Online, Offline and Transfer Learning for Decision-Making

Nicholas Menghi

School of Psychology
University of East Anglia
A thesis submitted in partial fulfilment of the requirements of the
University of East Anglia for the degree of
Doctor of Philosophy.

30 September 2021

Abstract

This thesis investigates online, offline and transfer learning for decision making tasks using a combination of behavioural experiments, computational modelling and Electroencephalography (EEG). Our experiments used a new set of decision-making tasks, in which the appropriate response depended on the linear or nonlinear combination of multiple stimulus features, and were developed to have better ecological validity than many previous tasks in the literature. The first study, in chapter 2, outlines the contextual settings in which representations of the environment can be learnt online. We manipulated the temporal structure of trials, and nature of stimulus-response mappings, and showed their effects on performance and declarative learning. We fitted a Latent Cause Model (LCM) of participants behaviour and derived measures that we used to gain insight into the representations formed. In chapter 3 we used EEG to identify the multiple successive stages of representation learning preceding decisions and following feedback. We used a Computational-EEG approach in which subject-specific LCM variables were used to predict a subject's EEG data, and found evidence of feature representation in sensory regions and more complex representations in frontal regions. In chapter 4 we shifted the focus to offline learning, by examining the effect of a period of quiet wakefulness on performance in the same task. We found that quiet wakefulness significantly improved the generalization of previously learnt associations. Finally, in the last study, in chapter 5, we investigated how knowledge acquired in one task can be transferred to another. We borrowed the concept of shared subspaces from the multitask learning literature and showed that this provides a useful framework for the study of human transfer learning. Taken as a whole, the thesis shows how humans form representations online and offline, and how extracted knowledge can be transferred to new tasks.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Table of contents

List of figures

List of tables

Acknowledgements

As a sports enthusiast, writing this thesis made me feel like *that* player receiving all the attention leaving in the shadow all the people deserving credits. The PhD is like a team game, and like every respected team, it needs a good franchise. I would first like to thank the University of East Anglia for drafting me as a freshman and supporting my endeavours.

Every team has a star player; maybe not *every* team, but mine as one. That player is also the head coach of the team, my supervisor, Will Penny. You did not just give me guidance, you walked me through the game, step by step, and got your hands dirty with me every single time it was needed. A special thanks for what you have done would never be enough.

I would also like to thank my secondary coaches, Louis Renoult and Tom FitzGerald, for their support.

A special thanks to all my teammates, the gang. Luca, for the countless hours spent in the EEG room gelling participants up, in the campus kitchen debating and in our kitchens preparing delicious dishes.

Tracuzzi, for the thousands of projects, started and never concluded, for being the handyman that saved us multiple times, for showing us a fun and adventurous life out of academia and for *never* playing in the edges in 0ad.

Laura, for being a good confident, a good ear, a person I could vent with. For being **always** on top of everything, every single useless detail. For putting the team, the gang, always on your priorities, without ever displeasing anyone.

Greta, in the last year you got to see the ups and, more than the others, the downs of my paths. I can't think of a single recent event, good or bad, in which you weren't there. Thanks for always giving me a *different* perspective, for all the discussions and for not giving up.

Francesco and Elio, I don't even know where to start. You are my blind sides, my battering rams. We shared so many things on so many different levels that it would be just impossible to list them. Living with you both has been an amazing adventure. I'm glad I can count on you.

Thanks to all the people who have walked in the field with me in these years: Alessia, Valeria, Prerna, Alvaro, Marco, Caterina, Giovanni and the ones I forgot to mention.

Finally, to my family and my friends back home I grew up with. My stadium. Thanks for being that solid and immutable presence I could shelter in. That is **un-valuable**.

Thanks.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others. The research presented in Chapter 2 has been previously presented as a conference paper and in two poster formats: (Menghi and Penny, 2019, 2021). The research presented in Chapter 4 has been submitted to a scientific journal and it's currently under revision for re-submission. The research presented in Chapter 5 has been accepted in a journal (Menghi et al., 2021).

Nicholas Menghi
30 September 2021

Chapter 1

Introduction

In our daily life, we make countless decisions. These decisions are based on internal stimuli, coming from the body, and external stimuli, coming from the world around us. To make sense of the environment, we scan the flow of information, choose the useful dimensions, integrate them into actionable representations and try to make sense of the environment. More formally, every time we are presented with new sensory inputs, based on the task at hand, we either predict their associated outcome (Gluck et al., 2002), categorize them (Ashby and Maddox, 2005) or select the action that, in our experience, is associated with the outcome with the highest value (Sutton and Barto, 1998). All these tasks require compact representations of the environmental features. These representations might include, for example, the integrated inputs, the actions we perform on them, and any consequent outcome. The process by which we form them is known as representation learning (Radulescu et al., 2019, 2021).

The thesis addresses the following broad set of questions. When and how do we form new representations? Does a period of quiet wakefulness help to consolidate these representations? Are representations re-organized during this state? Can we transfer the representation of one task to help learn another? If so, how does this take place? The thesis uses a set of experimental methodologies including behavioural studies, computational modelling and Electroencephalography (EEG). As part of our investigations, we developed novel linear and nonlinear decision-making tasks which are derived from the well studied "Weather Prediction Task" (WPT) (Knowlton et al., 1994). We have used these new tasks to probe representation learning in online and offline contexts, and how acquired representations can affect the learning of a new task (transfer learning). In what follows in this chapter we define the important constructs used in the thesis and briefly summarize the content of the different chapters.

1.1 Background

1.1.1 Investigating representation learning

Representations are models that mediate the relationship between sensory inputs from the environment and the actions you perform (or decisions you make). In this thesis, we explore how learners form these representations. Cognitive scientists have been using different decision-making tasks to investigate representation learning. For example, in reinforcement learning tasks participants choose between different stimuli and observe a reward outcome. For example, three different fractal patterns may be presented and the participant should choose the one they think will lead to a rewarding outcome. Their task is to maximize reward by learning which stimuli are associated with the highest value (expected reward)(Sutton and Barto, 1998). In category learning tasks participants learn how to group stimuli into different categories (Ashby and Maddox, 2005). Similarly, in the WPT participants are asked to learn the probabilistic associations between cues and a weather outcome (sun or rain) (Knowlton et al., 1994). The WPT has been used to investigate different topics such as; classification learning in amnesic patients (Knowlton et al., 1994), the interaction and competition of different memory systems (Li et al., 2016; Poldrack et al., 2001), the emergence of declarative knowledge (Kumaran et al., 2009; Rustemeier et al., 2013) and strategy learning (Gluck et al., 2002; Meeter et al., 2006).

Task and Strategy

A review of the various strategies employed by learners is provided in Radulescu et al. (2021). They highlight the role that attention has in defining the strategy used. The features we pay attention to, drive the process of representation learning. By attending to different features we can flexibly change and adjust the strategy we are using (Schuck et al., 2015).

In the context of the WPT, a recent investigation compared two main strategies: elemental (or linear feature-based) and configural (or object-based) (Duncan et al., 2018). In the former, the representation is composed of linear combinations of single elements. Rewards are then predicted from this linear sum. In the latter, a conjunctive and unique representation is created by binding all the elements together into a configuration. Rewards are then associated with each configuration. Duncan et al. (2018) investigated the relationship between task structure and strategy used by creating two different tasks. One where the outcome could be predicted only based on the configuration. The other, where the outcome could be predicted by the linear combination of the single elements of the configuration. They showed that participant's strategies changed based on the structure of the task. They also found that half of the participants used a configural strategy even when the structure of the task permitted a (computationally simpler - see next section) elemental strategy. The same strategies were compared in different contexts: Farashahi et al. (2017, 2020) fitted a trial-by-trial model to show the dynamics of the strategy used. They showed that participants, in a reinforcement learning task, initially used an elemental strategy and then switched to a configural strategy.

Nonlinear Decision Making

In this thesis, we describe new linear and nonlinear tasks and define linear and non-linear learning strategies. This is motivated by the limitations of elemental and configural strategies. Elemental learning assumes that outcomes (or more precisely, the "log-odds" of an outcome - see Chapter 2) are always predicted by a linear combination of stimuli, which is not always the case. A fully configural learning system, on the other hand, allows for mappings with complex non-linearities, but the number of configurations grows exponentially with the number of stimuli. In the real world, tasks are never purely linear nor configural but lie in between. The evidence for this comes from studies of real world data in the fields of pattern recognition and machine learning. These models can learn flexible mappings in which the degree of nonlinearity and complexity can be tuned to the data at hand. It is this nonlinear space of intermediate complexity that we examine in the thesis.

We used a revised version of the Weather Prediction Task (WPT) to investigate representation learning. As in the original WPT, participants learnt the probabilistic association between perceptual features and a weather outcome. Differently from the original WPT, which (in its original form - (Knowlton et al., 1994)) contained arbitrary stimulus-outcome associations, we put linear or nonlinear structures into this mapping (see Fig1.1). This allowed us to address a number of new questions. For example, in Chapter 4 we ask if a period of quiet wakefulness can improve generalisation. This would not be possible with a configural learning task because knowledge of the appropriate response for one configuration tells you nothing about the appropriate response for another.

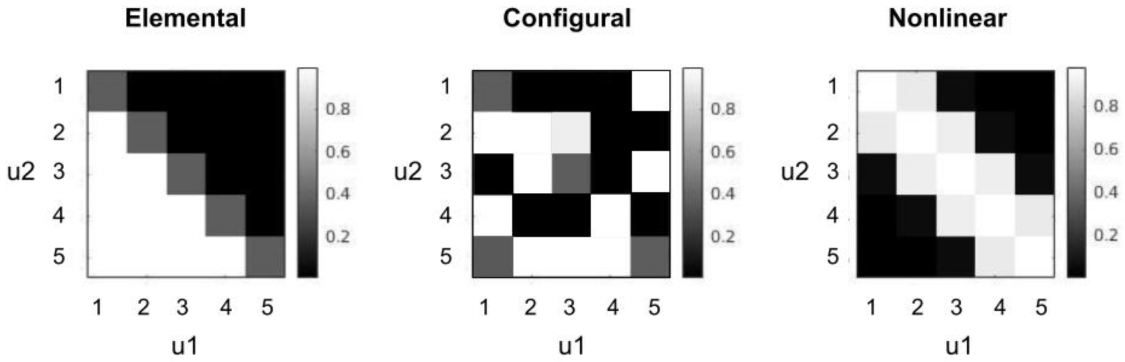


Fig. 1.1 **Task structures.** *Each image plots the reward probability given response "sun" as a function of stimulus, u. The variables u1 and u2 denote the number of perceptual features. The reward functions are generated such that the log-odds of an outcome is linear (Elemental), quadratic (Nonlinear), or maximally nonlinear such that the mapping contains no pattern (Configural).*

However, participant's performing these tasks may adopt a number of different strategies to learn these mappings. First, by directly learning a probabilistic mapping using a flexible nonlinear model (e.g. neural network). This fairly closely models the process used to define the task and is explored as a formal model in Chapter 5. Second, in the identification of homogeneous groups of exemplars, where an agent builds internal beliefs about a task, latent causes, in the shape of clusters. This is explored as a formal model in Chapters 2 and 3. Third, in the form of a verbalisable rule. This can be formalised in computational models such as COVIS (Ashby et al., 2011) and Bayesian rule learning (Ballard et al., 2018). In our behavioural experiments in chapters 2 to 5, we ask subjects if they are able to declare such a rule which would then imply that they'd been using such a strategy.

1.1.2 Online and Offline learning

Representation learning can happen online and offline. In this thesis, we define online learning as the knowledge about the task that participants acquire while doing the task. This can cover multiple stages of the decision-making process: feature perception, stimulus representation, action evaluation, feedback learning and encoding (Rangel et al., 2008). Conversely, we define offline learning as those processes that occur post encoding, during sleep or quiet wakefulness (also known as the offline wake state). Offline wake periods are those in which we disengage from the task at hand. The information we learn and encode while doing a task may be reactivated during these "offline" periods. These reactivations have been found to protect memories from forgetting (Craig and Dewar, 2018; Dewar et al., 2012), facilitate pattern recognition, leading so to faster learning and generalization (Tambini and Davachi, 2019; Wittkuhn et al., 2021). These improvements likely reflect a refinement of the representation of the task at hand.

The concepts of online and offline learning find parallels in the machine learning literature (Ng, 2017). The first refers to an algorithm where iteratively, at each trial, data inputs are used to make a prediction, and outcomes are used to update parameters of the model. Conversely, in offline learning, known also as batch learning, the model is updated using the inputs and outcomes of the entire training data set. Effectively, data and outcomes from each trial are repeatedly presented over and over (potentially for thousands or millions of iterations) until model parameters have converged to a satisfactory solution. It is this offline learning that has helped drive the recent success of Deep Learning approaches (Goodfellow et al., 2016).

1.1.3 Transfer Learning

Transfer learning, in psychology, is the learning process by which we make use of previously learnt representations, strategies and skills to learn a new task (Perkins and Salomon, 1992; Schubert et al., 2014; Steiner, 2001). There are two key concepts: positive versus negative and near versus far transfer. Positive and negative transfers refer to the effect that knowledge has on a new task in terms of performance. Positive transfer occurs when the knowledge acquired during a task improves performance in a different task. Negative, on the other hand, occurs when knowledge learnt during a previous task has a negative impact on a new task. Near and far transfer refers to the similarity of context between tasks. Near transfer occurs between two similar contexts. Far transfer, between contexts that seem very different.

The transfer learning field in psychology has homonyms in artificial intelligence. As in psychology, it focuses on improving learning in a task through the knowledge of a learnt related task (Torrey and Shavlik, 2010). This can bring a major advantage compared to the traditional machine learning algorithms. It offers the possibility of using the same trained set of parameters for multiple tasks so that parameters don’t have to be trained again. This topic is a very active area in the machine learning community (Flesh et al., 2018; Kirkpatrick et al., 2017) and integrates different approaches in artificial intelligence such as reinforcement learning (Sutton and Barto, 1998), deep-learning (LeCun et al., 2015) and multi-task learning (Caruana, 1998).

1.2 Overview of the Thesis

In Chapter 2, we look at the different contexts in which complex representations are built online. We manipulated task and temporal structure to investigate which favour the emergence of complex representations. The first define the relationship between stimuli and outcome. The second, in this thesis, define the temporal order in which stimuli are presented. For example, the same stimulus might be presented multiple times consecutively or it may be interleaved with other stimuli. We develop a computational model of online learning using a latent-cause approach that we use to gather insights into the representations that are built.

In Chapter 3, we describe an Electroencephalography (EEG) follow up study where we use measures derived from the latent-cause model to identify electrophysiological indices of representation learning. We break down the successive stages of representation learning preceding decision and following feedback. Our results are informed by findings from the number perception literature.

Chapter 4 describes a behavioural study in which we operationalized offline learning as a period of quiet wakefulness post-task. We tested participants performance before and after this period and were able to measure correct classification rates on old, previously seen, cues (an index of memorization) and new cues (an index of generalization). Differences in these measures were then compared between offline and a baseline ”active wake” condition.

Having observed how we learn complex representation online and offline, in Chapter 5 we investigate how representations can be transferred to new tasks. We borrowed concepts from the multitask learning literature (Ruder, 2017) to investigate performance in two consecutive tasks that either shared or did not share the same subspace.

1.2.1 Experimental Methods

We briefly discuss here the experimental methods used in this thesis. We used behavioural experiments where participants were instructed to learn to associate a set of visual stimuli, pairs of geometrical shapes or "pies" comprising a discrete number of "slices", to a weather outcome (sun or rain) through feedback learning, by pressing a button to express their decision. We recorded the choice they made, the reaction time and the feedback they received.

In the study described in chapter 3, we measured the electrical activity of the brain through EEG. EEG measures voltage fluctuations across multiple electrodes distributed over the scalp that are driven by changes in electrical currents in cortical pyramidal cells in the brain (Rugg and Coles, 1995). EEG data are typically described by time and frequency. A common analysis describing the time component is the Event-Related Potential (ERP) which is obtained by averaging the EEG activity time-locked to an event, like stimulus onset or feedback presentation (Cohen, 2014). Frequency analysis, on the other hand, describes rhythmic activity. These are usually obtained by running spectral analyses such as Fourier or Wavelet transforms (Cohen, 2014). We placed 64 channels using elastic caps on the scalp of participants and recorded brain activity while participants performed our behavioural task.

In chapters 2, 3 and 5 we used computational modelling to gain insight into the representation learning process. In chapters 2 and 3 we use a latent cause model. This model assumes that participants are learning the latent, hidden, causal structure of the environment (Gershman et al., 2010) (effectively grouping together similar trials into clusters). On each trial, the model either creates a new cluster or updates parameters of existing clusters (based on a probabilistic similarity measure) (Sanborn et al., 2010). This model gave us insight into the complexity of the representations formed.

In chapter 5 we compared our new behavioural findings on transfer learning to the behaviour of a neural network model. Neural networks are a framework of machine learning inspired by biological neurons. In cognitive neuroscience they are used to simulate, mimic, the brain processes through a set of algorithms (Saxe et al., 2020). They get inputs, train themselves to find patterns in data and predict the outputs (Yang and Wang, 2020). We build a neural network model able to do two tasks consecutively. These two tasks, as in the behavioural tasks, were divided into several blocks each. The neural network parameters were updated at the end of each block through Sequential Bayesian Learning, as defined by Kirkpatrick et al. (2017). This is a Bayesian algorithm (Bishop, 2006) in the sense that the prior distribution over parameters in one block is given by the posterior distribution from the previous block.

Chapter 2

The effects of temporal and task structure on representation learning

2.1 Introduction

Learning helps us creating representations of the world, which we use to select the right behaviour for each context. Representations are a function of abstracted sensory data, or "features". The representations we build might be as simple as the selection of an individual feature (and suppression of others), or more complex representations that allow separating and manipulating the features to abstract them into clusters (Badre et al., 2021). These representations are generally compressed (having a lower dimension than the original sensory input) and provide an efficient mapping between observations and outcomes (Radulescu et al., 2021). We can, for example, learn that there is a relationship between the colour and size of a fruit and its ripeness without having to learn every single combination of the two features and the related outcome. The complexity of the representation we build depends on several factors. Here we briefly review and test the separate and joint effect that the nature of the task at hand, and the temporal structure of the task, have on the representations that are created.

2.1.1 Task Structure and Declaration

A body of research shows how the complexity of the representations people employ adapts to the structure of the task at hand. Reinforcement learning studies showed how, based on the structure of the task, participants tended to use different strategies

(Duncan et al., 2018; Farashahi et al., 2017). Complex representations allow responses from previously encoded patterns to be used for new, similar patterns, thereby providing generalization. This benefited accuracy and learning efficiency in reinforcement learning task (Badre et al., 2010). The knowledge gained during representation learning can become explicit and declarable. Structure declaration has been found to improve performance in a category learning task (Kumaran et al., 2009) and reduce reaction time in a perceptual decision-making task (Schuck et al., 2015) and in a sequential associative learning task (Rose et al., 2010).

2.1.2 Temporal Structure

Another body of research focuses on the effect of the presentation order (temporal structure) of the stimuli (and outcomes) on performance during category learning tasks.

Interleaving stimuli by increasing delays between a stimulus and its repetition tend to cause the encoded stimulus to be forgotten, with a subsequent harder recall (Vlach, 2014). Interleaving stimuli allows for new stimuli to be encoded during this lag. Recalling and comparing encoded stimuli might be beneficial for pattern recognition (Carvalho and Goldstone, 2015; Zulkiply and Burt, 2013) hence for the creation of complex representations, useful for generalization, and performance (Kornell and Bjork, 2008; Kornell et al., 2010; Wahlheim et al., 2011).

Blocking stimuli together by reducing the repetition delay facilitates learning, encoding, recall of specific stimuli (Wulf and Shea, 2002; Xue et al., 2010) and increases familiarity (Dellarosa and Bourne, 1985). Since the stimuli encoded are repeated there is no variation. This leads to simpler representations being developed (Carvalho and Goldstone, 2015; Zulkiply and Burt, 2013) at the cost of worse pattern recognition performance (Carvalho and Goldstone, 2015; Wulf and Shea, 2002).

2.1.3 Learning Task

We designed an experiment to jointly assess the effect of task structure and temporal structure on category learning and declaration during learning. We used a revised version of the Weather Prediction Task (WPT) in which participants learnt the association between configurations of geometrical shapes and a weather outcome (sun or rain). Differently from the original WPT (Knowlton et al., 1994) which contained arbitrary stimulus-outcome associations we put a structure in this mapping. Participants could either build a simple representation learning all the one-to-one mappings for each stimulus-outcome pair or learn how to manipulate the features to build a more efficient representation of intermediate complexity.

All the participants did two tasks with two different structures: referred to as "subtraction" and "addition" - see below. We expected that the "subtraction" structure would be easy to declare, and that the other would not be. Three different groups of participants each learnt one of three different temporal structures: one where the order was interleaved, one where it was blocked and one in which it was mixed (blocked first, then interleaved).

2.1.4 Latent Cause Model

In the current study, we explain participants behaviour through an online latent cause model. The model compares the perceptual features of the new input to the encoded ones. It then picks the action with the highest probability of reward. We use the term reward to define correct feedback, as in Duncan et al. (2018). Stimuli are encoded in clusters defined along two continuous dimensions (the perceptual features of the stimuli) and two categorical (the decision made, and the reward received). The model is initiated with zero clusters. Trials with similar stimuli on which similar decisions were made and similar rewards received are then likely to be clustered together. If trials don't fit into any cluster, a new cluster is created. If participants learn a unique response for each unique stimulus configuration, the approach will resemble "configural reinforcement learning" (Duncan et al., 2018). If, however, participants are able to generalise across trials more useful representations of intermediate complexity (fewer, broader clusters) will arise. Our algorithm also defines a mechanism which prunes un-rewarded clusters. We operationalized the number of clusters created as an independent measure of the complexity of the representation. Furthermore, we derived two measures from the model: entropy and recognition. The recognition and entropy measures show, at each trial, respectively, the probability of a stimulus given the model and the uncertainty about which cluster a stimulus belongs to (Davis et al., 2012a).

2.1.5 Hypotheses

We hypothesized that performance and the representation formed will vary as a function of task and temporal structures (i). Task structure would have an impact on the formation of complex representations and structure declaration, hence a performance improvement. The difference between task structures would reduce when stimuli are presented in a blocked structure. This difference would be reduced also between declarative and non-declarative participants (ii). Finally, the mixed structure should show a difference in performance and recognition between declarative and non-declarative participants (iii). We expect them to be higher for declarative participants.

2.2 Materials and Methods

2.2.1 Participants

A total of sixty volunteers from the University of East Anglia (mean age = 20.83, SD = 3.18, 18 male) participated in the experiment. All of them were naive to the purpose of the experiment. All participants gave informed written consent, and the study was approved by the local institutional review board of the University of East Anglia, UK. At the end of the experiment, participants received course credits for their participation.

2.2.2 Apparatus and Stimuli

The experiment was performed in a dimly lit room in which participants were seated 60 cm away from the display with their head supported by a chin-rest. Stimuli were created and presented on a 23-inch HP Elite Display 240c monitor using the Psychophysics Toolbox (http://psychtoolbox.org/) (Brainard, 1997) for Matlab (Mathworks) running on Windows 7. Two geometrical shapes (1° x 1° visual angle) each having three, four, five, six or seven sides, for a total of twenty-five combinations, were displayed at 1° from the central fixation point (see Fig.2.1). The stimuli were presented on a dark grey background.

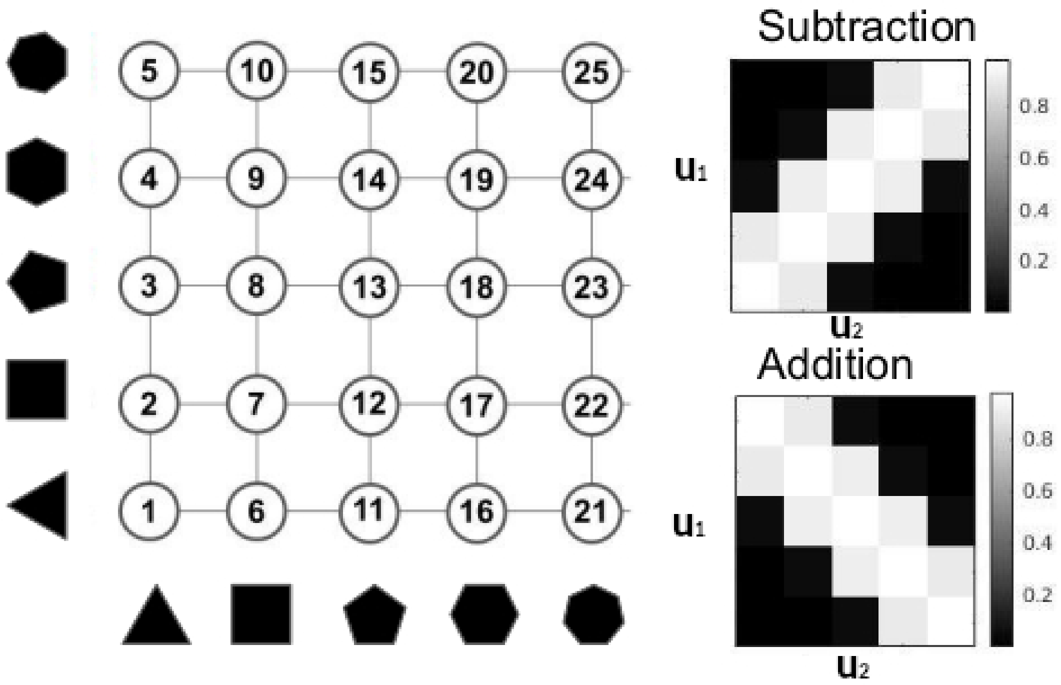


Fig. 2.1 **Experimental Stimuli and Stimulus Reward Mapping.** *The left panel depicts all the stimuli: each shape on the y-axis (presented to the participant on the left side of the computer screen) should be combined with each shape on the x-axis (right side of computer screen), creating 25 potential configurations. The right panel depicts gray scale image plots of the reward probability (given button press "sun"), as a function of stimulus, u. The variables* u_1 *and* u_2 *denote the number of sides in the left and right geometrical stimuli, respectively.*

2.2.3 Procedure

The experiment was composed of two consecutive tasks with two different mappings, which we refer to as "addition" or "subtraction", in counterbalanced order. Note that the response of the subject was not to report the sum or difference in the numbers of sides, but to base their decision on that feature, see Fig.2.1. As shown in Fig.2.2, each trial started with a black fixation cross presented at the center of the screen for 1000 ms. Afterwards, the stimuli appeared and stayed on screen for 2000 ms maximum or until response. Response was made on a standard keyboard, the "g" indicated sun prediction and "j" indicated rain. Responses not given within the required time constitute ”missed trials”. Right after button press, confirmation of the choice was given for 500 ms. Finally, feedback was provided, indicating "correct" if the prediction was correct, "incorrect" if it was not and "too slow" if they did not respond within 2000ms. In order to test participants knowledge about the task, at the end of each task, we asked them how they approached it and at which point in time they started approaching it that way. At the end of the experiment, we probed participants with two questionnaires, one per task, in order to assess their explicit knowledge. The experiment took about one hour to complete.

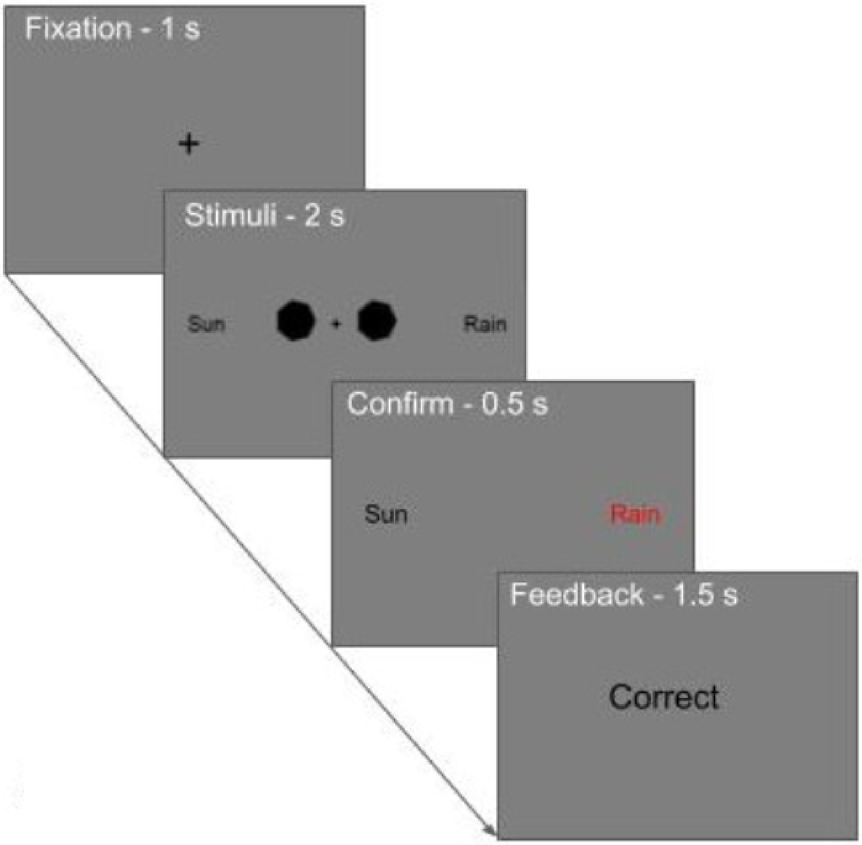


Fig. 2.2 **Trial structure.** *Each trial started with a fixation cross. Afterwards, two geometrical shapes appeared and participants had up to 2 sec to respond. Confirmation of the choice was then given and feedback was provided.*

2.2.4 Task Structure

The probabilistic structure of these two tasks was operationalized by making the log-odds of the outcome a quadratic function of stimulus characteristics, the number of sides.

$$\begin{aligned}
\log\left[\frac{p(y_t=1)}{p(y_t=0)}\right] &= (u_t-\mu)^T W(u_t-\mu)+w_0 \qquad (2.1)\\
W &= 2.4\times\begin{bmatrix} -0.71 & w_d \\ w_d & -0.71 \end{bmatrix}\\
\mu &= [3,3]^T\\
w_0 &= 2\\
u_t &= [u_t(1),u_t(2)]^T
\end{aligned}$$

Flipping the sign of w_d in this mapping produced either the addition or subtraction mapping where $w_d = 0.71$ produces the subtraction map and $w_d = -0.71$ produces the addition map. If the number of sides are the same it is sun. So, if by subtracting the number of sides in one shape from the other you get zero it is sun (subtraction). On the other hand, the addition task requires participants also identify a reference value, 10 (addition). Participants then needed to compare the sum of the number of sides to this value.

In terms of a verbalisable strategy, the subtraction mapping can be described using a single logical clause (are the number of sides equal?). The addition mapping also requires comparison to a reference value (is the sum of the number of sides equal to 10?). Thus, we might expect more participants to declare the "logically simpler" subtraction mapping.

2.2.5 Temporal Structure

Cues were presented with three different temporal structures. The first was generated as in the standard weather prediction task (WPT) (Knowlton et al., 1994) such that the probability of the kth cue pattern occurring was uniform over trials, $p(u_t = k) = 1/T$ where $T = 250$ is the number of trials per task. We refer to this as an interleaved structure. The second structure was generated to create a blocked design such that the probability of the kth cue pattern occurring was concentrated within an interval. We used the distribution $p(u_t = k) = \mathsf{N}(u_t;\mu_k,\tau)/\sum_j \mathsf{N}(u_t;\mu_j,\tau)$ where $\mathsf{N}(u;\mu,\tau)$ indicates a normal distribution with mean μ and precision τ. The third temporal structure was a mixture of the first two. The first 175 trials were generated from the blocked distribution, the last 75 from the interleaved one.

2.2.6 Experimental Design

We assess the effect of different stimulus-response mappings and temporal structures with a 2x3 mixed design. All the participants did two tasks, with both addition and subtraction stimulus-response mappings (within factor). Each task was composed of 250 trials (10 repetitions per configuration) divided into 5 blocks. Three groups of 20 participants each did the three different temporal structures (between factor). Given that participants are required to make Sun/Rain decisions and learn incrementally via feedback, this is reminiscent of the classic Weather Prediction Task. However, a major difference is that in our tasks there is a hidden structure in the stimulus-reward mappings that can be discovered by participants (See Chapter 1).

