Building synaptic plasticity rules that learn generalisation-apt neuronal architectures



Homero Barrocas Soares Esmeraldo

St John's College

University of Oxford

A thesis submitted for the degree of

Master of Science by Research

Michaelmas 2019/2020

Acknowledgements

Although I have always known that I was deeply interested in Science, I delayed serious dedication to a scientific career by exploring various other options. I was in doubt because I thought I could have a greater societal impact if I dedicated myself to more applied careers such as software engineering, investment banking or entrepreneurship; and also because there are too many interesting things to do. It was fun having first-hand experience with those, but not only do I enjoy more working towards understanding cognition, but I also hope that being so much more motivated, I will be able to contribute more. Science is a journey of a lifetime and mine has just begun.

This degree, even before it started, was very challenging and motivating. To get to Oxford I did a crowdfunding campaign which was one of the biggest crowdfunding campaigns for education in Brazil. I would like to thank the amazing financial and motivational support I received from all the donors of my campaign and my peers at Rede Mais Educação. In addition, I also thank the support I received for the CrowdEd project, a second crowdfunding, which I did in the UK. The funds from that also supported my studies. This self-funding quest has taught me the value of education and the inequality in its access.

At Oxford, I have been able to work to learn a lot about Neuroscience and also to improve myself as a person. Being so motivated with my research had the side-effect of making me work towards becoming the best version of myself I can. I am happy with what I have achieved and aware that there is a lot more to improve. I am eager to work on it.

I am grateful for the love and unconditional support of my parents Dudu and Grace, my sisters Lara and Michelle, my brothers-in-law Tio Chico and Barui and my lovely nephew Nicola. I would not have made it without you.

I am also very thankful for the love and patience of Ale. Your selfless heart, tolerance and yielding attitude empowered me to work towards becoming a better human being, one small improvement at a time. And you never asked me to do it!

I also thank my supervisors Tim Vogels and Sanjay Manohar for the intellectual support during the development of this research. Thanks, Tim, for your modelling insights and careful review of my draft, and thanks, Sanjay, for the availability amidst a tight schedule and for the insightful detail-oriented discussions.

My life has been mostly lived at the lab on the company of my peers at the Vogels lab and at the Centre for Neural Circuits and Behaviour. You made me happier and smarter. Thank you Raffaele Sarnataro, Alina Krebbers, Olof Rorsman, Bayar Menzat, Bill Podlaski, Victor Hernandez-Urbina, Friedemann Zenke, Georgia Christodoulou, Chaitanya Chintaluri, Everton Agnes, Panos Bozelos, Richard Lonsdale, Jake Stroud, David Jia, Yayoi Teramoto and all the other amazing people at CNCB.

At the Department of Physiology, Anatomy and Genetics, I thank Sarah Noujaim for the friendliness and all the administrative support. Sarah, you are the Queen of DPAG. At CNCB, I thank the admnistrative support of Fiona Woods as well.

My gratefulness to Jotun Hein and Kevin Sharp too. The perseverance and friendliness in our book and journal club discussions made me grow a better scientist by just committing to read and show up.

Finally, God, thank you!

Abstract

One of the hallmarks of intelligent behaviour is the ability to generalise, i.e. to make use of information from specific experiences to solve unencountered scenarios. Here I focus on the generalisation task of transitive inference, which tests the ability to infer that stimulus A is preferable to C, based solely on previous training that A is preferred over B and B is preferred over C. Such inference tasks are thought to be performed by hippocampal-entorhinal circuits and rely heavily on recurrent and conjunctive connectivity to bind associated stimuli. Notably, a 3-layer recurrent neural network model called REMERGE (Kumaran & McClelland, 2012) can solve the transitive inference task through recall dynamics in a circuit that also relies on recurrent and conjunctive connectivity. However, it is not clear how the architecture could be constructed by a learning rule. Here, I develop a biologically plausible non-linear Hebbian plasticity rule to build REMERGE's connectivity starting from an all-to-all architecture. I separate the problem of learning the complete architecture into two parts: 1 - how to learn input connections and 2 - how to learn output connections. For the first part, I design a learning rule that forms connections from coincidentally stimulated inputs to separated conjunctive units in the hidden layer. Learning the second set of weights requires a more complex rule that takes into account a reward signal if the choice was correct, allowing only connections that lead to reward to be strengthened. Ultimately, the results predict a reorganisation of recurrent activity flow across training by potentiating feedforward and feedback connections in hippocampal-entorhinal stimulus pathways, and the need of reward-dependent plasticity to form connections to response neurons. My model bridges the gap between conceptual models of generalisation and how to learn them through biologically plausible plasticity rules.

