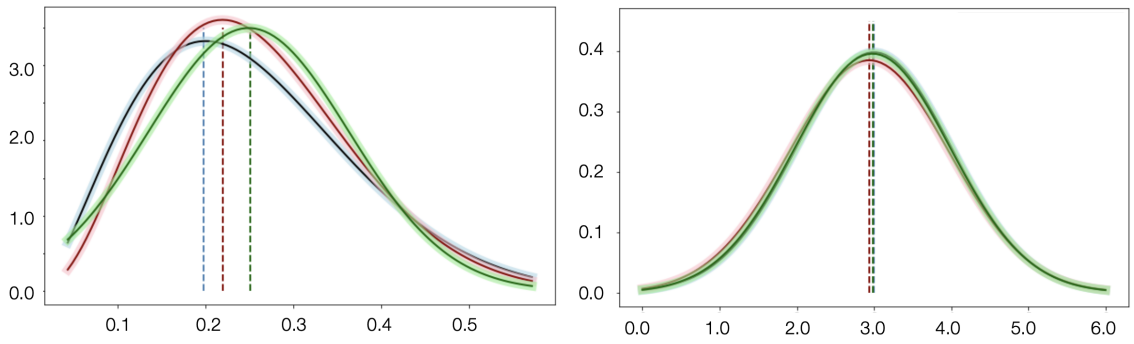
(A) $\alpha \sim Beta(a_\alpha, b_\alpha)$ (B) $\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta)$

FIGURE 4.10: Population models when taking the summary statistics approach. The blue curves detail the true (unknown) population distributions, for α and β respectively. As assumed in the data generating process, the individual parameters are drawn from these population priors. The red curves show the distribution fit to the *actual* (unknown) subject sample parameters. These values show the optimal model, as deviation from blue to red is a function of the stochasticity in the data generating process, and thus cannot be minimised. That is, if each individual subject's parameters were recovered perfectly, the red curve would be fit to the data. The green curves represent the models fit by the summary statistics approach. Both of which visually appear unbiased, and, while the β estimate appears to map extremely well to the optima, the α estimate does exhibit inflated variance. Note that the mean posterior estimate (shown as the vertical dotted lines for each distribution) are the parameter estimates. The summary statistics approach (green dotted line $\hat{\alpha} = 0.283, \hat{\beta} = 4.43$) very nearly corresponds to the best possible fit (red dotted line $\alpha = 0.284, \beta = 3.09$).

Though mathematically improper (as the variance in α is insufficiently captured, figure 4.10), it is easy to make a case for this approach. Its simplicity allows the statistician to readily *stack* subject models linearly and achieve reliable results. Additionally, the mean posterior estimates closely map to the true data generating values. These mean posterior estimates are used to point estimates when recovering parameters; therefore, if the research question at hand does not concern itself with variance this method should suffice. Many human-centred RL research questions are not concerned with variance but only learning rates α and the mapping to the exploratory trade-offs β. It is, nonetheless, statistically flawed. To illustrate why, we now contrast these results with a full Bayesian Hierarchical model.

Bayesian hierarchical model

Using the same data, we fit a full Bayesian Hierarchical model, as discussed in section 3.6.2. This technique assumes some joint data generating process allowing for shared variance across the population. The population level parameters serve as regularising priors that tighten the variation across subjects. Figure 4.11 shows the results obtained by specifying a full hierarchical model.



(A) $\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha)$ (B) $\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta)$

FIGURE 4.11: Similar to figure 4.10, the blue line indicates the true data generating distribution and the red line indicates the best possible model fit. The red line is achieved by fitting a distribution to the random samples. Discrepancy between red and blue is due to random sampling and cannot be minimised. The green line indicates the population level parameters achieved by the model procedure. It is evident that instantiating the hierarchical model nearly perfectly recovers the true (unknown) population parameters. There is only marginal deviation in the means (dotted lines) and great overlap in the variance as seen in the shape of the posteriors.

Despite uninformative priors, and assessed visually, the hierarchical model was able to adequately recover the population parameters. Note, that even the α parameters are recovered despite assuming a normal distribution - despite the fact that the data generating process is governed by a *beta* distribution, showing the reliability of the methodology. This fit is achievable by imposing an improper prior over the permissible range $\alpha \sim uniform(0, 1)$, and is favoured as it boasts great MCMC convergence properties (Gelman et al., 2004), as well as being less biased in practise (Knill and Pouget, 2004).

This simulation confirms the theoretical advantages of the hierarchical approach by offering empirical support for the methodology, and thus will be utilised moving forward. The simulation also illustrates how the Bayesian formulation correctly accounts for non-linear models (as with our RL model), whereas most theoretical applications focus on linear models.

4.5 Cognitive science RL models

Returning to our experimental data, we then fit the Bayesian hierarchical RL models discussed section 5.5. RL Models are compared using WAIC (a generalization of AIC that inherently accounts for parsimony detailed in section 3.7). The results are summarised in 4.3.

Model Results				
Parameters			WAIC	
Model	Biological Parameters	Psychological Parameters	Top 100	Random Set
null model	$\mu_a, \sigma_a, \mu_\beta, \sigma_\beta$		-10346.50	-12824.48
bio model 1	$\mu_{a_g}, \sigma_{a_g}, \mu_{a_l}, \sigma_{a_l}, \mu_\beta, \sigma_\beta$		-10002.27	-11695.95
bio model 2	$\mu_a, \sigma_a, \mu_\beta, \sigma_\beta, \mu^a_{bias}, \sigma^a_{bias}$		-9617.90	-11419.54
psy model 1	$\mu_a, \sigma_a, \mu_\beta, \sigma_\beta$	$\mu_{wcst_{rt}}, \sigma_{wcst_{rt}}$	-10345.50	-12760.20

TABLE 4.3: WAIC scores are reported for both sample sets (using the top scoring candidates and random set of candidates). Models are classified by their constituent covariates, either containing purely biological parameters or additional psychological covariates. The model producing the lowest WAIC in both samples is the simplest configuration - denoted the null model.

4.5.1 Model selection

Biologically inspired models: Contrasting the three biological models, WAIC scores are relatively close between models with the simplest null model producing the lower score (WAIC scores reported in table 4.3). The appropriate biological model is that which best captures the out of sample estimate and, consequently, is most likely to be indicative of the true data generating process. Highly intertwined with task complexity, the nature of our learning task appears not to solicit the additional, superfluous parameterisation that encode inertia, bias, or alternate learning rates - as discussed in section 5.5.

Parameter extensions: The null model - containing four population level parameters $\theta_{pop} : \{\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta\}$ was then amended to directly incorporate a psychological parameter to assess the utility of adding covariates directly into the learning model. The premise of this specification rests on the idea that the variation in subject psychological attributes may inform a subject's associative learning, driving choice behaviour.

As reported in table 4.3, the additional complexity of adding an additional hierarchical distribution did not produce a lower WAIC score than the null model. The *top* sample produced roughly the same score, and the *random* sampled produced a worse score. The null model is the best configuration and is used in the analysis moving forward.

4.5.2 Convergence properties

As with all model configurations tested, the parameters were recovered by implementing the NUTS algorithm (Homan and Gelman, 2014). Two independent Monte-Carlo Markov Chains of 2000 runs each (with a burn-in period of 1000 samples) were fitted and converged to stable cohesive results. Almost all parameter estimates converged to an $Rhat = 1$ - bar a few individual estimates that neared 1.1 - indicative of a stable Bayesian model. Visually, the recovered posteriors appear stationary, having converged to appropriate stable estimates. Importantly, these estimates coincide with theoretically plausible quantities: better performing subjects exhibit higher learning rates α^s, greater exploratory parameters β^s (showing the ability to adapt to new information) as well as greater consistency in their behaviour (reduced posterior variation).

4.5.3 Posterior checks

Population posterior distributions: The population posterior distributions are estimated by fitting the MCMC procedure to the data, plotted in figures 4.12. Consistent with theory, subjects with greater learning and exploratory rates perform better on the dynamic decision making task. They also exhibit less variation in their estimates, proving consistent over many trials and stability in their choice behaviour. The broader random sample not only achieved lower estimates, but also inflated and skewed variation in the posterior estimates.

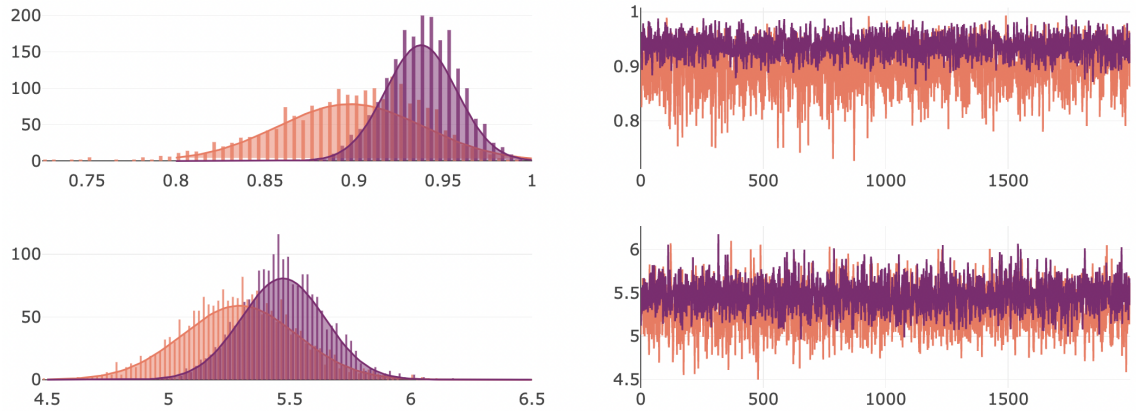
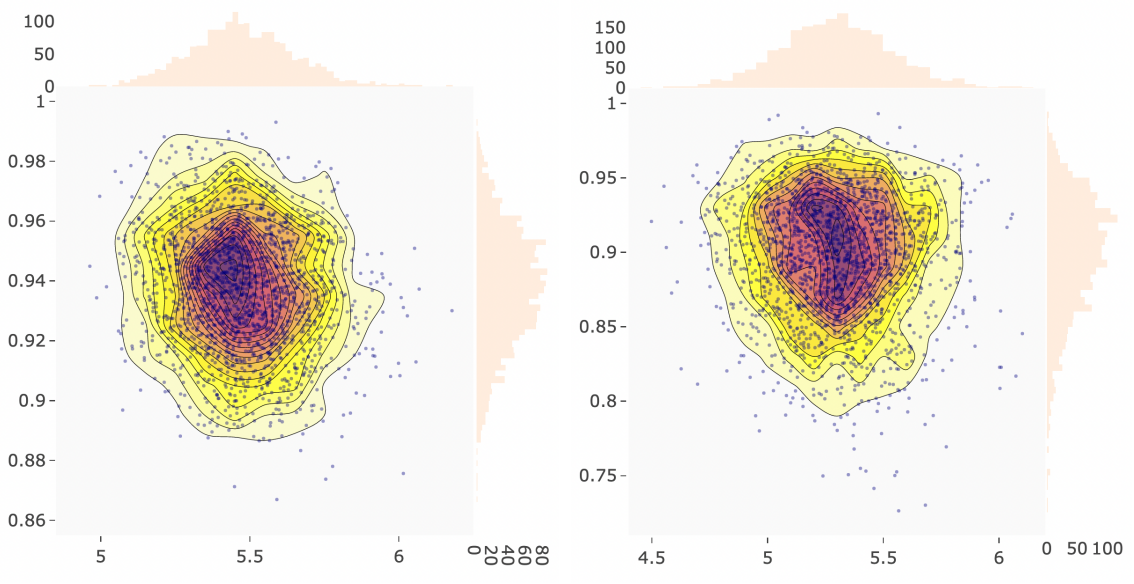


FIGURE 4.12: The recovered population posterior distributions. *Top* and *bottom* show the α and β population distributions respectively. The *orange* and *blue* showing samples from the random and best sub-sample respectively. Samples appear stationary with hierarchical parameters converging to stable distributions. As expected, the random set exhibits greater variance as well as having converged to a lower parameter estimate for both learning α and exploratory β parameters: the best performing candidates, by definition, exhibit faster learning (higher mean α values), so it is important to have reflected this in the model. Normal distributions are recovered from the population posterior samples in order to quantify the learning and exploratory sufficient statistics.

The population distributions are the model estimates of the generating process that produces the individual subjects' learning parameters. As discussed in section **??**, they serve as regularising priors over individual parameters. Before investigating the individual parameter estimates, it is important to examine any relationships between these population distributions - more specifically addressing the inflated variance and scrutinising the dependency between α and β.