2.3 A Latent Cause Model of Cues, Rewards and Actions

The model we used in this paper is a latent cause model in which new observations are viewed as originating from a latent cause (cluster) made of input features, actions and reward (Gershman et al., 2010). The number of clusters that the model creates is unbounded, like the rational model by Sanborn et al. (2010). The model starts with a single cluster and at each trial, it can flexibly create a new cluster or update the ones created before based on the similarity of the new event with the encoded ones (Niv, 2018). The algorithm we use is similar to the Fast Gaussian Mixture algorithm by Pinto and Engel (2015) but revised to accommodate for the multinomial observations (the various decision-reward combinations) and the removal of clusters. In our model, a cluster is removed whenever the probability of reward is lower than a fitted threshold so that encoded information with low reward probability is forgotten, thereby favouring the prioritization of frequently rewarded clusters (see Braun et al. (2018); Sterpenich et al. (2011) for the effects of reward on memory encoding).

2.3.1 The model

The joint density over rewards r_t, actions (decisions) a_t and inputs u_t is estimated using a mixture model

$$p(r_t, a_t, u_t) = \sum_{j=1}^{M} p(r_t, a_t, u_t|j)p(j) \tag{2.2}$$

where $p(j)$ is the prior probability of the jth cluster and the likelihood of data under the jth cluster is

$$\begin{aligned} p(r_t, a_t, u_t|j) &= p(r_t, a_t|j)p(u_t|j) \\ p(r_t = r, a_t = d|j) &= \beta^j_{rd} \\ p(u_t|j) &= \mathsf{N}(u_t; m_j, \Lambda_j) \end{aligned} \tag{2.3}$$

where $\Lambda_j = C_j^{-1}$. Thus, the joint probability of reward and actions given j is defined by a multinomial distribution and the probability of input stimuli given j by a Gaussian density. We therefore refer to this as a Multinomial-Gaussian Mixture (MGM). The multinomial distribution is parameterised by "count" variables as

$$\beta^j_{rd} = \frac{B^j_{rd}}{\sum_{r',d'} B^j_{r',d'}} \tag{2.4}$$

where the B^j_{rd} parameters count the number of times reward-action pairs occur (for trials that are soft-assigned to cluster j). This is equivalent to assuming a Dirichlet prior (Sanborn et al., 2010) over the multinomial variables which is recursively updated during online learning. Model parameters are $\theta = \{B, m, \Lambda\}$.

2.3.2 Online Learning

Here we describe an algorithm that estimates all relevant quantities online, except for the following. First, we must specify the "encoding precision", λ_0, which is the initial (isotropic) precision of the Gaussian clusters. Two further parameters, ϕ and ψ are fitted to behavioural data from each subject's decision time series using a Maximum Likelihood approach. Here, ϕ, governs the prior probability of creating a new cluster, with higher values leading to models with more clusters. The other parameter, ψ is a threshold for pruning clusters, with smaller values leading to less pruning.

The algorithm we describe is similar to the Fast Incremental Gaussian Mixture Model (IGMM) by Pinto and Engel (2015) except that we also include updates for the multinomial parameters. In the below, M is the number of clusters, N_j counts the number of trials (soft) assigned to cluster j the cumulative posterior distribution, and B^j_{rd} counts the number of trials for which the reward was $r_t = r$ and the decision was $a_t = d$, for trials assigned to cluster j.

The algorithm starts with an initialisation step

$$\begin{aligned} M &= 1 \\ B_0 &= 0.1 \times 1_{CK} \end{aligned} \tag{2.5}$$

where B_0 are the initial count parameters for reward-action pairs.

Online Decision

At each trial, we calculate the probability of reward given (potential) action and stimulus

$$\begin{aligned} \pi_{td} &\equiv p(r_t = 1 | a_t = d, u_t) \\ &= \frac{p(r_t = 1, a_t = d, u_t)}{\sum_{r'} p(r_t = r', a_t = d, u_t)} \end{aligned} \tag{2.6}$$

This can be evaluated as

$$\begin{aligned} \pi_{td} &= \frac{f_{1d}}{\sum_{r'} f_{r'd}} \\ f_{rd} &= \sum_{j=1}^{M} p(r_t = r, a_t = d | j) p(u_t | j) p(j) \end{aligned} \tag{2.7}$$

We then perform a softmax function

$$p(a = d) = \frac{\exp(\pi_{td})}{\sum_{d'} \exp(\pi_{td'})} \tag{2.8}$$

and take the action with the highest probability. At trial 1, the probabilities are set to $p(a = d) = 0.5$.

Cluster Creation

If none of the clusters has $p(u_t|j) > \phi p_m$ and $p(r_t, a_t|j) < .25$, (where p_m is the likelihood at the cluster mean), a new cluster is created. This is to ensure that the next cluster created is different enough from previous experiences. Naturally, a new cluster is always created for the first data point. When creating a new cluster we set

$$\begin{aligned} m_j &= u_t \\ N_j &= 1 \\ p(j) &= \frac{1}{\sum_{k=1}^{M} N_k} \\ C_j &= \lambda_0^{-1} I_D \\ B^j &= B_0 \\ B_{rd}^j &= B_{rd}^j + 1 \\ M &= M + 1 \\ v_j &= 1 \end{aligned} \tag{2.9}$$

Cluster Update

If a new cluster is not created then all existing clusters $j = 1..M$ are updated as follows.

$$\begin{aligned} \gamma_j &\equiv p(j|r_t, a_t, u_t) \\ &= \frac{p(r_t, a_t|j) p(u_t|j) p(j)}{\sum_{k=1}^{M} p(r_t, a_t|k) p(u_t|k) p(k)} \\ N_j &= N_j + \gamma_j \\ p(j) &= \frac{N_j}{\sum_{k=1}^{M} N_k} \\ \Delta m_j &= \frac{\gamma_j}{N_j}(u_t - m_j(t-1)) \\ m_j(t) &= m_j(t-1) + \Delta m_j \\ e_j &= u_t - m_j(t) \\ C_j(t) &= C_j(t-1) + \frac{\gamma_j}{N_j}\left[e_j e_j^T - C_j(t-1)\right] - \Delta m_j (\Delta m_j)^T \\ B_{rd}^j &= B_{rd}^j + \gamma_j \\ v_j &= v_j + 1 \end{aligned} \tag{2.10}$$

As N_j increases we see that the updates to both m_j and C_j both go towards zero, thus providing convergence. The computational bottleneck in IGMM is computing the likelihood $p(u_t|j)$ as this requires inverting C_j. Pinto and Engel (2015) have bypassed this bottleneck by deriving updates of precision rather than covariance matrices, so that no inversions are required. Our implementation of updates for m and Λ follow those derived in (Pinto and Engel, 2015) given by the Fast IGMM "Algorithm 2". This "Fast" implementation is only necessary if one uses full covariances for the Gaussian likelihood. For models with diagonal covariances the computation of matrix inversions and determinants is not demanding.

Cluster pruning

After cluster update, a cluster j is removed whenever the probability of reward

$$p(r=1|j) = \frac{\sum_{d'} p(r=1, a=d'|j)}{\sum_{d'} \sum_{r'} p(r=r', a=d'|j)} \tag{2.11}$$

is lower than the threshold ψ.

Fitting Model to a Subject's Behavioural Data

Here we use a Maximum Likelihood approach to fit the above model to a subject's time series of decisions (see discussion for a comparison with other approaches).

Subject-specific parameters $\theta = \{\phi, \psi\}$ are chosen in order to minimize the error between the action made by the model and action made by participants (or equivalently, to maximise the likelihood of a subjects behaviour under the model). The likelihood of a participants decisions, $Y = \{y_1, y_2, ..., y_T\}$, is

$$p(Y|\theta) = \prod_{t=1}^{T} a_{td}^{[y_t=d]} \tag{2.12}$$

The cost function J, the negative log likelihood, is then

$$\begin{aligned} J &\equiv -\log p(Y|\theta) && (2.13) \\ &= \sum_{t}^{T} [y_t = d] \log a_{td} \end{aligned}$$

Additional measures

Furthermore, we took two additional measures: entropy (E) and recognition (R) (Davis et al., 2012a). Recognition is the probability of a stimulus given the model:

$$\begin{aligned} R_{pre} &\equiv p(u_t) \\ &= \sum_j p(u_t|j)p(j) \end{aligned} \tag{2.14}$$

Entropy is the uncertainty about which cluster the new stimulus belongs to:

$$\begin{aligned} H_{pre} &= -\sum_j p(j|u_t)\log_2[p(j|u_t)] \\ p(j|u_t) &= \frac{p(u_t|j)p(j)}{\sum_{k=1}^{M} p(u_t|k)p(k)} \end{aligned} \tag{2.15}$$

Both are computed "pre-feedback" ie before the agent receives feedback.

2.4 Results

2.4.1 Empirical Results

In order to test our hypotheses, we divided the analysis into two parts. First, we looked at how participants performance varied as a function of our manipulations by performing a two-way mixed-design ANOVA with dependent variable accuracy and independent factors of task and temporal structure. Additionally, we were able to resolve each of the main effects and interactions as a function of time (trial number). We used a cluster permutation test (Maris and Oostenveld, 2007) to correct for the multiple comparisons over time.

Second, we calculated how performance varied based on whether participants declared a strategy or not. We restricted this analysis to the subtraction task because only a single participant declared the correct strategy in the addition task. We performed an independent sample t-test with dependent variable accuracy and independent factor declaration. Again, this was implemented as a function of time, and we used cluster permutation tests to correct for the associated multiple comparisons.

Effect of Task and Temporal Structures

Accuracy was computed as the correct rate over all 250 trials in each task. We performed a two-way mixed-design ANOVA with within factor task structure and between factor temporal structure. This revealed both main effects but not the interaction (Task structure, $F(1) = 21.462$, $p < 0.001$; Temporal structure, $F(2) = 3.978$, $p = 0.024$; Interaction, $F(2) = 1.026$, $p = 0.365$). As can be seen in Fig.2.3, participants performed better in the subtraction task compared to the addition task. Also, participants performance differed based on temporal structure. Participants in the blocked condition performed better than participants in the interleaved and mixed conditions. We then performed three separate paired t-tests to compare the accuracy in the addition and subtraction task for each temporal structure. Performance significantly differed for participants in both interleaved and mixed temporal structures (Interleaved, $t(19) = 3.0361$, $p = 0.007$; Mix, $t(19) = 4.086$, $p < 0.001$) but did not for participants with blocked trial presentations ($t(19) = 1.3613$, $p = 0.189$).

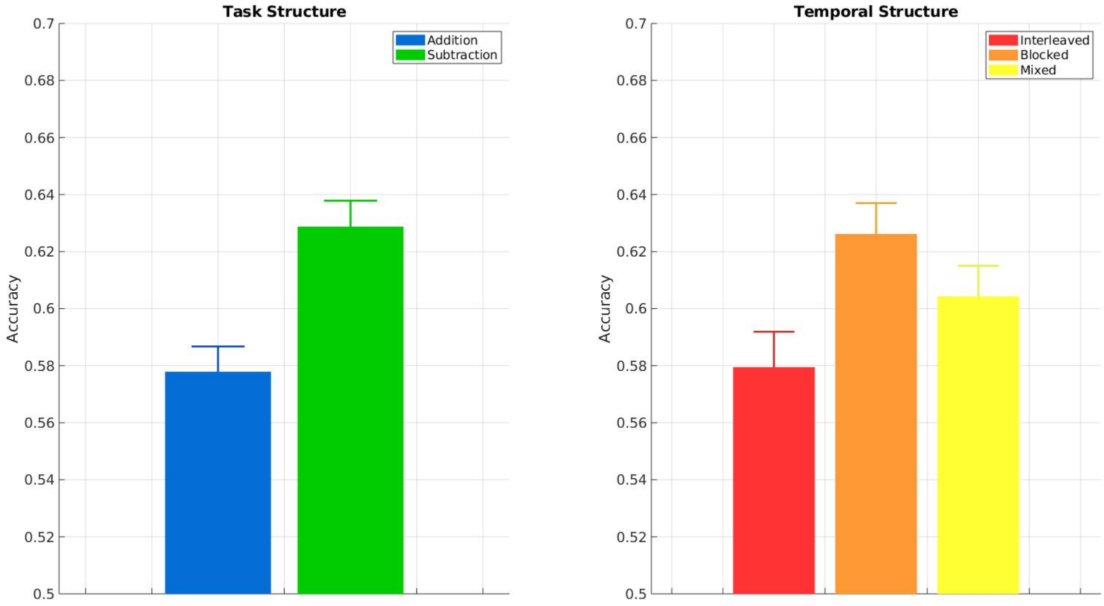


Fig. 2.3 **Accuracy per task structure.** *The left plot shows accuracy for the addition and subtraction task. Participants performed better in the subtraction task compared to the addition one. The right plot shows accuracy for each temporal structure. Participants performed better in the blocked temporal structure compared to the interleaved and mixed conditions. Error bars indicate the standard error of the mean.*

Accuracy over time

In order to further explore the difference in performance, we run a moving mean of participants accuracy with a window of 20 trials. We then performed a cluster-based permutation test to identify when participants performance differed between tasks, for each temporal structure. As shown in Fig.2.4, we found six significant clusters in the interleaved condition. Participants performed better for the subtraction task compared to the addition task. We found two small clusters in the blocked temporal structure at the beginning of the task, but otherwise performance was not different. Finally, we found three significant clusters in the mixed condition (in which the first 175 trials are blocked and the last 75 are interleaved). The biggest cluster we found was after the switch in temporal structure.

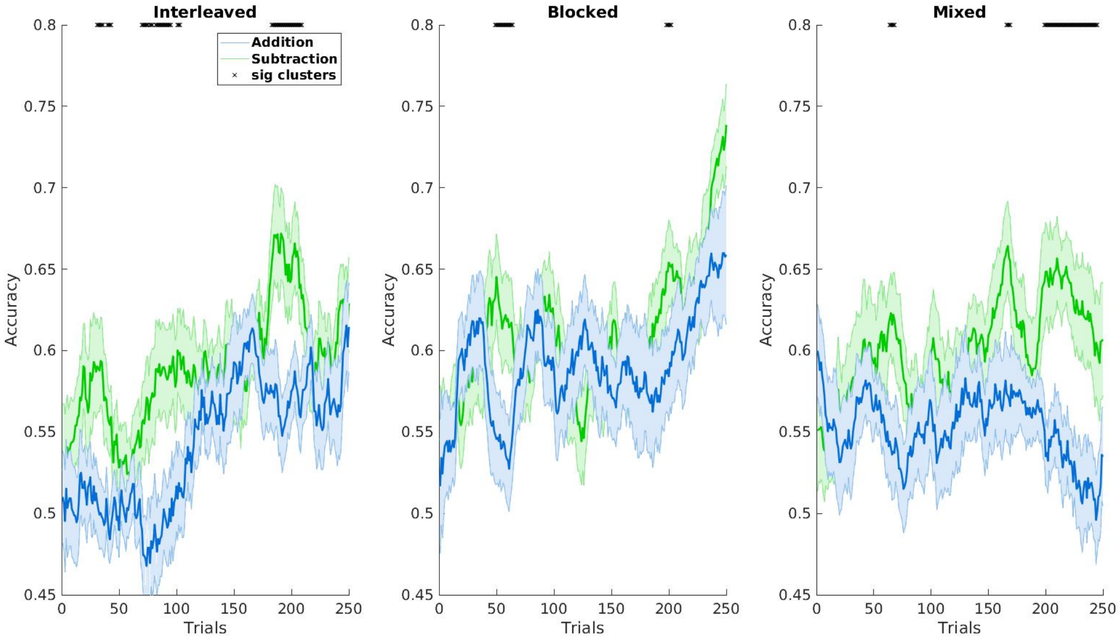


Fig. 2.4 **Accuracy over time.** *The plots show how accuracy evolves as a function of trial number for each combination of task (addition in green, subtraction in blue) and temporal structure (left, middle and right panels). Significant clusters are identified by asterisks at the top of each panel (many of which may merge together to form a continuous line). The right panel shows a significant difference in accuracy after the switch from blocked to interleaved trials (at trial 175).*

Declarations

Participants knowledge was categorized as explicit if participants were able to verbalize the hidden structure. In all other cases, it was classified as implicit. This criterion indicated that twenty-eight out of sixty participants declared the correct strategy in the subtraction task. Only one participant out of sixty made a correct declaration for the addition task - see table 2.1 for a full breakdown.

	Interleaved	**Blocked**	**Mixed**	
Subtraction	10	8	10	
Addition	0	0	1	

Table 2.1 ***Declaration*** *Number of participants who explicitly declared the correct strategy per task and temporal structures.*

Since only one participant declared the correct nature of the addition task, we limited the declaration analysis to the subtraction task. We performed an independent sample t-test comparing the performance of participants who declared the correct strategy against participants who did not. The t-test revealed that declarative participants performed better than non-declarative ones ($t(58) = 2.295$, $p = 0.025$) (See Fig.2.5).

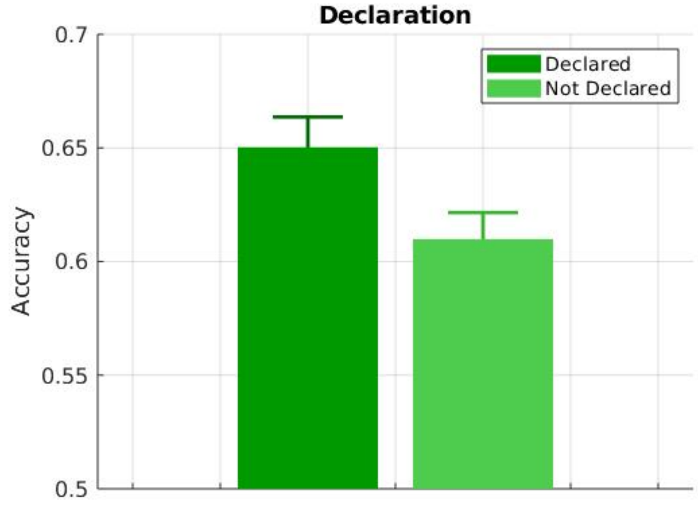


Fig. 2.5 **Accuracy per declaration.** *The plot shows accuracy difference between participants who declared and did not declare for the subtraction task. Error bars indicate the standard error of the mean.*

Declaration over Time and Temporal structure

In order to further explore the difference in performance between declarative and non-declarative participants, we run a moving mean of participants accuracy with a window of 20 trials. We then performed cluster-based permutation tests to identify when performance differed between declarative and non-declarative participants, for each temporal structure, the results are illustrated in Fig2.6.

We found four significant clusters in the interleaved temporal structure condition. Participants who declared the strategy performed better than participants who did not. We found a small significant cluster in the blocked temporal structure. Declarative knowledge of the structure of the task gave no significant advantage for blocked trials. Finally, we found a significant cluster in the mixed temporal structure condition. Performance differed between declarative and non-declarative participants around the time of the switch from blocked to interleaved trials (trial 175).

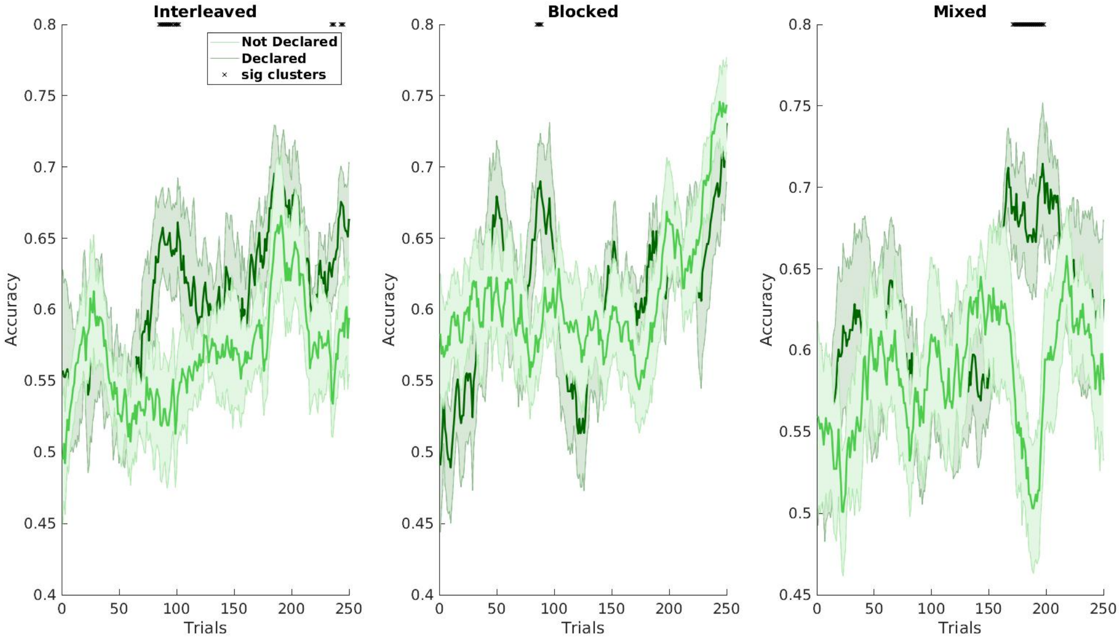


Fig. 2.6 **Declaration over time.** *The plot shows task accuracy as a function of declaration (dark green for declared, light green for non-declared) for each temporal structure (left, middle and right panels). Significant clusters are identified by asterisks on top of the each panel (many of which may merge together to form a continuous line). The right panel shows that performance differs between declarative and non-declarative participants around the time of the switch from blocked to interleaved trials (trial 175). Error bars indicate the standard error of the mean.*

2.4.2 Modelling Results

We used the latent cause model described earlier and fitted the parameters ϕ and ψ to each subjects behavioural data using the Maximum Likelihood approach. The parameters that maximised the Likelihood were identified using grid-search. The grid dimensions comprised 70 values for ϕ (linearly spaced from 10^{-9} to 10^{-3}) and 15 values for ψ (from 0 to 0.7).

We performed ANOVAs and t-tests to test if the total number of clusters, entropy and stimulus recognition were affected by our manipulations. Further analyses on the estimated parameters can be found in the appendix A.

Total Clusters

We took the number of clusters that the model identified for each participant and we performed a two-way mixed-design ANOVA with within-factor task and between-factor temporal structure, to investigate if the number of clusters varied with our manipulations. This revealed a significant main effect of temporal structure (Task structure, $F(1) = 3.309$, $p < 0.074$; Temporal structure, $F(2) = 25.445$, $p < 0.001$; Interaction, $F(2) = 0.166$, $p = 0.846$). As can be seen from Fig.2.7, participants create more clusters when trials are blocked.

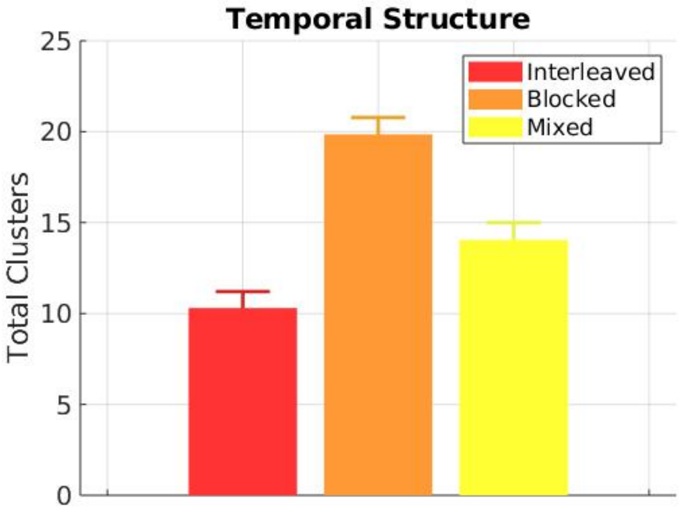


Fig. 2.7 **Total Clusters per Temporal Structure.** *The plot shows the average total number of cluster per temporal structure. Error bars indicate the standard error of the mean.*

Entropy and Recognition

We performed a two-way mixed design ANOVA with dependent variable recognition and independent within-factors of task and temporal structure. This revealed a significant main effect of temporal structure (Task, F(1) = 0.175, $p < 0.676$; Temporal structure, F(2) = 260.23, $p < 0.001$; Interaction, F(2) = 0.022, p = 0.978), see Fig.2.8. We repeated this analysis but with entropy as the dependent measure, but found no significant results (Task, F(1) = 3.489, $p < 0.066$; Temporal structure, F(2) = 2.406, p < 0.099; Interaction, F(2) = 1.557, p = 0.219).

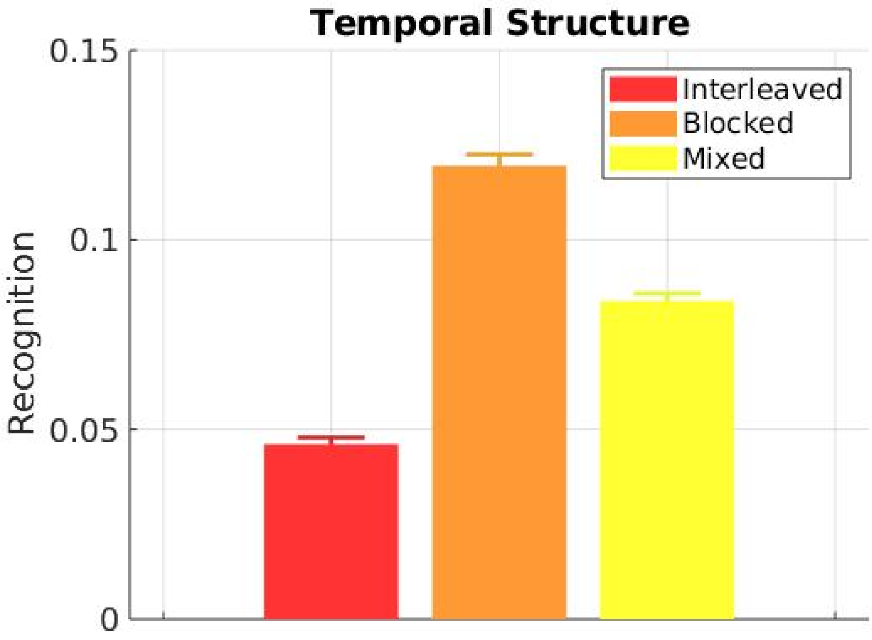


Fig. 2.8 **Recognition measure per Temporal Structure.** *The plot shows the recognition measure per temporal structure. Cues are more strongly recognised when the trials are blocked. Error bars indicate the standard error of the mean.*

Declaration and Recognition

We furthered explored the difference in recognition between declarative and non-declarative participants, also as a function of temporal structure. We run a moving mean of the recognition value with a window of 20 trials. We then performed a cluster-based permutation test to identify where recognition differed. We found four significant clusters in the blocked temporal structure. As illustrated in Fig.2.9, participants who declared the correct strategy had a higher recognition measure in the first half of the task. We also found two significant clusters in the Mixed temporal structure, one at the beginning of the task, and one after the switch to the interleaved temporal structure. We found no significant clusters in the interleaved temporal structure.

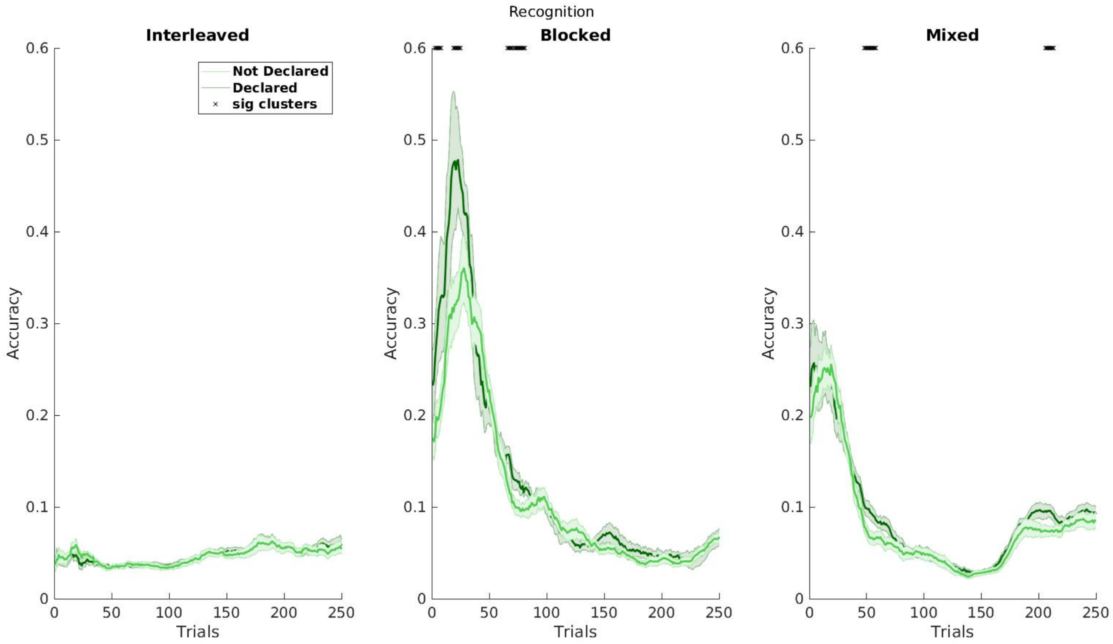


Fig. 2.9 **Difference between declarative and non-declarative participants in Recognition over time per Temporal Structure.** *The plot shows the difference in recognition between declarative (dark green) and non-declarative (light green) participants as a function of temporal structure. Significant clusters are identified by asterisks on top of the figure (which merge into lines in places). Cues are more strongly recognised by declarative subjects at the beginning of learning when trials are blocked (middle panel). Error bars indicate the standard error of the mean.*

2.5 Discussion

2.5.1 Task Structure

We found evidence in support of our hypotheses that participants performance and the number of declarative participants would be higher for the subtraction task compared to the addition task.

This was evident from several analyses. First, our ANOVA revealed a main effect of task structure driven by a better performance in the subtraction task as illustrated in figure 2.3. Second, a total of 28 participants declared the correct task structure in the subtraction task but only 1 participant did so for the addition task (see Table 2.1). Declarative participants performed better than non-declarative ones.

2.5.2 Temporal Structure

We found evidence in support of our hypothesis that participants performance differs based on the temporal structure. First, our ANOVA revealed a main effect of temporal structure driven by a better performance in the blocked condition compared to the mixed and interleaved ones.

We argue that, over the time scale of several trials, the reduced number of stimuli to encode and recall would benefit performance for non-declarative participants. This may be because these associations could be held in working memory (Wulf and Shea, 2002).

Second, the model revealed that the total number of clusters created was higher for the blocked temporal structure suggesting that blocking stimuli together favour the creation of a simple representation where each distinct cue (and its association with reward) is represented as a single cluster. This is clear in the example provided in Fig.2.10 where the clusters produced in the interleaved and blocked temporal structure are compared. This is consistent with the finding that recognition was also higher for the blocked structure. In the LCM, repeats of the same cue lead to a tightening of the Gaussian around the repeated stimulus values, consistent with the notion of sharpening of neuronal responses (Gerstner and Kistler, 2002).

The temporal structure did not affect the number of participants who declared the correct strategy. The mixed condition showed the difference in performance between declarative participants that used a complex representation and non-declarative participants who used a simple one. When the temporal structure switched from blocked to interleaved (trial 175) performance of non-declarative participants decayed showing how the information encoded during the first part was forgotten. This is probably due to the limited capacity of working memory, once new stimuli are presented they overwrite the previous ones. This appears to be confirmed by the recognition measures showing how, after the temporal structure switch in the mixed condition, declarative participants have a higher recognition.

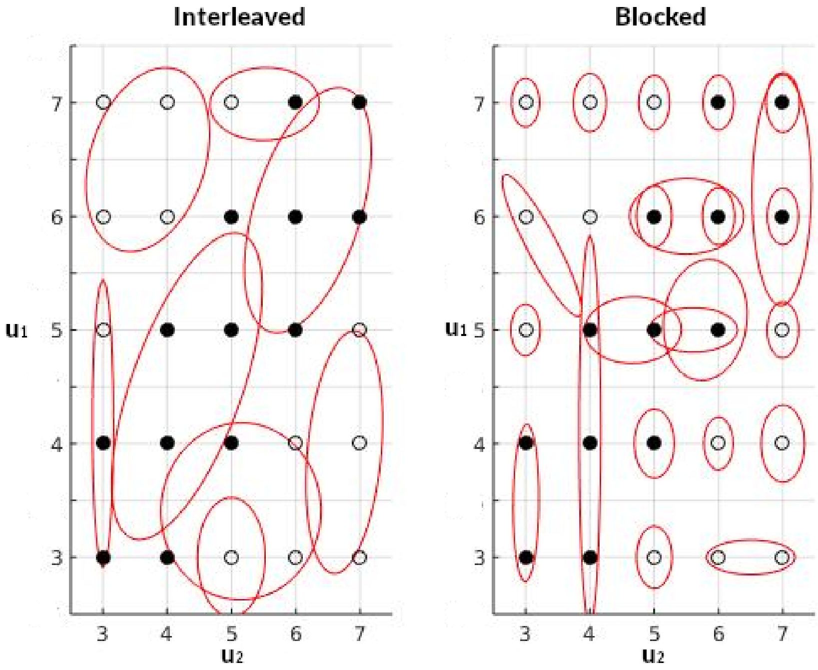


Fig. 2.10 **Number of clusters produced in interleaved and blocked structure.** *The plots show the clusters produced by the model in the subtraction task for two representative participants, one in the interleaved and the other in the blocked temporal structure condition. The dots represent the data points, black dots are associated with the response "sun", white dots with "rain".*

2.5.3 Interaction between Temporal and Task Structures

We found evidence of interaction between task and temporal structure. Our post-hoc tests revealed that participants performance between subtraction and addition structures differed in the interleaved and mixed conditions but not in the blocked one. We interpreted this as further evidence of the benefit of the blocked structure for non-declarative participants.