Contents

List of Figures **vii**

1 Introduction **1**
1.1 Theories of memory generalisation 2
1.2 Generalisation tasks 3
1.3 The transitive inference task 4
1.4 REMERGE model 6
1.4.1 Stimuli presentation 8
1.4.2 Dynamics 9
1.4.3 Choice 10
1.5 Current knowledge of generalisation, opportunities for improvement and my approach 12

2 Learning to associate **13**
2.1 Methods 14
2.1.1 Model inputs during learning 14
2.1.2 Plasticity 14
2.1.3 Initial connectivity 15
2.1.4 Measuring learning performance 16
2.2 Results 19
2.3 Discussion 27
2.4 Supplementary figures 31

3 Learning to respond **32**
3.1 Methods 33
3.1.1 Reward signal 34
3.1.2 Model-based reward complementarity through *switch* signal 34
3.1.3 Modulatory signal 35
3.1.4 Eligibility trace 35
3.1.5 Reward modulated learning rule 36
3.1.6 Integrating learning in all layers 36
3.1.7 Measuring behavioural accuracy 36

3.2 Results 38
3.2.1 Simpler model without switch signal 42
3.2.2 All plastic network 44
3.3 Discussion 47
3.4 Supplementary section: Learning rule performance in a network for 6 stimuli 48

4 Discussion **50**

Appendices

A Parameter values **53**

Bibliography **54**

List of Figures

1.1 Transitive inference task protocol 5
1.2 REMERGE model 7

2.1 Standard Hebbian learning rule disrupts REMERGE's conjunctive architecture and connects all features to a single conjunctive unit . 19
2.2 Synaptic weights between feature and conjunctive layers made plastic and clamped inputs 21
2.3 Learning rule trial-by-trial evolution of connection weights to form conjunctions 21
2.4 Learning performance and parameters robustness analysis via parameter sweep 26
2.5 Clamping stimuli to the feature layer impairs generalisation ability . 29
2.6 Example episodes in which learning took longer or failed 31

3.1 Learning synaptic weights between conjunctive and response layers 33
3.2 Model-based mechanism of reward-modulated learning rule 37
3.3 Intra-trial dynamics of reward modulated plasticity for conjunctive-to-response learning 39
3.4 Evolution of conjunctive-to-response weights across an entire episode 41
3.5 Only weights to one response unit are modified when not using switch 43
3.6 Successful learning without switch signal 44
3.7 Plasticity in all layers simultaneously 47
3.8 Testing designed plasticity rules to learn 6-feature REMERGE connectivity 49

1 Introduction

Contents

1.1 Theories of memory generalisation 2
1.2 Generalisation tasks 3
1.3 The transitive inference task 4
1.4 REMERGE model 6
1.4.1 Stimuli presentation 8
1.4.2 Dynamics 9
1.4.3 Choice 10
1.5 Current knowledge of generalisation, opportunities for improvement and my approach 12

The ability to generalise beyond the specifics of a single experience is one of the most important qualities of intelligence. It frees animals from the risks of making mistakes while learning, and allows them to use the limited learning experiences they undergo to infer what to do in new scenarios that entail related knowledge.

Generalisation implicates disparate realms of cognition such as, for example, spatial navigation and semantic knowledge. An example of spatial generalisation is the ability of animals to intelligently plan trajectories they have never taken before just by remembering previously traversed parts of routes experienced in the past. They mentally associate specific paths in such a way that enable them to get from where they are to the place where they want to go.