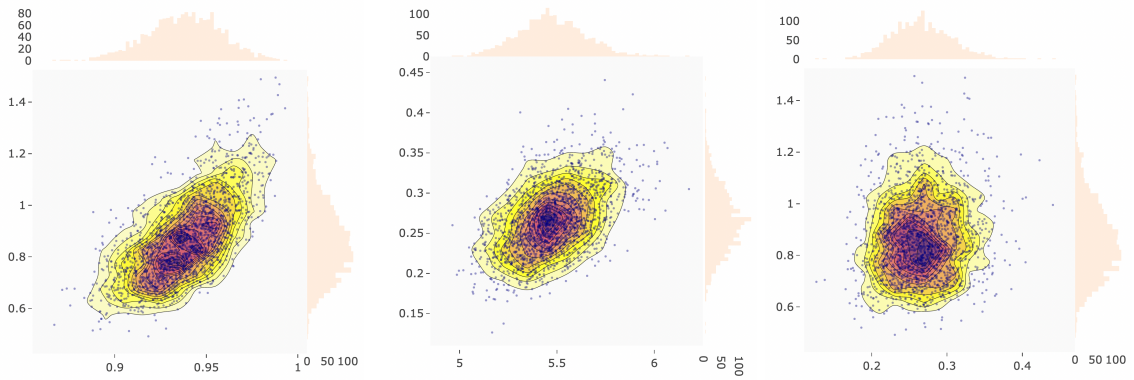
Interactions between learning parameters: The joint densities of the posterior estimates may allude to the interactive effects between parameter estimates. Figure 4.13 visualizes the joint posterior densities of the learning model parameters α and β on both the top (left) and random (right) sub-samples. Although we know that mathematically the variables encode some interaction - a consequence of the non-linear update equation - the sample from the best performing candidates appears relatively uniformly spread within some range of high learning rates ($0.88 \leq \alpha \leq 0.99$). Some marginal additional skewness towards higher α values is observable in the random set, as well as slightly higher β parameter values (possibly compensating for lower α pairings). Apart from this relatively minor skewness, there is no substantial visual relationship in the joint posterior samples. The statistical correlations between these parameters are discussed in table 4.4.



(A) $P_{top}(\mu_\alpha \wedge \mu_\beta)$ (B) $P_{random}(\mu_\alpha \wedge \mu_\beta)$

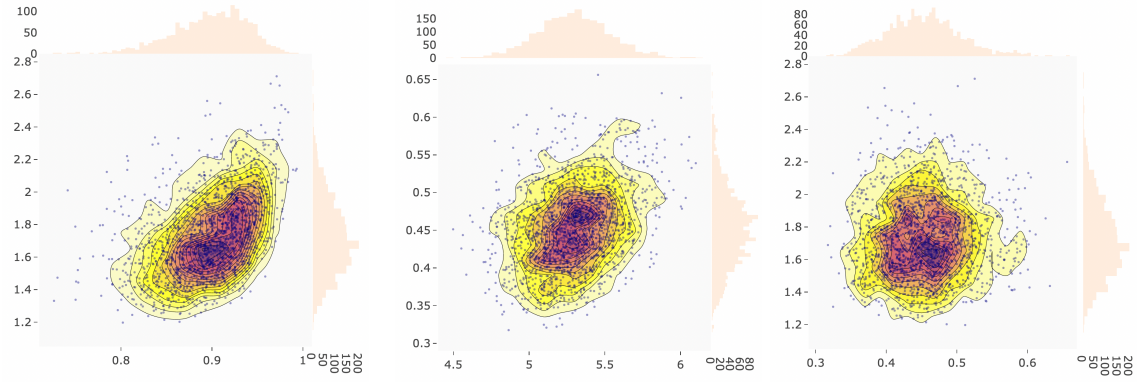
FIGURE 4.13: Joint posterior distributions of the mean learning model parameters μ_α (y-axis) and μ_β (x-axis) of model fit on the best performing sample (left) and random sub-sample (right) of candidates. The best performing set (left) appears very stable and somewhat uniform over a range of high learning rates $0.88 \leq \alpha \leq 0.99$. The random subject samples are positively skewed in α and generally more widespread over both α and β. The points represent samples from the posteriors and, therefore, there are 2000 samples in each plot.

If we now turn our attention to the posterior samples of not only the mean parameter estimates μ_α, μ_β but also the standard deviations of the learning parameters $\sigma_\alpha, \sigma_\beta$ more distinct patterns emerge. Figures 4.14 and 4.15 plot the posterior densities of a number of mean-standard deviation parameter combinations. Figure 4.14 provides plots of the model fit on the top WCST performing sample, while figure 4.15 provides the same plots but of the model fit to the random sample.



(A) $P_{top}(\sigma_\alpha \wedge \mu_\alpha)$ (B) $P_{top}(\sigma_\beta \wedge \mu_\beta)$ (C) $P_{top}(\sigma_\alpha \wedge \sigma_\beta)$

FIGURE 4.14: Using the model fit on the top performing sample we plot the posterior joint distributions of *(A)* μ_α and σ_α; *(B)* μ_β and σ_β; and *(C)* σ_α and σ_β.



(A) $P_{random}(\sigma_\alpha \wedge \mu_\alpha)$ (B) $P_{random}(\sigma_\beta \wedge \mu_\beta)$ (C) $P_{random}(\sigma_\alpha \wedge \sigma_\beta)$

FIGURE 4.15: The same plots as figure 4.14 are generated, but this time using the model fit to the random sample.

Inflated variance estimates: It is important to remember that these estimates are taken during the Monte Carlo sampling procedure. A point in figures 4.15, 4.14 and 4.13 represents a posterior sample, not a subject. Variance estimates σ_α and σ_β are estimated throughout this sampling procedure. The individual subject posterior means μ_α^s μ_β^s are used to recover the individual parameters α^s and β^s, and, therefore, the variance in the individual parameter posteriors are used to estimate variances. For this reason, the standard deviations estimated in figures 4.15 and 4.14 are inflated, but represent estimates during the sampling process.

Mean-standard deviation relationships: If we examine figures 4.14 and 4.15, similar distributions appear to emerge, although the linear relationships may be more refined in the *top* sample (figure 4.14). The statistical relationships between these parameters are quantified and tested in table 4.4.

Examining the top candidates, the posterior mean μ_α and standard deviation σ_α of the learning rate, α appears highly positively correlated. A positive correlation may also exist in the mean μ_β and standard deviation σ_β of the exploratory parameter β. The joint posterior of the parameter standard deviations σ_α and σ_{beta}, however, show no visible relationship.

Turning to the same plots on the random sample data (figure 4.15), similar relationships may be apparent but with much greater noise. In the next section, we extract the corresponding parameters and test these relationships statistically; however, it is worth highlighting that the strong correlation in mean-standard deviation pairs in the best candidate sample may coincide with subjects' confidence and consistency in their choice behaviours.

Now that we have examined the properties of the population (hierarchical) distributions during the Monte Carlo sampling procedure, we turn our attention to elucidating the differences in individual subjects' parameters.

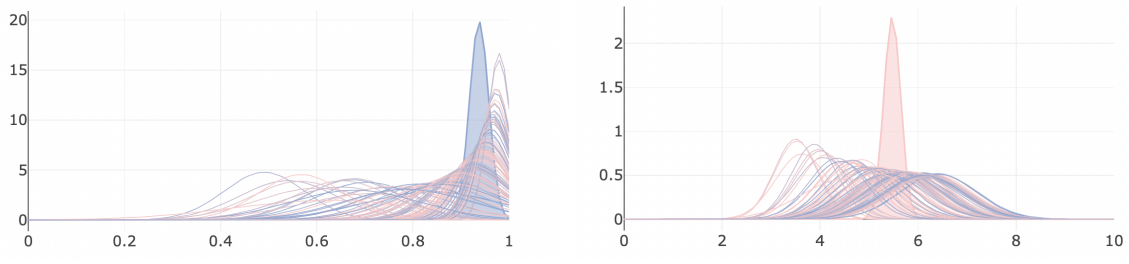
4.5.4 Recovering individual parameter estimates

The population parameters act as priors over the individual estimates, as detailed in section 3.6.2. We estimate the subject posteriors by fitting Gaussian distributions to the individual posterior traces - a viable approach given the apparent stationarity of the samples - that the process is the unconditional joint probability distribution that does not change as a function of time (Cover and Thomas, 2006). The Gaussians allow one to quantify the individual

learning characteristics $\mu_\alpha^s, \sigma_\alpha^s, \mu_\beta^s, \sigma_\beta^s$. Individual point estimates are taken as the mean of the posteriors.

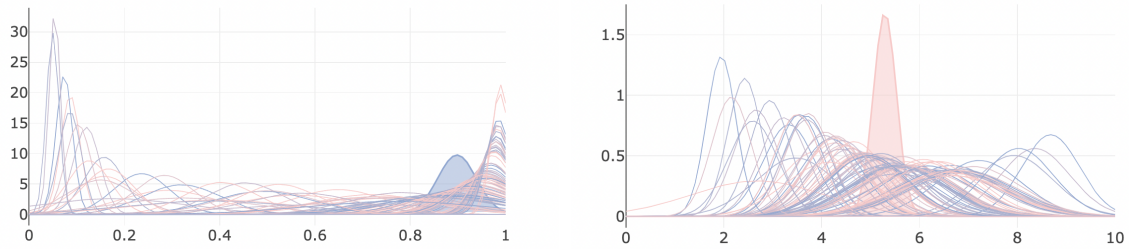
Regularising population distributions: Let's first examine the individual distributions with respect to the (regularising) population distributions. Figures 4.16 and 4.17 display the individual parameter distributions over the population distributions. In both samples there is substantial variation away from the population level parameters - indicating significant differences in individual task performance.

Whilst a spectrum of distributions emerge, the results could be further aggregated into a bi-modal or multi-modal distribution that captures variance in the cognitive learning process. This may indicate that a bi-modal or multi-modal Gaussian may be a more appropriate hierarchical prior (as in the case of Gaussian mixture models).



(A) $\alpha^s \sim \mathcal{N}^{top}(\mu_\alpha, \sigma_\alpha)$ (B) $\beta^s \sim \mathcal{N}^{top}(\mu_\beta, \sigma_\beta)$

FIGURE 4.16: Subject α^s, β^s posterior distributions in the best performing sample.



(A) $\alpha^s \sim \mathcal{N}^{random}(\mu_\alpha, \sigma_\alpha)$ (B) $\beta^s \sim \mathcal{N}^{random}(\mu_\beta, \sigma_\beta)$

FIGURE 4.17: Subject α^s, β^s posterior distributions in the random sample.

Compare figure 4.16 to figure 4.17. A much wider range of both α^s and β^s estimates are observed in the random set (figure 4.17). In both parameter sets, the hierarchical priors appear less confident (showing greater variance) in the random set; however, the hierarchical prior does not appear to regularise the individual parameters much.

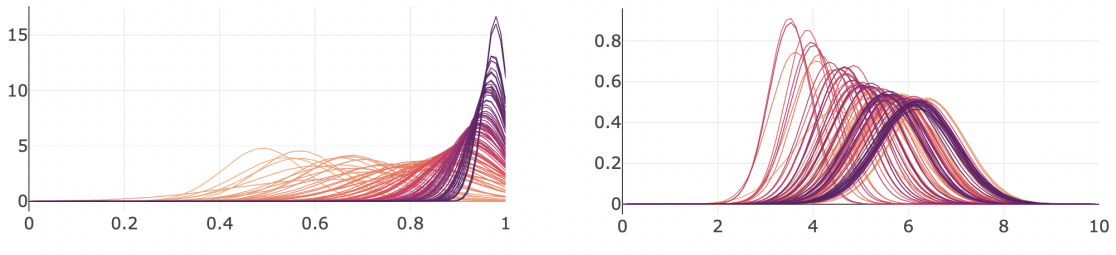
Individual parameter estimates: Now consider only the individual subject posterior distributions. In figure 4.19, we again plot the posteriors but this time colour the posterior by subject.

Recall that the parameters are fit by uninformative priors over the permissible range. The model appears to have converged to meaningful, high confidence/low variance, parameter estimates. The individual distributions capture the learning mechanics that govern a particular subject's decision making process. Quantifying the posteriors as a normal

distributions allows for the extraction of sufficient statistics to summarise an individual's cognitive process.