Our results extend the ones by Noh et al. (2016) who showed the different benefits of interleaving and blocking stimuli in different task structures during category learning. They found that blocking stimuli helped participants in identifying the relevant features whereas interleaving helped them to distinguish between categories.

2.5.4 Model insights and future directions

The latent cause model we used provided insight into the number of clusters created, recognition and entropy measures. We interpreted the total number of clusters as an index of the complexity of the representation created. The bigger the number of clusters the simpler the complexity of the representation. Recognition was higher for the blocked versus interleaved condition as participants observed repeated stimuli closely spaced in time. It was also different for declarative and non-declarative participants in the mixed temporal structure, showing a better recognition for the former compared to the latter. Our model results describe the context in which representation learning occurs, thereby providing a foundation for neuroimaging studies, where derived computationally-derived measures can be used as regressors (Love, 2020). Model-based analysis in neuroimaging offers the possibility of testing novel hypotheses as to the role of the medial temporal lobe in forming representations (Davis et al., 2012b), or the role of the hippocampus and ventral striatum in retrieving stored representations and assigning stimuli to clusters (Davis et al., 2012a).

In this chapter, we have taken a rather unsophisticated approach for fitting a computational model to behaviour, one based on grid search. This was motivated by the fact that we were fitting only two parameters so grid search was computationally feasible. More sophisticated methods, such as hierarchical Bayesian modeling (Daw, 2011; Lee and Wagenmakers, 2014), estimate parameters at the group and single subject levels, allowing group-level estimates to constrain subject-specific estimates, thereby reducing variance.

2.6 Conclusion

In this chapter, we tested the effects of temporal and task structure on representation learning and declaration. Blocking same-stimulus trials encourages participants to create a simple representation and benefits categorisation performance and stimulus recognition for both declarative and non-declarative participants. Interleaving trials (interval structure) discouraged simple representations and benefitted declarative participants only. We used the mixed temporal structure to show the differences in recognition and performances between declarative and non-declarative participants. We showed how, after the switch from blocked to interval temporal structure, declarative participants performance and stimulus recognition was higher than for non-declarative participants. This might be the result of the decay of information due to working memory limits. Many of these results are perhaps not too surprising but they do provide face validity of the computational modelling approach.

Chapter 3

The Electrophysiology of Representation Learning

3.1 Introduction

Decisions we make usually depend on abstract representations of the environment. Lower-level perceptual features are combined and manipulated in order to create, encode and recall abstract representations. This has the advantage of providing a compact and efficient mapping between observations and decisions (Radulescu et al., 2021). Our goal in this chapter is to decompose the successive stages of representation processing from stimulus onsets to feedback processing using electroencephalography (EEG) clarifying the temporal dynamics of lower level feature perceptions and more abstract representations (see also Hubbard et al. (2019)).

3.1.1 Multiple processes

Multiple cognitive processes unfold over time during decision making (Kikumoto and Mayr, 2020). One of the earliest stages of processing during associative learning tasks is to identify the stimulus, the dimension along which it varies and its properties. Different stimuli might be then combined and compared with existent representations in memory. In model-based perspectives, such as from Latent Causal Models, this abstraction can then be assigned to an existent cluster. Or based on decisions and feedback, a new cluster might be created. In the task presented in the previous chapter, participants had to combine two features into a single representation, in order to predict an outcome and make a decision. Here we refer to this representation as the feature subspace. More generally, a subspace is a mathematical concept capturing the idea that, although stimuli may be high-dimensional and be presented to a participant in this high-dimensional space, it is only some lower dimensional part of this space that will be useful for the task at hand. This is the subspace (Braun et al., 2010; Menghi et al., 2021). In this chapter the subspace is one-dimensional and comprises the single feature (combination of stimuli) that is useful for performing the nonlinear decision making task.

Analogously to the sequence of mental states in our associative learning task, previous studies have identified the multiple stages underlying elementary arithmetic tasks (Avancini et al., 2015; Fornaciai et al., 2017; Fornaciai and Park, 2018; Pinheiro-Chagas et al., 2019). Shortly after the presentation of operands, the visual cortex can locate each single operand and identify their numerical magnitude over the number line so that other areas can compute the operations between them (Mazza and Caramazza, 2010; Piazza, 2015).

3.1.2 Model-based EEG

This chapter uses a "model-based EEG" or "Computational EEG" approach to identify separable processes related to associative learning (Collins and Frank, 2018). This is a special case of the wider field of computational neuroimaging which fits computational models to behaviour to identify putative signals underlying neural computations and then uses imaging data (EEG, MEG, fMRI) to localise that activity (Huys et al., 2016; Stephan et al., 2017). Here, we use the latent cause model presented in the previous chapter to identify the various active processes during stimulus and feedback processing. We compute the probability of a stimulus under the model (Recognition) and uncertainty about which cluster a stimulus belongs to (Entropy) during stimulus presentation. These measures allow us to show when, in time, we recall and compare the new stimulus to the encoded ones, to make a decision. We then derive from the same model four new measures. We compute Recognition and Entropy after the onset of the feedback during encoding time, given action and reward. Finally, we also compute the updates from stimulus to feedback of these two measures. These measures allow us to show the encoding of the new stimuli after feedback is received and how the previous information is updated.

3.1.3 Learning Task

We designed an experiment that allowed us to assess the multiple cognitive processes that unfold over time: feature perception, extraction of a feature subspace, the recall of encoded elements, comparison with the new stimulus cues and the final encoding or update of a new cluster.

We used a revised version of the Weather Prediction Task (WPT) in which participants learnt the association between configurations of graphical pies and a weather outcome (sun or rain). However, a major difference is that in our tasks there is a hidden structure in the stimulus-reward mappings that can be discovered by participants (See Chapter 1) so that participants could build a complex representation out of the single features.

All the participants did two tasks with two different structures that required two different manipulations, hence subspace transformation, to produce the single discriminatory feature.

3.2 Materials and Methods

3.2.1 Participants

A total of 25 volunteers from the University of East Anglia (mean age = 20.88, SD = 4.94, 7 male) participated in the experiment. All of them were naive to the purpose of the experiment. Data from one participant became unavailable due to EEG-computer synchronization errors. A further participant was discarded because of performance below chance level in both tasks. We performed our analysis on the remaining sample of 23 participants (mean age = 20.86, SD = 5.15, 6 male). All participants gave informed written consent, and the study procedure was approved by the local institutional review board of the University of East Anglia, UK. At the end of the experiment, participants received course credits for their participation.

3.2.2 EEG Acquisition and Preprocessing

BrainProduct actiCAP was used to record EEG signals from 63 electrodes plus one additional electrode used as a horizontal electro-oculogram (hEOG). EEG electrodes were placed following the standard 64-channel arrangement, FT9 was used as hEOG and FT10 as Iz. All electrode impedances were kept below 25 $k\Omega$. EEG signals were recorded at a sampling rate of $1000Hz$. Preprocessing was carried out using Fieldtrip toolbox for MATLAB (Oostenveld et al., 2011).

Continuous data was highpass filtered at $0.1Hz$ and re-referenced to the common average. The data were epoched from 500 ms before the onset of the stimulus to 1.5 s following it and from 500 ms before the onset of feedback to 1.5 s following it. We visually inspected these epochs to remove trials containing muscular activity, electrical artifacts and to identify bad electrodes which were interpolated to the weighted average of neighbouring electrodes. A maximum of 2 non-neighbouring electrodes were interpolated per participant. We interpolated one electrode for 9 participants, two for 4 participants and none for the remaining 12.

Fast Independent Component Analysis (fastICA) (Comon, 1994) was then performed on the epoched data. ICA components were visually inspected to reject eye blinks, eye movements and sustained high-frequency noise. Furthermore, we performed baseline correction based on the average of the whole epoch, as the period pre stimulus onset was meaningful. EEG epochs were then low-pass filtered with a cut-off of $100Hz$, notch filtered at $50Hz$ to remove mains artefact, and downsampled to 500Hz. Finally, we visually reinspected the epochs to ensure no artifact remained. Rejected trials and EOG signal were excluded from all further analyses.

3.2.3 Apparatus and Stimuli

The experiment was performed in a dimly lit room with participants seated 60 cm away from the display with their head supported by a chin-rest. Stimuli were created and presented on a 23-inch HP Elite Display 240c monitor using the Psychophysics Toolbox (http://psychtoolbox.org/) (Brainard, 1997) for Matlab (Mathworks) running on Windows 7. Two virtual ”pies” (1 x 1 degrees of visual angle) were displayed at 1 degree from the central fixation point. Each pie was divided into six slices with from one up to five slices that could be filled with red colour, making a total of twenty-five combinations, as shown in Figure 3.1. The stimuli were presented on a dark grey background.

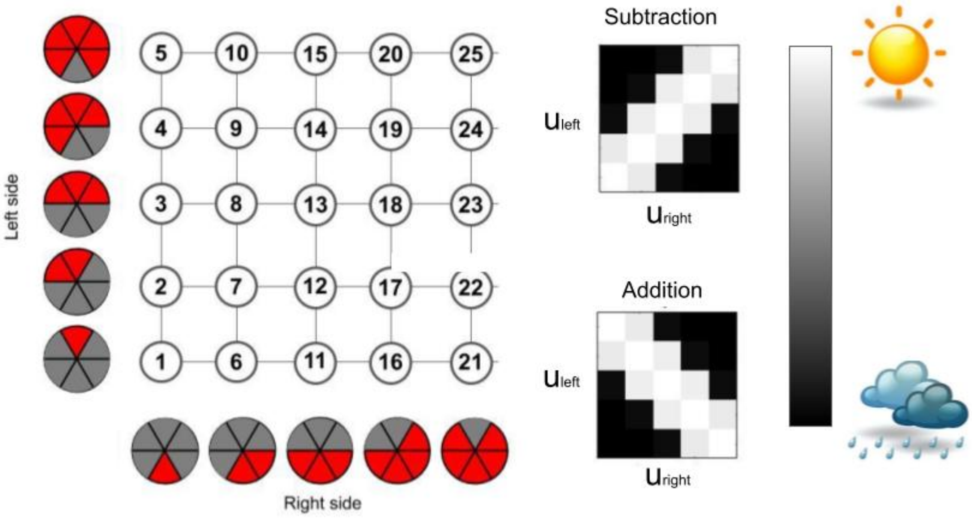


Fig. 3.1 **Experimental Stimuli and Stimulus Reward Mappings.** *The left panel depicts the stimuli: the "left side" pies were presented on the left side of the monitor, and "right side" pies on the right. To perform the task, participants needed to use information from both pies. Overall, there are* (5×5) *25 unique cues or 'configurations'. The right panel depicts reward probability maps (given button press "sun"), as a function of stimulus,* u. *The variables* u_{left} *and* u_{right} *denote the number of slices in the left and right pies.*

3.2.4 Procedure

The experiment was composed of two consecutive tasks with two different mappings, referred to as "addition" or "subtraction" (a term that describes the subspace - see earlier), in counterbalanced order. As illustrated in Fig.3.2, each trial started with a black fixation cross presented at the center of the screen for 1000 ms. Afterwards, the stimuli appeared and stayed on screen for 2500 ms maximum or until a response was made. Responses were made on a standard keyboard, the letter "g" indicating a prediction of sun and "j" predicting rain. Responses not given within the required time constitute ”missed trials”. Right after button press, confirmation of the choice was given for 500 ms. Finally, feedback was provided, saying "correct" if the prediction was correct, "incorrect" if it was not and "too slow" if they missed the trial (no response within 1500 ms). At the end of each block of trials participants were required to keep their eyes on a fixation cross for one minute. In order to test participants knowledge about the task, at the end of each task, we asked them how they approached it and at which point in time they started approaching it that way. At the end of the experiment, we probed participants with two questionnaires, one per task, to assess their explicit knowledge of the task. The experiment plus preparation took about one hour and a half to complete.

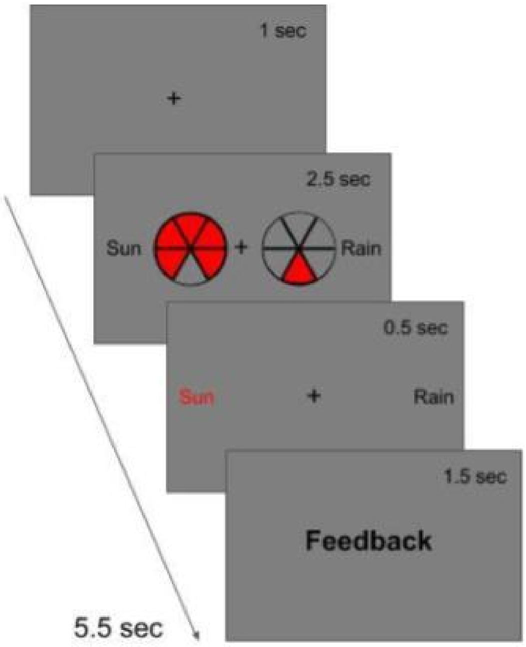


Fig. 3.2 **Trial Structure.** *Each trial started with a fixation cross. Afterwards, two pies appeared and participants had up to 2.5 sec to respond. Confirmation of the choice was then given and feedback was provided.*

3.2.5 Stimulus-Reward Mappings

The probabilistic structure of these two tasks was operationalized by making the log-odds of the outcome a quadratic function of stimulus characteristics, the number of slices.

$$
\begin{aligned}
\log\left[\frac{p(y_t=1)}{p(y_t=0)}\right] &= (u_t-\mu)^T W(u_t-\mu)+w_0 \\
W &= 2.4\times\begin{bmatrix} -0.71 & w_d \\ w_d & -0.71 \end{bmatrix} \\
\mu &= [3,3]^T \\
w_0 &= 2 \\
u_t &= [u_{left}, u_{right}]^T
\end{aligned}
\tag{3.1}
$$

Flipping the sign of w_d parameter in this mapping produced either the addition or subtraction task depicted in Figure 3.1, where $w_d = 0.71$ produces the subtraction map and $w_d = -0.71$ produces the addition map. In terms of the mathematical complexity of the mapping the two tasks should be of the same difficulty..

In terms of declarative learning (i.e. to explicitly learn a verbalisable rule), however, we expect the addition task to be more difficult, for the following reason. The subtraction task can be verbally described using a single logical clause. If the shapes are the same it is sun. So, if by subtracting one "number of slices" from the other you get zero it is sun (subtraction). On the other hand, for the addition task participants must add the number of slices together, but then additionally compare the sum to a numerical reference value, 6, or a verbal reference "make a full pie".

3.2.6 Experimental Design

We assess the effect of different task structures with a within-subject design with two levels of the factor task. All the participants did both addition and subtraction tasks. Each task was composed of 250 trials (10 repetitions per configuration) divided into 5 blocks. Given that participants are required to make Sun/Rain decisions and learn incrementally via feedback, this is reminiscent of the classic Weather Prediction Task. However, a major difference is that in our tasks there is a hidden structure in the stimulus-reward mappings that can be discovered by participants.

Model and measures

We fitted the Latent Cause Model to subjects behavioural data as described in the previous chapter, and then derived the same two measures of Recognition and Entropy

(Davis et al., 2012a). Recognition is the probability of a stimulus given the model:

$$\begin{aligned} R_{pre} &\equiv p(u_t) \\ &= \sum_j p(u_t|j)p(j) \end{aligned} \tag{3.2}$$

Entropy is the uncertainty about which cluster the new stimulus belongs to:

$$\begin{aligned} H_{pre} &= -\sum_j p(j|u_t)\log_2[p(j|u_t)] \\ p(j|u_t) &= \frac{p(u_t|j)p(j)}{\sum_{k=1}^{M} p(u_t|k)p(k)} \end{aligned} \tag{3.3}$$

Both are computed "pre-feedback" i.e. before the agent receives feedback. We also computed four further measures. The first two are the corresponding quantities after feedback

$$\begin{aligned} R_{post} &\equiv p(u_t|r_t, a_t) \\ &= \frac{p(u_t, r_t, a_t)}{\sum_{r'}\sum_{a'} p(u_t, r_t = r', a_t = a')} \\ p(u_t, r_t, a_t) &= \sum_j p(r_t, a_t|j)p(u_t|j)p(j) \end{aligned} \tag{3.4}$$

$$\begin{aligned} H_{post} &= -\sum_j p(j|r_t, a_t, u_t)\log_2[p(j|r_t, a_t, u_t)] \\ p(j|r_t, a_t, u_t) &= \frac{p(r_t, a_t|j)p(u_t|j)p(j)}{\sum_{k=1}^{M} p(r_t, a_t|k)p(u_t|k)p(k)} \end{aligned} \tag{3.5}$$

Finally, we can compute the Kullback-Liebler (KL) divergences (MacKay, 2003) between the pre- and post-feedback distributions, quantities that measure the distance between two probability distributions.

$$\begin{aligned} RelativeSurprise &= \log\left[\frac{p(u_t)}{p(u_t|r_t, a_t)}\right] \\ KL &= \sum_j p(j|u_t)\log\left[\frac{p(j|u_t)}{p(j|u_t, r_t, a_t)}\right] \end{aligned} \tag{3.6}$$

These six measures are the putative neuronal signals that we estimate by fitting the LCM to subjects time series of decisions. We then use EEG to localise these signals in peri-stimulus and peri-feedback time.

3.3 Results

We performed model-based EEG analysis. It offers the possibility of going behind the classic ERP (discussed briefly in Chapter 1) and tests novel hypotheses. We set up GLMs with different regressors and an offset. The Offset is a column of 1's (the associated regression coefficient will compute the mean of the EEG signal over trials and so corresponds to the standard ERP).

We divided the analysis in two parts. First, we set up a General Linear Model (GLM) (Dobson and Barnett, 2018; Friston et al., 2007) with dependent variable given by the Stimulus Epoch EEG signal and independent variables corresponding to the model measures (recognition and entropy). This was repeated in a second analysis with dependent variable given by the Feedback Epoch EEG signal and independent variables as before. These GLMs are referred to as the "Model Measure GLMs".

Second, we set up a General Linear Model (GLM) with dependent variable given by the Stimulus Epoch EEG signal and independent variables corresponding to the stimulus features (number of slices in left and right pies), and task (addition or subtraction). This was repeated in a second GLM with independent variables corresponding to the feature subspace (i.e. taking the sum or difference of number of pies) and task. We used a separate GLM analyses here because the feature subspace variable is collinear with the features themselves. Both of these analyses were then repeated with dependent variable given by the Feedback Epoch EEG signal and independent variables as before. These GLMs are referred to as the "Feature GLMs" and the "Subspace GLMs".

We ran each model for each participant, at each time point and each electrode. The dependent variables were $[N_{trials} \times 1]$ vectors of the EEG signal for each participant, time point and electrode. The corresponding GLM design matrices were of dimension $[N_{trials} \times P]$ where P is the number of regressors. All columns of the design matrix were set to have zero mean and unit variance except for the offset (constant) and any "tasks" columns.

The estimated regression coefficients for each subject were then entered into a group-level analysis using the summary-statistic approach (Friston et al., 2007). At the group level, a cluster-based nonparametric test was implemented, following the procedure described in the papers by Samaha and colleagues, Balestrieri and Colleagues and Maris and Ostenveld (Balestrieri and Busch, 2021; Maris and Oostenveld, 2007; Samaha et al., 2017). All the analysis were computed using the Fieldtrip Toolbox (Oostenveld et al., 2011) and Matlab built in functions (MATLAB, 2018).

3.3.1 Model Measures

In what follows X_i denotes the ith column of the design matrix used in the GLM analyses.

Stimulus Presentation

The regressors for this analysis were X_1 = Recognition, X_2 = Entropy and X_3 = Offset. The dependent variable was the Stimulus Epoch EEG signal. We used data from both tasks. Cluster permutation analysis revealed a significant negative cluster between 300ms and 500ms after stimulus onset for the entropy measure, this is shown in the Entropy row of Figure 3.3. We did not find any significant clusters for the recognition measure.

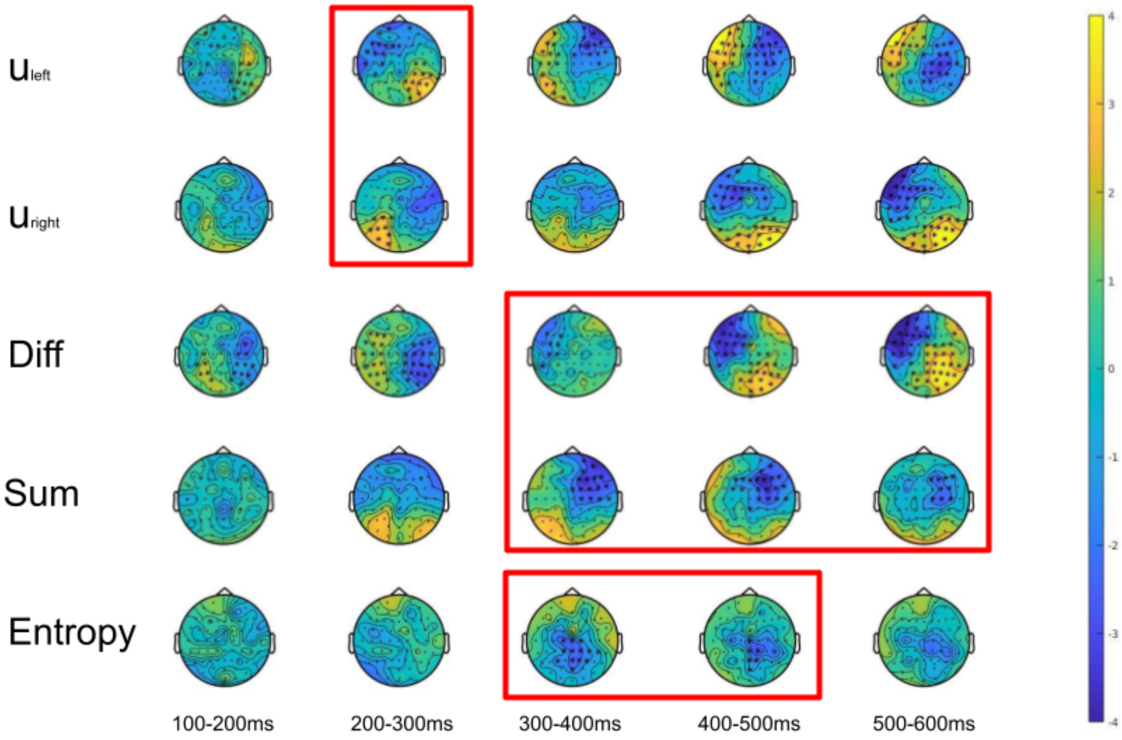


Fig. 3.3 **Results after Stimulus onset.** *The plot shows topographies with t-values on a scale from -4 to 4 for different contrasts between 100 and 600 msec after stimulus onset. Significant electrodes are shown by asterisks. The results are obtained from three different GLM analyses. The first two rows represent results from the "Feature GLM", the third and fourth rows from the "Subspace GLM" and the last row from the "Model Measures GLM".* u_{left} *and* u_{right} *show the linear relationship between the EEG signal and the number of slices of the left and right pies. Diff and Sum show the relationship between EEG signal and the subtraction and addition subspaces. The last row shows the relationship between EEG signal and entropy measure.*

Feedback Presentation

The regressors of this GLM were X_1 = Recognition, X_2 = Entropy, X_3 = Relative Surprise, X_4 = KL, X_5 = Reward and X_6 = Constant. Reward was defined as 1 for correct and 0 for incorrect feedback. The dependent variable was the Feedback Epoch EEG signal. We used data from both tasks.

Cluster permutation analysis revealed two significant negative clusters for the recognition measure, one between 300ms and 500ms and an occipital one between 500ms to 700ms after feedback onset (second cluster is not highlighted with a red box). The analysis also revealed a significant negative cluster for the reward predictior between 400ms and 600ms after feedback onset. Both Recognition and Reward results are depicted in Figure 3.4. We did not find any significant clusters for the entropy, KL or RelativeSurprise measures. The significant effect of the offset term (i.e. the standard ERP from the Feedback Epoch) is reported in the appendix B.

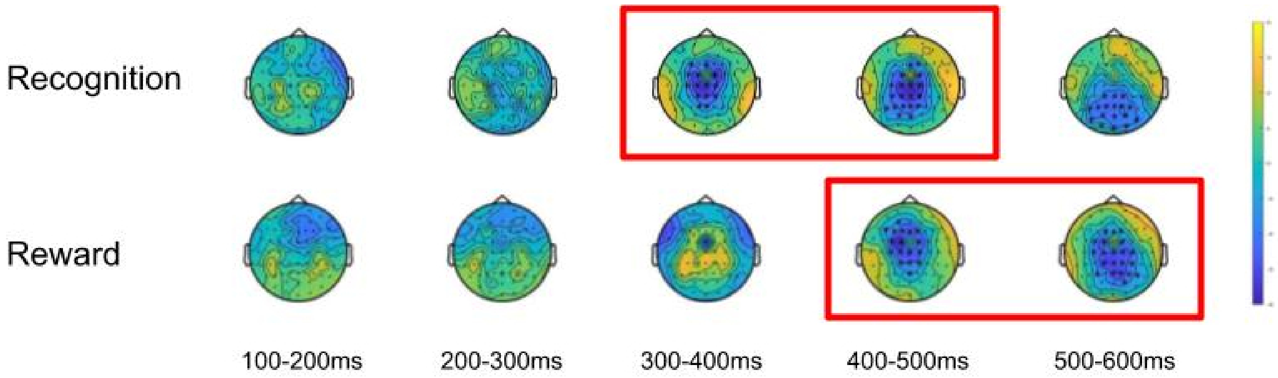


Fig. 3.4 **Results after Feedback onset.** *The plot shows topographies with t-values on a scale from -4 to 4 for different contrasts between 100 and 600 msec after stimulus onset. Significant results are represented by asterisks. The results are obtained from the feedback presentation GLM analysis. The first row shows the relationship between EEG signal and recognition measure. The second row shows the relationship between EEG signal and the reward that participants received.*

3.3.2 Stimulus and Subspace

Feature GLM

The regressors of this GLM were $X_1 = u_{left}$, $X_2 = u_{right}$, $X_3 =$ Task1, $X_4 =$ Task2, $X_5 = (X_3 - X_4). * X_1$, $X_6 = (X_3 - X_4). * X_2$ where Task1 and Task 2 are "one-hot vectors" coding for task, and X_5 and X_6 comprise the interaction terms. The dependent variable was the Stimulus Epoch EEG signal. We used data from both tasks.

Cluster permutation analysis revealed two significant positive clusters and three significant negative clusters for the left perceptual feature as shown in the first row of Figure 3.3. A positive occipitoparietal, controlateral to the stimulus, cluster between 200ms and 300ms and a positive frontolateral cluster between 300ms and 600ms. A negative fronto-controlateral cluster between 300ms and 500ms. A negative fronto-ipsilateral cluster between 100ms and 300ms. A negative central between 500ms and 600ms.

Cluster permutation analysis revealed two significant positive clusters and one significant negative cluster for the right perceptual features as shown in the second row of Figure 3.3. A positive occipitoparietal cluster, controlateral to the stimulus, between 200ms and 300ms and a positive occipitocentral cluster between 400ms and 600ms. A negative fronto-controlateral cluster between 400ms and 600ms.

We did not find any significant clusters for the interaction between perceptual features and tasks. The significant effect of offset (Task1 and Task2) is reported in the appendix. No difference was found between the two tasks.

Subspace GLM

The regressors of this model were $X_1 = (u_{left} + u_{right})$, $X_2 = (u_{left} - u_{right})$, $X_3 =$ Task1, $X_4 =$ Task2, $X_5 = (X_3 - X_4) * X_1$, $X_6 = (X_3 - X_34). * X_2$. Task1 and Task 2 were one-hot vectors coding for task. We used the Stimulus Epoch EEG signal for both tasks as the dependent variable.

Cluster permutation analysis revealed three positive clusters and two negative clusters for the addition task, Figure 3.3, fourth column. A positive central between 900ms and 1200ms. A positive central between 1100ms and 1300ms. A positive central between 1200ms and 1300ms. A negative occipito-temporal between 1200ms and 1400ms. A negative fronto-lateral between 300 and 600 ms.

Cluster permutation analysis revealed two positive clusters and two negative clusters for the subtraction subspace as shown in the third row of Figure 3.3. An early positive and an early negative between 100ms and 300ms. A negative fronto-lateral between 300ms and 600ms. A positive occipital between 400ms and 600ms.

We did not find any significant clusters for any of the interactions between task and subspace. The significant effect of offset (i.e. average effect over Task1 and Task2) is reported in the appendix B. This is equivalent to the standard Feedback ERP. No significant EEG differences were found between the two tasks.

3.4 Discussion

3.4.1 Stimulus presentation

Our models allowed us to identify the separate and successive cognitive processes during learning and executon of a nonlinear decision making task.

We found evidence of an early processing of stimulus features in occipital electrodes 200 ms after stimulus presentation. This activity was controlateral to the feature presented and scaled to the magnitude of the feature. This finding fits well with the number processing literature which suggest an involvement of visual and parietal cortices in magnitude processing. Subsequent to this we found frontolateral activity between 300 and 600 ms that was linearly correlated with the feature subspace variable - reflecting the way in which features are combined (either subtracted or added together). This is coherent with previous findings showing how abstract representations of task structures are reflected in activation of prefrontal cortex (Badre and D'esposito, 2008; Collins et al., 2014).

We did not find any difference in use of features/feature combinations between tasks. A potential explanation is that all the participants were doing both tasks consecutively so that maybe representations from the first task are automatically evoked during performance of the second task. Another explanation could be that participants tried both strategies, and so created representations of both feature subspaces. A third explanation is that our stimuli may have helped in automatically eliciting both representations - see the arithmetic case in Dasgupta and Gershman (2021).

3.4.2 Model Measures

Simultaneously to the activity related to feature and feature combinations, parieto-central activity reflected the uncertainty (entropy) with which the current cue could be assigned to one of the pre-existing encoded clusters. To the extent we can relate this EEG sensor-space activity with parietal cortex, this is consistent with previous literature as follows. The parietal cortex has been previously linked to the integration of bottom-up inputs and top-down predictions (Seghier, 2013) and proposed to integrate prior knowledge with new information (Tomov et al., 2018). Furthermore, Akrami et al. (2018) found that rats posterior parietal cortex was sensitive to the history of the stimuli, more than the current stimulus presented. Our results fit well with these ideas, indeed, this is the same area we found to associated with the entropy measure. Our study adds a temporal timeline to previous this previous literature: activity related to the uncertainty of the new input as compared to previous knowledge, takes place after the new input features have been processed.

3.4.3 Feedback presentation

After a decision has been made, and 300 ms after feedback presentation we found activity reflecting encoding of the configuration. The encoded configuration contains information regarding its value, the action associated and the feedback received. During encoding and slightly afterwards, we found evidence of reward processing in central electrodes. Reward is thought to improve task representation by strengthening the encoding process (Aberg et al., 2020; Daw and Shohamy, 2008; Yee and Braver, 2018).

3.5 Limitations

During this study, we encountered limitations that could be explored in future studies. For example, no jitter was present during fixation cross and feedback presentation. This likely resulted in participants keeping track of time and predicting the sequence in which stimuli, fixation cross and feedback were presented. This ended up in a meaningful period pre-stimulus onset and post feedback presentation. Furthermore, future analysis should be directed to the recording of the minute of break after each block of trials, in which participants disengaged from the task. The information that participants encoded in the block before the break might be reactivated favouring so pattern recognition. The brain activity could so be associated, at the same time, with the stimuli presented in the previous block and to the performance in the next one.

Chapter 4

Learning During Quiet Wakefulness Promotes Generalization

4.1 Introduction

Recent experiments in human learning have shown that a period of Quiet Wakefulness, also known as “Quiescence” or the “Offline Wake State”, has beneficial effects across a broad range of cognitive tasks. We briefly review these effects, separating them into effects on memory, and effects on other cognitive processes. This perspective then motivates an experimental design which allows us to separately examine memory effects, which are likely supported by memory stabilisation processes (Squire et al., 2021), and generalization effects, which are likely supported by the learning of new representations (Niv, 2019).