In the domain of semantic cognition, many examples exist in which generalisation can be done both by making use of information learned across experiences pursued in different days or just a few minutes, in what is called *rapid generalisation*. When someone gradually learns the semantic structure of concepts in a new field of study, the process of internalising the regularities in semantic knowledge is developed over days, weeks or even months and years. On the other hand, rapid generalisation happens during university lectures, in which simple isolated facts are explained briefly and students can solve exercises relating and extending the explained concepts minutes after they were presented for the first time.

While the cortex is involved in the generalisation process happening over days (McClelland et al., 1995), evidence implicates the hippocampus in rapid generalisation, the ability to do generalisation in short time-frames (Eichenbaum, 2004; Zeithamova et al., 2012).

1.1 Theories of memory generalisation

The involvement of the hippocampus in episodic memory and rapid generalisation is explained by two influential theories of learning. Complementary Learning Systems theory (CLS theory) (McClelland et al., 1995; Kumaran et al., 2016) argues that the hippocampus stores memories using pattern-separated (non-overlapping) representations to reduce interference between similar memories and increase memory capacity. On the other hand, the Relational theory of memory (Cohen and Eichenbaum, 1993; Eichenbaum, 1999; Eichenbaum et al., 1999) proposes that the hippocampus performs generalisation by linking related episodes in a memory space. And in order to perform generalisation, these memorised episodes (episodic memories) would be linked through overlapped representations, contrary to the pattern-separated representations proposed by CLS theory. The role of shared representations for generalisation across different episodes in the Relational theory is at odds with the sparse and non-overlapping representations of CLS theory.

A model called REMERGE (recurrency and episodic memory results in generalisation) was recently proposed to combine these conflicting computational

views of hippocampal function (Kumaran and McClelland, 2012). It reconciles Complementary Learning Systems theory and Relational theory by demonstrating that a neural system with completely pattern-separated representations can perform generalisation, if it has recurrent connections (which is consistent with recent anatomical and functional evidence of a big-loop recurrency between the hippocampus and neocortex; Van Strien et al., 2009; Koster et al., 2018). REMERGE model contributes to the literature of memory generalisation by accounting for behavioural choices and neural activity recordings in several generalisation tasks.

1.2 Generalisation tasks

Many different animals can perform a variety of tasks involving generalisation. Among these tasks are transitive inference (humans, Ryan et al., 2009; monkeys, Brunamonti et al., 2016; rats, Dusek and Eichenbaum, 1997; and pigeons, Wynne, 1997), paired associate inference (humans, Zeithamova and Preston, 2010; and rats, Bunsey and Eichenbaum, 1996), acquired equivalence (humans, Shohamy and Wagner, 2008; rats, Coutureau et al., 2002; and pigeons Urcuioli, 1996), recognition memory (humans, Wixted and Squire, 2010; monkeys, Logothetis et al., 1995; and rats Bevins and Besheer, 2006), and categorisation (humans, Ashby and Maddox, 2005, 2011; monkeys, Orlov et al., 2000; and rodents, Wetzel et al., 1998; Ohl et al., 2001).

All of these tasks require the use of knowledge acquired previously and the ability to make inferences on stimulus-reward associations to make decisions that lead to rewards. In the laboratory they are initiated with a training phase. Afterwards, animals undergo a test-phase in which they are presented with novel stimuli configurations, dissimilar to the ones presented for the choices made during training. To succeed, the test phase requires the subjects to perform generalisation.

The REMERGE model can be used to model how all tasks mentioned above are solved. Here, I focus on the learning process in the model for transitive inference (TI) task. I will introduce the transitive inference task in section 1.3, explain how

REMERGE model works for this task in section 1.4 and introduce my approach and its relation with other theoretical treatments of this task in section 1.5.