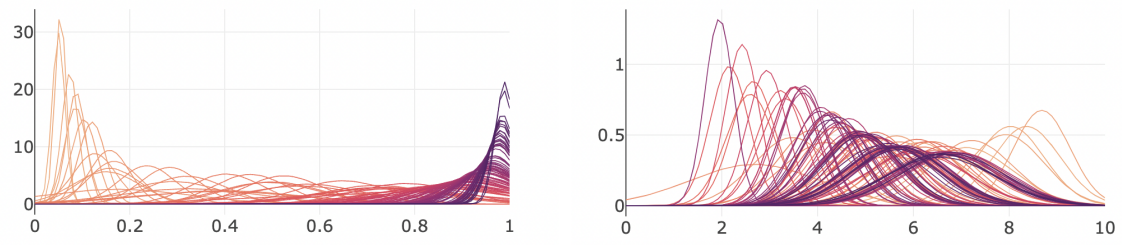
In figure 4.19 we sort and colour the individual distributions by their (mean) α^s estimates - colouring the posteriors by subject.

Plotted in figure 4.19 in the sample of top performing subjects, we observe a clean monotonic relationship between α^s and β^s. That is, larger α^s values (coloured in dark purple) are associated with larger β^s values. In the random set, however, the distributions are more variable and less stable. They appear consistent within the range of higher α^s values; however, a small subset clearly exhibits very low α^s values but very high β^s. This is undoubtedly a consequence of the interaction between parameter estimates (larger exploratory rates compensating for smaller learning rates - producing identical behaviour, though, through a very different parameterisation). Because parameters are connected non-linearly, the same behaviour can be modeled by different combinations of α^s and β^s values. We may wish, in future, to circumvent this plurality by specifying the priors (particularly of α) over a tighter permissible range. Alternatively, one may wish to suppress or penalise large β^s values as their range is relatively arbitrary.



(A) $\alpha^s \sim \mathcal{N}^{random}(\mu_\alpha, \sigma_\alpha)$ (B) $\beta^s \sim \mathcal{N}^{random}(\mu_\beta, \sigma_\beta)$

FIGURE 4.18: An examination of individual learning rates α^s (left) and exploratory coefficients β^s (right) in the best performing sub sample. Distributions are coloured by subjects, demonstrating the relationship between the learning parameter α^s and exploratory parameter β^s. There appears to be an increasing relationship in the parameters as individuals with higher α^s values exhibit higher β^s values. The high learning rates ^{s} are indicative of the subjects' readiness to adapt to new information.



(A) $\alpha^s \sim \mathcal{N}^{random}(\mu_\alpha, \sigma_\alpha)$ (B) $\beta^s \sim \mathcal{N}^{random}(\mu_\beta, \sigma_\beta)$

FIGURE 4.19: Similar to figure 4.18, here we examine individual learning rates α^s (left) and exploratory coefficients β^s (right) in the random sample. Not only revealing some relationship in parameters, but also the algorithm's tendency to converge to alternative parameter combinations, and yet still achieve similar data generating properties (as the effects of α and β may be offset). A notable caveat of fitting the non-linear update equation. This can be observed as some subjects have very low α^s values that are offset by extremely high β^s values.

4.5.5 Recovering the data generating process

It is also important to sense check the results from an intuitive perspective. Although we have demonstrated the models' convergence properties and chosen the optimal fit by approximating leave-out-one cross validation, it is useful to explicitly illustrate the model's ability to capture data generating properties.

To this end, we recover the individual model parameters (as shown above), and; subsequently, use these parameters to generate a sequence of actions performing the WCST. We can then examine how tightly these simulated results correspond to the observed data to assess how well the model captures the data, providing a proxy for model accuracy.

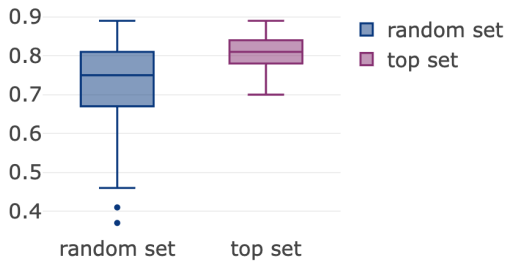


FIGURE 4.20: The distribution of model accuracy on both the random (left) and best (right) samples. The box-plots represent the model accuracy percentage of each subject in the sample, showing the distribution of prediction accuracy over subjects. With an average score of 75% and 81% respectively, the model appears to accurately capture the data generating process. Note that the 75% is severely impeded by a few outliers - visible in the long tail of the left plot. On average, the model fits the data well and can be used reliably.

When tested on both the random and best datasets, shown in figure 4.20, the recovered parameters were able to accurately predict the subjects' choice behaviours 75% of the time. In the absence of some severe outliers, this average would be much greater. In light of these estimates, we can confidently rely on the model to adequately represent the data.

Visualising an individual's learning process: For the purpose of illustration we visualise the choice behaviour of a single subject by their modeled parameters.

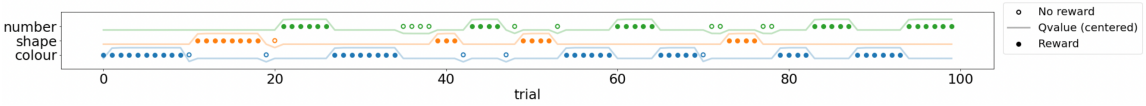


FIGURE 4.21: Illustrating the Q-learning processing by plotting the estimated $Q_t^s(a)$ values of a subject s. Circles represent actions taken by a subject, with coloured and empty circles representing positive and negative feedback. The lines, centered around each vertical choice access, depict the $Q_t^s(a)$ approximations generated by the model parameters. It is clear that the subject rapidly adjusts their state-value estimates in light of new information. The figure illustrates the underlying data generating process that produces the subjects observable behaviour.

Displayed in figure 4.21, the state value estimates $Q_t^s(a)$ are clearly a function of the reward mechanism. This demonstrates how we have modeled the learning process and are, therefore, able to offer real-time estimates of a subject's beliefs.

4.6 Covariate Analysis

Finally, we are interested in the relationships between the extracted learning parameters and the additional psychological and demographic covariates. Addressing the question: *Under what psychological constructs is learning supported?*

Table 4.4 summarises a suite of correlation metrics over parameter estimates. The columns are separated by three quantities to measure the strength of relationships in the data:

1. **Pearson Correlation**: measuring linear dependency.
2. **F-test (p-value)**: Testing linear dependency for statistical significance.
3. **Mutual Information**: measuring non-linear dependency.

Extracting individual parameters: We opted to use the random sample as it is more conservative and generalisable. The table is structured to contrast all variables (rows) with the biological parameters (columns). Individual subject's parameters are extracted by fitting Gaussian distributions to the (stationary) posterior traces, this assumes:

$$\alpha^s \sim \mathcal{N}(\mu_\alpha^s, \sigma_\alpha^s)$$

and similarly:

$$\beta^s \sim \mathcal{N}(\mu_\beta^s, \sigma_\beta^s).$$

Therefore the learning parameter data for this component of the analysis consists of a matrix of four vectors: $\{\mu_{alpha}^s, \sigma_{alpha}^s, \mu_{beta}^s, \sigma_{beta}^s\}$.

If we examine the lower half of figure 4.19 (representing the distributions from the individual parameters) we notice that the model may converge to very different α^s, β^s combinations and still achieve results that are able to regenerate the data. More specifically, we notice

that some subjects α^s values are extremely low (coloured in light orange) and offset with very high β^s values (the corresponding colour in the adjacent graph).

It is likely that more robust correlations would be observed if we reduced the permissible range over either α or β (by imposing tighter priors). This is likely to be more realistic as these extremely low α^s values from such a simple task are not supported by theory.

Covariate Correlation Analysis													
		Pearson Correlation				F-test				Mutual Information			
	Covariate	μ_α	σ_α	μ_β	σ_β	μ_α	σ_α	μ_β	σ_β	μ_α	σ_α	μ_β	σ_β
Bio	μ_α												
	σ_α	-0.03				0.76				1.00			
	μ_β	-0.26	-0.50			< 0.01	< 0.01			0.29	0.20		
	σ_β	-0.09	-0.21	0.67		0.35	0.03	< 0.01		0.17	0.03	1.17	
Psy	wcst RT	0.04	-0.16	0.11	0.05	0.72	0.12	0.27	0.59	0.07	0.00	0.06	0.00
	Fitts	0.06	0.06	-0.06	-0.07	0.54	0.55	0.55	0.46	0.00	0.00	0.02	0.03
	NBack	-0.02	-0.12	0.01	-0.06	0.86	0.23	0.93	0.56	0.01	0.00	0.04	0.02
	NB RT	0.20	-0.26	0.01	-0.06	0.04	0.01	0.90	0.57	0.04	0.06	0.07	0.14
	Navon	-0.13	0.04	-0.04	-0.09	0.20	0.71	0.71	0.35	0.18	0.00	0.00	0.00
	Nv RT	0.08	-0.11	0.06	0.16	0.42	0.26	0.53	0.12	0.00	0.00	0.00	0.02
	Corsi	0.07	0.11	-0.09	-0.11	0.49	0.27	0.39	0.28	0.00	0.00	0.00	0.04
Dem	Dem RT	0.07	-0.13	0.12	0.20	0.51	0.19	0.22	0.05	0.00	0.00	0.05	0.18
	Income	0.17	-0.02	-0.01	0.05	0.10	0.84	0.93	0.63	0.00	0.00	0.03	0.14
	Comp	-0.11	0.01	0.14	0.28	0.28	0.89	0.16	< 0.00	0.00	0.00	0.00	0.00
	Age	0.08	-0.04	-0.03	-0.10	0.45	0.69	0.77	0.32	0.00	0.04	0.14	0.09

TABLE 4.4: Covariate correlation analysis across all variable classes. Significant correlative relationships are coloured in green. The parameters are categorised by *biological (bio) parameters*: learning parameters extracted from the model, *psychological (psy) parameters* (representing neuropsychological covariates), and *demographic (dem) covariates* (representing demographic data).

Learning parameter analysis: The first section of table 4.4 contrasts all of the extracted learning parameters with only a single set of off-diagonals presented to mitigate redundancies. The mean values of individual posteriors provide their point estimates; however, we are equally interested in the confidence (variability) in these estimates.

In accordance with our theoretical discussion, μ_{α^s} and μ_{β^s} are highly correlated across all metrics, showing a correlation of -0.26, that is statistically significant, $(p - value \leq 0.01)$ and a relatively high mutual information 0.29.

Our extracted learning rates μ_α do not linearly correlate with either variance metric $\sigma_\alpha, \sigma_\beta$ but does, however, exhibit great mutual information with both metrics. Intuitively, this matches with figure 4.19 as the multi-modal distributions break any linear relationship;

however, there still appears to be structure on the spread of distributions (captured by mutual information).

The mean exploratory parameters μ_β, however, correlate highly with both variance metrics $\sigma_\alpha, \sigma_\beta$ when examined by both linear and non-linear metrics. Again, observable in figure 4.19, one can clearly see an increase in variance as the mean β^s estimates increase. These relationships allude to the confidence and stability of a subject's actions: more confidence/stable behaviours are reflected in reduced variance.

Psychological covariate analysis: Examining linear correlation, the NBack task is correlated with both the mean μ_α and variance σ_α of the learning parameter α. The WCST RT (reaction time) is also negatively correlated with variance in learning rates σ_α. Both of these reaction time metrics indicate that lower (possibly less ergonomically comfortable or less attentive) subjects show greater variance in learning. Interestingly, the Navon task RT shows a positive correlation with the variance in exploratory parameter σ_β. No other psychological covariates show meaningful linear correlations with the learning parameters. Of these relationships, only the NBack task RT appears statistically significant after conducting an F-test.

There is very little shared mutual information between learning parameters and neurospychological covariates, with only the Navon task score reporting a notable MI with μ_α. This indicates a relationship between one's learning rate and Navon performance - recall that the Navon task measures one's attentiveness and ,therefore, logically correlates with one's learning update efficiency α^s.

Demographic covariate analysis: Turning our attention to demographic covariates, both *computer hours* and *demographic RT* (the time taken to complete the demographic form) are significantly linearly correlated with the variance in exploratory parameters σ_β. *demographic RT* show a further non-linearity, captured in the MI between the same variables. *Income*, perhaps indicative of other social-economic factors, shows a mild correlation with mean learning rates μ_α, suggesting that subjects with greater income learn more efficiently.

4.7 GAMs

We have thus far modeled the learning process of each subject, recovering their learning parameters, and, subsequently, shown the raw relationship (linear and nonlinear) between these biological approximations and both psychological and demographic information pertaining to the subjects. This offers great insight into the potential causal relationships between different psychological metrics, social-economic information, and dynamic learning under uncertainty.