4.1.1 Effects on Memory

One body of work has focussed on the effects of Quiescence on memory. Here, memory performance is compared to that achieved in a baseline group who are assigned to an "Active" rather than Offline wake condition. Benefits have been found, for example, in the number of memories recalled (Dewar et al., 2012) or the fine detail of new memories (Craig and Dewar, 2018). Here, the benefit is with respect to the active wake condition. Overall, memory performance degrades over time but less so for participants assigned to Offline verse Active wake groups. This body of work extends previous studies which have shown that periods of sleep benefit memory when compared to typical waking activities (Axmacher et al., 2008; Graveline and Wamsley, 2017; Lewis and Durrant, 2011; Schapiro et al., 2018).

Potential mechanisms underlying the memory stabilisation afforded by quiescence have been recently revealed using functional imaging experiments. These studies have, for example, found that neuronal activation patterns detected during encoding are reactivated during Offline Wake states (Tambini and Davachi, 2019). This reactivation is thought to be isomorphic with the "pattern replay" (a temporally ordered sequence of reactivations) observed in rodent studies (Foster, 2017) that promote synaptic plasticity.

4.1.2 Effects on Other Cognitive Processes

A more recent body of work investigates the effects that Quiescence has on cognitive tasks beyond memory studies (Tambini and Davachi, 2019; Wamsley, 2019). Reactivation of encoded elements during quiescence is thought to facilitate feature selection, similarity extraction and pattern recognition, thereby promoting generalization and improvement in performance (Tambini and Davachi, 2019). These improvements are supported by the learning representations that are useful for the task at hand, for example, a new discriminatory feature (Craig et al., 2018a), a new cognitive map (Craig et al., 2018b), or a new higher-order rule (Quentin et al., 2020). Neuroimaging studies have shown that memory reactivation during quiescence increases connectivity between cortical areas which is thought to distribute and reorganize memory representations across hippocampal and neocortical networks (Schlichting and Preston, 2014; Tompary and Davachi, 2017).

4.1.3 Hypotheses and learning task

We designed an experiment to separately assess the effects of an offline wake period on generalization and memorization. We used a revised version of the Weather Prediction Task (WPT) in which participants learnt the association between configurations of virtual pies and a weather outcome (sun or rain) as shown in Figure 4.1A. The original WPT (Knowlton et al., 1994) contains arbitrary associations from stimulus configurations to outcome which therefore forces participants to memorise outcomes for each stimulus pair. In our revised version, however, there is a structure in this mapping. Figure 4.1B shows that the probability of "Sun" is high if the number of pies are similar.

Correct generalization in this task can be achieved by learning a new representation which could take the form of (i) a logical or verbal rule (Ballard et al., 2018), (ii) identification of a discriminatory feature (u_1-u_2) (Menghi et al., 2021), or (iii) identification of homogeneous clusters of exemplars (one along the diagonal, and one on either side) (Sanborn et al., 2010).

Participants were trained on a set of pies and then tested on the same set plus a new one (Figure 4.1A). They went through two testing blocks, one after the training, and one after a wake condition (either offline or active wake where they were doing a 2-back task - Figure 4.1C). During these testing blocks, they received no feedback. Since participants viewed configurations of pies they were trained on ("old cues with feedback"), and new configurations of pies they had not seen before, without receiving feedback ("new cues"), it was possible to separately assess both memorization and generalization performance. We hypothesized that the active wake condition would disrupt the representation learning process that results in generalization. We also expected that performance on old cues ("memorisation") would be less degraded for the offline wake condition.

4.2 Materials and Methods

4.2.1 Participants

A total of one hundred volunteers from the University of East Anglia (mean age = 20.8, SD = 5.99, 15 male) were recruited through SONA System (https://uea-uk.sona-systems.com/) and redirected to Pavlovia (https://pavlovia.org/) website where the experiment was hosted. All of them were naive to the purpose of the experiment. At the end of the experiment, participants received credits for their participation. The study was approved by the University of East Anglia School of Psychology Research Ethics Committee (PSY-REC) and all participants consented to participate through an online consent form at the beginning of the experiment. The data from this group were then analysed with two separate approaches. First, with a standard data analysis approach using an Analysis of Variance (ANOVA) and a set of exclusion criteria to remove participants who didn't perform the tasks satisfactorily. For this analysis we excluded participants whose:

1. Performance in either training, pre-test or post-test was under 50%.
2. Performance in a Fowlkes–Mallows index criterion (FM) (see Fowlkes and Mallows (1983)) in the 2-back task was under 50%.

3. Declared the correct strategy unless they improved in both old and new configurations between pre-test and post-test. This was to exclude participants who learned the rule before the wake period and therefore had no chance to improve.

We performed our ANOVA on the remaining sample of 58 participants (mean age = 20.27, SD = 3.87, 9 male). In our second data analysis approach we ran a nonparametric regression on data from all 100 participants, regardless of performance.

4.2.2 Apparatus and Stimuli

Stimuli were created and presented using the PsychoPy3 Toolbox for Python (Peirce, 2007), (https://www.psychopy.org/). Two virtual ”pies”, equidistant from the fixation point, were displayed. Each pie was divided into six slices with from one up to five slices that could be filled with red colour; the combination with three slices filled in both pies was excluded, making a total of twenty-four combinations, see figure 4.1 (panel A). The stimuli were presented on a dark grey background.

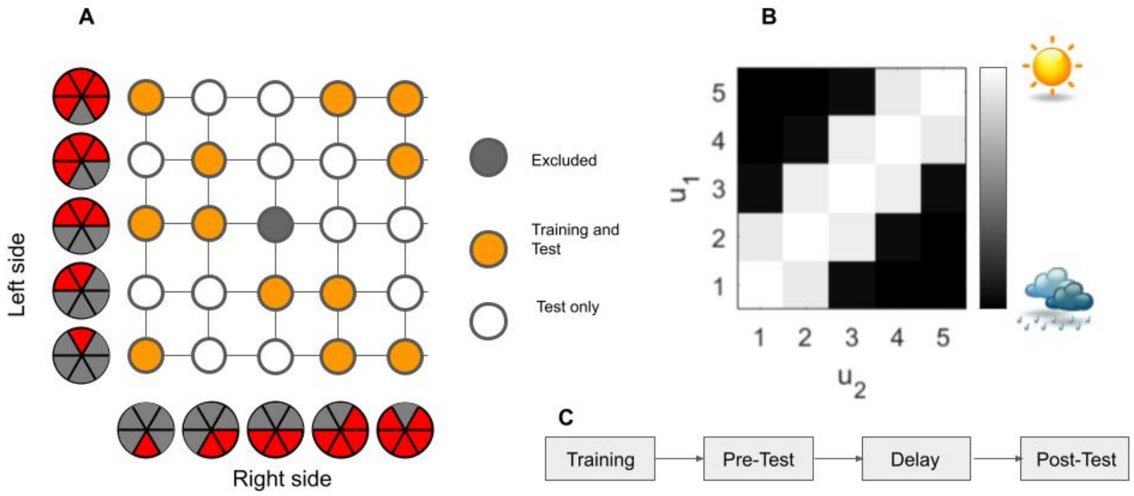


Fig. 4.1 **(A) Experimental Stimuli (B) Stimulus Reward Mapping (C) Overview of the procedure.** *(A) Each pie on the left side should be combined with each pie on the right side, creating 25 potential configurations. One pie was excluded, 12 were shown during both training and test blocks (pre-test and post-test), the remaining 12 were shown during the test blocks only (B) The gray scale image plots the reward probability (given button press "sun"), as a function of stimulus, u. The term reward is used to define correct feedback. The variables* u_1 *and* u_2 *denote the number of slices in the left and right pie stimuli, respectively. (C) All participants completed one block of training, one block of testing ("pre-test") followed by the wake block (either 2-back or offline wake) and finally a second testing block ("post-test").*

4.2.3 Procedure

The experiment was composed of one training block (120 trials), a pre-test block (120 trials), a wake block and a post-test block (120 trials), see figure 4.1 (panel C). During the wake block half of the participants were assigned to the distractor task ("Active Wake"), the other half to a period of quiescence ("Offline Wake").

As we can see in figure 4.2 (panel A), in the training block, each trial started with a black fixation cross presented at the center of the screen for an interval of 1000 ms. Afterwards, the stimuli appeared and stayed on screen for 2500 ms maximum, or until response. Response was made on a standard keyboard, the "a" indicated sun prediction and "l" indicated rain. Responses not given within the required time constitute ”missed trials”. Right after button press, confirmation of the choice was given for 500 ms. Finally, feedback was provided, indicating "correct" if the prediction was correct, "incorrect" if it was not and "too slow" if they missed the trial. In the pre-test and post-test blocks, the trial structure was identical to the training block but without feedback (see figure 4.2 panel B).

The distractor task was a 2-back task, composed of four blocks of 35 trials each, lasting 5 minutes. A letter (A to G) was presented at the center of the screen for 500 ms. A fixation cross was then shown for 1500 ms. Participants were required to press the space-bar every time the letter presented was the same as the letter presented two items before. Subjects in the quiescence condition were asked to close their eyes and relax for 5 minutes. Their monitor was also made darker to reduce ambient lighting (Craig et al., 2018a).

After the wake block participants were presented with the instructions for the final test block (post-test) and were asked to press a button to start. At the end of the experiment, participants in the quiescence condition were asked to rate how much they managed to relax and avoid external stimulation.

At the end of the experiment participants were probed with a questionnaire to test their knowledge of the task. First, they were asked to select from a list of strategies the one that most resembled their own strategy (See appendix C for the questionnaire). Second, they were asked at which point in time they started using this strategy with a questionnaire (See appendix C for the questionnaire). The overall experiment took about thirty minutes to complete.

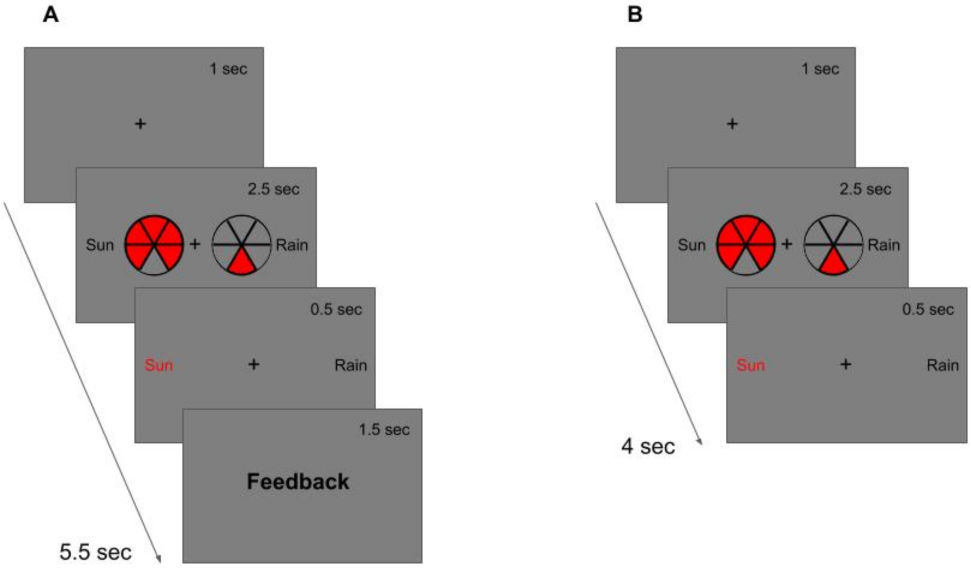


Fig. 4.2 **(A) Training Trial structure.(B) Testing Trial structure.** *(A) Each trial started with a fixation cross. Afterwards, two pies appeared and participants had up to 2.5 sec to respond. Confirmation of the choice was then given and feedback was provided. (B) The only difference with the training was the feedback was not provided during testing blocks.*

4.2.4 Stimulus-Reward Map

The probabilistic structure of the task was operationalized by making the log-odds of the outcome a quadratic function of stimulus characteristics (number of red slices) (see figure 4.1 panel B). Like in similar studies, we used the term reward to define correct feedback (Duncan et al., 2018).

$$\begin{aligned} \log\left[\frac{p(y_t=1)}{p(y_t=0)}\right] &= (u_t-\mu)^T W (u_t-\mu)+w_0 \qquad (4.1) \\ W &= 2.4\times\begin{bmatrix} -0.71 & 0.70 \\ 0.70 & -0.71 \end{bmatrix} \\ \mu &= [3,3]^T \\ w_0 &= 4 \\ u_t &= [u_t(1),u_t(2)]^T \end{aligned}$$

If, for each cue, subjects choose the option with the highest probability, then the correct classification rate would be 95 per cent. Additionally, this map could be approximately described the verbal rule: "Choose Sun if the difference in number of pie slices is zero".

4.2.5 Experimental Design

We assess the effect of Offline versus Active Wake conditions using a between-subjects design in which 50 participants were assigned to the Active condition and 50 participants to the Offline condition. During training, 12 configurations of stimuli were presented and subjects received feedback about whether their decisions were correct or not. There were 10 repetitions of each trial type making 120 trials in all. During testing (both pre-test and post-test), these same stimuli were again presented ("old stimuli") along with 12 new stimuli that had not been presented during training ("new stimuli") in a random order. Each trial type was repeated 5 times making 120 trials in all. A "memorisation" score could then be computed based on the correct decision rate over old stimuli categorised with feedback in the initial training phase, and a "generalization" score on the correct decision rate over new stimuli.

4.3 Results

Our main hypothesis is that an offline wake period, compared with a period of active wake, will facilitate consolidation, so promoting memory and generalization. This facilitation is measured by "improvement scores" which are the post-test minus the pre-test scores, computed separately for memorisation and generalization. In order to test this effect, we divided the analysis into two parts; first, a more standard data analysis approach using performance-based selection criteria and analysis of variance, second a nonparametric regression of improvement score onto training performance using data from all participants, regardless of (sometimes poor) task performance, with this latter analysis providing an additional perspective on our data.

4.3.1 Analysis of Variance

We performed a two-way mixed-design ANOVA with dependent variable improvement score and independent factors of (i) stimulus novelty (new/old) and (ii) wake condition (offline/active wake). Out of the 58 participants included in the analysis, 24 were from the active group and 34 from the quiescence group.

Overall, we found only the main effects of wake (Novelty, F(1,56)=2.466, $p = 0.121$; Wake, F(1,56)=5.049, $p = 0.028$; Interaction, F(1,56)=1.952, $p = 0.167$). We first performed four post-hoc dependent t-tests to check in which condition participants performance improved. We found, no significant effect in the active condition for neither the old (t(23)= -0.934, p=0.359) nor the new condition(t(23)= -0.786, p=0.439). We found, a significant result for the offline condition for the new stimuli (t(33)= 2.994, p=0.005) but not for the old ones (t(33)= -0.037, p=0.97). We then performed two post-hoc independent and two dependent t-tests to compare the effect of novelty within and between wake conditions. We found a significant difference between offline and active wake when the configurations were new (t(56)= 2.536, p=0.014) but not when they were old (t(56)=0.722, p=0.473). Furthermore, there was a difference in improvement score between new and old stimuli in the offline condition (t(33)=2.316, p=0.026) but not in the awake condition (t(23)=0.112, p=0.911). Quiet rest, therefore, improves generalization (but not memorization) as compared to an equivalent duration active wake period.

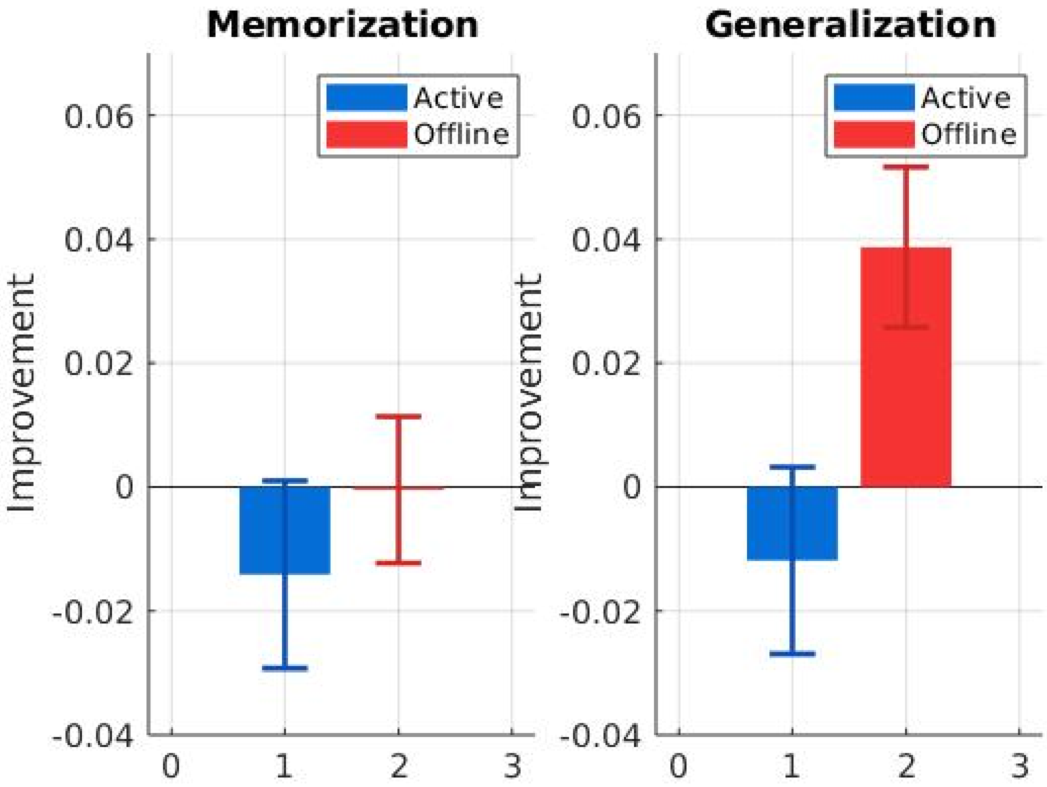


Fig. 4.3 **Improvement Scores by Condition.** The barplots show the improvement scores for each wake/novelty combination, with memorisation scores computed from old stimuli, and generalization scores from new stimuli. We observe an increase in generalization score for the offline wake condition only (right panel). The error bars indicate the standard error of the mean.

4.3.2 Nonparametric Regression

We performed a k-nearest neighbour nonparametric regression (Altman, 1992) of improvement scores onto training performance using data from all 100 participants (50 in each group). We found a significant relationship between generalisation improvement and training score for participants belonging to the offline wake, but not online wake, group. As can be seen in Figure 4.4, this improvement was manifest for training scores between 50% and 60%. We found no significant effects in memorisation improvement for either offline or active wake groups. The results in Figure 4.4 were obtained with k=20 neighbours but very similar results were found over a range of k. Confidence intervals were derived using N=10,000 bootstrap samples (Efron and Tibshirani, 1993).

Similar results (but with stronger effects) were found with the subset of participants selected using the previously described exclusion criteria (see appendix C for the figure).

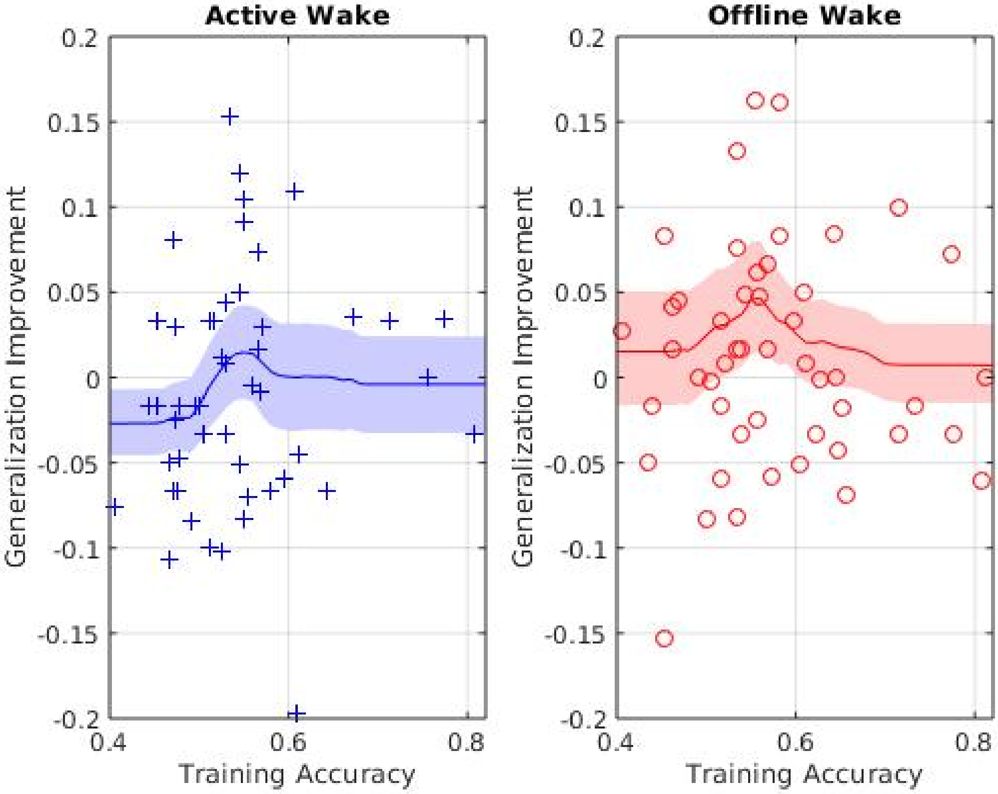


Fig. 4.4 **Nonparametric Regression of Generalisation Improvement onto Training Accuracy.** The plots show the relationship between improvement in generalisation and training accuracy for both active and offline wake groups, with data points shown as crosses/circles, a line indicating the mean, and bands showing 90 percent bootstrap confidence intervals. We observe a significant relationship between generalisation improvement and training score in the offline wake condition for training scores between 50 and 60 percent (where the lower confidence interval is above zero).

4.4 Discussion

4.4.1 Generalization Effects

We found evidence in support of our hypothesis that generalization would be facilitated by an offline wake period. This was evident from several analyses. First, our ANOVA and post-hoc analysis revealed an improvement in accuracy for novel stimuli for offline versus active wake conditions. This was driven by a significant improvement in the offline group and no improvement in the active group. Second, relationship between training accuracy and improvement for new stimuli was found for the offline wake group only. Participants in the offline wake group who had a performance between 50% and 60% in training, had a greater improvement between test sessions, thus benefiting from the period of quiescence.

This latter analysis suggests that, for the learning task we have studied, a period of quiescence is most beneficial for the subset of participants who hit the "sweet spot"- they learn a little from the training data, but not too much. Interestingly, this same analysis shows that there is a sweet spot for participants in the Active Wake condition too, but that this is masked by an overall downwards shift in generalisation improvement, thus potentially revealing a consolidation effect that is overwhelmed by an actively interfering working memory task (see left panel of Figure 4.4).

4.4.2 Memorisation Effects

We did not find evidence of improved memory retention for the Offline versus Active groups. This is inconsistent with previous work where it benefited memory recall (Craig and Dewar, 2018; Dewar et al., 2012). This is likely due to the difference in the length of the offline period and in the time intervals in which participants were tested. Previous studies used an offline wake period of 10 minutes (twice as long as ours) (Craig and Dewar, 2018; Dewar et al., 2012) and tested participants at longer intervals, 15 to 30 minutes and a week after the first test (Craig et al., 2015; Martini et al., 2018). Our study was a behavioural experiment in which data was acquired on the internet and therefore had tighter constraints on experimental design. Additionally, we emphasise that what is "memorised" in our task is rather different to these previous studies. Here, a memory-based decision would be implemented by recalling the (feedback) label associated with an old cue item, which is analogous to a paired associates learning task (Poldrack et al., 2001).

4.4.3 Active Wake Condition

One characteristic of Offline versus Active wake studies that warrants further discussion is the nature of the Active Wake condition. One perspective is that brain resources are limited so that when you are doing an engaging task, these resources have to be reallocated from one task to another based on priorities (Raichle and Gusnard, 2002). Another perspective is that wake activities are not disruptive per se, but only when novel encoding uses the same resources that would otherwise be engaged in consolidation (Peigneux et al., 2006). Supporting results for the latter view come from Varma and colleagues (Varma et al., 2018, 2017). They found that a cognitively demanding n-back task (lasting between 9 and 12 minutes) created no interference compared to an offline period, whereas autobiographical thinking impaired episodic memory consolidation after a memorization task. Autobiographical thinking depends on similar neurocognitive resources that would be otherwise allocated to consolidation of memorized elements (Craig et al., 2014; Varma et al., 2018). Our results are also consistent with these findings in that engaging in an n-back task after training did not interfere with memorization. It did, however, interfere with generalization suggesting that the n-back task uses neuronal resources that would otherwise be supporting generalization.

4.4.4 Neurobiology of Generalization

One view of how the brain generalizes is that this is instantiated in striatal (Poldrack et al., 2001) and cortical (Kumaran et al., 2016; McClelland, 2013; McClelland et al., 1995) systems which slowly and incrementally learn statistical regularities, with cortical learning supported by experience replay from a hippocampal system that rapidly encodes conjunctions of events (episodes) (Lewis and Durrant, 2011; Richards et al., 2014; Sekeres et al., 2018). This replay process may well be under way during Offline Wake states (Dewar et al., 2012; Wamsley, 2019).

However, it is also known that cortical systems can also learn quickly (Kumaran et al., 2016; Sharon et al., 2011), and that the hippocampus supports a variety of functions beyond episodic encoding and replay (Gaesser et al., 2013; Gavert et al., 2017; Moser et al., 2017; Renoult et al., 2019). Indeed, a recent review summarises the evidence that the hippocampus itself is part of a network of regions that together support generalization across experiences (Zeithamova and Bowman, 2020).

We propose that our new experimental paradigm be used to probe the neurobiological basis of learning during quiet wakefulness, separating out memorisation and generalization effects. This could take place in combination with appropriate computational models of behaviour (Menghi et al., 2021; Sanborn et al., 2010) and model-based functional imaging studies (Davis et al., 2012a; Tomov et al., 2018).

Chapter 5

Multitask Learning over Shared Subspaces

5.1 Introduction

Recent advances in machine learning have delivered human-like levels of performance across a variety of domains from speech and image recognition (LeCun et al., 2015) to language understanding (Radford et al., 2018) and game-playing (Silver and Hassabis, 2016). These advances have been achieved, in the main, using neural network models with very large numbers (e.g. millions) of parameters that are estimated using very large numbers (e.g. millions) of data points. The requirement for such a huge amount of training data places limits on the tasks that can be learnt and is at odds with much of the psychology literature on human learning which suggests that concepts can be learnt using very few examples. One way of achieving such "data-efficient" learning is to leverage information learnt on one task to more efficiently learn another. Subfields of machine learning that have been using this approach include Multitask Learning (learning multiple tasks simultaneously) (Caruana, 1998; Ruder, 2017), Transfer Learning (learning tasks sequentially) (Goodfellow et al., 2016; Ng, 2017) and Continual Learning (Hadsell et al., 2020; Parisi and Wermter, 2019) (learning an indefinite number of tasks sequentially). This paper uses constructs from the machine learning literature to better understand how humans learn across multiple tasks.

Our starting point is the original Multitask Learning architecture proposed by Caruana (1998) in which generalisation across tasks is achieved using shared parameters. This architecture comprises a feature module, which can be shared across tasks, and an output module which is task-specific. In the original "hard-parameter sharing" architecture (Ruder, 2017) the parameters defining the feature model are identical across tasks. Mathematically, this feature model defines a subspace that is shared across tasks. The idea that shared subspaces are useful for learning over multiple tasks has previously been highlighted, for example, under the term "structure learning" (Braun et al., 2010).

This chapter uses an experimental design in which participants learn a pair of tasks that either do or do not share a common subspace. We investigate how learning proceeds with the hypothesis that learning will be facilitated for tasks that share a common subspace. Facilitation of learning could be manifested as faster and/or more accurate learning. We restrict ourselves to linear subspaces so that the shared features are a reduced-dimension linear projection of the input space, leaving nonlinear subspaces to subsequent experiments.

In additional modelling work we make use of a second construct from the Multitask Learning literature - that of "soft-parameter sharing" (Ruder, 2017). Here, a second task does not share exactly the same feature model, but parameters determining the features are constrained to be similar. We use a Sequential Bayesian learning algorithm for neural network training, also known as Elastic Weight Consolidation (EWC) (Kirkpatrick et al., 2017), in which the prior over feature parameters for a second task is given by the posterior over feature parameters from the first. This is implemented by having two parameters for each network connection, a "mean" and a "precision", which together specify a Gaussian probability distribution. Bayesian estimation results in high precisions for those connections that have strongly adapted to data, and lower precisions for those that have not. Having a high precision makes connections more resilient to being overwritten on subsequent tasks. This is the mechanism for preventing so-called "catastrophic interference" (see Kirkpatrick et al. (2017) and Aitchison et al. (2021) for discussion of potential neurobiological substrates). In this paper we use Sequential Bayesian learning over tasks and over mini-batches of data within a single task. This produces learning dynamics both within and between tasks, and the model predicts facilitation of learning (or "positive transfer" (Perkins and Salomon, 1992)) in tasks that share a common subspace. We compare these simulation results to empirical findings.

Overall, the chapter presents a novel experimental task, empirical results on behavioural data, theoretical results from computer simulation, highlights similarities between them, and discusses ideas for future work in this area. We propose that the concept of shared subspaces provides a useful framework for the experimental study of human multitask and transfer learning.

5.2 Materials and methods

Our model-based analysis (see below) is described mathematically using the notation defined here. We use $\mathsf{N}(x; m, \Lambda)$ to denote a multivariate Gaussian density over x with mean m and precision matrix Λ. The transpose of vector x is written x^T. 1_{RK} denotes an R-by-K matrix of ones, $A_{k\bullet}$ is the kth column of A and $A_{\bullet k}$ is the kth row of A. The delta function Δ_{ab} takes the value 1 if $a = b$ and zero otherwise, $\mathsf{vec}(A)$ vectorises the matrix A into a column vector and the sigmoid function is given by

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{5.1}$$

5.2.1 Participants

A total of ninety-six volunteers from the University of East Anglia (mean age = 19.90, SD = 1.36, 17 male) participated in the experiment. Data from seven participants became unavailable due to computer network synchronization errors. A further nine participants were discarded because they performed below chance level in both tasks. We performed our analysis on the remaining sample of 80 participants (mean age = 19.80, SD = 1.34, 13 male). All of them were naive to the purpose of the experiment. All participants gave informed written consent, and the study procedure was approved by the local institutional review board of the University of East Anglia, UK. At the end of the experiment, participants received course credits for their participation.

5.2.2 Apparatus and Stimuli

The experiment was performed in a dimly lit room with participants seated 60 cm away from a computer display with their head supported by a chin-rest. Stimuli were presented on a 23-inch HP Elite Display 240c monitor using the Psychophysics Toolbox (Brainard, 1997) for Matlab (Mathworks) running on Windows 7.

Two virtual "pies" (1° x 1° visual angle) were displayed at 1° from the central fixation point. Each pie was divided into six slices with from one up to five slices that could be filled with red colour, making a total of twenty-five combinations. The slices of the two pies were filled in a mirrored way as shown in Fig 5.1. The stimuli were presented on a dark grey background.

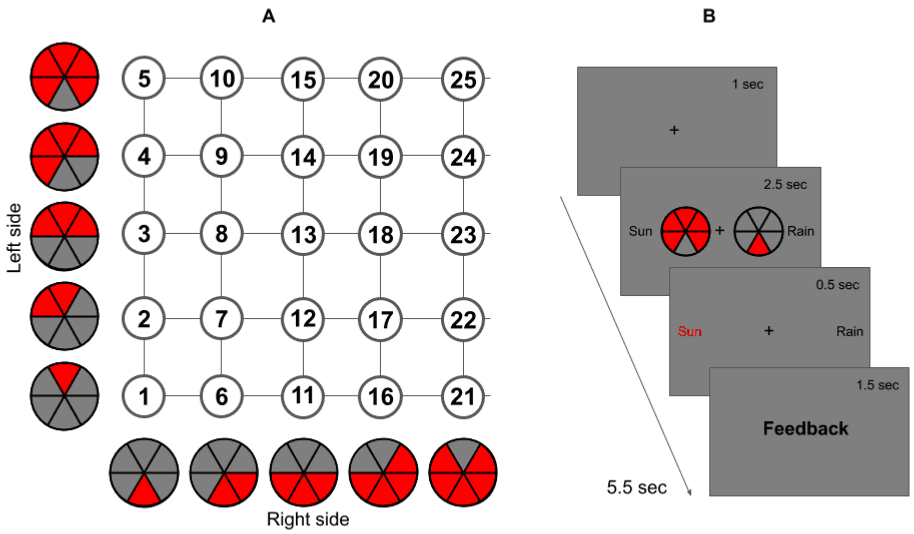


Fig. 5.1 **(A) Experimental Stimuli.(B) Trial Structure.** (A) Each pie on the left/right side was combined with each pie on the right/left side, creating 25 potential configurations. (B) Each trial started with a fixation cross. Afterwards, two pies appeared and participants had up to 2.5 sec to respond. Confirmation of the choice was then given and feedback was provided.