1.3 The transitive inference task

In a transitive inference experiment a subject is presented with pairs of stimuli (e.g., A and B) and then prompted to choose one of them (Figure 1.1A). The experiment is divided in a training phase and a test phase. During the training phase, the subject's choices may be followed by a reward or by a delay period with no reward. There are at least 5 stimuli options, denoted here by A, B, C, D and E (some studies use even more; Moses et al., 2006), and they are always shown in pairs of contiguous stimuli from the predefined ordered sequence ABCDE (figure 1.1B). The trained reward contingencies are set as $[A^+B]$, $[B^+C]$, $[C^+D]$ and $[D^+E]$ (rewarded choice marked with [+]; figure 1.1C).

During training, the subject learns the reward contingencies so to consistently choose the correct (rewarded) options. Afterwards, the test phase probes generalisation capacity, the ability to use acquired knowledge to infer new reward contingencies. A choice has to be made between pairs of stimuli never presented together before, like BD. Correct transitive inference means choosing B over D, generalised from previously trained pairs $[B^+C]$ and $[C^+D]$.

Not all animals are capable of performing transitive inference tasks. Humans, for instance, are known to develop that ability between the ages of 7 and 11 years old, when we reach the so-called "concrete operational stage", the third of four stages of cognitive development in Piaget's theory (Piaget, 1928). Animals for which there is evidence of capacity of transitivity include monkeys, pigeons and rats, but they might use different strategies from the ones used by humans (Moses et al., 2006).

Even in animals which are capable of transitive inference reasoning, the ability to perform transitive inferences develops over the course of the task. It takes extended exposure to the task until the participants are able to choose the correct options (Ryan et al., 2009). Solving transitive inference involves learning the

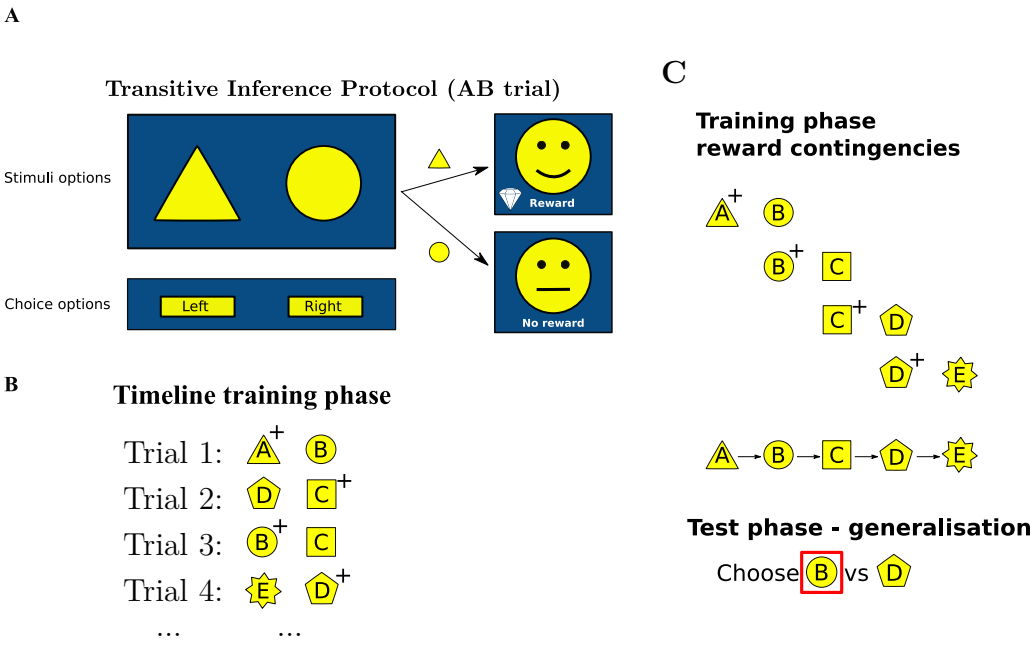


Figure 1.1: A. *Transitive inference task protocol. Two stimuli are shown and the subject has to choose one of them, either the one on the left or on the right (stimuli are randomly shown in each position). The subject can receive a reward or just a delay period with no reward. The task is composed of several trials similar to the exemplified AB trial, but using different stimuli from the possible pairs (AB, BC, CD or DE).* **B.** *Example sequence of presented pairs of stimuli during training phase. Stimuli that, if chosen, would lead to reward are marked with a '+' (plus) sign at the top right.* **C.** *Structure of the task according to the rewards delivered after each choice. The stimuli can be organised in an order based on their reward contingencies. During the test phase of the protocol, the subject is tested for the ability to infer the correct choice (marked with a red square) on pairs of stimuli never presented together before.*

reward contingencies used during training and extrapolating from them when presented with novel action options.