In light of more robust correlations, GAMs would offer the perfect tool to model the latent learning parameters. Offering flexibility, interpretability, theoretical bounds, and, applicable to both small and large datasets, the model class could offer generalisable technique to uncover the relationship between latent learning parameters and other physiological and psychological metrics. GAMs have an inherent ability to tune hyper-parameters, avoid over-fitting and suppressing parameters that fail to exhibit meaningful predictive power.

Unfortunately, the absence of strong statistical relationships between our psychological, demographic and biological covariates prohibits such a model. We tested a series of GAMs, with minimal covariates, and no significant parameters were found. It appears the relationships between the psychological covariates in question are spurious - or at least unobserved

in these experiments. We feel as if this finding warrants reporting as it speaks to the relationships (or lack there of) in the data.

Chapter 5

Discussion

In this chapter we place our results in the broader context of the literature, highlighting previous studies that our work supports - as well as that which it does not support - with the ultimate objective of laying a foundation for further enquiry.

The discussion loosely follows the same structure as the results chapter, elucidating each component of our findings in the context of the literature.

5.1 Sample demographics

Examining section 4.1, the two (best and random) samples are contrasted throughout much of the analysis. Their demographics, however, appear very similarly distributed (shown in figure 4.1). This, coupled with the simplicity of the task, may suggest that the sample of the best participants is not necessarily biased, but rather indicative of the candidates that were most attentive and engaged with the task or showed natural inclination towards this type of associative learning.

Subject *Age* is roughly centred around age 31 and positively skewed up towards 70, given the over 18 mandate this is a reasonable spread, although we may have underrepresented older individuals.

The *income*, although scored on a relative scale, appears to be roughly normally distributed which is a positive sign for the reliability of this self-reported metric. The distribution of *computer hours*, however, are largely clustered around low values but display an elongated positive tail. Intuitively, this may represent a natural population as one might imagine a subset of individuals spend exponentially more time on computers than others (gamers, software engineers etc).

We highlight the similarity between the two sample sets as it demonstrates the reliability of the best performing WCST sample by indicating that the underlying characteristics in the best sample are very similar to that of a broader sample.

This demographic distribution analysis was followed by data pre-processing, removing of outliers and variable compression.

5.2 Data pre-processing and cleaning

5.2.1 WCST Outlier removal

Individuals who achieved an average WCST performance under $\lambda = 0.4$ were considered outliers and removed from our analysis.

The top sample of candidates can, of course, be considered a subset with a much higher threshold. The fact that there is no meaningful difference in the sample demographics in the two sub samples may suggest that a higher minimum threshold could be justified. There is limited research to suggest the expected amount of outliers when using MTurk (Lu et al., 2021; Crowston, 2012). Despite the limited information, the chosen threshold does not dramatically influence the results, seen in similarities across the two samples.

5.2.2 Navon task data compression

After demonstrating the strong correlative relationship between global and local attention (section 4.2.1), we were able to compress the Navon covariates into a single attention metric. *Global* and *local* Navon performance are highly correlated with a Pearson correlation of 0.73 and mutual information 0.44, suggesting minimal difference between the two scores within subjects.

To mitigate the potential of losing relevant information, we extracted two variables from this compression: the first is the average performance score in both global and local Navon instances, the second measures the difference between these performance scores (global − local performance). This was chosen as the emphasis in much of the literature is to examine the discrepancy between global and local performance (Navon, 1977a; Navon, 1977b; Wen and Kawabata, 2018; Tan, Lim, and Manalo, 2017).

As discussed in section 5.3 we found no significant correlation, suggesting that the difference in global and local may not significantly (or at least directly) correlate with associative learning.

5.3 Covariate prioritisation

In order to assess the statistical dependencies between the WCST and the covariates, we scrutinised the relationships between the average WCST performance scores and each covariate independently. Although this negates the data generating process - failing to account for the state value updates as captured in the RL model - this offers an intuition about covariate relevance. It also offers a valid approach to modeling the WCST (or its underlying processes such as working memory capacity (WMC)) if the research question of interest is not concerned with the underlying learning mechanics.

The table reported in section 4.3 examines each covariate with respect to linear correlation, non-linear mutual information and relative variable importance.

Demographic covariates: The demographic covariate shows less significant relationships with WCST than their neuropsychological counterparts. *Reaction times, computer hours, education and income* all show significant linear correlation with WCST scores. To the best of our knowledge there is no WCST analysis that examines these demographic relationships.

Demographic covariate analysis: Although less correlated with WCST than the psychological covariates, some statistical relationships are observable in the demographic covariates. Handedness, education level, gender and age are not significant in the F-test nor produce meaningful MI results. Computer hours, demographic reaction time and Income appear to have a significant linear relationship with the WCST, whereas *gender, age and handedness* fail to show any linear relationships. *Reaction time* showing additional congruency with the WCST, was the only demographic covariate to exhibit some mutual information with the aggregate WCST scores.

Although *handedness* and *gender* served as control covariates - unsurprisingly producing no significant relationship - one would have expected age to correlate with the WMC driven task in some way, as many studies show the decay of WMC throughout life (Huizinga, Dolan, and van der Molen, 2006; Jaeggi et al., 2010; Kane et al., 2007).

It is likely that *computer hours* and *demographic reaction times* capture similar underlying generative processes: the former being a self-reported account of how comfortable one may be with their computer (alluding to the fluency with which they undergo the experiment), and the latter (the actual time taken to answer the demographic questions) likely constitutes attentiveness, alertness and how comfortable one is with the machine on which they took the task. A further coupling between covariates may exist between *income* and *education*, as these often correlate significantly (Tan, Lim, and Manalo, 2017).

Nonetheless, we clearly observe that more highly educated, more highly earning computer literate individuals score higher (on average) in the WCST than their respective counterparts, and, thus, the same individuals may exhibit greater WMC, operant/associative learning, set-shifting (cognitive flexibility) and cognitive inhibition (self-regulation) (D'Alessandro et al., 2020), (Miyake et al., 2000). These relationships should caveat the neuropsychological interpretations to follow.

Neuropsychological covariates: In support of the above Navon variable compression, there is no observable linear nor non-linear relationship with the *global-local* Navon scores and the WCST aggregate performance. This means that we do not observe a meaningful difference in WCST scores in individuals with greater global and local attention. Notably, the aggregate Navon score does linearly correlate with the WCST, revealing the importance of attentiveness and visual perception in performing associative learning (just not at the granularity of distinguishing between global and local effects that is highlighted in the literature) (Navon, 1977a; Wen and Kawabata, 2018).

All other neuropsychological covariates (*WCST reaction times (RTs), Fitts scores, NBack performance, NBack RTs*, and *Corsi block spans*) show significant linear relationships with aggregate WCST performance. Additionally, the *WCST reaction times (RTs), Fitts scores, NBack performance*, and *NBack RTs* show join Mutual Information (MI) with the aggregate WCST statistic, further illustrating a dependency between the covariates.

All the reaction time metrics (RTs) are significantly linear correlated, assessed across an F-test, ML and mRMR strategies, they consistently show stronger statistical relation to the WCST aggregate performance. All the RT mechanisms are likely to be highly correlated, capturing much of the same information (attentiveness, computer literacy etc).

Fitts scores are also highly significant linear correlations, eluding to motor skills and, once again, computer familiarity and attention. It was been shown that motor skills correlate with working memory when performing cognitive tasks (Fitts, 1954a; Fitts, 1954b). This attention mechanism appears to consistently predict associative learning.

The *NBack* and *Corsi block span* - both measuring working memory capacity - are also significantly correlated with aggregate WCST performance, again this is previously observed in the literature (Jaeggi et al., 2010). This was to be expected as working memory is thought to be the primary executive function underlying associative learning (Humann, Fischer, and Ullsperger, 2020).

After examining the correlation between covariates and average WCST performance, we then simulated population based hierarchical learning to investigate the behaviour of a population drawn from a shared underlying latent process.

5.4 Simulated Reinforcement Learning sequences

5.4.1 Single subject

We began by simulating the action sequence of a single individual, assuming a Rescorla-Wagner RL model for a data generating process, to illustrate the model's ability to recover the nonlinear parameters. Shown in section 4.4, the generating parameters were recovered by fitting a Monte Carlo procedure to the action sequence, with estimated values $\hat{\theta} : \{\alpha = 0.36, \beta = 11.22\}$ very nearly matching their data generating counterparts $\theta : \{\alpha = 0.4, \beta = 10\}$.

Although showing promise, these initial generating values are well within the theoretical bounds ($0 \leq \alpha \leq 1;\ 0 \leq \beta \leq \infty$). To test the robustness of this approach, we then repeated the same experiment over a wide range of α and β values.

Despite employing uniform priors, the model appears to bias results towards more mid-range conservative estimates (Daw, 2011b), consistently overestimating very low α or β values and underestimating very high α or β values. Boundaries are enforced through the prior; however, this appears to hold true within the permissible range.

It is worth remembering that the non-linearity of the model allows the same data to be produced with different α, β couplings, which may encourage the model to fit more conservative estimates yet still adequately capture the observed sequence of action. Therefore, the model was able to regenerate the data, capturing a very similar underlying process, but may slightly bias for more centred parameter convergence.

5.4.2 Many subjects

Thereafter, we simulated an entire population that draw their individual parameters from a pooled hierarchical process. The objective was to assess the dangers of inflating the variance in the model when using the summary statistics approach. It is known that in many model architectures, applying the summary statistics method greatly inflates variance, failing to adequately account for mutual information (Cover and Thomas, 2006).

It is clear from our experiments that the summary statistics approach does indeed inflate the hierarchical variance - revealing the this shortcoming still applies in the non-linear model domain.

Because the mean parameter estimates are near perfectly recovered in both approaches, one may opt to employ the summary statistics method if one wishes to test the discrepancy in learning rates in many sub-populations. To address our research question, however, we require the variance estimates and thus a full Bayesian hierarchical model.

Approach validity: The consequence of the uniform prior undoubtedly affects the model fit. Both issues highlighted here, may be circumvented by a more refined, well specified prior distribution (Gelman et al., 2004). Additionally, the expert knowledge available in the realm of cognitive science may be used to inform the prior, producing more cohesive results.

Although the analysis of information-theoretic quantities, as illustrated by the Bayesian brain hypothesis (Holmes and Nolte, 2019), is beyond the scope of our work, we were able to demonstrate how hierarchical Bayesian methods sufficiently regulate individual RL parameters, thus supporting the approach of coupling Bayesian methods with nonlinear functions when measuring cognitive abilities (D'Alessandro et al., 2020), (Steinke, Lange, and Kopp, 2020).

The convergence, stability, and diagnostics of the model support predictive processing - directly analogous to RL - as a plausible theoretical model for dynamic, stimulus driven, learning (Euler, 2018), (Whyte, 2019).

After testing these model architectures, we then fit the same Bayesian hierarchical architecture to the experimental data we observed, with the objective of making inference about the latent biologically inspired processes governing associative learning.

5.5 Cognitive Science RL models

The above simulations give us confidence in the chosen model architecture. We fit four RL models to capture the data generating process and (as often done in the hierarchical Bayesian literature) compared the models on the basis of WAIC, which approximates leave-one-out cross validation (Gelman et al., 2004).

We fit three variants of neuropsychologically inspired model parameterisations, aiming to capture different levels of granularity in the cognitive process. We further, fit a forth model after selecting the optimal neuropsychologically inspired model to assess the utility of encoding covariates directly into the learning process.

5.5.1 Selecting the model architecture

Biologically inspired parameterisation: Our findings, shown in section 4.5, show that the parsimonious null model best fits the experimental data. This model contains four population level parameters, consisting of a mean and standard deviation for both learning and exploratory parameters $\theta_{pop} : \{\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta\}$. Each individual subject s samples their learning parameters from these population priors:

$$\alpha^s \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha), \quad \beta^s \sim \mathcal{N}(\mu_\beta, \sigma_\beta).$$

The WAIC scores are purely relative and do not offer independent interpretation (Hastie, 2001). Much of the literature would have suggested that a more complex, nuanced model would be appropriate in recapitulating the data, as many predictive processing studies have shown variable learning rates for positive and negative feedback or state selection bias (Euler, 2018), (Barcelo, 2020).

Our experiment, however, does not necessitate this additional complexity, with the lower WAIC scores indicating that adding parameters to capture these cognitive processes are largely superfluous. This finding is consistent in both sample sets. The predictive processing literature strongly supports the additional parameters (Humann, Fischer, and Ullsperger, 2020); however, it is likely that our task instance is just too simple to necessitate this level of granularity.