5.2.3 Procedure

As we can see in Fig 5.1 each trial started with a black fixation cross presented at the center of the screen for an interval of 1000 ms. Afterwards, the stimuli appeared and stayed on screen for 2500 ms maximum or until a response was made. Responses were made on a standard keyboard, the "g" indicated sun/heads prediction in Task 1/2 and "j" indicated rain/tails in Task 1/2. Responses not given within the required time constitute "missed trials". Right after the button press, confirmation of the choice was given for 500 ms. Finally, feedback was provided, saying "correct" if the prediction was correct, "incorrect" if it was not and "too slow" for a missed trial.

The experiment took about one hour to complete and was composed of two tasks, comprising 250 trials each (10 repetitions per configuration) divided into 5 blocks, each of 50 trials. For the first task subjects had to make sun/rain decisions, as in the classic Weather Prediction Task (Knowlton et al., 1994), and for the second task they made heads/tails decisions ("Coin Prediction Task"). The mappings from stimulus to reward (correct/incorrect) were specified as described in the following section.

At the end of each task, we probed participants knowledge. We first asked them to describe the way they approached the task. We then gave them a list of six strategies (where only one was correct) and asked them to tick the one that resembled the most the one they used. Finally, we presented them with a timeline of the task asking to mark the point in time in which they started using that strategy.

5.2.4 Stimulus-Reward Maps

Four different Stimulus-Reward Maps, or "Reward Functions", were used over the course of the experiment (but only two per subject), as shown in Fig 5.2. The underlying subspaces were operationalized by defining a common feature that, when represented, reduced the task to an approximate rule. In one case this was addition, in the other subtraction. Mathematically, the reward functions were generated using log-quadratic (Sub1, Add1) or log-linear (Sub2, Add2) mappings as follows. For the log-quadratic maps (Sub1, Add1), the probabilistic structure was specified by making the log-odds of the outcome a quadratic function of stimulus characteristics. Flipping the sign of a single parameter in this mapping changes the Sub1 map to the Add1 map. That is

$$\begin{aligned}
\log\left[\frac{p(y_t=1)}{p(y_t=0)}\right] &= (u_t-\mu)^T W(u_t-\mu)+w_0 \qquad (5.2)\\
W &= 2.4\times\begin{bmatrix} -0.71 & w_d \\ w_d & -0.71 \end{bmatrix}\\
\mu &= [3,3]^T\\
w_0 &= 4
\end{aligned}$$

where $w_d = -0.71$ produces the Sub1 map and $w_d = 0.71$ produces the Add1 map. If, for each cue, subjects choose the option with the highest probability, then the correct classification rate would be 95 per cent. This is the maximum possible for the Sub1 and Add1 tasks.

We also defined tasks using a log-linear model which can produce, for example, the Sub2 map shown in Figure 5.2. Although generated from different models (log-linear versus log-quadratic), from a multitask learning perspective this task is similar to the Sub1 task in that the relevant feature for both tasks is $x = u_2 - u_1$ ie. subtraction. The Add2 map was similarly defined. The maximum performance levels for the Sub2 and Add2 maps were both 93 per cent.

Additionally, these maps could be approximately described using the following rules: Sub1 - "Choose Sun if the difference in pie slices is zero"; Sub2 - "Choose Heads if there more are slices on the left than right"; Add1 - "Choose Sun if the sum of slices makes a full pie"; Add2 - "Choose Heads if the sum of slices is greater than six".

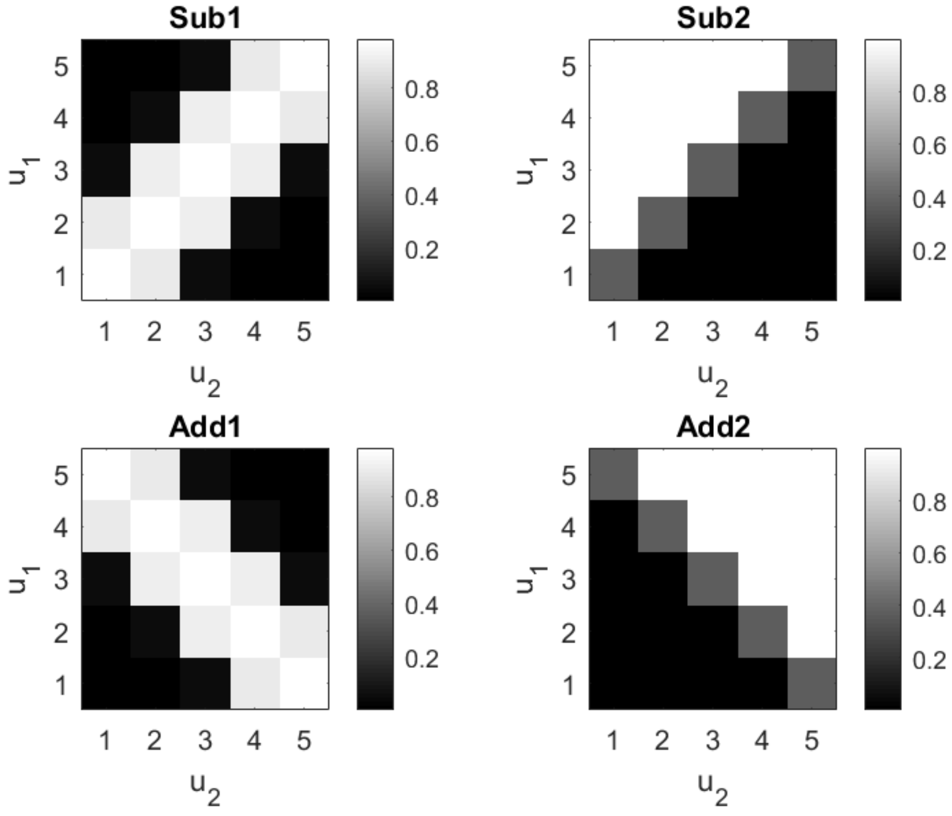


Fig. 5.2 **Stimulus-Reward Maps.** Each gray scale image plots the reward probability (given button press "g" i.e. choosing Sun for Task 1 and Heads for Task 2) as a function of stimulus, u. The variables u_1 and u_2 denote the number of slices in the left and right pie stimuli, respectively. Each task can be implemented using two-stages of processing. For example, for the Sub1 and Sub2 maps the first stage requires extraction of a feature, $x = u_1 - u_2$. For Add1 and Add2 the required feature is $x = u_1 + u_2$. Tasks which use the same stimulus to feature space function (ie. subtraction or addition) are said to share the same subspace.

5.2.5 Experimental Design

Each participant did two tasks, in the first one they had to learn the association between stimuli and weather outcome (sun or rain); in the second one they had to learn the association between stimuli and a coin toss outcome (heads or tails). Subjects were also explicitly instructed that the mapping in the second task was different. These two tasks were carried out on the same day in a 1 hour long experiment. The stimulus to outcome mapping in task 1 was specified by either the Sub1 or Add1 map. Task 2 was specified by either the Sub2 or Add2 map. Participants were assigned to either a

"Same-Subspace" (Same) or "Different-Subspace" (Diff) group according to the logic of table 1. There are 20 subjects per "condition" and 40 subjects per group.

Table 5.1 **Subjects and Groups.**

Condition	Task 1	Task 2	Subspace	Subjects
1	Add1	Add2	Same	20
2	Sub1	Sub2	Same	20
3	Add1	Sub2	Diff	20
4	Sub1	Add2	Diff	20

Subjects were assigned to one of Same or Different Subspace Groups in a between-subjects design. Each of the Same/Different groups comprises data from two conditions e.g. data from the same subspace group is from both Add1-Add2 and Sub1-Sub2 conditions. There were 20 subjects assigned to each condition. For all subjects, Task 1 was presented as a weather prediction task and Task 2 as a coin prediction task.

Additionally, orthogonal subgroups of participants had a minimum 12 seconds break between one learning block and another whereas another orthogonal subgroup had minimum 120 seconds break between one block and another, in a two-by-two between-subject design (with factors of subspace and break-length). However, the break-length factor is ignored in the data analyses presented in this chapter.

Given that participants are required to make Sun/Rain decisions and learn incrementally via feedback, Task 1 is reminiscent of the classic Weather Prediction Task (WPT) (Knowlton et al., 1994; Poldrack et al., 2001). However, a major difference is that in our tasks there is a hidden structure in the stimulus-reward mappings that can be discovered by subjects. Further, Task 1 is also similar to the Configural and Elemental Learning tasks defined by Duncan et al. (2018), with elemental tasks containing a hidden structure (the log-odds of an outcome being a linearly separable function of stimuli). However, the hidden structure we have specified is a linear subspace lying within a non-linear (quadratic) mapping. Task 1 also shares similarities with the Feature-based Multi-Armed Bandit (FMAB) task of Stojic et al. (2020) in that the reward probability is a function of bivariate stimuli. However, FMAB uses a linear function and participants make a multi-way (rather than binary) decision on each trial.

5.2.6 Neural Network Model

This section describes a Neural Network model that we hope provides insight into some of the computational processes that may be engaged when solving Multitask learning problems. Learning in this model uses a sequential Bayesian estimation algorithm, similar to the Elastic Weight Consolidation approach (Kirkpatrick et al., 2017), in which the prior over feature parameters for a second task is given by the posterior over feature parameters from the first. Bayesian learning for neural networks was first proposed by Mackay (Mackay, 1992a), and Bishop's textbook (Bishop, 2006) provides a comprehensive introduction to the methodology. A novel aspect of our modelling work is that we implement sequential Bayesian learning over both tasks and mini-batches of trials within tasks, allowing the model to predict learning dynamics at the time scale of tens of trials. The neural network models are exposed to exactly the same stimuli and stimulus-reward maps provided to experimental participants, and in the results section we compare simulations from these models with empirical findings.

In the machine learning literature, Multitask Learning means training a neural network simultaneously on data with multiple output labels but where the inputs are of the same type, for example, learning to detect multiple types of object from the same visual images (Caruana, 1998; Ng, 2017). Whereas, Transfer Learning means training a network sequentially on data from task A and then task B, but only tuning the final layer or layers using data from task B (Goodfellow et al., 2016; Ng, 2017). We have designed our neural network model to accommodate both types of learning (using a mini-batch buffer to potentially store trials from multiple tasks) although our empirical data is from a transfer task.

Figure 5.3 shows our neural network model. It has a dynamic structure in which new output subnetworks are added as new tasks are encountered. New connections are created from units trained on previous tasks to units created for the new task. It is via these "Transfer Connections" that transfer of knowledge from one task to another is possible.

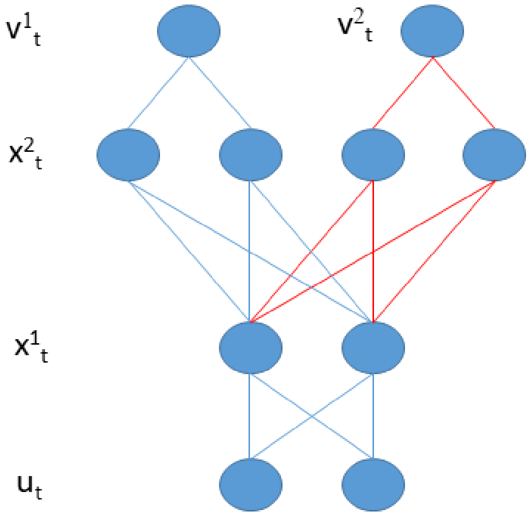


Fig. 5.3 **Neural Network Architecture.** Sensory inputs, u_t (where t indexes trial number), map onto feature detectors in the first hidden layer, x_t^1, according to equation 5.6 . The corresponding weight matrix W^1 defines the feature subspace. Hidden units in a second hidden layer, x_t^2, further transform these (equation 5.5) so that the output unit for the nth task, v_t^n, can provide task-specific value estimates for decision making (equation 5.4). Here we depict two output networks, one for each task (weather prediction and coin prediction). Connections in blue exist when learning task 1 and are augmented by those in red when learning task 2. For the modelling results in this paper we used a *minimal capacity* network, having a single unit in the first hidden layer, and an *increased capacity* network having two units in the first hidden layer.

Value Network

Let r_t be a Bernoulli reward signal received after taking decision $d_t = k$ where $k = \{1,2\}$ e.g. $\{Sun, Rain\}$ in Task 1, and $\{Heads, Tails\}$ in Task 2. A neural network is used to estimate the value (defined as the "expected reward" or "reward probability" (Sutton and Barto, 1998)) to be obtained when choosing $k = 1,2$

$$\begin{aligned} v_{t1}^n &\equiv p(r_t = 1 | d_t = 1) \\ v_{t2}^n &\equiv p(r_t = 1 | d_t = 2) \end{aligned} \tag{5.3}$$

An artificial agent making decisions using these values takes a decision on trial t for task n by sampling from the Bernouilli distribution v_t^n. Here we assume that the task variable $s_t = n$ is known (i.e. agent performs task s_t on trial t), that is, we have no task ambiguity.

We start our description of the neural net model at the output stage (top of Figure 5.3). In what follows x variables denote the hidden unit output values, w and W connection strengths, b biases, and a the activations before entering the activation function that produces the output of each node. Superscripts 1, 2 and n denote first and second hidden layers and nth output subnetwork. For each of $n = 1..N$ output subnetworks we have

$$\begin{aligned} v_{t1}^{n} &= \sigma(\tilde{a}_t^n) \\ v_{t2}^{n} &= 1 - \sigma(\tilde{a}_t^n) \\ \tilde{a}_{tk}^{n} &= \sum_{j=1}^{H} \tilde{w}_j^n \tilde{x}_{tj}^n + \tilde{b}^n \\ \tilde{x}_t^n &= P_n x_t^2 \end{aligned} \tag{5.4}$$

where P_n is a selection matrix which selects those H second layer units that belong to the nth output subnetwork. As our experimental paradigm involves binary decisions the above formulation with sigmoid functions suffices. More generally, with $K > 2$ potential actions, as with multi-armed bandits, values would need to be defined using softmax functions (Bishop, 2006). For $k = 1..H_2$ nodes in the second hidden layer we have

$$\begin{aligned} x_{tk}^{2} &= f_2(a_{tk}^2) \\ a_{tk}^{2} &= \sum_{j=1}^{H_1} W_{kj}^2 x_{tj}^1 + b_k^2 \end{aligned} \tag{5.5}$$

where $f_2()$ is the activation function of the second-layer units and W_k^2 specifies the dependency of the layer two hidden units on the layer one units. This reflects the structure shown in Figure 5.3. For $k = 1..H_1$ nodes in the first hidden layer we have

$$\begin{aligned} x_{tk}^{1} &= f_1(a_{tk}^1) \\ a_{tk}^{1} &= \sum_{j=1}^{D} W_{kj}^1 u_{tj} + b_k^1 \end{aligned} \tag{5.6}$$

where $f_1()$ is the activation function of the first-layer units, and D is the dimension of the input vector (i.e. number of inputs). A number of choices are available for the activation functions including Gaussian Error Linear Units (GELUs), $f(x) = x\Phi(x)$ where Φ is the Cumulative Density Function of the Gaussian distribution, Rectified Linear Units (RELUs), $f(x) = \max(0, x)$, Cosine Units, $f(x) = \cos(x)$ and linear units,

$f(x) = x$. See (Goodfellow et al., 2016) for a discussion of their relative merits. In this paper, for the first hidden layer we use linear units, and for the second hidden layer we use GELU units for Task 1 (as the mapping to output is nonlinear) and linear units for Task 2 (as the mapping to output is linear).

We augmented the inputs with a third input, $u_{t3} = 6$, reflecting the maximum number of slices in a pie, a variable readily available to human subjects. Thus we have $D = 3$ and the Add subspace can be represented with the weights $W^1_{k\bullet} = [1,1,-1]^T$ (sum of number of slices is maximal) and the Sub subspace with the weights $W^1_{k\bullet} = [1,-1,0]^T$ (difference in number of slices is zero).

We write the weights and biases that parameterise the neural network as $\{W,b\}$. Optimisation and statistical inference on these parameters is best described (and implemented in generic code) by first transforming them into a vector format (Nabney, 2003). We write this transformation generically as $\theta = \mathsf{Pack}[W,b]$. For example, given a single task this Pack function is

$$\theta = [\mathsf{vec}(W^1);\mathsf{vec}(W^2);\tilde{w}^1;b^1;b^2;\tilde{b}^1] \tag{5.7}$$

Given two tasks we have

$$\theta = [\mathsf{vec}(W^1);\mathsf{vec}(W^2);\tilde{w}^1;\tilde{w}^2;b^1;b^2;\tilde{b}^1;\tilde{b}^2] \tag{5.8}$$

After parameter estimation, we use the UnPack function to recover $\{W,b\}$.

Sequential Bayesian learning

We update model parameters not after each trial, but rather after a "mini-batch" or "block" of training trials. In this paper we use Sequential Bayesian learning (SBL) over tasks where separate blocks contain data from different tasks, and over blocks of learning trials within each task. We define the jth block of training data, R_j, to comprise the input and task variables along with the decisions made by an agent and the rewards received. We write this as $R_j = \{r_t, d_t, s_t, u_t\}$ for $t \in \tau_j$ where τ_j is the set of all trials in the jth block.

In this paper, once a block of training data has been used for offline learning it is then discarded. To make best use of this data we use SBL so that information is efficiently propagated from one block to the next. We define Y_j to denote all blocks of data up to and including block j. That is $Y_j = \{R_1, R_2, ..., R_j\}$. Bayesian estimation of θ proceeds over blocks such that the prior over θ is updated to a posterior using

Bayes rule

$$p(\theta|Y_j) = \frac{p(R_j|\theta)p(\theta|Y_{j-1})}{p(R_j)} \tag{5.9}$$

The likelihood of the jth block of data, $p(R_j|\theta)$, is defined in the following section (on "Model Likelihood"). We use a Laplace approximation to compute the posterior density, $p(\theta|Y_j)$, (see "Posterior Distribution" section below) which does not require explicit computation of the denominator term $p(R_j)$. We use a Gaussian prior over θ that factorises over parameters

$$p(\theta|Y_{j-1}) \quad = \quad \prod_{i=1}^{P} \mathsf{N}(\theta_i; m_{j-1}(i), \lambda_{j-1}(i)) \tag{5.10}$$

where m_{j-1} is the prior mean, λ_{j-1} is the prior precision and P is the number of network parameters. For the first learning episode on the first task the prior is initialised with mean, $m_{j-1} = 0_P$, and prior precision λ_{j-1} set so that hidden units with more inputs have smaller weights(Nabney, 2003). Given data R_j from the first learning episode, SBL is used to compute the posterior distribution. This is also chosen to factorise over parameters

$$\begin{aligned} m_j &= \mathsf{MAP}(R_j, m_{j-1}, \lambda_{j-1}) \qquad (5.11) \\ \lambda_j(i) &= \lambda_{j-1}(i) + \sum_{t \in \tau} v_{tk}(1 - v_{tk})\eta_t(i)^2 \\ \eta_t(i) &\equiv \frac{d\tilde{a}_t^n}{d\theta_i} \end{aligned}$$

Here MAP refers to a gradient-based offline algorithm (see "MAP Estimation" section below) that finds the maximum-a-posterior parameters. That is, the parameters that are a-posteriori most likely. A fully factorised Laplace approximation is used to estimate the posterior precisions (see "Posterior Distribution" section below). The quantity $\eta_t(i)$ is referred to as the "output sensitivity" (the variable $\tilde{a}_t^n$ produces the network output as shown in equation 5.4). Intuitively, network parameters θ_i that cause larger changes in the output will be better determined by the data and so be estimated more precisely.

The following sections on "Model Likelihood", "Prior Distribution", "Joint Distribution" , "MAP Estimation" and "Posterior Distribution" break down each of the above steps into more detail, but can be skipped if technical details are not of interest.

In SBL, as with all dynamic Bayesian models (such as the HMM or Kalman Filter), the posterior from one learning episode becomes the prior for the next, as shown by equation 5.9 which is applied recursively. If we were working with linear Gaussian models then SBL over J mini-batches would be exactly equivalent to Bayesian learning from a single batch (comprising all exemplars) (Bishop, 2006). However, as we are using a fully factorised Laplace approximation in a nonlinear model, its an empirical matter as to whether this procedure works well. The SBL approach, also known as Elastic Weight Consolidation (EWC), has previously been used for multitask learning of high-dimensional pattern recognition problems in the machine learning literature (Kirkpatrick et al., 2017). Here we apply SBL over mini-batches of data, as well as over tasks.

Model Likelihood

Let r_t be a Bernoulli reward signal received after taking action $d_t = k$. This paper employs an offline learning approach (similar in concept to offline Reinforcement Learning (Levine et al., 2020)), in which data is stored in a memory buffer. This buffer contains all inputs observed, task variables specified, decisions made and rewards received over a given set of trials, $R_j = \{u_t, s_t, d_t, r_t\}$. The likelihood over the jth batch of data is then given by

$$\begin{aligned} p(R_j|\theta) &= \prod_{t \in \tau_j} p(r_t|d_t, s_t, u_t) \\ p(r_t|d_t, s_t, u_t) &= [v_{tk}^n]^{r_t}[1 - v_{tk}^n]^{(1-r_t)} \end{aligned} \tag{5.12}$$

where $k = d_t$ is the selected action, $n = s_t$ is the selected task, u_t is the sensory input on trial t, and v_{tk}^n is the output of the value network. The Log Likelihood is

$$\begin{aligned} \log p(R_j|\theta) &= \sum_{t \in \tau_j} L_t \\ L_t &= r_t \log v_{tk}^n + (1 - r_t) \log(1 - v_{tk}^n) \end{aligned} \tag{5.13}$$

We refer to the quantity L_t as the sample log likelihood as it is based on a single data sample. The gradient, g_t, of the sample log likelihood is derived in the paper Menghi et al. (2021) and computed using backpropagation. The Hessian (curvature) matrix is given by

$$H(i, i') = \frac{d^2 \log p(R_j|\theta)}{d\theta_i d\theta_{i'}} \tag{5.14}$$

As in (Kirkpatrick et al., 2017) we compute the Hessian using an outer-product approximation (Mackay, 1992b).

$$\begin{aligned} H &= -\sum_{t \in \tau_j} v_{tk}^n (1 - v_{tk}^n) \eta_t \eta_t^T \\ \eta_t(i) &= \frac{d\tilde{a}_t^n}{d\theta_i} \end{aligned} \tag{5.15}$$

where k and n index the decisions made and tasks undertaken on trial t. The output sensitivity, η_t, can be computed using back-propagation (see Menghi et al. (2021)).

Prior Distribution

The log prior is given by

$$\begin{aligned} \log p(\theta|Y_{j-1}) &= \sum_{i=1}^{P} \log \mathsf{N}(\theta_i; m_{j-1}(i), \lambda_{j-1}(i)) \\ &= -\frac{P}{2}\log(2\pi) + \frac{1}{2}\sum_{i=1}^{P}\left(\log \lambda_{j-1}(i) - \lambda_{j-1}(i)[\theta_i - m_{j-1}(i)]^2\right) \end{aligned} \tag{5.16}$$

with gradient and curvature given by

$$\begin{aligned} \frac{d \log p(\theta)}{d\theta_i} &= -\lambda_{j-1}(i)[\theta_i - m_{j-1}(i)] \\ \frac{d^2 \log p(\theta)}{d\theta_i^2} &= -\lambda_{j-1}(i) \end{aligned} \tag{5.17}$$

Joint Distribution

We can then define the log joint density and its gradient as

$$\begin{aligned} J &= \log p(R_j|\theta) + \log p(\theta|Y_{j-1}) \\ \frac{dJ}{d\theta_i} &= \frac{d \log p(R_j|\theta)}{d\theta_i} + \frac{d \log p(\theta)}{d\theta_i} \\ &= \left(\sum_{t \in \tau_j} g_t\right) - \lambda_{j-1}(i)[\theta_i - m_{j-1}(i)] \end{aligned} \tag{5.18}$$

where g_t is the gradient of the sample log likelihood derived can be found in the paper by Menghi et al. (2021) and computed using backpropagation. Bayesian learning from data set R_j can then proceed by ascending the gradient of the log joint to reach a local maximum of the posterior density. Inclusion of the prior term ensures that parameter estimates are constrained to be similar to values found useful for previous blocks of data or for previous tasks (see last term in above equation).

Importantly, the prior precision λ_{j-1} controls the strength of this effect, and this quantity increases in proportion to the number of data samples so far observed (in sequential Bayesian Learning for linear Gaussian models the posterior precision equals the prior precision plus the data precision and therefore always increases - see "Posterior Distribution" below). This leads to the desirable property that the connection parameters converge to high precision solutions and is the mechanism described by Kirkpatrick et al. Kirkpatrick et al. (2017) for protecting previously learnt representations.

MAP Estimation

Offline learning proceeds using gradient ascent. For the implementation in this chapter, rather than using fixed step size updates we use a line search algorithm (Press et al., 1992). Specifically, on iteration *it* of batch learning we use

$$\theta(it+1) \quad = \quad \theta(it) + \alpha \frac{dJ}{d\theta} \tag{5.19}$$

where $\frac{dJ}{d\theta}$ is the gradient of the log-joint. Optimal values for α are found using a single-variable bounded nonlinear function minimisation (implemented using fminbnd.m in Matlab (Mathworks, Inc) with step sizes bounded between 0 and 1) to minimise the negative Log Joint. If the above does not result in a decreased cost function (increased Log Joint) the maximum step size is reduced by a half and the process repeated. This can occur for a further three halvings of the maximal step size.

All θ values are initialised by sampling from the prior. This is a stochastic process which leads to different results on each simulation run. Other than this sampling process, the optimisation is deterministic. Additionally, we found that the posterior landscape contains local maxima. We therefore implemented a multistart optimisation procedure in which optimisation is re-initialised (with a different sample from the prior) until a satisfactory solution was found (see e.g. (Hickernell and Yuan, 1997) for alternative multistart approaches). This was defined as a solution with an average trial likelihood of at least $pc_T = 0.60$. This is computed by dividing the log likelihood (equation 5.13) by the number of trials and then exponentiating, and is also equivalent to the average probability of being correct (Daw, 2011) . If no such solution is found within a maximum of `maxstarts` starts the best solution is returned. For the results in this paper we used `maxstarts` $= 3$, the motivation for which is described in the results section.

Posterior Distribution

We compute the posterior distribution over θ given data, Y_j, from blocks 1 to j. We use an approximate posterior based on a factorised Laplace approximation

$$\begin{aligned} p(\theta|Y_j) &= \prod_i^P p(\theta_i|Y_j) \\ p(\theta_i|Y_j) &= \mathsf{N}(\theta_i; m_j(i), \lambda_j(i)) \end{aligned} \tag{5.20}$$

where the posterior mean m_j is the MAP estimate of θ found on the jth block of data (see above section). From the Laplace approximation we have

$$\begin{aligned} \lambda_j(i) &= \Lambda_j(i,i) \\ \Lambda_j &= -H + \Lambda_{j-1} \end{aligned} \tag{5.21}$$

where the posterior precision, Λ_j, is the sum of the data precision $(-H)$ and the prior precision. Given we only need the diagonal elements of the Hessian we can write

$$\lambda_j(i) = \lambda_{j-1}(i) + \sum_{t \in \tau_j} v_{tk}^n (1 - v_{tk}^n) \eta_t(i)^2 \tag{5.22}$$

5.3 Results

5.3.1 Behavioural Results

Our main hypothesis is that learning will be facilitated for tasks that share a common subspace where facilitation of learning could be manifested as faster and/or more accurate learning. In order to test the effect of subspace on participants performance, we divided the analysis into two parts. First, we calculated how participants performance in the second task was correlated to that in the first, and then tested whether these correlations differed as a function of subspace. Second, we performed a two-way mixed ANOVA with dependent variable accuracy and independent variables of task (first or second) and subspace (same or different).

Positive versus Negative Correlations in Same versus Different Subspace

Participants performance in Task 2 correlated with their performance in Task 1 when the second task was in the same ($r(38) = 0.42$, $p = 0.007$) but not different ($r(38) = -0.089$, $p = 0.584$) subspace. The two correlations were significantly different from each other (Fisher's z-transform $z = 2.31$, $p = 0.021$).

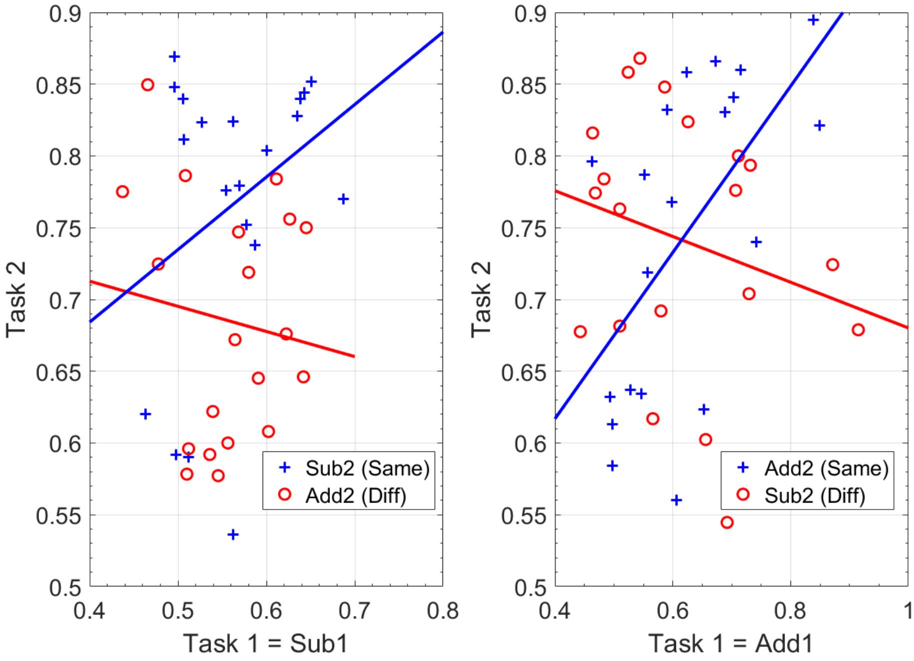


Fig. 5.4 **Correlations over Subjects.** For Task 1 = Add1 (right panel), performance in Task 2 is significantly positively correlated with performance in Task 1 when the second task is in the same subspace (blue line and crosses), but negatively when Task 2 is in a different subspace (red line and circles), and the difference in these correlations is significant. For Task 1 = Sub1 (left panel), these effects are not significant although the pattern is similar.

We then tested whether this effect depended on the Task 1 subspace (Add/Sub) with data and lines of best fit shown in Figure 5.4. For addition, participants performance in Task 1 significantly correlated with their Task 2 performance for the same ($r(38) = 0.583$, $p = 0.007$) but not different ($r(38) = -0.236$, $p = 0.317$) subspace. These two correlations were significantly different from each other (Fisher's z-transform, $z = 2.647$, $p = 0.008$). The signs of the effects and significant inferences are consistent with the overall picture.

In the subtraction condition, participants performance in Task 1 was not significantly correlated with their Task 2 performance for the same ($r(38) = 0.317$, $p = 0.174$) or different ($r(38) = -0.129$, $p = 0.588$) subspace. These two correlations were not significantly different from each other (Fisher's z-transform, $z = 1.333$, $p = 0.182$). The signs of the effects are consistent with the overall picture but there were no significant inferences.

We do not know why there would be no significant correlations for the Sub subspace but note that the variance of Task 1 accuracies is significantly lower for Sub1 than Add1 (Std Dev=0.06 for Sub, 0.11 for Add, Levene's test $F(1,78) = 10.94$,$p = 0.001$). Generally, lower variances make it more difficult to detect co-variances/correlations.

Increases in Task 2 accuracy for Same versus Different Subspace

We then performed a two-way mixed design ANOVA with dependent variable accuracy and independent factors of (i) task (1/2) and (ii) subspace (same/different). Overall, we found only a main effect of task (Task, F(1,78) = 103.71, $p < 0.001$; Subspace, F(1,78) = 2.043, $p = 0.156$; interaction, F(1,78) = 1.674, $p = 0.199$). We then performed two separate two-way mixed design ANOVAs for the Add and Sub subspaces. For Add, only the main effect of task was significant (Tasks, F(1,38) = 29.845, $p < 0.001$; Subspace, F(1,38) = 0.024, $p = 0.876$; interaction, F(1,38) = 0.001, $p = 0.974$). For Sub, we found all the main effects and the interaction to be significant (Tasks, F(1,38) = 100.86, $p < 0.001$; Subspace, F(1,38) = 5.669, $p = 0.022$; interaction, F(1,38) = 5.195, $p = 0.028$). The increases in Task 2 accuracy for Same versus Different Subspace were 4.3% overall, 8.2% for Task 1=Sub1, and 0.4% for Task 1=Add1. Figure 5.5 plots the mean accuracies for each combination of Task 1/2.