Lesion studies indicate an importance of the hippocampus in transitive inference reasoning. If the hippocampus is lesioned, although the memory for previous knowledge remains intact (animals remember trained premises), they lose the ability to do correct inference decisions (Dusek and Eichenbaum, 1997; Buckmaster et al., 2004; for review, see Eichenbaum, 2004).

The main focus of this research is to analyse the learning process that happens during learning how to perform the transitive inference task.

1.4 REMERGE model

The REMERGE architecture is composed of 3 layers: feature, conjunctive and response ones (see Figure 1.2A). Each layer is composed of units that model neuronal populations in the brain. Feature units code for the stimuli presented in the protocol (A, B, C, D, in this case; see below for discussion of simplification from 6 to 4 stimuli). The feature layer is recurrently connected to the conjunctive layer, which comprises one unit for every association to be learned, i.e., AB, BC and CD. Finally, the response layer is composed of read-out units whose activities are used to represent behavioural choices (A, B, C or D; figure 1.2B).

Feature units model neurons that could be present in the entorhinal cortex or in some upstream cortical sensory area, responding to stimulus identity. The conjunctive layer represent neurons in the hippocampus on subregions such as Dentate Gyrus and/or CA3. Finally, the response layer units could represent neurons that send signals to motor-planning areas.

The algorithm underlying the computation that REMERGE model does to perform generalisation is called *recurrent similarity computation*. When inputs are given to the feature layer, similarity with previously stored associations (AB, BC, CD) is calculated through the feature-to-conjunctive connections and the activation function of conjunctive units (hedged softmax; cf. section 1.4.2). Then, conjunctive units influence back feature-unit activities through the bidirectional (i.e. recurrent) connections.

The hedged softmax activation function (a softmax with an extra constant in the denominator) computes the similarity between the momentaneous activity at the feature layer and the stored associations. The associations are coded through the hard-wired feature-to-conjunctive connections - note that the conjunctive units' labels are exactly the features that are presented together during training.

In the generalisation trial, the network is presented with, for example, A and C as choice options. Correct generalisation is choosing A: if A is preferable to B and B is preferable to C, then A should be preferred over C.