Encoding additional covariates: The fourth model, using this chosen null model as a base, added an additional hierarchical distribution to demonstrate the possibility and potential utility of adding covariates directly into the learning model. Given a sufficiently complex task, it is possible that one's sequential associative learning process may be explained by some psychological covariates - justifying a direct covariate encoding (D'Alessandro et al., 2020).

The *WCST Reaction time (RT)* was chosen as an additional covariate as it showed the greatest correlation with the aggregate WCST performance, as detailed in section 5.3. Two population level parameters were added $\mu_{wcst_{rt}}, \sigma_{wcst_{rt}}$, and an additional subject

level covariate was sampled from these hierarchical priors $rt^s_{wcst} \sim \mathcal{N}(\mu_{wcst_{rt}}, \sigma_{wcst_{rt}})$. This parameter was then used to inform the observational model, acting as a weight in computing the action sampling probabilities π^s_t.

Applied to both the best and random samples, and compared with the WAIC, the additional complexity offers no benefit. The WAIC scores are essentially unchanged when modeling the best sample (WAIC scores of -10346.50 and -10345.50 for the null and RT model respectively), and degrade significantly when modeling the random sample (WAIC scores of -12824.48 and -12760.20 for the null and RT model respectively).

Potential extensions: The above recursive process can be implemented into arbitrary complex settings. In theory, we could repeat the procedure by adding additional psychological or demographic covariates to capture any reproducible differences across subjects. It is unlikely that the model will improve past the psychological implementation, as the nearest examples from the literature (Slooten, Jahfari, and Theeuwes, 2019; D'Alessandro et al., 2020) employ models similar to those discussed above (discussed at depth in section 2.3.1 and 2.6.6). Further, adding $s+2$ parameters adds substantial computational overhead, and should be avoided in the absence of evidence of its efficacy.

As an illustration for completeness, consider the case of adding demographic covariates to the RL model. Letting the chosen psychological learning model be denoted $\xi+\phi$ (constituting the biological ξ and psychological ϕ parameters respectively) added to the observation model as done previous. Demographic information is then encoded in the following 4 additional models:

1. **Demo-model 1: Demographic RT**: $\psi = \xi + \phi + rt^s_{demo}$
2. **Demo-model 2: Income**: $\psi = \xi + \phi + rt^s_{demo} + inc^s$
3. **Demo-model 3: Computer hours**: $\psi = \xi + \phi + rt^s_{demo} + inc^s + cs^s$
4. **Demo-model 4: Age**: $\psi = \xi + \phi + rt^s_{demo} + inc^s + cs^s + age^s$

The additional parameterisation was not justified as the single covariate did not improve the model.

Our experimentation favours simpler biologically inspired models and, therefore, these additional covariate models were not tested beyond this. Secondly, we are only interested in explanatory power in these covariates and their quantities as the nonlinear nature of RL updates translate to little to no interpretability of the covariate coefficients. Downstream analysis, as discussed in the following section, may offer more naturally interpretable and intuitive results. The selected, null model was then subject to much more meticulous analysis.

5.5.2 Population level distributions

The population posterior distributions were examined in section 4.5.3. The best performing sample exhibited greater average learning rates α and exploratory parameter β, consistent with both the literature and intuition (D'Alessandro et al., 2020). A faster learning rate in particular would be indicative of timely state-value updating, quickly adapting to new information.

The broader random sample also showed greater variation in the parameter point estimates, thus indicating less consistent choice behaviour.

We also observe a relationship in the variance and mean estimates of the learning rates throughout the Monte Carlo sampling procedure. This may well be a consequence of the

nonlinear update, as suggested by Euler, 2018. This phenomenon is not well understood (Daw, 2011b), but should caveat inflated learning rates. Population level distributions serve as regularising priors over the individual subject level parameters.

5.5.3 Recovering individual level parameters

Individual subject parameters are fit in the same procedure - discussed in section 4.5.4. When examining the individual posteriors α^s, β^s bi- (or possibly multi-) modal distributional peaks emerge, which may indicate that a bi-modal Gaussian mixture model prior would be more appropriate for the hierarchical regularising distribution (Daw, 2011b).

A very interesting dichotomy appears when one contrasts the individual parameters of the two respective sub-samples. If we examine the best performing sample, a monotonic relationship exists between the learning rate α^s and exploratory parameter β^s, whereas when we examine the same parameters fit to the broader random sample one observes a coupling between α^s and β^s. More specifically, some very low α^s correspond to very high β^s values. The non-linearity of the model results in a peculiar case whereby the same choice behaviour could be generated by different α^s, β^s combinations. One may wish to suppress the permissible range over α^s to circumvent this instability; however, we have encountered little theoretical guidance about this. Furthermore, one should caution overfitting by greatly biasing the permissible range.

5.5.4 Recovering the data generating process

To assess the model fit when dealing with generative models, one may wish to examine the model's ability to reproduce the data. Conducted in section 4.5.5, we simulated choice behaviour from the recovered individual parameters to contrast the simulated choice behaviour with observed choice behaviour.

It is both intuitive and tempting to use this discrepancy as some *mean error* estimate of model fit; however, the literature warns against this approach for model selection (Daw, 2011b). Nonetheless, it serves as evidence of the model's ability to recover the data.

Our models were able to generate action sequence that correspond to the observed data 0.75% and 0.82% of the time when applied to the random and best sample respectively. This offers great confidence in our design choices as the latent process is clearly replicated in some way.

Although it is impossible to know if a better modeling paradigm exists, the fact that the RL model was able to successfully fit and, to a certain extent, regenerate the data, supports using RL to model learning and predictive processing, as widely advocated in the literature (Rusanen et al., 2021; Myin and Hutto, 2015; Stepp, Chemero, and Turvey, 2011; Hutto and Myin, 2020; Chemero and Silberstein, 2008). This offers an intuitive mechanism by which agents maximise expected future rewards when performing associative learning or operant conditioning (Rusanen et al., 2021; Hutto and Myin, 2020; Steinke, Lange, and Kopp, 2020).

Regenerating the subject behaviour can also be used to both assess state-value estimates in real-time and simulated choice behaviour in another analogous setting.

As demonstrated by Van Slooten et al., 2018 and Slooten, Jahfari, and Theeuwes, 2019, the validity of leveraging an RL paradigm appears self-evident in our study; in particular, it is fascinating to assess not only how average learning rates and feedback modulating parameters differ across individuals, but also how the variation and stochasticity vary in

demographics and neuropsychological functions. The complexity of the RL specification, however, may be greatly limited by task complexity and data availability. While both Van Slooten et al., 2018 and Slooten, Jahfari, and Theeuwes, 2019 were able to fit more biologically accurate RL models (capturing different learning rates for positive and negative feedback), this added complexity was largely superfluous when used in our experiment. This may purely be a function of task complexity, and data richness; however, we see no evidence to support the additional parameterisation.

5.6 Covariate analysis

Our final component of the analysis assessed the extracted learning parameters α^s, β^s with respect to the subject covariates. The convincing model fittings provide confidence in the models ability to capture the latent cognitive attributes of the subjects - or at least loose abstractions of these attributes - and therefore we can assess the statistical relationship between learning rates α^s, exploratory tendencies β^s and subject covariates.

We took a further interest in assessing not only the point estimates but also the standard deviation of these estimates, resulting in the parametersation:

$$\alpha^s \sim \mathcal{N}(\mu_{\alpha^s}, \sigma_{\alpha^s}), \quad \beta^s \sim \mathcal{N}(\mu_{\beta^s}, \sigma_{\beta^s}).$$

Learning parameters: examining the learning parameters (both point estimates and standard deviations) reveals strong correlations among various mean-variance couplings. Detailed in section 4.6, we see a statistical relationship (both linear and nonlinear) in the mean learning rates μ_{α^s} and exploratory parameter μ_{β^s}. This statistically articulates the monotonic relationship described in section 4.5.4 and is expected in the literature (Nyhus and Barceló, 2009). This suggests that individuals with higher learning rates are more likely to explore the parameter space more aggressively.

Learning rates μ_{α^s} also show high mutual information with the standard deviation in both learning rates σ_{α^s} and exploratory parameters σ_{β^s}. This relationship indicates that increased learning rates reduce variability in actions. That is, individuals with faster learning rates show more consistent behaviour, showing less variation in the underlying data generating parameters.

The mean exploratory parameter μ_{β^s} shows a strong negative correlation with the same standard deviation metrics σ_{α^s}, σ_{β^s}, again indicative of individuals with more exploratory nature (higher μ_β) to be more consistent in their behaviour.

Note that this part of the analysis uses the random sample, as it is more likely representative of the broader population, but we would expect the more pronounced correlation if the best sample set was used given the duality observed in section 5.5.4.

Psychological covariates:

The *WCST RT* and *NBack RT* are negatively correlated with learning rate standard deviation σ_α. This suggests that subjects with slower reaction times, who are possibly more deliberate and cautious in their responses, show less variation in their learning rates and therefore act more consistently.

The *NBack RT*, however, also shows a positive correlation with learning rates μ_α which may allude to the fact that faster/more attentive/ergonomically familiar individuals learn faster on average, updating their state believes more aggressively.

Despite not showing any linear correlation with the learning parameters, the *Navon* task shows mutual information (distributional overlap) with learning rates μ_α. Once again linking attention to faster associative learning updates.

We would have anticipated a greater link between working memory metrics such as the *NBack* task or *Corsi block span* and learning ability, as observed in the literature (Barcelo, 2020). Recall that we do observe a statistically significant relationship between these metrics and aggregate WCST performance, so perhaps the learning rates and exploratory parameter have more nuanced links with working memory capacity, or similarly only become observable in the latent data generating process during more elaborate experiments.

It is widely accepted that working memory forms a large component of learning (Unsworth and Engle, 2005); however, our work suggests the interaction and causal dependency may be more subtle than previously believed. Our task components that measure working memory capacity show little to no relationship (linear or nonlinear) with the subject's ability in the associative learning task. This may be a consequence of the specific task employed, or other confounding effects; however, we do not observe the direct causality (or tight relationship) between WMC and feedback based learning as suggested by Humann, Fischer, and Ullsperger, 2020.

In support of Panichello and Buschman, 2021, however, attention does appear to exhibit some statistical relationship with associative learning. Although linear correlations are meager, mutual information (MI) - capturing distributional dependency - shows some relationship between an individual's average learning rate μ_α and attention (measured by the Navon task). Despite this, there's no evidence to suggest that global attention is more present in predictive control, as suggested by Tan, Lim, and Manalo, 2017, or even that any discrepancy between global and local attention is observable in predictive processing assessments.

When making inferences on these results one should caution conflating attentiveness with computer familiarity and ergonomics. We observe that reaction times and (self-reported) computer hours are correlated with various learning metrics, thus attentiveness may simply be a consequence of familiarity and comfort with one's device.

In agreement with much of the neighbouring literature (Chang, 2021), we observe mild effects of aging on particular motor and executive functions. Although we do not observe any linear correlations between age and associative learning, MI alludes to distributional overlap between age and both the magnitude of exploratory tendencies μ_β (how readily one explores the action space) and (to a lessor extent) the variance of the same metric σ_β when updating state-value estimates.

Now that we have placed our results in the context of the literature, the following, final chapter recapitulates our findings with reference to our research objectives, and lays out potential future directions.

Chapter 6

Conclusion

This research project set out to investigate an ambitious inquiry into the associative learning process. First, we were able to show how to model predictive processing with Bayesian inference and non-linear Reinforcement Learning modelling. Our model sufficiently captures the data generating process, as it is able to simulate similar data instances, as well as incorporate biologically inspired mechanisms that loosely map to our current understanding of predictive processing during associative learning.

Beyond this, the framework allowed us to test varying levels of complexity accounting for greater nuances and details in the cognitive process. We were able to show that additional parameterisation (accounting for biological nuance such as different learning rates for positive and negative feedback) may be unnecessary in simple tasks, but may be effectively incorporated when modeling more complex assessments.

After demonstrating the utility of incorporating neuropsychological covariates directly into the learning model, we found it to be largely superfluous. We then illustrated another method to examine the relationships between subject neuropsychological and demographic covariates and their learning strategies by extracting individual learning characteristics.

We were able to extract statistical quantities that represent subjects' learning abilities, measuring the individuals' speed of knowledge acquisition (learning rate α) and tendency to try new actions to gain information (exploratory coefficient β). After assessing the distributional qualities of these learning characteristics, we illustrated how to examine learning coefficients with respect to neuropsychological and demographic covariates. We scrutinised the results through a range of statistical tools that capture both linearities and distributional overlap in the data, fully accounting for potential relationships between covariates.