Overall, the empirical subspace effects are a significant correlation difference for the Add subspace, and a significant interaction (improvement in mean accuracy) for the Sub subspace.

5.3.2 Modelling Results

We used the neural networks described earlier with the following model and optimisation parameters: 4 hidden units per output sub-network, convergence tolerance = 0.001, accuracy threshold pcT=0.60, MaxIterations=64, and GELU activation functions in the output sub-networks for Task 1. GELU were preferred over RELU activation functions as, in preliminary work, they produced more similar performance levels on Sub1 and Add1 mappings. For the Task 2 output sub-networks we used linear activations. This was motivated by the fact that these tasks are linear functions of the first layer hidden units and, empirically, this led to better performance on Task 2.

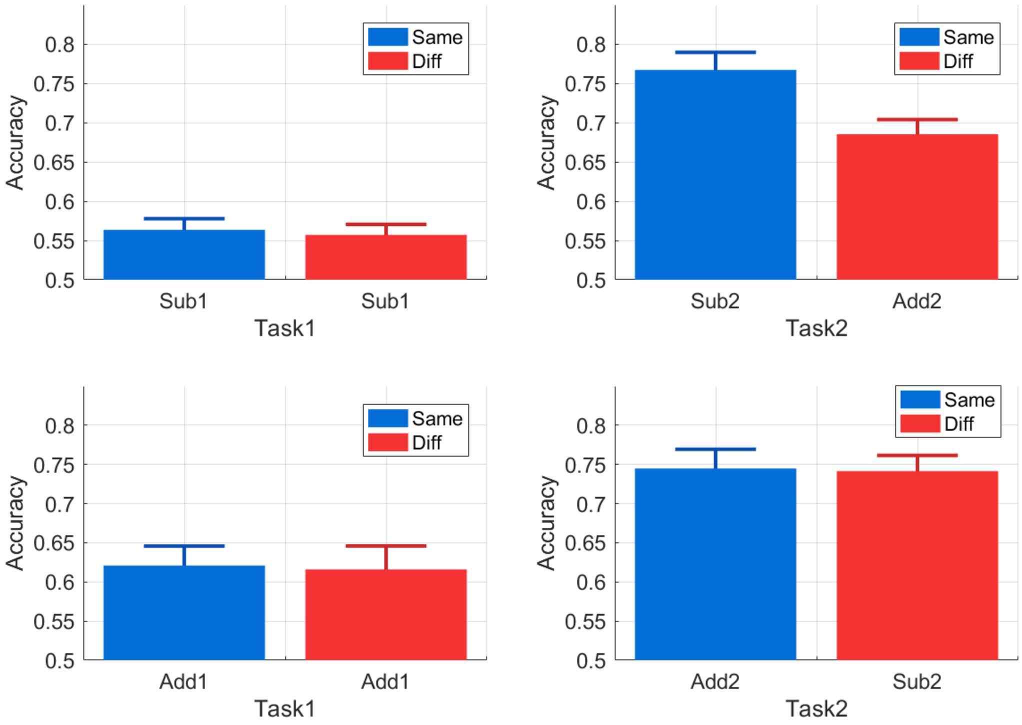


Fig. 5.5 **Increases in Task 2 Accuracy for Same versus Different Subspace.** The barplots show the mean accuracies for Tasks 1 and 2 as a function of whether the second Task is in the same subspace as the first. These results are shown separately for Task 1 = Sub1 (top row) and Task 1 = Add1 (bottom row). For Task 1 = Sub1 there is a significant increase in Task 2 accuracy (of 8.2%) for same versus different subspace (top right). For Task 1 = Add1, mean Task 2 performances are not significantly affected by subspace (bottom right). The error bars indicate the standard error of the mean.

We then ran Sequential Bayesian Learning in one of two modes: SBL over tasks, and SBL over blocks (and tasks). The following sections on "Minimal Capacity Network", "Increase Capacity Network" and "Reduced Precision Representation" are based on the SBL over Tasks approach.

Sequential Bayesian Learning over Tasks

First, we created a data set for each mapping as follows. We used 100 input stimuli, u_t, drawn from a uniform distribution covering input space. These inputs were presented to a neural net model whose parameters were sampled from their prior distribution (see equation 5.10 and the following paragraph). This network then made decisions, d_t, by sampling from neural net outputs (see equation 5.3 - highly stochastic decisions as none of its parameters were yet tuned) and received rewards r_t according to one of the mappings from the behavioural experiment (Sub1, Add1, Sub2, Add2 - see section on "Stimulus-Reward Maps"). This created a data set, $R_j = \{u_t, s_t, d_t, r_t\}$ with $t = 1..100$, for that mapping. This was repeated to create a data set for each mapping. The accuracy of a model was then measured using the average probability of being correct (also known as the average trial likelihood - see section on "MAP Estimation") as computed over the training data.

We then tuned the accuracy of Task 1 learning to broadly match the behavioural data, by changing the maximum number of "starts", `maxstarts`, of the multistart optimisation algorithm (see MAP Estimation section). We obtained average task accuracies of 0.56, 0.63, 0.67, 0.69 and 0.73 for `maxstarts` equal to 1, 2, 3, 4 and 8 respectively. This parameter helps the optimiser avoid local maxima by restarting the optimisation with a different initialisation. In what follows we used `maxstarts`=3 and ran 40 simulations per Task 1/2 combination as per the human experiments. Because we were using SBL over tasks, the prior over network parameters for Task 2 was the posterior from Task 1 (see equation 5.9). We emphasise that no model parameters were specifically tuned to the particular subspace (Add or Sub) or to individual subject data. The models were simply provided with the above parameter settings, and the same stimuli and reward functions provided to the participants.

Minimal Capacity Network

Here we present results obtained with a *minimal capacity* neural network model having only a single hidden unit in the first layer.

For Task 1 = Sub1, Task 1 performance was 0.66, and Task 2 performance was 0.77 for same and 0.64 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies ($t(39) = 4.76, p < 0.001$) and Task 2 accuracies were significantly higher for same versus different subspace ($t(39) = 3.46, p = 0.001$). The correlation between Task 2 and Task 1 performance was significantly positive ($r = 0.38, p = 0.017$) for same subspace and negative ($r = -0.78, p < 0.001$) for different subspace.

For Task 1 = Add1, Task 1 performance was 0.64, and Task 2 performance was 0.76 for same and 0.65 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies ($t(39) = 4.79, p < 0.001$) and Task 2 accuracies were significantly higher for same versus different subspace ($t(39) = 2.57, p = 0.014$). The correlation between Task 2 and Task 1 performance was significantly positive

$(r = 0.47, p = 0.002)$ for same subspace and negative $(r = -0.68, p < 0.001)$ for different subspace.

Thus, these modelling results show transfer effects of the sort exhibited in the behavioural data i.e. both increases in Task 2 performance, and correlations between Task 1 and Task 2 performance.

Quantitatively, the standard deviations of hidden unit parameters were 24 times smaller in the posterior (after learning Task 1) than the prior (before Task 1). The figure of 24 is an average over all weights in the hidden unit and over both Add and Sub Tasks. The precisions were thus $24^2 = 576$ times higher in the posterior (after Task 1) than the prior (before Task 1). As the prior (before Task 2) is set to the posterior (after Task 1) this strongly constrains the Task 2 solution to be close to the Task 1 solution (see last term in last row of equation 5.18).

Increased Capacity Network

We then repeated the simulations but this time with an *increased capacity* neural network model having two hidden units in the first layer. All transfer effects disappeared.

For Task 1 = Sub1, Task 1 performance was 0.65, and Task 2 performance was 0.79 for same and 0.79 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies $(t(39) = 6.87, p < 0.001)$ but Task 2 accuracies were not significantly higher for same versus different subspace $(t(39) = -0.023, p = 0.982)$. Correlations between Task 2 and Task 1 performance were not significant for same $(r = 0.05, p = 0.743)$ or different $(r = 0.26, p = 0.110)$ subspace.

For Task 1 = Add1, Task 1 performance was 0.61, and Task 2 performance was 0.78 for same and 0.78 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies $(t(39) = 6.67, p < 0.001)$ but Task 2 accuracies were not significantly higher for same versus different subspace $(t(39) = 0.118, p = 0.907)$. Correlations between Task 2 and Task 1 performance were not significant for same $(r = -0.31, p = 0.053)$ or different $(r = 0.04, p = 0.786)$ subspace.

These results show that no transfer effects were evident in the increased capacity network, suggesting that a minimal capacity network may be an important factor underlying the behavioural results.

Quantitatively, the standard deviations of hidden unit parameters were 19 times smaller in the posterior (after learning Task 1) than the prior (before Task 1). The figure of 19 is an average over all weights in both hidden units and over both Add and Sub Tasks. The precisions were thus $19^2 = 361$ times higher in the posterior (after Task 1) than the prior (before Task 1). This is a smaller increase than for the minimal capacity network, thus rendering Task 2 solutions somewhat less constrained to be similar to Task 1 solutions (see last row of equation 5.18). However, we expect that the main factor in the loss of transfer effects is the increased representational capacity of the network (the required subspace for the Task 2 mapping can be implemented by either hidden unit or distributed over both).

Reduced Precision Representation

We also repeated the simulations with the minimal capacity network but this time resetting the posterior precision of network parameters from Task 1 to their prior precision at the beginning of learning. All transfer effects disappeared.

For Task 1 = Sub1, Task 1 performance was 0.62, and Task 2 performance was 0.78 for same and 0.77 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies ($t(39) = 6.43, p < 0.001$) but Task 2 accuracies were not significantly higher for same versus different subspace ($t(39) = 0.307, p = 0.761$). Correlations between Task 2 and Task 1 performance were not significant for same ($r = 0.17, p = 0.302$) or different ($r = -0.30, p = 0.063$) subspace.

For Task 1 = Add1, Task 1 performance was 0.67, and Task 2 performance was 0.77 for same and 0.77 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies ($t(39) = 3.55, p = 0.001$) but Task 2 accuracies were not significantly higher for same versus different subspace ($t(39) = 0.086, p = 0.932$). Correlations between Task 2 and Task 1 performance were not significant for same ($r = 0.07, p = 0.670$) or different ($r = -0.00, p = 0.983$) subspace.

These results show that no transfer effects were evident with reduced precision representations, suggesting that Bayesian estimation may be an important factor underlying the behavioural results.

Reducing the posterior precision effectively removes the protection afforded by Sequential Bayesian Learning to the newly learnt representation, thus allowing it to be overwritten when learning Task 2 (quantitatively, the precision variable, λ_{j-1}, in the last row of equation 5.18 is on average 576 times smaller than for the minimal capacity network - see above section). This results in Task 2 solutions being only very weakly constrained to be similar to Task 1 solutions, thereby eliminating the subspace effect.

Sequential Bayesian Learning over Blocks and Tasks

We now report results using Sequential Bayesian Learning over blocks and tasks for the minimal capacity model. We chose our block size to be 25 trials as preliminary analysis (see paper by Menghi et al. (2021) on "Within-versus-Between Block Learning") found there was demonstrable learning within the 50 trial blocks in the empirical data. In SBL over blocks and tasks, the prior over network parameters for learning from data block j is the posterior from block $j-1$ (see equation 5.9). Decisions on data from block j were made by the network before training on that data. The accuracy of a model was assessed using the average probability of being correct (also known as the average trial likelihood - see section on "MAP Estimation"), as computed over the test data set (we refer to this as "test data" as the model has not yet been trained on it). Learning accuracies were then averaged over neighbouring 25-trial blocks to present the model learning curves in Figure 5.6 (right panel). The equivalent learning curves for the behavioural data are shown in the same Figure (left panel). The empirical data show averages over 80 subjects, 40 in each group (same/different subspace). The simulated data are from a minimal capacity neural net as described above, with 40 simulations per group. The simulated data exhibit similar transfer effects to the behavioural data.

For Task 1 = Sub1, Task 1 performance was 0.62, and Task 2 performance was 0.67 for same and 0.54 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies ($t(39) = 3.21, p = 0.003$) and Task 2 accuracies were significantly higher for same versus different subspace ($t(39) = 3.61, p < 0.001$). The correlation between Task 2 and Task 1 performance was significantly positive ($r = 0.74, p < 0.001$) for same subspace and negative ($r = -0.56, p < 0.001$) for different subspace. These were significantly different from each other (Fisher's Z transform: $p < 0.001, z = 6.77$).

For Task 1 = Add1, Task 1 performance was 0.62, and Task 2 performance was 0.68 for same and 0.54 for different subspaces. Same-subspace Task 2 accuracies were significantly higher than Task 1 accuracies ($t(39) = 2.90, p = 0.006$) and Task 2 accuracies were significantly higher for same versus different subspace ($t(39) = 3.42, p = 0.002$). The correlation between Task 2 and Task 1 performance was significantly positive ($r = 0.81, p < 0.001$) for same subspace and negative ($r = -0.58, p < 0.001$) for different subspace. These were significantly different from each other (Fisher's Z transform: $p < 0.001, z = 7.64$).

The behavioural and neural net data are therefore similarly matched in terms of the positive versus negative correlations for same versus different subspace, and relative increases in Task 2 performance for same versus different subspace. But there are also a number of discrepancies. For example, accuracies at the beginning of the second task experience a sudden drop for the model but not for behaviour, and increases in performance are rather sudden for the model but more gradual for behaviour. These discrepancies are addressed in the Discussion.

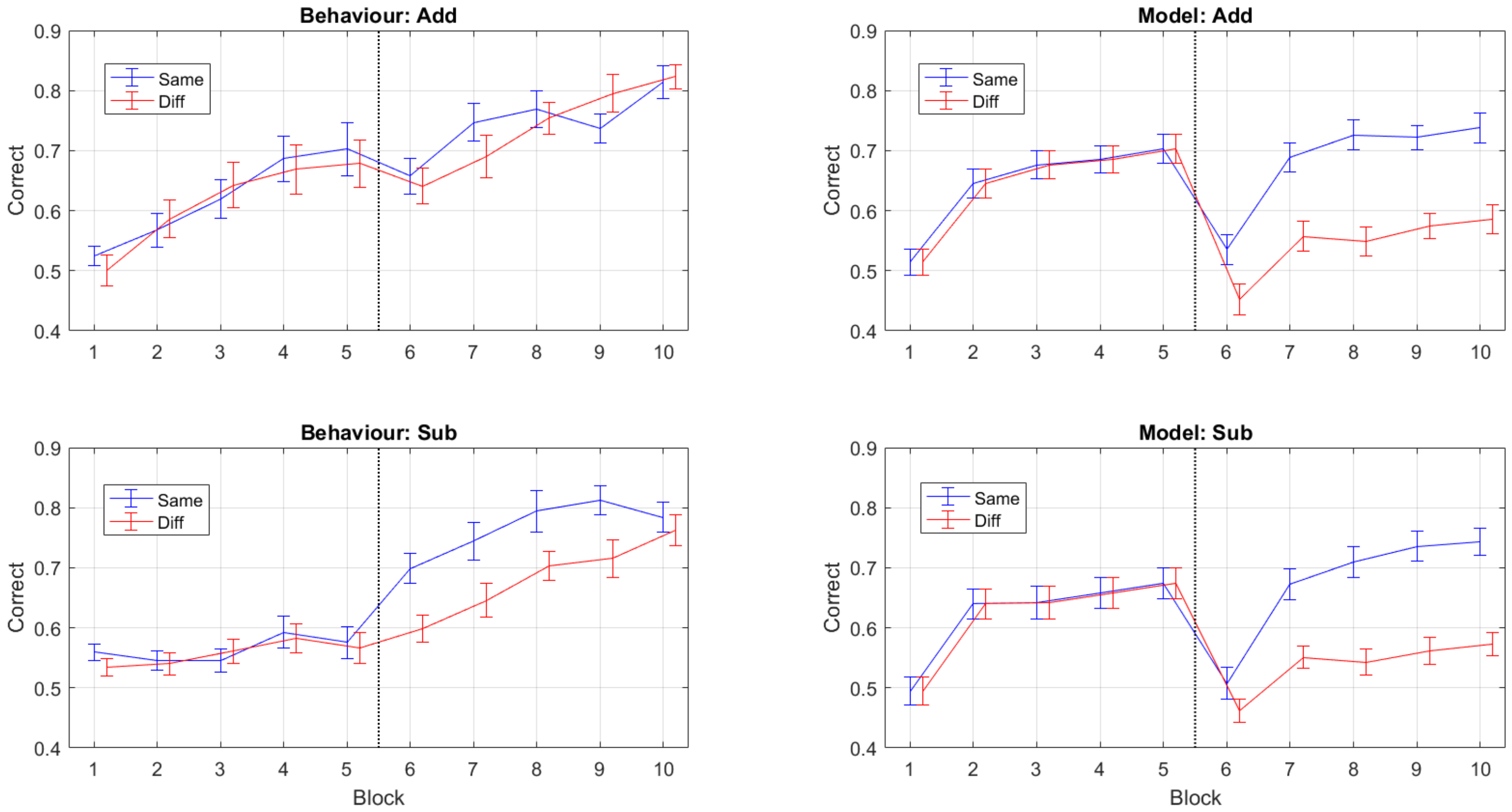


Fig. 5.6 **Behavioural and Model Learning Trajectories.** The behavioural learning trajectories are averaged over 40 subjects for each of the Sub and Add subspaces. The model learning trajectories are 40 simulations from the minimal capacity neural net model for each of the Sub and Add subspaces. Each block comprises 50 trials and the vertical line denotes that blocks 1 to 5 are from Task 1, and 6 to 10 from Task 2. All error bars indicate the standard error of the mean and the red and blue curves have been offset to improve readability.

5.4 Discussion

We found evidence from our behavioural results in support of our main hypothesis that learning would be facilitated (positive transfer) for tasks that share a common subspace. However, the nature of these transfer effects depended on the subspace. For the Add subspace, transfer effects manifested as positive correlations between Task 2 and Task 1 accuracy (Fig.5.4, right panel). Whereas for the Sub subspace, transfer effects manifested as higher average accuracy in Task 2 (Fig.5.5, top right). We do not have an explanation as to why these transfer effects should be different. Clearly, more empirical data is required over a larger number of paired tasks to investigate further.

In our modelling work we found that a minimal capacity neural net model, with a single unit in the first hidden layer, trained using sequential Bayesian learning produced transfer effects that were broadly consistent with the empirical data (Fig.5.6). This model produced positive correlations between Task 1 and Task 2 accuracy, and differences in Task 2 accuracy, for both subspaces. We then investigated two variants of this approach. First, an increased capacity model with two units in the first hidden layer. Second, a reduced precision model in which sequential Bayesian learning was interfered with by reducing the posterior precision after learning Task 1 to its prior level. All transfer effects disappeared for both of these variants indicating that both a minimal capacity "bottleneck" and sequential Bayesian learning are necessary mechanisms for replicating behavioural findings in this modelling framework.

5.4.1 Negative Transfer Effects

In our behavioural experiments participants only performed Task 1 once, and were not told that they would be performing it again. This may have led to an expectation that they would not need to retain Task 1 representations. If this were the case, an increased capacity network would not be needed; a minimal capacity representation would suffice that could be overwritten during the second task. However, we do not know the subject's expectations because they were not explicitly manipulated. This does however motivate a future experiment. If participants were to be told that they will be tested again on Task 1, after learning Task 2, then the prediction is that we will not see negative transfer effects when Task 2 is in a different subspace to Task 1. Such an experiment would speak to a distinction in the neural network literature (Parisi and Wermter, 2019) in which two of the major approaches in the area of transfer/continual learning are to use (i) dynamic architectures in which new representational capacity is added for each new task to be performed and (ii) regularisation approaches in which a fixed architecture is used but regularisation prevents forgetting.

5.4.2 Future Modelling Work

There are a number of discrepancies between the model and behavioural learning trajectories. First, accuracies at the beginning of the second task experience a sudden drop for the model but not for the behavioural trajectories. This is likely an artefact of using a modelling approach in which parameters are updated after each block of trials rather than after each trial. This necessarily means that accuracies on the first block of the second Task will be close to chance level.

Second, for the Sub subspace data, same/different trajectories appear to converge towards the end of Task 2 for the behavioural data but not for the model. This might suggest, for example, that there is a recovery mechanism in place in which tight inappropriate priors (inappropriate for different subspace subjects) are replaced by vaguer priors, allowing a more standard learning trajectory to evolve. In preliminary work we had proposed a mechanism in which the prior precision is gradually reduced if learning does not go well. This could be replaced by more formal models, for example with mixture model priors allowing switching from one prior to another during learning. However, as this empirical effect is only evident for one of the experimental groups (Sub not Add) we have decided to postpone further modelling until more data is available.

The model we have presented employs a within-block multistart optimisation procedure such that if estimated parameters do not provide a sufficiently good solution, the estimation is repeated, with a total of $\mathsf{maxstarts} = 3$ model fits allowed per block. This serial model fitting process is biologically implausible but could potentially be implemented using parallel architecture and may fit in with evidence that up to three or four decision making strategies can be simultaneously updated and monitored (Collins and Koechlin, 2012). One possibility for future modelling, however, would be to use a moving window of samples to which the model is fitted, rather than splitting the samples into non-overlapping blocks. This would remove the "blockiness" of the results referred to above, and the inherent stochasticity of the approach may remove the need for multistart optimisation. Such an approach has been used to good effect in recent work on dynamical models (Fitzgerald et al., 2020).

5.4.3 Elemental and Configural Learning

One distinction in the category learning literature is between elemental (or linear feature-based) learning and configural (or object-based) learning. Duncan et al. (2018) showed that people switch between learning styles as a function of the empirical contingencies in the data (e.g. elemental if reward functions are indeed a linear function of features). They also noted, however, that a proportion of participants persisted with a configural strategy even when a more efficient elemental strategy would have sufficed.

In a similar experiment, Farashahi et al. (2017) showed that people shift from elemental to configural representations as they learn. They describe an elemental RL agent in which values are learnt for each discrete setting (out of M settings) of each input variable (out of D inputs) - thus requiring up to $D \times M$ values to be learnt. The overall value of a stimulus is then given by a linear combination of feature values. This is to be contrasted with an "object-based" RL agent which learns a value for each object (or "configuration"). Given that there are M^D possible objects, this requires learning a potentially much larger set of parameters. Empirical results demonstrated that people initially employed a (linear) feature based strategy and later switched to an object-based one. This took place even when the true contingencies were not linear.

One of the goals of the current paper was to explore mechanisms underlying learning of rather general nonlinear mappings (which are in turn composed of an input to latent space function (subspace) and a latent space to output function). Both elemental and configural learning approaches are, however, highly suboptimal for these tasks, the elemental strategy because it is linear, and the configural strategy because it is statistically inefficient (having a number of parameters that grows exponentially). More specifically, an elemental learning strategy could be applied for the Sub2 and Add2 linear value mappings in the current study, but would be unsuccessful for the Sub1 and Add1 reward functions which are nonlinear. A configural strategy would be highly inefficient for any of the tasks due to the large number of configurations, $M^D = 25$.

5.4.4 Learning accuracy across mappings

There is a long-standing debate in the category learning literature about whether and how humans can learn non-linearly separable categories (Levering et al., 2020; Medin and Schwanenflugel, 1981; Shepard et al., 1961). Medin and Schwanenflugel (1981) and Levering et al. (2020) both find that non-linearly separable categories are easier to learn than linearly separable ones. Their experiments used three binary input features (resulting in only 8 unique stimulus vectors, 6 of which were shown during learning) and binary classification labels. Importantly, the input vectors were chosen so that the "well-formedness" of the categories (and therefore, presumably, the maximum achievable classification rates) were matched across linear and nonlinear tasks.

For the tasks in the current paper, the maximum achievable classification rates were closely matched across the linear (93 per cent) and nonlinear (95 per cent) tasks. In contrast to previous work, we found that the linear mappings were easier to learn than the non-linear mappings (mean accuracy = 73.5% for linear, 58.9% for nonlinear, see Appendix D for further details). However, as the linear tasks were always performed after the nonlinear tasks, this could be due to an order effect, or indeed the transfer effects that are the main interest of this paper. The linear/nonlinear issue could be addressed in a future experiment in which participants learn just a single task.

There is a literature on "human function learning" that presents participants with data points sampled from one-dimensional functions and asks them to predict where future samples will be drawn from (Griffiths et al., 2009; Lucas et al., 2015). This literature shows that people have a preference for linearly increasing rather than decreasing completions i.e. positive rather than negative functions. These findings on 1-dimensional functions are perhaps difficult to extrapolate to the 2-dimensional functions used in the current paper. Empirically, we did not find that (collapsing across task 1 and 2) the add subspace functions were learnt more accurately than the subtract subspace functions (see Appendix D for details). One difference we did find was that more participants correctly declared the rule underlying the Sub2 map than the Add2 map (see Appendix D).

5.4.5 Rule-based Learning

Subjects who performed the Sub1 and Sub2 or Add1 and Add2 tasks did better (than those who performed Sub1 and Add2 or Add1 and Sub2), but was this really because the tasks were in the "same subspace". Are there not other similarities among these tasks ? For example, that both required the same logical operation or rule-based operation as an intermediate step ? This speaks to a body of work in rule-based learning. One approach to this topic is the "Rational-Rules (RR)" model (Goodman et al., 2018) which formalizes a statistical learner that operates over the space of Boolean propositional logic expressions e.g. "A or B", "A and B", "A or (B and C)". In an fMRI study, Ballard et al. (2018) found that the pattern of striatal responses was more consistent with prediction errors derived from such a rule-learning model than a Reinforcement Learning model. We accept therefore that there could be an ambiguity in interpretation here and that resolution of this issue requires further empirical work, perhaps with experiments using nonlinear and/or multivariate subspaces that are not readily expressible using rational rules.

5.4.6 Declarative Learning

In additional statistical analysis presented in Appendix D we show that subjects who were able to declare a correct rule-based strategy also showed a stronger subspace effect. We also show, however, that subjects who performed better in the first task also showed a stronger subspace effect. Further analysis then showed these two effects to be moderately collinear (as those who declared a correct rule-based strategy also performed better on the first task). Therefore, with the current data, we are unable to infer which of these factors (declarative learning or accurate learning) drives the subspace effect. Again, further experiments are required perhaps using nonlinear and/or multivariate subspaces.

5.4.7 Creation or Selection of Representations ?

Are new representations created i.e. features learnt ? Or, are pre-existing representations prioritized as potentially useful and selected from, as proposed by Collins and Koechlin (2012). For example, there may be representations in brain regions encoding for numerosity (Harvey and Dumoulin, 2017) that already encode differences and sums over numbers of items (see also Chapter 3). An additional component in the model proposed in (Collins and Koechlin, 2012) is a process that creates new stimulus-response mappings from old ones. It could be that the offline learning algorithm we have described in this paper, or some similar process, plays this creative role.

5.4.8 Structure Learning

This paper fits in more broadly with previous studies of structure learning which show that people take advantage of shared structure across tasks. For example, Costa et al. (2015) studied rhesus monkeys taking part in a probabilistic two-armed bandit reversal learning task in which monkeys were exposed to a distribution of reversal times and were able to make use of this information during decision making. Tomov et al. (2019) studied people engaged in a novel two-step decision making task, finding evidence that human subjects use a multitask learning strategy that maps previously learned policies to novel scenarios. As with our paper, the rewards were a function of multiple input features and this changed across tasks. In theoretical and simulation work, Franklin and Frank (2018) address the problem of transfer learning in a Markov Decision Process context by designing a non-parametric Bayesian agent that can generalise across state-transition functions, reward functions or both.

Radulescu et al. (2019) review recent research which suggests that, in complex learning tasks, human behaviour is consistent with an integrative model in which approximate Bayesian inference acts as a source of selective attention, allowing Reinforcement Learning (RL) to focus on the relevant dimensions for decision making. Within the Bayesian approaches, Latent Causal Models (also known as Non-Parametric models - See Chapter 2) organise experience into similar episodes, and Probabilistic Programming allows rules based on logical operations to be inferred. In principle, it may be possible to adapt the Latent Causal Model framework to the study of transfer learning, for example, by allowing for common causes among tasks but adapting contingencies between causes and outputs. This is an avenue to be explored in future work.

5.4.9 Transfer Learning

The study of transfer learning has a long history in psychology (Perkins and Salomon, 1992), and more recently in the fields of cognitive training and cognitive neuroscience (Schubert et al., 2014). A key qualitative concept here is the notion of near versus far transfer where distance reflects how similar the different learning contexts are. This may naturally map onto the quantitative measures defined in Bayesian learning e.g. the probability density of task-two feature parameters under the task-one posterior. Noack et al. (2014) propose a theory-driven approach to studying transfer effects in cognitive training research. They argue that data should be analysed within the context of theoretically motivated (using hierarchical cognitive process models) and/or latent factor analysis methods, so that inferences can be made at the level of latent processes. The work in this paper concurs with this latent and hierarchical perspective, but whereas Noack et al. deconstruct existing batteries of cognitive tasks, our goal is to design new tasks with better defined relationships among latent and observed variables.

Building on long established models of cognitive control, Musslick and Cohen (2019) present a three-layer neural network architecture with stimulus layer, hidden layer and output layer but augmented with task units that affect the hidden and output layers. Learning in these networks allows a mapping between task and hidden units such that irrelevant hidden units are inhibited. The network is trained on multiple tasks with simulations showing interference between tasks that required activation of common hidden units (representations). By adding temporal persistence to the hidden and output layer activations (reflecting the dynamics observed in biological networks) they were able to explain well-established phenomena such as the psychological refractory period. This important issue of task switching and maintenance has been neglected in our paper. We have instead assumed that only the relevant output subnetwork is engaged while the other is inhibited, without providing a mechanism for this.

Flesh et al. (2018) compared human learners and neural net learners in transfer learning tasks involving categorisation of naturalistic images of trees. As expected, the neural network suffered from catastrophic forgetting when samples of each task were blocked rather than interleaved. Conversely, human performance was better if the samples were blocked rather than interleaved. They showed that neural net performance on blocked data could be improved by pre-training using a generative model approach. This was implemented using an autoencoder in which a two-dimensional "bottleneck" layer enabled learning of the appropriate two-dimensional subspace. This subspace comprised the two relevant features that predicted reward across tasks and is analogous to the one-dimensional subspaces studied our paper.

Wu et al. (2020) studied the transfer of knowledge between spatial and conceptual domains. They specified a series of two-dimensional reward maps which were identical over both domains and found transfer effects from spatial to conceptual domains but not vice-versa. Impressively, transfer was examined using eighty different reward maps (rather than the four examined in the current paper). Subject's behaviour was well described by Gaussian Process (GP) models (as in Stojic et al. (2020)). GPs are an ideal choice for the goals of their study but do not break down mappings compositionally as in the current paper (such that mappings can share a subspace but have different subspace to reward functions).

Wang et al. (2018) present simulations of a meta-reinforcement learning agent in which a recurrent neural network, posited to reside in prefrontal cortex, has adjustable parameters that are trained using RL, not on a single task, but instead in a dynamic environment comprising a series of related tasks. The activation dynamics of this network then manifest a second within-task RL algorithm that is automatically tuned to the task at hand. The model explains a broad variety of well-established phenomena including an updated version Harlow's original learning to learn (multitask learning) paradigm in which, after a series of learning episodes, monkeys (and the Meta-RL agent) exhibit single-shot learning.

Yang et al. (2019) also present simulations of a recurrent neural net model of frontal cortex showing how it can learn twenty different cognitive tasks. Interestingly their model employs a regularisation approach, similar to the EWC method used in this paper, to prevent parameters of 'older' tasks being overwritten during learning. They also analyse the representations formed noting that transfer can be mediated either by clustering of parameters over tasks or by the development of compositional representations (of the sort investigated in the current paper).

5.4.10 Machine Learning

The role of task units examined in Musslick and Cohen (2019), in which task units can inhibit hidden units, has been examined as a potential mechanism for aiding multitask learning by Masse et al. (2018). Their studies, using high-dimensional pattern recognition problems, also examined an alternative "gating" strategy in which task units could directly "gate" hidden unit activations (thus mimicking neuromodulation in the brain), such that a proportion of hidden units are gated (set to zero) for any given task. Both of these proposals were examined in combination with EWC (Kirkpatrick et al., 2017) with the findings being that the gating strategy produced better empirical results.

The starting point of our paper was to leverage recent conceptual and algorithmic progress in machine learning to define new experimental psychology tasks and computational models, with the longer term goal of better understanding human multitask and transfer learning. To do this we made use of a sequential Bayesian regularization approach to prevent catastrophic forgetting. This literature, however, is rich with other quantitative ideas about how to define relationships among tasks which could inform the design of future experiments. These include, for example, "sluice" and "cross-stitch" networks (Ruder, 2017) which automatically infer how to share subspaces at multiple hierarchical levels and across multiple tasks. Sequential Bayesian learning for neural networks is also being applied to the more challenging problem of continual learning and is producing state-of-the-art performance on benchmark problems (Nguyen et al., 2018).