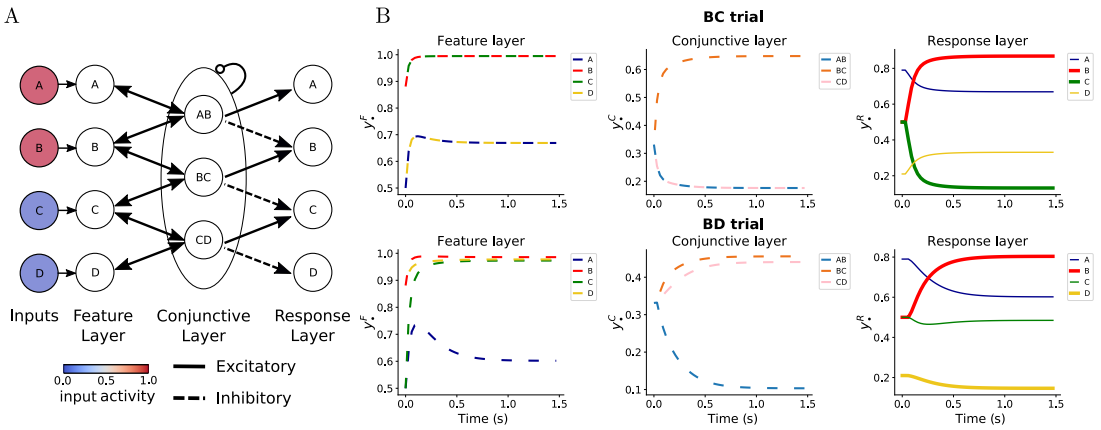


Figure 1.2: A. *Neural network architecture that can solve the transitive inference task (simplified architecture adapted from (Kumaran and McClelland, 2012); see text for discussion of the simplification). Solid lines are excitatory connections and dashed lines are inhibitory connections. Curved connection ended in open-circle represents the use of the hedged softmax function in the conjunctive layer. Inputs are 1 at units A and B and 0 at C and D as they would be in an AB trial.* **B.** *Plots of the activities of all the units in feature (left), conjunctive (centre) and response layers (right), when given inputs in feature units B and C (so called BC trial) - top row; and when given inputs in feature units B and D (so called BD trial) - bottom row.*

How does the network solve this? Conjunctive units' activities interactively (through recurrence) quantify the similarity with the activities in feature units until the activities converge. The response is then computed simply by feedforward conjunctive-to-response connections learned during the training phase of the task. In sum, REMERGE's generalisation capacity results from feedforward response activation made by conjunctive similarities of recurrently activated features with stored associations.

In this study, on chapter 2, I will include synaptic plasticity to a simplified (smaller) REMERGE-like neural network. Thus, the connection weights between units can change over time. As plasticity introduces complexity to the model, I decided to reduce the number of stimuli from six (in the original REMERGE study; Kumaran and McClelland, 2012) to four in order to simplify the dynamics of the model circuit and facilitate learning.

But five is the minimum number of stimuli for the task to require transitivity to be solved. With five stimuli, choosing between B and D requires transitivity, because they are rewarded and not rewarded the same amount of times, only depending

on what other stimulus they are paired with. The choice depends on the subject having learned the linear hierarchy of the reward contingencies (figure 1.1C).

The reduction of the model to only 4 units has an important implication: the task can be solved by associating each feature with its history of reward. The terminal stimuli (A and D) are either always rewarded (A) or never rewarded (D). So any trials involving these stimuli cannot be used to test TI ability, as there is no need to use transitivity to know what choice to make. But when using 4 stimuli (A, B, C and D), the non-adjacent stimulus pairs available for generalisation trials are AC and BD, which involve the terminal stimuli. Still, I will use this simplification for most of this work because the connectivity is similar to the case with more units. Further research should extend this model to use five or more stimuli; I have performed a short exploration of the case for 6 stimuli in the supplementary section 3.4.

The results of the simulation of this simplified REMERGE model are shown in figure 1.2B. My model, similarly to the original REMERGE, makes correct choices, which are based on the activity of response units, both in premise trials (AB, BC and CD - called premise because they are the base for the inference) and in test trials with inference pairs (AC, BD).

The simplified REMERGE model and TI task were simulated using the mathematical formulations and parameters described below.

1.4.1 Stimuli presentation

A TI experiment is simulated through several *episodes* in which the subject learns the reward contingencies of the stimuli choices, without any prior knowledge (network weights are re-initialised).

Each episode is divided in trials in which a pair of contiguous stimuli (AB, BC or CD) is randomly chosen as inputs to the network. Each trial lasts 3 seconds, with 2.5s of stimulus presentation and 500ms interval. The network's stimulus choice is read out from the activity in the response units at the end of the trial stimulation period.

Note that I am using only 4 stimuli, rather than the minimum of 5 in transitive inference experiments. The reason is to facilitate learning, which will be analysed

in the following chapters. Reducing the number of stimuli to four facilitates pattern separation: having different conjunctive units win the competition when stimulated by different feature units. With more feature units, this is harder because more feature units can be randomly initialised having connections that associates two pairs to the same conjunctive unit.