Examining the parameters that govern associative learning reveals a number of notable statistical relationships. Individuals with faster learning rates are more likely to explore the state-space in light of new information. They also exhibit more consistency (less variability) in their actions, showing less random fluctuations and strategy shifting during learning. Individuals with a greater tendency to explore the state-space also exhibit more consistent actions.

Our work suggests that, at least in simple tasks, attentiveness may supersede working memory as a predictor of associative learning. We also observe greater action consistency in individuals with slower reaction times, indicative of more diligent, attentive task taking.

Given the strong correlation between working memory capacity and associative learning observed in the literature (Barcelo, 2020), we anticipated a greater statistical relationship

between the *NBack* and *Corsi block span* task and one's associative learning abilities. Interestingly, we observe strong correlations (both linear and nonlinear) between these parameters and associative learning on aggregate. That is, they do predict associative learning, however, not at the level of granularity of the individual learning parameters (which postulate the underlying generative process). This may allude to a more nuanced relationship between working memory capacity and associative learning.

We were also able to show that the *Navon* task (a measure of attentiveness) shows a nonlinear distributional overlap with the latent associative learning parameters, highlighting the importance of attentiveness and spatial awareness in operant conditioning. The discrepancy between global and local attention, however, did not predict associative learning or working memory capacity.

It is clear that the relationship between working memory capacity, attention and associative learning may be more nonlinear and nuanced than previously imagined.

6.1 Extensions and future work

It is our hope that our efforts lay the foundations for future scientific inquiry into the underlying process behind many cognitive activities and executive functions. The nature of the project lends itself to a myriad of natural extensions.

Bayesian cognition: In the context of Bayesian theories of the mind, a natural addition would be to quantify the relationship between Bayesian estimates of abstract cognitive properties and learning under uncertainty. As discussed in the literature review, Gershman, 2016 provides a Bayesian formalisation of a series of neuropsychological phenomena:

1. *State-value priors*: decomposed as a weighted function of previous posterior beliefs of transition dynamics to form a predictive probability over hidden states (Gershman, 2016).
2. *State transition matrices*: quantifying the belief about state transition probabilities- more suitable for probabilistic tasks where the state is not guaranteed to coincide with the feedback.
3. *Bayesian surprise*: the divergence between the current and previous state probability estimates, directly analogous to predictive processing (Schultz, Dayan, and Montague, 1997).
4. *Shannon surprise*: marginal information gained by evidence.
5. *Entropy*: quantify uncertainty in the agents' internal model (Barcelo, 2020).

Not only can we rely on statistical theory to reliably estimate these quantities, we can, in theory, then model the relationship between these cognitive properties and individual neuropsychological characteristics and learning parameters.

Statistical methods: another direct avenue of potential extension would be to exploit statistical properties to either scale the model to larger datasets or improve the model.

The reliance on Monte Carlo methods sets significant scalability limitations, particularly in temporal settings where each subject represents multiple action samples. Exacerbated by the nonlinear nature of the Reinforcement Learning update equation, prohibiting full vectorisation, iteration through large datasets becomes impractical as datasets grow. This can, in theory, be circumvented to some degree though the application of Variational Bayes (an

alternative optimisation procedure). Noting, however, that this may suffer from alternative drawbacks such as challenges quantifying uncertainty (Gelman et al., 2004).

In the face of exponentially growing datasets, perhaps when applying this technology in a commercial setting, variational inference would too likely fail (Hastie, 2001). The same situations, however, may solicit the possibility of estimating state-values by neural networks or other scalable nonlinear function approximators; frequently used in neighbouring RL applications (Silver, 2015).

Finally, embedding better theoretical knowledge of the problem could greatly reduce the space of possible models: as demonstrated by Gershman, 2016 when leveraging proper prior distributions. It is possible to greatly improve sampling efficiency through more carefully curated Bayesian priors (Gelman et al., 2004). Moreover, our simulations suggest that models may converge to parameter estimates that are over-regularised towards the mean of the permissible range. Intelligent prior specification allows for more deliberate encoding of neuropsychological information.

fMRI, EEG and MEG: As demonstrated in the literature, when it is possible to collect EEG, MEG or fMRI data one can more precisely map these behavioural neuropsychological findings to biological processes.

Extracting the variance around each RL parameter allows us to quantify uncertainty, offering a metric to measure the frequency with which the subjects vary their update parameters. The uncertainty of state-value updates may also represent the distinction between incremental learning (categorised by slow state-value updates) and one-shot learning (where an individual's internal state-value estimate is approximated after a single data sample) as one-shot learning would manifest as low variance, high expected value learning parameters $\theta : \{\alpha, \beta\}$ - with all other configurations representing incremental learning. Lee, O'Doherty, and Shimojo, 2015 were able to demonstrate how different neurological activity follows these alternative learning paradigms: with the hippocampus activating during one-shot learning and the ventrolateral prefontal cortex encoding uncertainty about causal associations (observed through incremental learning). It would be fascinating to examine the robustness of these findings by measuring the relationship between neurological activity on the spectrum between one-shot and incremental learning, by taking fMRI scans or EEG during the WCST.

The commonality across studies is to contrast neurological activity during different tasks or distinct stages of a task. It is, therefore, natural to examine the relationship between psychological or neurological changes during tasks and the learning rates that (possibly) generate subject behaviour. EEG has revealed an increase of $\sim 4-8$Hz and decrease of $\sim 8-25$Hz spectral power bands during cognitive memory tasks (Palomäki et al., 2012; Humann, Fischer, and Ullsperger, 2020). On replicating these findings it would be interesting to see if the same activity occurs during associative learning and how they relate to the RL learning or exploratory parameters.

Similarly, it has been shown that working memory capacity is observable during instrumental learning (Humann, Fischer, and Ullsperger, 2020; De Renzi, Faglioni, and Previdi, 1977; MILNER, 1971). Reinforcement Learning parameters may allow us to examine how this relationship may depend on an individual's learning attributes (learning rate, variability and exploratory nature) (MILNER, 1971).

Although fMRI, EEG and MEG are the most reliable (medically graded) tools to examine real-time physiology (Mele et al., 2019), wearables are beginning to play a larger role in medical research (Lee, O'Doherty, and Shimojo, 2015). At the expense of (marginally)

decreased quality, wearables offer cheap, accessible physiological readings. A natural progression may be to perform similar tasks but collect an extra dimension of wearable data (the most frequently used being movement information and photoplethysmogram) (Mele et al., 2019) - potentially using this data to map physiological readings to cognitive attributes.

Predictive processing: Extracting RL model parameters gives direct insight into the learning rates and variability around learning. Put in another way, we are able to examine one's reaction to task induced stimuli, and, thus, gain insight into the predictive processing behaviour observable when an individual's expectations are not met (Euler, 2018). The learning parameters may be used to build upon predictive processing theory by examining the discrepancies between neurological activity (via EEG, MEG or fMRI) in individuals with faster (more aggressive) versus slower (more passive) state-value update parameters.

Alpha-band neural oscillations (Sherman et al., 2016), frontoparietal cortical activity (Barcelo, 2020), and visual cortex activity (Sherman et al., 2016) have all be linked to predictive processing. Might a difference be observable in the operant conditioning patterns governing individuals with high-vs-low performance on the WCST?

Cognitive disorders: Our learning analysis could offer information about the processes associated with cognitive disorders, psychiatric illnesses and/or mental illnesses. Considering higher order abstraction, correlation across executive functions have been linked to mental dysfunction (Miyake and Friedman, 2012). Central executive processing has been shown to negatively correlate with Alzheimer's, Dementia, Schizophrenia and Parkinson's (Carlesimo et al., 1994), (Dalrymple-Alford et al., 1994), (Chey et al., 2002). Perhaps we could reveal discrepancies in the associative learning abilities of individuals who suffer from these conditions. More broadly, perhaps we could quantify the discrepancies between individual's suffering from a range of pathologies, offering a non-invasive mechanism to better understand the cognitive processes or cognitive decay in individuals suffering from different conditions.

Broader theories of cognition: We may wish to examine the relationship between learning characteristics and alternative executive functions. For example, Unsworth and Engle, 2005 showed that the relationship between WMC and fluid intelligence (abstract reasoning that is independent of prior knowledge) is only invariant over a certain range and then displays irregularity. Perhaps the learning parameters may reveal some causal relationship between associative learning and cognitive fluidity.

It has been shown that attentional bias can significantly influence the mapping of sensory inputs to motor outputs (and thus decision making) (Deco and Rolls, 2005) - understanding the individuals' learning parameters may offer another layer of explanatory power to describe this causality, by revealing the process that generates choice behaviour. The same study was able to parameterise the effects of pharmacological agents: how might the effects of particular pharmaceuticals effect learning and might our learning parameters better elucidate this?

Shared neurological activity occurs in the prefrontal cortex during attention, WMC, and cognitive control (Panichello and Buschman, 2021). Following the same theme, the learning parameters may offer an increased granularity to detail these correlations.

The field of computational cognitive science is very much in its infancy, and it is likely that many revolutionary ideas come to fruition in the near future. Through the many scientific and engineering applications are yet to be explored - each with incredible societal and scientific impact potential - it is an exciting time to study the most complex system known to man: human cognition.

Bibliography

Adams, Rick, Quentin Huys, and Jonathan Roiser (July 2015). "Computational Psychiatry: Towards a mathematically informed understanding of mental illness". In: *Journal of neurology, neurosurgery, and psychiatry* 87. DOI: 10.1136/jnnp-2015-310737.

Ahonen, L., M. Huotilainen, and E. Brattico (2016). "Within- and between-session replicability of cognitive brain processes: An MEG study with an N-back task". In: *Physiology Behavior* 158, pp. 43–53. ISSN: 0031-9384. DOI: https://doi.org/10.1016/j.physbeh.2016.02.006. URL: https://www.sciencedirect.com/science/article/pii/S0031938416300506.

Baker, David A. (Apr. 2012). "Handbook of Pediatric Neuropsychology". In: *Archives of Clinical Neuropsychology* 27.4, pp. 470–471. ISSN: 0887-6177. DOI: 10.1093/arclin/acs037. eprint: https://academic.oup.com/acn/article-pdf/27/4/470/13551/acs037.pdf. URL: https://doi.org/10.1093/arclin/acs037.

Ball, Tali M. and Andrea N. Goldstein-Piekarski (2017). "Computational Psychiatry: New Perspectives on Mental Illness". In: *American Journal of Psychiatry* 174.7. PMID: 28669207, pp. 698–699. DOI: 10.1176/appi.ajp.2017.17030328. eprint: https://doi.org/10.1176/appi.ajp.2017.17030328. URL: https://doi.org/10.1176/appi.ajp.2017.17030328.

Barcelo, Francisco (2020). *A predictive processing account of card sorting: Fast proactive and reactive frontoparietal cortical dynamics during inference and learning of perceptual categories*. DOI: 10.31234/osf.io/zsw3t. URL: psyarxiv.com/zsw3t.

Barcelo, Francisco et al. (Oct. 2006). "Task Switching and Novelty Processing Activate a Common Neural Network for Cognitive Control". In: *Journal of Cognitive Neuroscience* 18.10, pp. 1734–1748. ISSN: 0898-929X. DOI: 10.1162/jocn.2006.18.10.1734. eprint: https://direct.mit.edu/jocn/article-pdf/18/10/1734/1935541/jocn.2006.18.10.1734.pdf. URL: https://doi.org/10.1162/jocn.2006.18.10.1734.

Barceló, Francisco (Aug. 2021). "A Predictive Processing Account of Card Sorting: Fast Proactive and Reactive Frontoparietal Cortical Dynamics during Inference and Learning of Perceptual Categories". In: *Journal of Cognitive Neuroscience* 33.9, pp. 1636–1656. ISSN: 0898-929X. DOI: 10.1162/jocn_a_01662. eprint: https://direct.mit.edu/jocn/article-pdf/33/9/1636/1956140/jocn_a_01662.pdf. URL: https://doi.org/10.1162/jocn_a_01662.

Baudot, Pierre et al. (2019). "Topological Information Data Analysis". In: *Entropy* 21.9. ISSN: 1099-4300. DOI: 10.3390/e21090869. URL: https://www.mdpi.com/1099-4300/21/9/869.

Boyle, Gregory John, Donald H Saklofske, and Gerald Matthews (2012). *Psychological assessment: Four volume set*. SAGE Publications Ltd.