Chapter 6

Discussion

In this thesis, we used a novel set of nonlinear decision-making tasks to investigate representation learning. We started with a behavioural experiment outlining in which context new representations are formed. We then divided representation learning into multiple stages, pre and post-decision and examined the underlying electrophysiology using EEG. We presented the advantages that a brief period of quiescence has on representation learning. Finally, we showed how acquired representations can affect the learning process for a new task. We review these findings in more detail below, we discuss them in relation to the broader literature and consider future directions.

6.1 Summary of Findings

Chapter 2 starts by investigating the different contexts in which complex representations are built. We manipulated task and temporal structures and examined their effect on rule declaration and classification performance in a novel decision-making task. We used an online Latent Cause Model (LCM) to gather further insights into the representations that participants built. This model builds representations out of "clusters" which are defined as groups of trials with similar stimulus features and a similar stimulus-outcome mapping. We found that both task and temporal structure influenced participants classification performance and representation learning. Task structure was the only variable that affected the number of participants who correctly declared the rule. The latent cause model gave us further insights into the complexity of the representation created and how, based on this complexity, participants encoded and processed stimuli in homogeneous clusters.

In Chapter 3, we unfolded the successive stages of representation learning, from stimulus to feedback processing. We ran an EEG study where participants learnt the same stimulus-reward mappings as in Chapter 2, but with a minor difference in how the stimuli were presented (pies with variable numbers of slices instead of shapes with variable numbers of sides). We used measures derived from the latent cause model described in chapter 2 as regressors and looked for the electrophysiological indices of representation learning. We found evidence of feature representations in sensory cortices and found the more complex representations of feature subspaces (combinations of features) in frontal regions. Furthermore, we found that parietal electrodes reflected the uncertainty about which cluster a new stimulus belongs to. Finally, we showed central areas to be involved in reward processing and encoding after feedback.

In Chapter 4, we operationalized offline learning as a period of quiet wakefulness and examined its effect on the performance of a probabilistic decision making task. An initial feedback-based learning period using a subset of ("old") stimuli, was followed by (i) a "pre-test" period using a full set of ("old and new") stimuli but without feedback, (ii) a period of quiet wakefulness, and finally (iii) a "post-test" period, again without feedback. By computing correct classification rates separately for old and new stimuli, and assessing differences in these from pre-test to post-test, we were able to separately identify the effect of quiet wakefulness on memory and generalization. Correct classification rate was compared between an "offline" and an "active wake" condition. We found that quiet wakefulness significantly improved generalization but found no evidence for improvements in memory.

In Chapter 5 we investigated how old representations can be useful for learning new tasks (transfer learning). We borrowed concepts from the multitask literature to define pairs of learning tasks that either shared or did not share a common subspace (set of useful features). Human subjects then learnt two consecutive tasks using a feedback-based approach. We hypothesised that learning would be boosted for tasks that shared subspaces. Our findings broadly supported this hypothesis with either better performance on the second task if it shared the same subspace as the first or positive correlations over task performance for shared subspaces. These empirical findings were compared to the behaviour of a Neural Network (NN) model trained using sequential Bayesian learning and human performance was found to be consistent with a minimal capacity variant of this model. Networks with an increased representational capacity, and networks without Bayesian learning, did not show these transfer effects. We proposed that the concept of shared subspaces provides a useful framework for the experimental study of human multitask and transfer learning.

6.1.1 Relation to the broader literature

In Chapter 2, we found evidence supporting the hypothesis that representation learning varies as a function of temporal structure (blocked or interleaved cues). This was evident from different analyses: participants performance, rule declaration, and analysis of measures derived from the LCM. This is consistent with previous work showing how blocking stimuli favoured the formation of a one-to-one mapping between stimulus and outcome (Wulf and Shea, 2002; Xue et al., 2010), and repeating stimuli favoured the creation of a complex representation (Carvalho and Goldstone, 2015). We extended the previous literature by showing the effect of the interaction between temporal structure and task mapping on participants performance, declaration and estimated model measures. In Chapter 2 we also proposed a LCM in which new stimuli are viewed as originating from clusters made of input features, decisions and rewards (Gershman et al., 2010). As in the model proposed by Sanborn et al. (2010) the model started with a single cluster and flexibly created new ones based on the similarity with new trial data (Niv, 2018).

In Chapter 3, we investigated the multiple processes underlying representation learning. We compared our results to the various stages of processing underlying elementary arithmetic. We found that occipital electrodes were associated with the processing of stimulus features, and noted similarites with magnitude processing in visual cortex (Fornaciai et al., 2017; Pinheiro-Chagas et al., 2019). Consistent with the task representation literature (Collins et al., 2014), we found that frontal EEG activity reflected operations on these features. Finally, aided by the LCM, we found that parieto-central activity reflected the uncertainty with which cues belonged to clusters. This is consistent with literature suggesting that the parietal cortex carries out the job of integration between prior knowledge and new information (Seghier, 2013; Tomov et al., 2018).

In Chapter 4, we found evidence supporting the idea that a period of quiet wakefulness promotes generalization. This is consistent with previous work showing how representations formed during a task are re-organized during quiet wakefulness (Craig et al., 2018a,b; Quentin et al., 2020). We failed to show evidence of improved memory retention, this is however likely due to the brief length of the offline period (5 minutes), as compared to what has been used in earlier studies (Craig et al., 2018a; Dewar et al., 2012) (10 minutes).

Chapter 5 investigated far versus near and positive versus negative transfer (Schubert et al., 2014; Steiner, 2001). We highlighted the roles played by subspaces in defining how similar different tasks are. In our study participants did two consecutive tasks in which the subspaces were shared or not. Shared subspaces benefited learning and non-shared subspaces were detrimental. The two consecutive tasks used were, in order, a nonlinear and a linear task. It has been a matter of debate in the literature as to whether humans perform beter with linearly or non-linearly separable categories (Griffiths et al., 2009; Levering et al., 2020; Lucas et al., 2015; Medin and Schwanenflugel, 1981). Our results, however, clearly showed that participants performed better for the linear task. But due to our study design we cannot disambiguate this finding from an order effect.

Finally, starting from the multitask learning concepts (Caruana, 1998), we develop a neural network model, similar to the one proposed by Kirkpatrick (Kirkpatrick et al., 2017). The parameters of the model are constrained to be similar across tasks. Information from older tasks is protected by the prior benefiting new tasks similar to old ones; but, when combined with a limited capacity network, leading to negative transfer.

6.1.2 Neurobiology of representation learning

A prominent view on how the brain learns representations is the Complementary Learning System (CLS) theory (Kumaran et al., 2016; McClelland et al., 1995). Learning is instantiated by two interacting learning systems: cortico-striatal (Poldrack et al., 2001) and hippocampal/MTL ones (Kumaran et al., 2016). The first that slowly and incrementally learns task-specific structured knowledge, supported by the second, which rapidly encodes conjunctions of events. The latent cause model presented in Chapter 2 might embody the model-based behaviour of the hippocampus/MTL system (Davis et al., 2012a). The hippocampus stores specific configurations of events, clusters. These configurations share elements that could inform a way of representing them together in single representations (Zeithamova and Bowman, 2020), or bigger clusters. The process of identifying the common elements and creating a generalized representation could be carried out by the interaction between the hippocampus and the neo-cortex (Bowman and Zeithamova, 2018; Preston et al., 2004; Zeithamova et al., 2012). This process of generalization might happen during a period quiescence (see Chapter 4), where overlapping, or nearby, clusters might be re-organized into a single representation.

The neural network presented in Chapter 5, on the other hand, might embody neocortical and striatal activity. Block by block our model incrementally learns a subspace, a way of integrating inputs (McClelland et al., 1995).

6.1.3 Neurobiology of declarative learning

In psychology, learning has been traditionally divided into declarative and non-declarative forms. The declarative form is commonly associated with the hippocampus, whereas the non-declarative is associated with the basal ganglia (Moscovitch, 1995). Similar to this distinction, Ashby et al. (2011) developed a model of category learning (COVIS) that assumes two separate and competing systems, a rule-based and a procedural-learning one. The first one is hypothesised to be mediated by hippocampus (Cohen and Lefebvre, 2016), which learns declarative associations, the second, based in the basal ganglia (Cohen and Lefebvre, 2016) that gradually learns associations. The model assumes that people begin learning categories using the rule-based system. It selects stimulus features and uses them to associate the stimulus to a category, a rule that can be described verbally. If the performance of the model with the rule is not accurate, the model might switch to a different rule or the procedural system.

However, recent studies showed how the hippocampus is involved in encoding events independently of their declarable nature (Carpenter et al., 2016; Ortu and Vaidya, 2013). Declarative knowledge can not take place without an event being encoded first meaning that the hippocampus is necessary but not sufficient for declarative learning to happen.

The model proposed by Ashby et al. (2011) is somewhat similar to the model described in Chapter 2 where stimuli are clustered together based on similar stimulus features and stimulus-outcome mapping. Differently from the COVIS model, our model begins by encoding a single conjunctive event and trial by trial learns how to cluster similar stimuli in representations.

6.2 Limitations and Future Directions

Several limitations arose during these studies that could be explored in future studies. For example, in Chapter 2, we limited our investigation to feedback-based online learning. Future studies could investigate how the representations created with different temporal structures can influence generalization and memorization. Blocking stimuli together favours the creation of one cluster per cue-outcome combination (thus a larger number of clusters if there are a large number of combinations - 25 in our tasks), which should resulting in poor generalization. On the other hand, interleaving stimuli together may favour the creation of bigger clusters thereby promoting generalization to new unseen stimuli.

In Chapter 2 we also proposed a model, the LCM, that identified clusters of exemplars giving us insight into the complexity of the structure of the representation created. This LCM had three parameters: one that we hand-tuned (or "supplied"), the encoding precision $\lambda 0$, and two that were fitted to participant's behavioural data - the threshold for cluster creation ϕ, and the threshold for pruning ψ. These parameters were fitted to each participant's behavioural data and kept fixed for the whole task. However, we expect these parameters may vary during the task, therefore future studies might explore this. Furthermore, the LCM created partially overlapping clusters, so that the same stimulus belonged to multiple clusters. It might be beneficial to implement an algorithm that merges overlapping clusters into a single cluster.

In chapter 3, we followed up the experiment presented in the second chapter with an EEG study. The experimental design included both task structures used in chapter 2, but only a single temporal structure. Future investigations might compare activity related to different temporal structures. Further attention might be also directed to the minute break between one block and the next one. Activity during this offline period might show patterns of replay and be driven by stimuli presented in the previous block, or might correlate with performance in the block that follows.

In chapter 4, we found that a period of quiet wakefulness can benefit the generalization process. The study was a behavioural experiment in which data was acquired on the internet and therefore had several drawbacks, such as low-performance scores and an uncontrolled environment. It would be important therefore to replicate this experiment in a labortatory setting. That would allow also to increase the offline period from five to ten minutes, as in other studies, without risking participants doing other tasks unchecked. Furthermore, as discussed, the nature of the active wake condition itself should be the subject of future studies. Is the sensory domain used to encode the information important? If so, an acoustic task might create no interference with a visual one, and vice versa. Are different tasks creating interferences for different processes? As generalization and memorization rely on different neural substrates (Schacter et al., 2012; Tambini and Davachi, 2019; Varma et al., 2017), different tasks might create interference for one but not the other.

In chapter 5, we proposed the concept of subspace from the multitask learning literature as a useful framework for the study of transfer learning. Participants did two tasks consecutively. And as we discussed, they were not instructed that they would be performing the same task twice. This might have led participants to overwrite the knowledge gained during the first task. Future studies might therefore test if that is the case by creating a design in which participants are tested on two different structures and then re-tested on one of those. This would allow testing if what was learnt in the first task was overwritten during the second. Furthermore, future studies could move from a design with two consecutive tasks towards a design in which transfer learning is explored while participants do two tasks simultaneously, like driving a car and paying attention simultaneously to other cars on the road, pedestrians and road signs (Ng, 2017).

Recently, new paradigms using up-to-date analysis started investigating the CLS using fMRI, bringing new possibilities in the field of cognitive neuroscience (Schuck and Niv, 2019; Wittkuhn and Schuck, 2020). fMRI experiments using a paradigm similar to the one presented in Chapter 4 and model-based analysis, as in Chapters 2 and 5, might provide insights on which computations, what part of the brain, is processing and at which stage of learning. The latent cause model presented in chapter 2 could be instantiated during learning, representing so the configurations presented. Its computations might be associated with the activity of the hippocampus storing conjunctive information. During the quiescence period (Chapter 4) the hippocampus and the neocortex interacting might be re-organizing the encoded events in a common representation. This would result in the creation of subspaces, as presented in the neural network model in Chapter 5, that would be then used during testing. The same paradigm could answer other timely questions such as: do representations of generalized information and specific events coexist (Banino et al., 2016)? is the hippocampus representing both (Poppenk et al., 2013)? if so where?

6.3 Conclusion

This thesis offered an overview of the representation learning topic. We presented four experimental chapters where we investigated representation learning through behavioural and EEG experiments and computational models. We discussed it through different perspectives: online, offline and transfer learning. We proposed two different computational models, LCMs and NNs, each of which provides insight into how representations can be created and transferred. Also, we examined the electrical brain activity related to one of these models. We showed how over time, features are perceived, integrated, compared with existing schema, and incorporated into them or encoded anew.

References

Aberg, K., C., Kramer, E., E., and Schwartz, S. (2020). Interplay between midbrain and dorsal anterior cingulate regions arbitrates lingering reward effects on memory encoding. *Nature Communications*, 11:1–14.

Aitchison, L., Jegminat, J., Menendez, J., Pfister, J., Pouget, A., and Latham, P. (2021). Synaptic plasticity as Bayesian inference. *Nature Neuroscience*.

Akrami, A., Kopec, C., D., Diamond, M., E., and Brody, C., D. (2018). Posterior parietal cortex represents sensory history and mediates its effects on behaviour. *Nature Letter*, 554:368–372.

Altman, N., S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46:175–185.

Ashby, F. G., Paul, E. J., and Maddox, W. T. (2011). COVIS. *Formal approaches in categorization*, pages 65–88.

Ashby, F., G. and Maddox, W., T. (2005). Human category learning. *Annual Review Psychology*, 56:149–178.

Avancini, C., Soltesz, F., and Szucs, D. (2015). Separating stages of arithmetic verification: An ERP study with a novel paradigm. *Neuropsycholgia*, 75:322–329.

Axmacher, N., Elger, C., E., and Fell, J. (2008). Ripples in the medial temporal lobe are relevant for human memory consolidation. *Brain*, 131:1806–1817.

Badre, D., Bhandari, A., Keglovits, H., and Kikumoto, A. (2021). The dimensionality of neural representations for control. *Current Opinion in Behavioral Sciences*, 38:20–28.

Badre, D. and D'esposito, M. (2008). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, 19:2082–2099.

Badre, D., Kayser, A., S., and D'esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, 66:315–326.

Balestrieri, E. and Busch, N., A. (2021). Spontaneous alpha-band oscillations bias subjective contrast perception. *Biorxiv*.

Ballard, I., Miller, E., M., Piantadosi, S., T., and Goodman, N., D. (2018). Beyond reward prediction errors: Human striatum updates rule values during learning. *Cerebral Cortex*, 28:3965–3975.

Banino, A., Koster, R., Hassabis, D., and Kumaran, D. (2016). Retrieval-based model accounts for striking profile of episodic memory and generalization. *Scientific Reports*, 6(1):1–15.

Bishop, C. (2006). *Pattern Recognition and Machine Learning.* Springer, New York.

Bowman, C. R. and Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *Journal of Neuroscience*, 38(10):2605–2614.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, 10:433–436.

Braun, D., Mehring, C., and Wolpert, D. (2010). Structure learning in action. *Behavioural Brain Research*, 206:157–165.

Braun, E., K., Wimmer, G., E., and Shohamy, D. (2018). Retroactive and graded prioritization of memory by reward. *Nature Communications*, 9:1–12.

Carpenter, K. L., Wills, A. J., Benattayallah, A., and Milton, F. (2016). A comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human Brain Mapping*, 37(10):3557–3574.

Caruana, R. (1998). Multitask learning. *Autonomous agents and multi-agent systems*, 27:95–133.

Carvalho, P., F. and Goldstone, R., L. (2015). What you learn is more than what you see: what can sequencing effects tell us about inductive category learning? *Frontiers in Psychology*, 6:1–12.

Cohen, H. and Lefebvre, C. (2016). *Handbook of categorization in cognitive science.* Elsevier, New York.

Cohen, M., X. (2014). *Analyzing Neural Time Series Data: Theory and Practice.* The MIT Press, Massachussetts.

Collins, A. and Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS Biology*, 10:e1001293.

Collins, A., G., Cavanagh, J., F., and Frank, M., J. (2014). Human EEG uncovers latent generalizable rule structure during learning. *The Journal of Neuroscience*, 26:4677–4685.

Collins, A., G. and Frank, M., J. (2018). Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. *Proceedings of the National Academy of Sciences*, 115:2502–2507.

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, 36:287–314.

Costa, V., Tran, V., Turchi, J., and Averbeck, B. (2015). Reversal learning and dopamine: A Bayesian perspective. *The Journal of Neuroscience*, 35:2407–2416.

Craig, M., Della Sala, S., and Dewar, M. (2014). Autobiographical thinking interferes with episodic memory consolidation. *PLos ONE*, 9:1–9.

Craig, M. and Dewar, M. (2018). Rest-related consolidation protects the fine detail of new memories. *Scientific Reports*, 8:6857–6866.

Craig, M., Dewar, M., Della Sala, S., and Wolbers, T. (2015). Rest boosts the long-term retention of spatial associative and temporal order information. *Hippocampus*, 25:1017–1027.

Craig, M., Ottaway, G., and Dewar, M. (2018a). Rest on it: Awake quiescence facilitates insight. *Hippocampus*, 29:862–868.

Craig, M., Wolbers, T., Strickland, S., Achtzehn, J., and Dewar, M. (2018b). Rapid improvement of cognitive maps in the awake state. *Cortex*, 109:205–214.

Dasgupta, I. and Gershman, S., J. (2021). Memory as a computational resource. *Trends in Cognitive Sciences*, 25:240–251.

Davis, T., Love, B., C., and Preston, A. (2012a). Striatal and hippocampal entropy and recognition signals in category learning: Simultaneous processes revealed by model-based fmri. *Journal of Experimental Psychology*, 38:821–839.

Davis, T., Love, B., C., and Preston, A., R. (2012b). Learning the exception to the rule: Model-based fmri reveals specialized representations for surprising category members. *Cerebral Cortex*, 22:260–273.

Daw, N., D. and Shohamy, D. (2008). The cognitive neuroscience of morivation and learning. *Social Cognition*, 5:593–620.

Daw, N. (2011). *Decision Making, Affect and Learning: Attention and Performance XXIII*, chapter Trial by trial data analysis using computational models. Oxford University Press.

Dellarosa, D. and Bourne, L., E. (1985). Surface form and the spacing effect. *Memory and Cognition*, 13:529–537.

Dewar, M., Alber, J., Butler, C., Cowan, N., and Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological Science*, 23:955—-960.

Dobson, A., J. and Barnett, A., G. (2018). *An Introduction to Generalized Linear Models*. Chapman and Hall.

Duncan, K., Doll, B., B., Daw, N., D., and Shohamy, D. (2018). More than the sum of its parts: A role for the hippocampus in configural reinforcement learning. *Neuron*, 98:645–657.

Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman and Hall.

Farashahi, S., Rowe, K., Aslami, A., Lee, D., and Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. *Nature Communications*, 8:1768.

Farashahi, S., Xu, J., Wu, S., and Soltani, A. (2020). Learning arbitrary stimulus-reward associations for naturalistic stimuli involves transition from learning about features to learning about objects. *Cognition*, 205:104425.

Fitzgerald, T., Bonnici, H., Penny, W., and Adams, R. (2020). Retrospective inference as a form of bounded rationality, and its beneficial influence on learning. *Frontiers in Artificial Intelligence*, 3.

Flesh, T., Balageur, J., Dekker, R., Nili, H., and Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proc. Nat. Acad. Sci.*, 115:E10313–E10322.

Fornaciai, M., Brannon, E., M., Woldorff, M., G., and Park, J. (2017). Numerosity processing in early visual cortex. *NeuroImage*, 157:429–438.

Fornaciai, M. and Park, J. (2018). Early numerosity encoding in visual cortex is not sufficient for the representation of numerical magnitude. *Journal of Cognitive Neuroscience*, 30:1788–1802.

Foster, D. (2017). Replay comes of age. *Annual Reviews of Neuroscience*, 40:581–602.

Fowlkes, E., B. and Mallows, C., L. (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78:553–569.

Franklin, N. and Frank, M. (2018). Compositional clustering in task structure learning. *PLoS Computational Biology*, 14(4):e1006116.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W., editors (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images.* Academic Press.

Gaesser, B., Spreng, R., N., McLelland, V., C., Addis, D., R., and Schacter, D., L. (2013). Imagining the future: Evidence for a hippocampal contribution to constructive processing. *Hippocampus*, 23:1150–1161.

Gavert, M., M., Dolan, R., J., and Behrens, T., E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, 6:e17086.

Gershman, S., J., Blei, D., M., and Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, 117:197–209.

Gerstner, W. and Kistler, W. M. (2002). *Spiking neuron models: Single neurons, populations, plasticity.* Cambridge university press.

Gluck, M., A., Shohamy, D., and Myers, C. (2002). How do people solve the “weather prediction” task?: Individual variability in strategies for probabilistic category learning. *Learning and Memory*, 9:408–418.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning.* MIT Press.

Goodman, N., Tenenbaum, J., Feldman, J., and Griffiths, T. (2018). A rational analysis of rule-based concept learning. *Cognitive Science*, 32:108–154.

Graveline, Y., M. and Wamsley, E., J. (2017). The impact of sleep on novel concept learning. *Neurobiology of Learning and Memory*, 141:19–26.

Griffiths, T., Lucas, C., Williamns, J., and Kalish, M. (2009). Modeling human function learning with gaussian processes. *NIPS 21*.

Hadsell, R., Rao, D., Rusu, A., and Pascanu, R. (2020). Embracing change: Continual learning in deep neural networks. *Trends in Cognitive Sciences*, 24:1028–1040.

Harvey, B. and Dumoulin, S. (2017). A network of topographic numerosity maps in human association cortex. *Nature Human Behaviour*, 1:0036.

Hickernell, F. and Yuan, Y. (1997). A simple multistart algorithm for global optimization. *OR Transactions*, 1.

Hubbard, J., Kikumoto, A., and Mayr, U. (2019). EEG decoding reveals the strength and temporal dynamics of goal-relevant representations. *Scientific Reports*, 9:1–11.

Huys, Q. J., Maia, T. V., and Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature neuroscience*, 19(3):404–413.

Kikumoto, A. and Mayr, U. (2020). Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. *Proceedings of the National Academy of Sciences*, 117:10603–10608.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Jumaran, D., and Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *PNAS*, 114:3521–3526.

Knowlton, B., J., Squire, L., R., and Gluck, M., A. (1994). Probabilistic classification learning in amnesia. *Learning and Memory*, 1:106–120.

Kornell, N. and Bjork, R. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, 19:585–592.

Kornell, N., Castel, A., D., Eich, T., S., and Bjork, R. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging*, 25:498–503.

Kumaran, D., Hassabis, D., and McClelland, J., L. (2016). What learning systems do intelligent agents need? complementary learning systems theory updated. *Trends in Cognitive Sciences*, 20:512–534.

Kumaran, D., Summerfield, J., J., Hassabis, D., and Maguire, E., A. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron*, 63:889–901.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature*, 521:436–444.

Lee, M., D. and Wagenmakers, E. (2014). *Bayesian Cognitive Modeling*. Cambirdge University Press.

Levering, K., COnway, N., and Kurtz, K. (2020). Revisiting the linear separability constraint: new implications for theories of human category learning. *Memory and Cognition*, 48:335–347.

Levine, S., Kumar, A., Tucker, G., and Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *ArXiv:2005.01643v1*.

Lewis, P., A. and Durrant, S., J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, 15:343–350.

Li, K., Fu, Q., Sun, X., Zhou, X., and Fu, X. (2016). Paired-associate and feedback-based weather prediction tasks support multiple category learning systems. *Frontiers in Psychology*, 7:1–10.

Love, B., C. (2020). Model-based fmri analysis of memory. *Current Opinion in Behavioral Sciences*, 32:88–93.

Lucas, C., Griffiths, T., Williams, J., and Kalish, M. (2015). A rational model of function learning. *Psychon Bull Rev.*

Mackay, D. (1992a). A practical Bayesian framework for backprop networks. *Neural Computation*, 4(3):448–472.

Mackay, D. (1992b). The evidence framework applied to classification networks. *Neural Computation*, 4(5):720–736.

MacKay, D., J. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Maris, E. and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164:177–190.

Martini, M., Martini, C., Maran, T., and Sachse, P. (2018). Effects of post-encoding wakeful rest and study time on long-term memoryperformance. *Journal of Cognitive Psychology*, 30:558–569.

Masse, N., Grant, G., and Freedman, D. (2018). Alleviating catastrophic forgetting using context dependent gating and synaptic stabilization. *PNAS*, 115:10467–10475.

MATLAB (2018). *version 9.7.0.1190202 (R2019b)*. The MathWorks Inc., Natick, Massachusetts.

Mazza, V. and Caramazza, A. (2010). Multiple object individuation and subitizing in enumeration: a view from electrophysiology. *Trends in Cognitive Sciences*, 14:542–551.

McClelland, J., L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, 142:1190—-1210.

McClelland, J., L., McNaughton, B., L., and O'Reilly, R., C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102:419–457.

Medin, D. and Schwanenflugel, P. (1981). Linear separability in classificiation learning. *Journal of Experimental Psychology: Human Learning and Memory*, 7:355–368.

Meeter, M., Myers, C., E., Shohamy, D., Hopkins, R., O., and Gluck, M., A. (2006). Strategies in probabilistic categorization: Results from a new way of analyzing performance. *Learning and Memory*, 13:230–239.

Menghi, N., Kacar, K., and Penny, W. (2021). Multitask learning over shared subspaces. *PLoS Computational Biology*, 17.

Menghi, N. and Penny, W. (2019). Configural learning depends on task complexity and temporal structure. Conference Paper at Cognitive Computation Neuroscience (CCN) Conference.

Menghi, N. and Penny, W. (2021). Outlining contextual settings for rule learning through a probabilistic category learning task. Poster at Organization for Computational Neurosciences (OCNS) Conference.

Moscovitch, M. (1995). Recovered consciousness: a hypothesis concerning modularity and episodic memory. *Journal of clinical and experimental neuropsychology*, 17(2):276–290.

Moser, E., I., Moser, M., McNaughton, B., and L. (2017). Spatial representation in the hippocampal formation: a history. *Nature Neuroscience*, 20:1448–1464.

Musslick, S. and Cohen, J. (2019). A mechanistic account of constraints on control-dependent processing: Shared representation, conflict and persistence. *Annual Meeting of the Cognitive Science Society (COGSCI'19).*

Nabney, I. (2003). *NETLAB: Algorithms for Pattern Recognition.* Springer.

Ng, A. (2017). Deep learning specialization. Technical report, https://www.coursera.org/specializations.

Nguyen, C., Li, Y., Bui, T., and Turner, R. (2018). Variational continual learning. *International Conference on Learning Representations (ICLR).*

Niv, Y. (2018). Learning task-state representations. *Nature Neuroscience*, 22:1544–1553.

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, 22:1544–1553.

Noack, H., Lovden, M., and Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. *Psychological Research*, 78:773–789.

Noh, S., M., Yan, V., X., Bjork, R., A., and Maddox, W., T. (2016). Optimal sequencing during category learning: Testing a dual-learning systems perspective. *Cognition*, 155:23–29.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011.

Ortu, D. and Vaidya, M. (2013). A neurobiology of learning beyond the declarative non-declarative distinction. *Frontiers in Behavioral Neuroscience*, 7:161.

Parisi, G. and Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71.

Peigneux, P., Orban, P., Balteau, E., Degueldre, C., Luxen, A., Laureys, S., and Maquet, P. (2006). Offline persistence of memory-related cerebral activity during active wakefulness. *Plos Biology*, 4:647–658.

Peirce, J., W. (2007). Psychopy—psychophysics software in python. *Journal of Neuroscience Methods*, 162:8–13.

Perkins, D. and Salomon, G. (1992). Transfer of learning. In *International Encyclopedia of Education*. Pergamon Press.

Piazza, M. (2015). Neurocognitive start-up tools for symbolic number representations. *Frontiers in Human Neuroscience*, 162:1–7.

Pinheiro-Chagas, P., Piazza, M., and Dehaene, S. (2019). Decoding the processing stages of mental arithmetic with magnetoencephalography. *Cortex*, 114:124–139.

Pinto, R., C. and Engel, P., M. (2015). A fast incremental gaussian mixture model. *Plos One*, pages 1–12.

Poldrack, R., A., Paré-Blagoev, R., A., Shohamy, D., Moyano, J., C., Myers, C., and Gluck, M., A. (2001). Interactive memory system in the human brain. *Nature*, 414:546–550.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., and Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in cognitive sciences*, 17(5):230–240.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C (Second Edition)*. Cambridge, Cambridge.

Preston, A. R., Shrager, Y., Dudukovic, N. M., and Gabrieli, J. D. (2004). Hippocampal contribution to the novel use of relational information in declarative memory. *Hippocampus*, 14(2):148–152.

Quentin, R., Fanuel, L., Kiss, M., Vernet, M., Vékony, T., Janacsek, K., Cohen, L., and Nemeth, D. (2020). Statistical learning occurs during practice while high-order rule learning during rest period. *Biorxiv*, 1:1–15.

Radford, A., Narasimham, K., Salimans, T., and Sutskever, I. (2018). Improving language understanding by generative pre-training. Technical report, Open AI.

Radulescu, A., Niv, Y., and Ballard, I. (2019). Holistic reinforcement learning: The role of structure and attention. *Trends in Cognitive Sciences*, 23(4):278–292.

Radulescu, A., Shin, Y., S., and Niv, Y. (2021). Human representation learning. *Annual Review of Neuroscience*, 13:253–273.

Raichle, M., E. and Gusnard, D., A. (2002). Appraising the brain's energy budget. *PNAS*, 99:10237–10239.

Rangel, A., Camerer, C., and Montague, P., R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9:545–556.

Renoult, L., Irish, M., Moscovitch, M., and Rugg, M., D. (2019). From knowing to remembering: The semantic–episodic distinction. *Trends in Cognitive Sciences*, 23:1041–1057.

Richards, B., A., Xia, F., Santoro, A., Husse, J., Woodin, M., A., Josselyn, S., A., and Frankland, P., W. (2014). Patterns across multiple memories are identified over time. *Nature Neuroscience*, 17:981—-990.

Rose, M., Haider, H., and Buchel, C. (2010). The emergence of explicit memory during learning. *Cerebral Cortex*, 20:2787–2797.

Ruder, S. (2017). An overview of multitask learning in deep neural networks. Technical report, BioArchive.

Rugg, M. and Coles, M., G. (1995). *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition.* Oxford Psychology Series, Oxford.

Rustemeier, M., Schwabe, L., and Bellebaum, C. (2013). On the relationship between learning strategy and feedback processing in the weather prediction task—evidence from event-related potentials. *Neuropsychologia*, 51:695–703.

Samaha, J., Iemi, L., and Postle, B. (2017). Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. *Consciousness and Cognition*, 54:47–55.

Sanborn, A., N., Griffiths, T., L., and Navarro, D., J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117:1144–1167.

Saxe, A., Nelli, S., and Summerfield, C. (2020). If deep learning is the answer, what is the question? *Nature Reviews Neuroscience*, 22:55–67.

Schacter, D., L., Addis, D., R., Hassabis, D., Martin, V., C., Spreng, R., N., and Szpunar, K., K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, 76:677–694.

Schapiro, A., C., McDevitt, E., A., Rogers, T., T., Mednick, S., C., and Norman, K., A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications*, 9:1–11.

Schlichting, L., M. and Preston, A., R. (2014). Memory reactivation during rest supports upcominglearning of related content. *PNAS*, 111:15845–15850.

Schubert, T., Strobach, T., and Karbach, J. (2014). New directions in cognitive training: on methods, transfer, and application. *Psychological Research*, 78:749–755.

Schuck, N., G., R., Wenke, D., Heinzle, J., Frensch, P., Haynes, J., and Reverberi, C. (2015). Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron*, 86:331–340.

Schuck, N., W. and Niv, Y. (2019). Sequential replay of non-spatial task states in the human hippocampus. *Science*, 364:1–24.

Seghier, M., L. (2013). The angular gyrus. *Neuroscientist*, 19:43–61.

Sekeres, M., J., Winocur, G., and Moscovitch, M. (2018). The hippocampus and related neocortical structures in memory transformation. *Neuroscience Letters*, 680:39–53.

Sharon, t., Moscovitch, M., and Gilboa, A. (2011). Rapid neocortical acquisition of long-term arbitraryassociations independent of the hippocampus. *PNAS*, 108:1–6.

Shepard, R., Hovlan, C., and Jenkins, H. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, 75(13).

Silver, D. and Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529:484–489.

Squire, L., R., Genzel, L., Wixted, J., T., and Morris, R., G. (2021). Memory consolidation. *Cold Spring Harbor Laboratory Press.*

Steiner, G. (2001). Transfer of learning, cognitive psychology of. In *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier Science.

Stephan, K. E., Schlagenhauf, F., Huys, Q. J., Raman, S., Aponte, E. A., Brodersen, K. H., Rigoux, L., Moran, R. J., Daunizeau, J., Dolan, R. J., et al. (2017). Computational neuroimaging strategies for single patient predictions. *Neuroimage*, 145:180–199.

Sterpenich, V., van Schie, M., K.-M., Catsiyannis, M., Ramyead, A., Perrig, S., Yang, H., Van De Ville, D., and Schwartz, S. (2011). Reward biases spontaneous neural reactivation during sleep. *Nature Communication*, 12:1–11.

Stojic, H., Schulz, E., Analytis, P., and Speekenbrink, M. (2020). It's new, but is it good? how generalization and uncertainty guide the exploration of novel options. *Journal of Experimental Psychology*, In Press.

Sutton, R. and Barto, A. (1998). *Reinforcement Learning: An Introduction.* MIT Press, Cambridge.

Tambini, A. and Davachi, L. (2019). Awake reactivation of prior experiences consolidates memories and biases cognition. *Trends in cognitive sciences*, 23:876–890.

Tomov, M. S., Schulz, E., and Gershman, S. J. (2019). Multi-task reinforcement learning in humans. *bioRxiv.*

Tomov, M., S., Dorfman, H., M., and Gershman, S., J. (2018). Neural computations underlying causal structure learning. *The Journal of Neuroscience*, 38:7143–7157.

Tompary, A. and Davachi, L. (2017). Consolidation promotes the emergence of representational overlap in the hippocampus and medial prefrontal cortex. *Neuron*, 96:228–241.

Torrey, L. and Shavlik, J. (2010). Handbook of research on machine learning applications and trends: Algorithms, methods, and techniques. chapter Transfer Learning. IGI Global.

Varma, S., Daselaar, S., M., Kessels, R., P., and Takashima, A. (2018). Promotion and suppression of autobiographical thinking differentially affect episodic memory consolidation. *PLoS ONE*, 1:1–15.

Varma, S., Takashima, A., Krewinkel, S., van Kooten, M., Fu, L., Medendorp, W., P., Kessels, R., P., and Daselaar, S., M. (2017). Non-interfering effects of active post-encoding tasks on episodic memory consolidation in humans. *Frontiers in behavioral neuroscience*, 11:1–13.

Vlach, H., A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Developmental Perspectives*, 8:163–168.

Wahlheim, C., N., Dunlosky, J., and Jacoby, L., L. (2011). Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging. *Memory and Cognition*, 39:750–763.

Wamsley, J., E. (2019). Memory consolidation during waking rest. *Trends in Cognitive Sciences*, 23:171–173.

Wang, J., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Lelbo, J., Hassabis, D., and Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, 21:860–871.

Wittkuhn, L., Chien, S. ans Hall-McMaster, S., and Schuck, N., W. (2021). Replay in minds and machines. *Neuroscience and Biobehavioral Reviews*, 129:367–388.

Wittkuhn, L. and Schuck, N., W. (2020). Dynamics of fmri patterns reflect subsecond activation sequences and reveal replay in human visual cortex. *Nature Communications*, 12:1–22.

Wu, C., Schulz, E., Garvert, M., Meder, B., and Schuck, N. (2020). Similarities and differences in spatial and non-spatial cognitive maps. *PLoS Comp. Biol.*, 16(9):e1008149.

Wulf, G. and Shea, C., H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin and Review*, 9:185–211.

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J., A., and Poldrack, R., A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, 330:97–101.

Yang, G., Joglekar, M., Song, H., Newsome, W., and Wang, X.-J. (2019). Task representations in neural networks trained to perform many cognitive tasks. *Nature Neuroscience*, 22:297–306.

Yang, G., R. and Wang, X., J. (2020). Artificial neural networks for neuroscientists: A primer. *Neuron*, 107:1048–1070.

Yee, D., M. and Braver, T., S. (2018). Interactions of motivation and cognitive control. *Behavioral Sciences*, 19:83–90.

Zeithamova, D. and Bowman, C. (2020). Generalization and the hippocampus: More than one story? *Neurobiology of Learning and Memory*, 175:1–10.

Zeithamova, D., Dominick, A. L., and Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, 75(1):168–179.

Zulkiply, N. and Burt, J., S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory and Cognition*, 41:16–27.

Appendix A

Chapter 2

The following analyses show the difference per task and temporal structure of the cluster pruning threshold ψ and the prior probability of creating a new cluster ϕ.

Modelling parameters

We performed a two-way mixed design ANOVA with within factor task structure and between factor temporal structure to investigate if the estimated parameters, ϕ and ψ varied with our manipulations. This revealed no significant effects for the ϕ, threshold for cluster creation, parameter (Task structure, $F(1) = 3.489$, $p < 0.066$; Temporal structure, $F(2) = 3.489$, $p = 0.099$; Interaction, $F(2) = 1.557$, $p = 0.219$) and both main effects but no interactions for the ψ, threshold for pruning, parameter (Task structure, $F(1) = 4.611$, $p < 0.036$; Temporal structure, $F(2) = 0.033$, $p = 0.099$; Interaction, $F(2) = 0.322$, $p = 0.725$). The threshold for pruning a cluster was higher for addition compared to subtraction stimulus-outcome mapping and for interleaved compared to the blocked temporal structure. A higher threshold required the clusters to have a higher probability of reward in order to not be pruned.

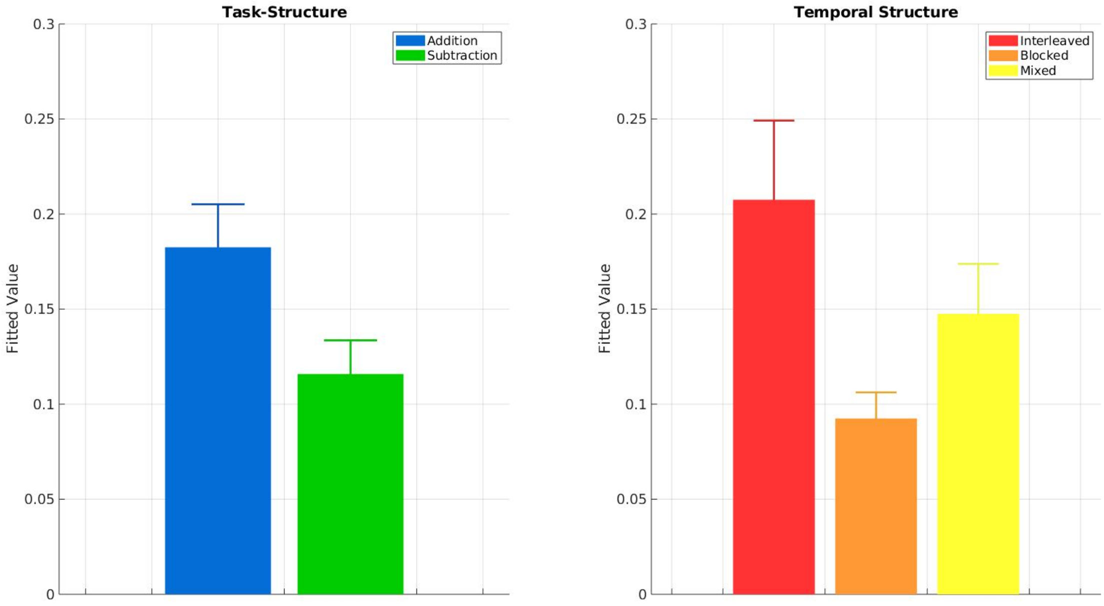


Fig. A.1 **Threshold for Cluster Pruning.** *The plot shows the difference for the fitted threshold for cluster pruning in task complexity and temporal structure.*

Appendix B

Chapter 3

The following figures show the significant effects of the offset term for the feedback GLM and the feature GLM.

Feedback Offset

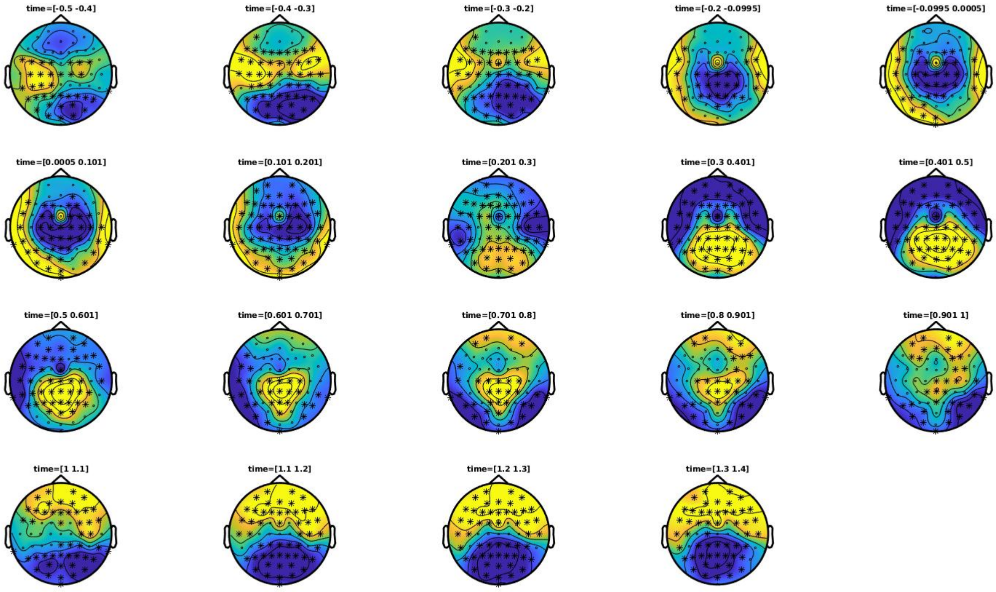


Fig. B.1 **Feedback Offset.** *The plot shows topographies with t-values on a scale from -4 to 4 for different contrasts between -500 and 1500 msec after feedback presentation. Cluster permutation analysis revealed two positive clusters and a negative cluster as a residual activity not captured by the other GLM regressors. These clusters represent activity associated with the button press starting 500 msec before feedback onset in the motor strip, button press confirmation in the visual areas and the visual presentation of the feedback message. Significant results are represented by asterisks.*

Subspace and Feature Offset

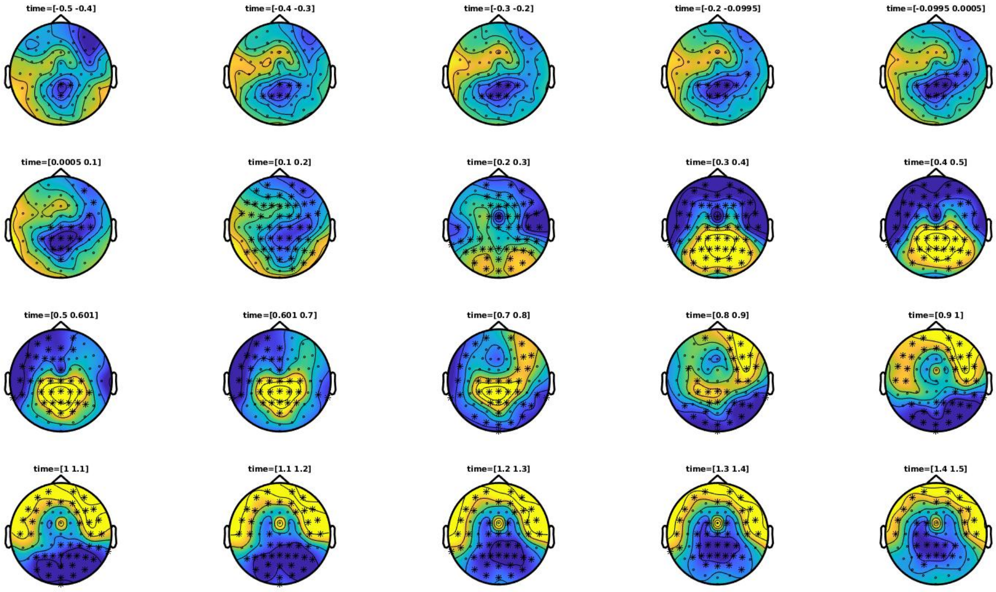


Fig. B.2 **Results after Stimulus onset.** *The plot shows topographies with t-values on a scale from -4 to 4 for different contrasts between -500 and 1500 msec after stimulus onset. Cluster permutation analysis revealed one positive cluster and two negative clusters as a residual activity not captured by the other GLM regressors. These clusters represent activity associated with the presentation of the fixation cross and preparation to stimulus presentation 500 msec before stimulus onset in the central areas, activity associated with the presentation of the stimulus in the visual areas and the motor preparation. Significant results are represented by asterisks.*

Appendix C

Chapter 4

In this appendix there is a link to the Github Archive containing the script of the behavioural analyses, a figure showing the main effect of wake and stimulus novelty conditions, the questionnaires used and further analysis to exclude confounding effects.

Data and Software Archive

Matlab scripts for the nonparametric analysis, and participants behavioural data, are available from a Github Archive https://github.com/Nich0Me/Quiescence.

ANOVA Main Effects

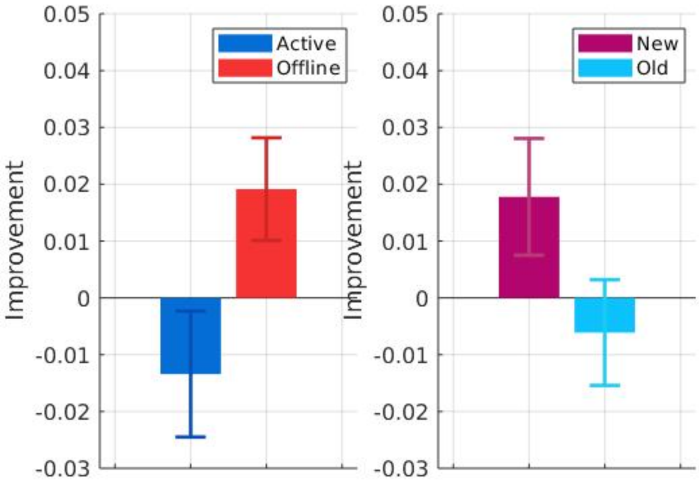


Fig. C.1 **Main Effects of Anova.** The barplots show the accuracies for the main effects of the ANOVA with wake condition (Active/Offline) in the left panel and stimulus novelty (Old/New) on the right panel. The error bars indicate the standard error of the mean. We observe an increase in accuracy for the offline versus active wake condition.

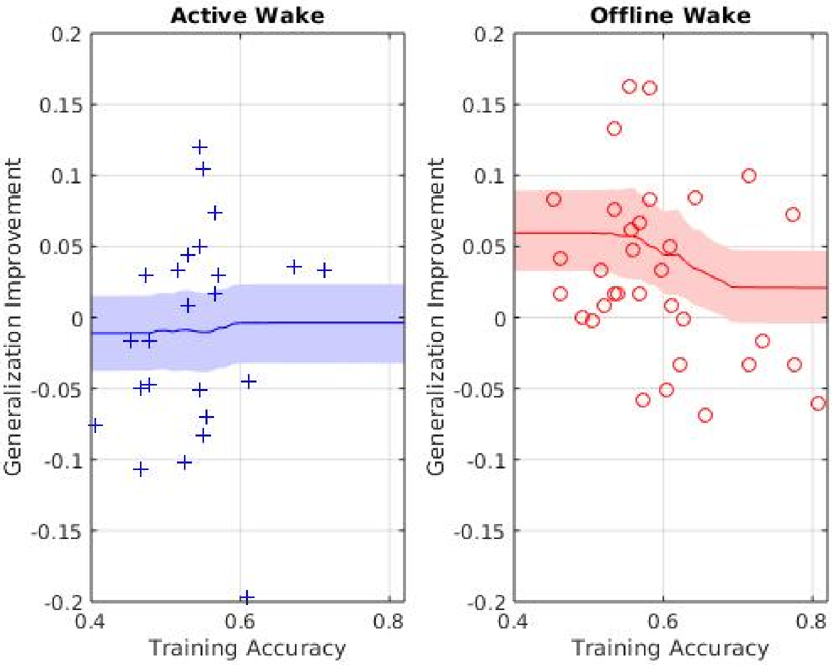


Fig. C.2 **Nonparametric Regression after applying exclusion criteria.** The plots show the relationship between training and improvement in generalisation for the different wake conditions (Active/Offline). The plot shows the 90 percent confidence intervals. We observe a significant relationship between training and generalisation improvement in the offline wake condition.

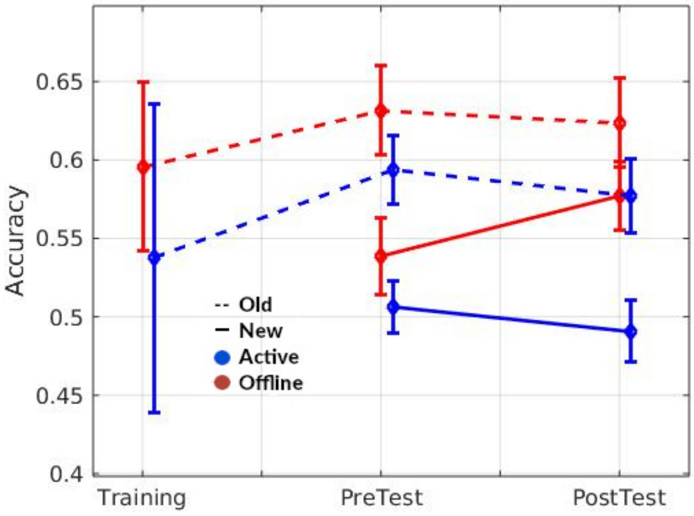


Fig. C.3 **Accuracy per wake and novelty conditions.** The plot shows the accuracies for the wake condition (Active/Offline) and stimulus novelty (Old/New) per time interval (Training/Pre-test/Post-test). The error bars indicate the standard error of the mean. We observe an increase in accuracy for the offline versus active wake condition.

Relaxation and 2-Back task

At the end of the experiment, participants in the quiescence condition rated how much they managed to relax and avoid external stimulation on a scale from 1 to 7. The average relaxation reported is 4.655, with a standard deviation of 1.157. The relaxation reported did not correlate with either the improvement in memorization ($r_s = -0.177, p = 0.316$) or the improvement in generalization ($r_s = -0.278, p = 0.110$). Participants in the active wake condition had an average Fowlkes-Maddox score on the 2-back task of 76% with a standard deviation of 0.148. These scores did not correlate with either with the improvement in memorization ($r_s = -0.005, p = 0.98$) or in generalization ($r_s = 0.365, p = 0.079$).

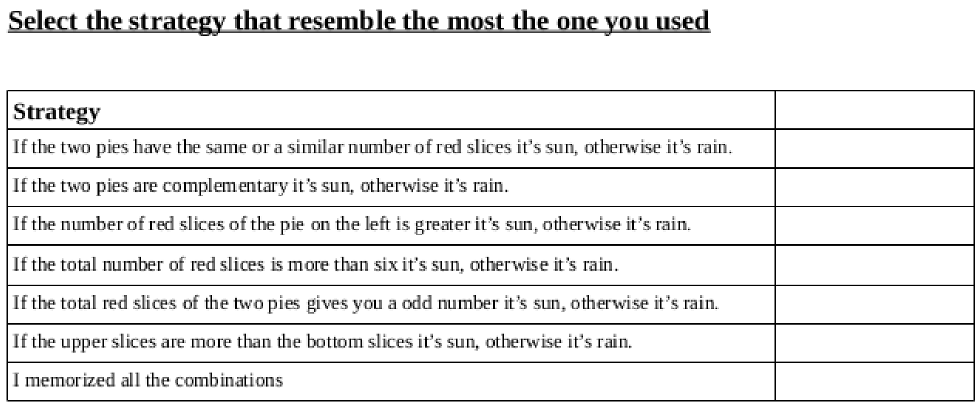
Break Length and Improvement

We also measured the time taken for the break that participants took between the end of the wake condition and the beginning of the post-test session. No difference was found in the break length between the two wake conditions (t(56) = 1.606, p = 0.113).

Putative task-switching cost

We tested if participants in the active wake group incurred a task switching cost by comparing generalization and memorisation scores on the first versus second half of the post-test epoch. No difference was found in memorization (t(23) = 0.278, p = 0.783), nor generalization (t(23) = 0.359, p = 0.722).

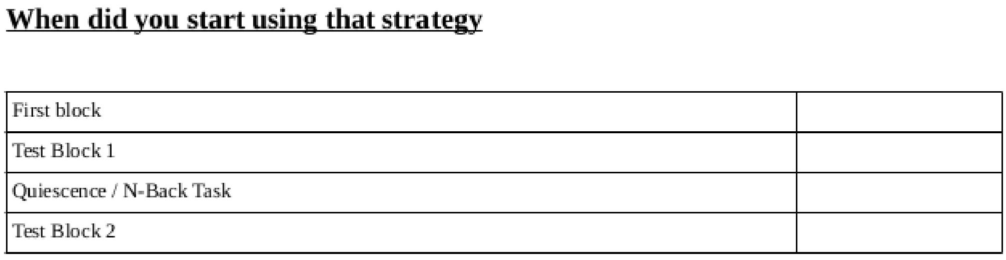
Questionnaires



Select the strategy that resemble the most the one you used

Strategy	
If the two pies have the same or a similar number of red slices it's sun, otherwise it's rain.	
If the two pies are complementary it's sun, otherwise it's rain.	
If the number of red slices of the pie on the left is greater it's sun, otherwise it's rain.	
If the total number of red slices is more than six it's sun, otherwise it's rain.	
If the total red slices of the two pies gives you a odd number it's sun, otherwise it's rain.	
If the upper slices are more than the bottom slices it's sun, otherwise it's rain.	
I memorized all the combinations	

Fig. C.4 **List of strategies.** Participants were asked to select from this list of strategies the one that most resembled their own strategy.



When did you start using that strategy

First block	
Test Block 1	
Quiescence / N-Back Task	
Test Block 2	

Fig. C.5 **When they started using this strategy.** Participants were asked at which point in time they started using this strategy with a questionnaire.

Appendix D

Chapter 5

The following contains further analyses of behavioural data showing the difference in accuracy across mappings, the improvement in accuracy between and within blocks and the effects of declaration.

Further Analyses of Behavioural Data

Learning Accuracy Across Mappings

Participants performed better in the linearly separable tasks (addition and subtraction in Task 2) compared to the nonlinearly separable tasks. A one tail, dependent sample t-test revealed the effect to be significant (mean linear = 0.735, mean nonlinear=0.589, $t(158) = -9.234$, $p < 0.001$). However, this result is biased by both order and subspace effects. Collapsing across Tasks 1 and 2, we found no significant effect of subspace (addition versus subtraction) with mean addition = 0.667, mean subtraction = 0.657 ($t(158) = 0.491$,$p = 0.624$).

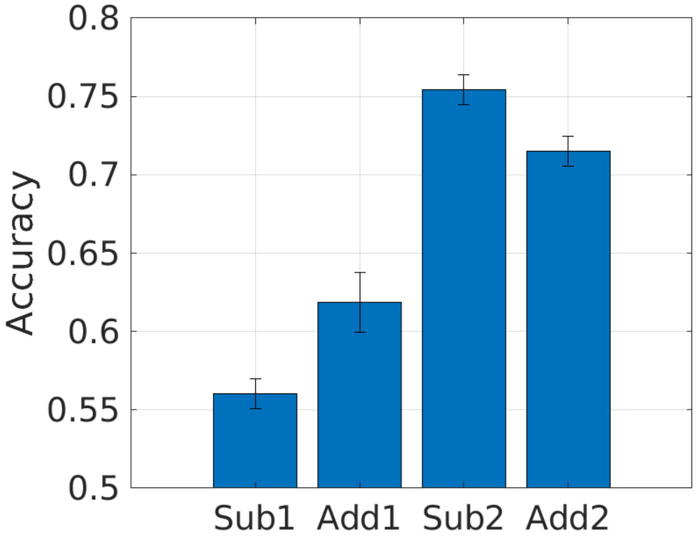


Fig. A **Effect of Mapping on Accuracy.** The plot shows the mean accuracy for each mapping with error bars indicating the standard error of the mean.

Within- versus Between-Block Learning

To identify at which time scale it would be most appropriate to run Sequential Bayesian learning within a learning task, we tested to see how much learning takes place within versus between blocks as subjects learn the first task.

We first look at learning within blocks by comparing the average accuracy increase (averaged over all 80 subjects) from the first 25 to the last 25 trials. This increase was significant (one-sided) for all five learning blocks (paired t-tests, average increase and two-sided p-values for Block 1: Mean=0.038, p=0.029; Block 2: Mean=-0.036, p=0.021; Block 3: Mean=0.051, p=0.00004; Block 4: Mean=0.023, p=0.100; Block 5: Mean=0.033, p=0.029).

We now look at learning between blocks by comparing the average accuracy increase (proportion correctly classified) from one block of trials to the next. This increase was significant for the first 3 transitions between blocks (paired t-tests, average increase and two-sided p-values for Block 1 to Block 2: Mean=0.033, p=0.021; Block 2 to Block 3: Mean=0.031, p=0.028; Block 3 to Block 4: Mean=0.038, p=0.001; Block 4 to Block 5: Mean=-0.003, p=0.824). Note that no significant learning takes place in the last transition.

We then looked to see if the increase between blocks was significantly greater than within - it was not (paired t-test, Mean Between = 0.025, Mean Within = 0.0218, p=0.17). Finally, we compared the first 25 trials of one block to the last 25 trials of the previous block to find out if there was an increase during the break between blocks. This difference, Mean Between = 0.0065, was less than Mean Within = 0.0218 (paired t-test, p=0.17) so there is no benefit of the break (if anything there's a startup cost).

Overall, this analysis shows that learning takes place within blocks and there is no beneficial effect of breaks. This motivated us to choose a block size of 25 trials for the simulations of within task learning (see subsection on Sequential Bayesian Learning over Blocks and Tasks in the Results section). Each point in Fig 6. (right panel) shows mean accuracy averaged over two subsequent blocks (i.e. 50 trials).

Self-Reports on Task 1

We first define a declarative learner as a subject who correctly declared their strategy for the first task. This was based on the participant's response to the open-ended question posed after each task: "How did you approach the task?". A correct declaration of strategy was inferred if their verbal report contained the following phrase (i) for Sub1 "if the number of pie slices is the same/similar in both pies, then choose Sun" or , (ii) Add1 "if the total number of slices on the left and right makes a complete pie, then choose Sun". Phrases that were deemed logically or semantically equivalent were also assessed positively. All assessments were made by author NM.

A two-way between-subjects ANOVA with dependent variable accuracy and independent factors of declaration (declared/not declared) and subspace (same/different) showed a main effect of subspace (F(1,76)=7.66,p=0.007), no main effect of declaration (F(1,76)=2.7, p=0.104) and a significant declaration by subspace interaction (F(1,76)=7.51, p=0.008). Note that the size of the subspace effect in the declarative group is 12 per cent, much larger than the subspace effect computed over all subjects (4.3 per cent - see main text). For the declarative learners in the same subspace group (N=14) the mean accuracy in task 2 is 82 per cent, whereas for declarative learners in the different subspace group (N=15) the mean accuracy in task 2 is only 70 per cent. For the non-declarative learners the corresponding figures are both 72 per cent.

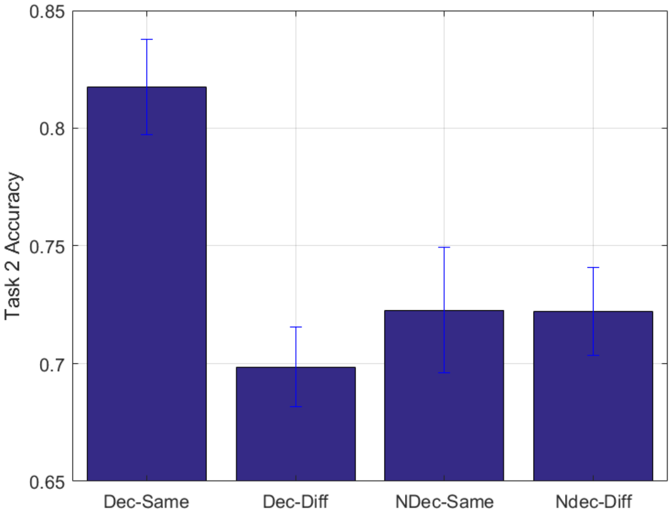


Fig. B **Declaration by Subspace Interaction.** The left two bars are for declarative learners and the right two for non-declarative.

However, we cannot conclude from this analysis that the subspace effect is driven by declarative learners. This is because declarative learners also performed well on the first task. On the first task overall, average correct rates were 66 and 55 per cent for participants who declared or could not declare their strategy ($t(78) = 5.45, p < 10^{-6}$). On the last block of the first task, these correct rates were 76 and 56 per cent ($t(78) = 6.22, p < 10^{-7}$). We then defined a "good learner" as being in the group of those 29 subjects with highest performance in the last block of the first task (we chose 29 subjects to match the number of declarative learners). By this definition 19 of the 29 good learners also declared. A two-way between-subjects ANOVA with dependent variable accuracy and independent factors of learner (good/bad) and subspace (same different) showed a main effect of subspace (F(1,76)=5.46,p=0.022), a main effect of learner (F(1,76)=6.32, p=0.014) and a significant learner by subspace interaction (F(1,76)=4.01, p=0.049). For good learners in the same subspace group (N=16) the mean accuracy in task 2 is 81 per cent, whereas for good learners in the different subspace group (N=13) the mean acccuracy in task 2 is only 72 per cent. For bad learners the corresponding figures are 72 and 71 per cent.

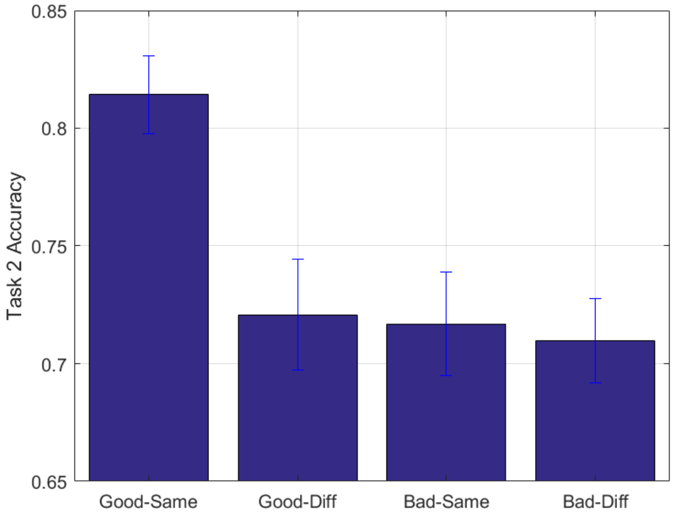


Fig. C **Learner by Subspace Interaction.** The left two bars are for good learners and the right two for bad learners.

But, again, we can't conclude that being a "good learner" drives the subspace effect as many of these subjects (19/29) also declared. To shed further light on this matter we ran a 3-way between-subjects ANOVA with factors of subspace, declaration and learner, and neither of the above two-way interactions (declaration by subspace and learner by subspace) were significant. Furthermore, we calculated the Variance Inflation Factor between these predictors, finding a moderate collinearity of VIF=1.715. We therefore cannot tell which of these factors is driving the subspace effect.

Self Reports on Task 1 and 2

More participants declared the correct strategy in the second task compared to the first one ($x^2(1) = 4.949$, $p = 0.026$). No difference was found between subtraction and addition conditions in task 1 ($x^2(1) = 2.650$, $p = 0.103$). Significantly more participants declared the subtraction strategy in task 2 compared to the addition one ($x^2(1) = 4.073$, $p = 0.043$).

	Subtraction	**Addition**
Task 1	11/40	18/40
Task 2	26/40	17/40

Table A ***Declarations in task 1 and 2*** *The table shows a breakdown of the number of participants who declared the correct strategy in task 1 and task 2 as a function of Add/Sub subspace.*