Braver, Todd S. (2012). "The variable nature of cognitive control: a dual mechanisms framework". In: *Trends in Cognitive Sciences* 16.2, pp. 106–113. ISSN: 1364-6613. DOI: https://doi.org/10.1016/j.tics.2011.12.010. URL: https://www.sciencedirect.com/science/article/pii/S1364661311002610.

Brunetti, Riccardo, Claudia Del Gatto, and Franco Delogu (2014). "eCorsi: implementation and testing of the Corsi block-tapping task for digital tablets". In: *Frontiers in Psychology*

5. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.00939. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2014.00939.

Capon, Alison, Simon Handley, and Ian Dennis (2003). "Working memory and reasoning: An individual differences perspective". In: *Thinking & Reasoning* 9.3, pp. 203–244. DOI: 10.1080/13546781343000222. eprint: https://doi.org/10.1080/13546781343000222. URL: https://doi.org/10.1080/13546781343000222.

Carlesimo, G. A. et al. (1994). "Verbal and spatial memory spans in Alzheimer's and multi-infarct dementia". In: *Acta Neurologica Scandinavica* 89.2, pp. 132–138. DOI: https://doi.org/10.1111/j.1600-0404.1994.tb01648.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0404.1994.tb01648.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0404.1994.tb01648.x.

Chang, Erik Chihhung (2021). "Information-Theoretic Quantification of Dedifferentiation in the Aging of Motor and Executive Functions". In: *Frontiers in Aging Neuroscience* 13. ISSN: 1663-4365. DOI: 10.3389/fnagi.2021.634089. URL: https://www.frontiersin.org/article/10.3389/fnagi.2021.634089.

Chemero, Anthony and Michael Silberstein (2008). "After the Philosophy of Mind: Replacing Scholasticism with Science*". In: *Philosophy of Science* 75.1, 1–27. DOI: 10.1086/587820.

Chey, Jeanyung et al. (2002). "Spatial working memory span, delayed response and executive function in schizophrenia". In: *Psychiatry Research* 110.3, pp. 259–271. ISSN: 0165-1781. DOI: https://doi.org/10.1016/S0165-1781(02)00105-1. URL: https://www.sciencedirect.com/science/article/pii/S0165178102001051.

Cover, Thomas M. and Joy A. Thomas (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience. ISBN: 0471241954.

Crowston, Kevin (2012). "Amazon Mechanical Turk: A Research Tool for Organizations and Information Systems Scholars". In: *Shaping the Future of ICT Research. Methods and Approaches*. Ed. by Anol Bhattacherjee and Brian Fitzgerald. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 210–221. ISBN: 978-3-642-35142-6.

D'Alessandro, Marco et al. (2020). "A Bayesian brain model of adaptive behavior: an application to the Wisconsin Card Sorting Task". In: *PeerJ* 8, e10316. DOI: 10.7717/peerj.10316. URL: https://doi.org/10.7717%2Fpeerj.10316.

Dalrymple-Alford, J C et al. (1994). "A central executive deficit in patients with Parkinson's disease." In: *Journal of Neurology, Neurosurgery & Psychiatry* 57.3, pp. 360–367. ISSN: 0022-3050. DOI: 10.1136/jnnp.57.3.360. eprint: https://jnnp.bmj.com/content/57/3/360.full.pdf. URL: https://jnnp.bmj.com/content/57/3/360.

Davidoff, Jules, Elisabeth Fonteneau, and Joel Fagot (2008). "Local and global processing: Observations from a remote culture". In: *Cognition* 108.3, pp. 702–709.

Daw, Nathaniel (Mar. 2011a). "Trial-by-trial data analysis using computational models". In: *Affect, Learning and Decision Making, Attention and Performance XXIII* 23. DOI: 10.1093/acprof:oso/9780199600434.003.0001.

– (Mar. 2011b). "Trial-by-trial data analysis using computational models". In: *Affect, Learning and Decision Making, Attention and Performance XXIII* 23. DOI: 10.1093/acprof:oso/9780199600434.003.0001.

De Renzi, E., P. Faglioni, and P. Previdi (1977). "Spatial Memory and Hemispheric Locus of Lesion". In: *Cortex* 13.4, pp. 424–433. ISSN: 0010-9452. DOI: https://doi.org/10.1016/S0010-9452(77)80022-1. URL: https://www.sciencedirect.com/science/article/pii/S0010945277800221.

Deco, Gustavo and Edmund T. Rolls (2005). "Attention, short-term memory, and action selection: A unifying theory". In: *Progress in Neurobiology* 76.4, pp. 236–256. ISSN: 0301-0082. DOI: https://doi.org/10.1016/j.pneurobio.2005.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0301008205000912.

Duncan, John (2010a). *How Intelligence Happens*. Yale University Press. ISBN: 9780300168730. DOI: doi:10.12987/9780300168730. URL: https://doi.org/10.12987/9780300168730.

– (Feb. 2010b). "The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour". In: *Trends in cognitive sciences* 14, pp. 172–9. DOI: 10.1016/j.tics.2010.01.004.

D'Alessandro, Marco et al. (Nov. 2020). "A Bayesian brain model of adaptive behavior: an application to the Wisconsin Card Sorting Task". In: *PeerJ* 8, e10316. ISSN: 2167-8359. DOI: 10.7717/peerj.10316. URL: https://doi.org/10.7717/peerj.10316.

Euler, Matthew J. (2018). "Intelligence and uncertainty: Implications of hierarchical predictive processing for the neuroscience of cognitive ability". In: *Neuroscience Biobehavioral Reviews* 94, pp. 93–112. ISSN: 0149-7634. DOI: https://doi.org/10.1016/j.neubiorev.2018.08.013. URL: https://www.sciencedirect.com/science/article/pii/S0149763418302045.

Farrell Pagulayan, Kathleen et al. (2006). "Developmental normative data for the Corsi Block-tapping task". In: *Journal of clinical and experimental neuropsychology* 28.6, pp. 1043–1052.

Fitts, Paul M (1954a). "The information capacity of the human motor system in controlling the amplitude of movement." In: *Journal of experimental psychology* 47.6, p. 381.

Fitts, Paul M. (1954b). "The information capacity of the human motor system in controlling the amplitude of movement." In: *Journal of experimental psychology* 47 6, pp. 381–91.

Friedman, Naomi P. and Akira Miyake (2017). "Unity and diversity of executive functions: Individual differences as a window on cognitive structure". In: *Cortex* 86. Is a "single" brain model sufficient?, pp. 186–204. ISSN: 0010-9452. DOI: https://doi.org/10.1016/j.cortex.2016.04.023. URL: https://www.sciencedirect.com/science/article/pii/S0010945216301071.

Gagniuc, Paul (May 2017). *Markov Chains: From Theory to Implementation and Experimentation*. ISBN: 978-1-119-38755-8. DOI: 10.1002/9781119387596.

García-Molina, A. (2012). "Phineas Gage and the enigma of the prefrontal cortex". In: *Neurología (English Edition)* 27.6, pp. 370–375. ISSN: 2173-5808. DOI: https://doi.org/10.1016/j.nrleng.2010.03.002. URL: https://www.sciencedirect.com/science/article/pii/S2173580812001198.

Gazzaniga, Michael S. (2009). *Cognitive neuroscience : the biology of the mind*. eng. 3rd ed. / Michael S. Gazzaniga, Richard B. Ivry, George R. Mangun with Megan S. Steven. New York ; W. W. Norton. ISBN: 9780393111361.

Gelman, Andrew and Jennifer Hill (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press. DOI: 10.1017/CBO9780511790942.

Gelman, Andrew et al. (2004). *Bayesian Data Analysis*. 2nd ed. Chapman and Hall/CRC.

Gershman, Samuel J. (2016). "Empirical priors for reinforcement learning models". In: *Journal of Mathematical Psychology* 71, pp. 1–6. ISSN: 0022-2496. DOI: https://doi.org/10.1016/j.jmp.2016.01.006. URL: https://www.sciencedirect.com/science/article/pii/S0022249616000080.

Hastie Trevor, Tibshirani Robert Friedman Jerome (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY, USA: Springer New York Inc.

Hazy, Thomas, Michael Frank, and Randall O'Reilly (Nov. 2009). "Neural mechanisms of acquired phasic dopamine responses in learning". In: *Neuroscience Biobehavioral Reviews* 34, pp. 701–720. DOI: 10.1016/j.neubiorev.2009.11.019.

Holmes, Jeremy and Tobias Nolte (2019). ""Surprise" and the Bayesian Brain: Implications for Psychotherapy Theory and Practice". In: *Frontiers in Psychology* 10, p. 592. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2019.00592. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2019.00592.

Homan, Matthew D. and Andrew Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". In: *J. Mach. Learn. Res.* 15.1, 1593–1623. ISSN: 1532-4435.

Hooker, Davenport (1960). "Plans and the structure of behavior. By George A. Miller, Eugene Galanter and Karl H. Pribram 1960. Henry Holt and company, New York. 226 pp". In: *Journal of Comparative Neurology* 115.2, pp. 217–217. DOI: https://doi.org/10.1002/cne.901150208. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.901150208. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.901150208.

Huizinga, Mariëtte, Conor V. Dolan, and Maurits W. van der Molen (2006). "Age-related change in executive function: Developmental trends and a latent variable analysis". In: *Neuropsychologia* 44.11. Advances in Developmental Cognitive Neuroscience, pp. 2017–2036. ISSN: 0028-3932. DOI: https://doi.org/10.1016/j.neuropsychologia.2006.01.010. URL: https://www.sciencedirect.com/science/article/pii/S0028393206000224.

Humann, Jil, Adrian G Fischer, and Markus Ullsperger (2020). *The Dynamics of Feedback-based Learning is Modulated by Working Memory Capacity*. DOI: 10.31234/osf.io/8qzu2. URL: psyarxiv.com/8qzu2.

Hutto, Daniel and Erik Myin (Dec. 2020). "Deflating Deflationism about Mental Representation". In: pp. 79–100. ISBN: 9780190686673. DOI: 10.1093/oso/9780190686673.003.0004.

Huys Michael Moutoussis, Jonathan Williams (2011). "Are computational models of any use to psychiatry?" In: *Neural Networks* 24.6. Special Issue: Neurocomputational Models of Brain Disorders, pp. 544–551. ISSN: 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2011.03.001.

Huys, Quentin (2013). "Computational Psychiatry". In: *Encyclopedia of Computational Neuroscience*. Ed. by Dieter Jaeger and Ranu Jung. New York, NY: Springer New York, pp. 1–10. ISBN: 978-1-4614-7320-6. DOI: 10.1007/978-1-4614-7320-6_501-2. URL: https://doi.org/10.1007/978-1-4614-7320-6_501-2.

Huys, Quentin J. M. et al. (2015). "Decision-Theoretic Psychiatry". In: *Clinical Psychological Science* 3.3, pp. 400–421. DOI: 10.1177/2167702614562040. eprint: https://doi.org/10.1177/2167702614562040. URL: https://doi.org/10.1177/2167702614562040.

Jaeggi, Susanne M et al. (2010). "The concurrent validity of the N-back task as a working memory measure". In: *Memory* 18.4, pp. 394–412.

Jonides, John and Derek Nee (Nov. 2005). "Assessing Dysfunction Using Refined Cognitive Methods". In: *Schizophrenia bulletin* 31, pp. 823–9. DOI: 10.1093/schbul/sbi053.

Joue, Gina et al. (Oct. 2021). "Sex Differences and Exogenous Estrogen Influence Learning and Brain Responses to Prediction Errors". In: *Cerebral Cortex*. bhab334. ISSN: 1047-3211. DOI: 10.1093/cercor/bhab334. eprint: https://academic.oup.com/cercor/advance-article-pdf/doi/10.1093/cercor/bhab334/40655266/bhab334.pdf. URL: https://doi.org/10.1093/cercor/bhab334.

Kane, Michael J et al. (2007). "Working memory, attention control, and the N-back task: a question of construct validity." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33.3, p. 615.

Kessels, Roy et al. (July 2008a). "The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span". In: *Assessment* 15, pp. 426–34. DOI: 10.1177/1073191108315611.

Kessels, Roy PC et al. (2000). "The Corsi block-tapping task: standardization and normative data". In: *Applied neuropsychology* 7.4, pp. 252–258.

Kessels, Roy PC et al. (2008b). "The backward span of the Corsi Block-Tapping Task and its association with the WAIS-III Digit Span". In: *Assessment* 15.4, pp. 426–434.

Knill, David C. and Alexandre Pouget (2004). "The Bayesian brain: the role of uncertainty in neural coding and computation". In: *Trends in Neurosciences* 27.12, pp. 712–719. ISSN: 0166-2236. DOI: https://doi.org/10.1016/j.tins.2004.10.007. URL: https://www.sciencedirect.com/science/article/pii/S0166223604003352.

Konstantakopoulos, George (2019). "Insight across mental disorders: A multifaceted metacognitive phenomenon." In: *Psychiatrike = Psychiatriki* 30 1, pp. 13–16.

Kopp, Bruno (2012). "A simple hypothesis of executive function". In: *Frontiers in Human Neuroscience* 6. ISSN: 1662-5161. DOI: 10.3389/fnhum.2012.00159. URL: https://www.frontiersin.org/article/10.3389/fnhum.2012.00159.

Lee, Sang Wan, John P. O'Doherty, and Shinsuke Shimojo (Apr. 2015). "Neural Computations Mediating One-Shot Learning in the Human Brain". In: *PLOS Biology* 13.4, pp. 1–36. DOI: 10.1371/journal.pbio.1002137. URL: https://doi.org/10.1371/journal.pbio.1002137.

Lezak, M.D. et al. (2012). *Neuropsychological Assessment*. OUP USA. ISBN: 9780195395525. URL: https://books.google.co.za/books?id=meScZwEACAAJ.

Lu, Lu et al. (2021). "Improving Data Quality Using Amazon Mechanical Turk Through Platform Setup". In: *Cornell Hospitality Quarterly* 0.0, p. 19389655211025475. DOI: 10.1177/19389655211025475. eprint: https://doi.org/10.1177/19389655211025475. URL: https://doi.org/10.1177/19389655211025475.

Macrae, C Neil and Helen L Lewis (2002). "Do I know you? Processing orientation and face recognition". In: *Psychological Science* 13.2, pp. 194–196.

Mammarella, Irene Cristina and Cesare Cornoldi (2005). "Sequence and space: The critical role of a backward spatial span in the working memory deficit of visuospatial learning disabled children". In: *Cognitive Neuropsychology* 22.8, pp. 1055–1068.

McKone, Elinor et al. (2010). "Asia has the global advantage: Race and visual attention". In: *Vision research* 50.16, pp. 1540–1549.

Mele, Giulia et al. (2019). "Simultaneous EEG-fMRI for Functional Neurological Assessment". In: *Frontiers in Neurology* 10. ISSN: 1664-2295. DOI: 10.3389/fneur.2019.00848. URL: https://www.frontiersin.org/article/10.3389/fneur.2019.00848.

MILNER, BRENDA (Sept. 1971). "INTERHEMISPHERIC DIFFERENCES IN THE LOCALIZATION OF PSYCHOLOGICAL PROCESSES IN MAN". In: *British Medical Bulletin* 27.3, pp. 272–277. ISSN: 0007-1420. DOI: 10.1093/oxfordjournals.bmb.a070866. eprint: https://academic.oup.com/bmb/article-pdf/27/3/272/743531/27-3-272.pdf. URL: https://doi.org/10.1093/oxfordjournals.bmb.a070866.

Miyake, Akira and Naomi P. Friedman (2012). "The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions". In: *Current Directions in Psychological Science* 21.1. PMID: 22773897, pp. 8–14. DOI: 10.1177/0963721411429458. eprint: https://doi.org/10.1177/0963721411429458. URL: https://doi.org/10.1177/0963721411429458.

Miyake, Akira et al. (2000). "The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis". In: *Cognitive Psychology* 41.1, pp. 49–100. ISSN: 0010-0285. DOI: https://doi.org/10.1006/cogp.1999.0734. URL: https://www.sciencedirect.com/science/article/pii/S001002859990734X.

Myin, Erik and Daniel Hutto (June 2015). "REC: Just radical enough". In: *Studies in Logic, Grammar and Rhetoric* 41. DOI: 10.1515/slgr-2015-0020.

Navon, David (1977a). "Forest before trees: The precedence of global features in visual perception". In: *Cognitive psychology* 9.3, pp. 353–383.

– (1977b). "Forest before trees: The precedence of global features in visual perception". In: *Cognitive Psychology* 9.3, pp. 353–383. ISSN: 0010-0285. DOI: https://doi.org/10.1016/0010-0285(77)90012-3. URL: https://www.sciencedirect.com/science/article/pii/0010028577900123.

Nyhus, Erika and Francisco Barceló (2009). "The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update". In: *Brain and Cognition* 71.3, pp. 437–451. ISSN: 0278-2626. DOI: https://doi.org/10.1016/j.bandc.2009.03.005. URL: https://www.sciencedirect.com/science/article/pii/S0278262609000451.

O'Doherty, John P., Alan N. Hampton, and Hackjin Kim (2007). "Model-Based fMRI and Its Application to Reward Learning and Decision Making". In: *Annals of the New York Academy of Sciences* 1104.

Owen, Adrian M et al. (2005). "N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies". In: *Human brain mapping* 25.1, pp. 46–59.

Palomäki, Jussi et al. (2012). "Brain oscillatory 4–35 Hz EEG responses during an n-back task with complex visual stimuli". English. In: *Neuroscience Letters* 516.1, pp. 141–145. ISSN: 0304-3940. DOI: 10.1016/j.neulet.2012.03.076.

Panichello, Matthew and Timothy Buschman (Apr. 2021). "Shared mechanisms underlie the control of working memory and attention". In: *Nature* 592, pp. 1–5. DOI: 10.1038/s41586-021-03390-w.

Parr, Thomas, Geraint Rees, and Karl J. Friston (2018). "Computational Neuropsychology and Bayesian Inference". In: *Frontiers in Human Neuroscience* 12, p. 61. ISSN: 1662-5161. DOI: 10.3389/fnhum.2018.00061. URL: https://www.frontiersin.org/article/10.3389/fnhum.2018.00061.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Poletti, Céline et al. (2017). "Strategic Variations in Fitts' Task: Comparison of Healthy Older Adults and Cognitively Impaired Patients". In: *Frontiers in Aging Neuroscience* 8. ISSN: 1663-4365. DOI: 10.3389/fnagi.2016.00334. URL: https://www.frontiersin.org/article/10.3389/fnagi.2016.00334.

Riddell, Allen, Ari Hartikainen, and Matthew Carter (Mar. 2021). *pystan (3.0.0)*. PyPI.

Rusanen, Anna-Mari et al. (Dec. 2021). "Action control, forward models and expected rewards: representations in reinforcement learning (open access)". In: *Synthese*. DOI: 10.1007/s11229-021-03408-w.

Salminen, Tiina and Torsten Schubert (June 2012). "On the impacts of working memory training on executive functioning. Front. Hum". In: *Frontiers in human neuroscience* 6, p. 166. DOI: 10.3389/fnhum.2012.00166.

Schultz, Wolfram, Peter Dayan, and P. Read Montague (1997). "A Neural Substrate of Prediction and Reward". In: *Science* 275, pp. 1593 –1599.

Schönberg, Tom et al. (Dec. 2007). "Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Nonlearners during Reward-Based Decision Making". In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27, pp. 12860–7. DOI: 10.1523/JNEUROSCI.2496-07.2007.

Sherman, Maxine T. et al. (Sept. 2016). "Rhythmic Influence of Top–Down Perceptual Priors in the Phase of Prestimulus Occipital Alpha Oscillations". In: *Journal of Cognitive Neuroscience* 28.9, pp. 1318–1330. ISSN: 0898-929X. DOI: 10.1162/jocn_a_00973. eprint: https://direct.mit.edu/jocn/article-pdf/28/9/1318/1951736/jocn_a_00973.pdf. URL: https://doi.org/10.1162/jocn_a_00973.

Silver, David (2015). *Lectures on Reinforcement Learning*. URL: "https://www.davidsilver.uk/teaching/".

Slooten, Joanne van, Sara Jahfari, and Jan Theeuwes (Nov. 2019). "Spontaneous eye blink rate predicts individual differences in exploration and exploitation during reinforcement learning". In: *Scientific Reports* 9, p. 17436. DOI: 10.1038/s41598-019-53805-y.

Smyth, Mary and Keith Scholey (Feb. 1994). "Interference in immediate spatial memory". In: *Memory cognition* 22, pp. 1–13. DOI: 10.3758/BF03202756.

Steinke, Alexander, Florian Lange, and Bruno Kopp (Sept. 2020). "Parallel Model-Based and Model-Free Reinforcement Learning for Card Sorting Performance". In: *Scientific Reports* 10. DOI: 10.1038/s41598-020-72407-7.

Stepp, Nigel, Anthony Chemero, and Michael T. Turvey (2011). "Philosophy for the Rest of Cognitive Science". In: *Topics in Cognitive Science* 3.2, pp. 425–437. DOI: https://doi.org/10.1111/j.1756-8765.2011.01143.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2011.01143.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2011.01143.x.

Stoet, G. (Nov. 2010). "PsyToolkit: a software package for programming psychological experiments using Linux". In: *Behav Res Methods* 42.4, pp. 1096–1104.

Sutton, Richard S. and Andrew G. Barto (2018). *Reinforcement Learning: An Introduction*. Second. The MIT Press. URL: http://incompleteideas.net/book/the-book-2nd.html.

Szepesvari, Csaba (2010). *Algorithms for Reinforcement Learning*. Morgan and Claypool Publishers. ISBN: 1608454924.

Tan, Elvis W. S., Stephen Wee Hun Lim, and Emmanuel Manalo (2017). "Global-local processing impacts academic risk taking". In: *Quarterly Journal of Experimental Psychology* 70.12. PMID: 27778753, pp. 2434–2444. DOI: 10.1080/17470218.2016.1240815. eprint: https://doi.org/10.1080/17470218.2016.1240815. URL: https://doi.org/10.1080/17470218.2016.1240815.

Thomson, Paula and S. Jaque (Dec. 2017). "Self-regulation, emotion, and resilience". In: pp. 225–243. ISBN: 9780128040515. DOI: 10.1016/B978-0-12-804051-5.00014-7.

Toepper, M et al. (2010). "Functional correlates of distractor suppression during spatial working memory encoding". In: *Neuroscience* 165.4, pp. 1244–1253.

Unsworth, Nash and Randall W. Engle (2005). "Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven". In: *Intelligence* 33.1, pp. 67–81. ISSN: 0160-2896. DOI: https://doi.org/10.1016/j.intell.2004.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0160289604000959.

Van Slooten, Joanne C. et al. (Sept. 2017). "Individual differences in eye blink rate predict both transient and tonic pupil responses during reversal learning". In: *PLOS ONE* 12.9, pp. 1–20. DOI: 10.1371/journal.pone.0185665. URL: https://doi.org/10.1371/journal.pone.0185665.

– (Nov. 2018). "How pupil responses track value-based decision-making during and after reinforcement learning". In: *PLOS Computational Biology* 14.11, pp. 1–24. DOI: 10.1371/journal.pcbi.1006632. URL: https://doi.org/10.1371/journal.pcbi.1006632.

Vandierendonck, André et al. (2004). "Working memory components of the Corsi blocks task". In: *British journal of psychology* 95.1, pp. 57–79.

Watanabe, Sumio (2010). "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory". In: *CoRR* abs/1004.2316. arXiv: 1004.2316. URL: http://arxiv.org/abs/1004.2316.

– (2012). "A Widely Applicable Bayesian Information Criterion". In: *CoRR* abs/1208.6338. arXiv: 1208.6338. URL: http://arxiv.org/abs/1208.6338.

Wen, Wen and Hideaki Kawabata (2018). "Impact of Navon-Induced Global and Local Processing Biases on the Acquisition of Spatial Knowledge". In: *SAGE Open* 8.2, p. 2158244018769131. DOI: 10.1177/2158244018769131. eprint: https://doi.org/10.1177/2158244018769131. URL: https://doi.org/10.1177/2158244018769131.

Whyte, Christopher J. (2019). "Integrating the global neuronal workspace into the framework of predictive processing: Towards a working hypothesis". In: *Consciousness and Cognition* 73, p. 102763. ISSN: 1053-8100. DOI: https://doi.org/10.1016/j.concog.2019.102763. URL: https://www.sciencedirect.com/science/article/pii/S1053810019300595.

Zhang, Lei and Jan Gläscher (2020). "A brain network supporting social influences in human decision-making". In: *Science Advances* 6.34, eabb4159. DOI: 10.1126/sciadv.abb4159. eprint: https://www.science.org/doi/pdf/10.1126/sciadv.abb4159. URL: https://www.science.org/doi/abs/10.1126/sciadv.abb4159.

Zhao, Zhenyu, Radhika Anand, and Mallory Wang (2019). "Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform". In: *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 442–452. DOI: 10.1109/DSAA.2019.00059.