1.4.2 Dynamics

The units are rate neuron models with dynamics as follows(Dayan and Abbott, 2001):

$$\tau_n \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} w_{ij} y_j(\boldsymbol{x}) + \alpha \cdot ext_i(t) \tag{1.1}$$

The evolution of the synaptic input x_i to unit i in time t is calculated through the integration of equation 1.1. $\tau_n = 70$ ms is the time constant of the neurons, w_{ij} is the strength of connection from a unit j connected to unit i, y_j's are the firing rates of each unit j among the N units connected to unit i (calculated using the activation function $y_j(\bullet)$ on a vector of units' synaptic inputs $\boldsymbol{x}$; here, $\bullet$ is a placeholder for the synaptic input as independent variable of the function). α is a fixed parameter (set to 0.5) to scale the external input $ext_i(t)$ given to unit i. The external inputs represent the stimuli presented to the subject in each trial, thus ext_i is non-zero only at units in the feature layer at certain times.

The activation function $y_i(\bullet)$, which describes the relation between the unit's synaptic input x_i (or of a set of units $\boldsymbol{x}$) and its firing rate y_i can be different in each layer. A logistic function was used in the feature and response layers:

$$y_i^F(\boldsymbol{x}) = y_i^R(\boldsymbol{x}) = \frac{1}{1 + e^{-x_i/\theta}} \tag{1.2}$$

The superscript $\bullet^F$ denotes quantities related to the feature layer, and, similarly, $\bullet^C$ and $\bullet^R$ for conjunctive and response layers. $\theta = 0.25$ is a temperature parameter that controls how much a change in the synaptic input affects the unit's output activity. With this logistic function, the activation is at 0.5 for null input and approaches 1 for high inputs. The conjunctive layer used a different activation function called *hedged softmax*:

$$y_i^C(\boldsymbol{x}) = \frac{e^{x_i/\theta}}{\kappa^{1/\theta} + \sum_{k=1}^{N_l} \mathrm{e}^{x_k/\theta}} \tag{1.3}$$

N_l is the number of units in the conjunctive layer. This function is similar to a softmax function, except for the "hedge" term $\kappa^{1/\theta}$ in the denominator, which enables the conjunctive layer to have low activity in all units when there is low input from the feature layer. The hedged softmax implements competitive inhibition in the conjunctive layer, similar to a winner-take-all mechanism. The temperature parameter θ controls the level of competition: when θ is small there is high competition, with mostly only one unit with high activity.

The competition arises from the term $\sum \mathrm{e}^{x_k/\theta}$ in the denominator: an increase in the inputs to only one of the conjunctive units affects not only its own activity y_i (increasing it), but it also reduces the activity of the other conjunctive units. Furthermore, the sum of all y_i's is at max 1 (when $\kappa = 0$), so no matter how big are the inputs, the units compete for the activity without substantially increasing the overall activity in the layer.

1.4.3 Choice

The experiment uses a set of stimuli $\mathcal{S} = \{A, B, C, D\}$, for which there is an ordered linear hierarchy $\mathcal{H} : h1 \rightarrow h2 \rightarrow h3 \rightarrow h4$, with $h1, h2, h3, h4 \in \mathcal{S}$. The experimenter arbitrarily sets $\mathcal{H} : A \rightarrow B \rightarrow C \rightarrow D$, which is not known a priori by the subject performing the task. The goal of the task is to test if the subject is able to infer the hierarchy $\mathcal{H}$, using the limited training experience acquired during the training phase.

In the task setting for humans (Ryan et al., 2009) and monkeys (Brunamonti et al., 2016), the stimuli are presented side by side in a screen. Their position, left or right, is randomised across trials. Thus, at each trial, the subject is presented with a pair of stimuli mathematically represented here by an ordinate pair (p_1, p_2), with p_1 on the left and p_2 on the right. Furthermore, during training p_1 and p_2 are always contiguous stimuli from the linear hierarchy, i.e. $p_1 = h_i, p_2 = h_{i+1}$ or $p_1 = h_{i+1}, p_2 = h_i$, with

$1 \leq i < 4, i \in \mathbb{Z}$. As the position is randomised and is irrelevant for the task, the neural network model does not take into account left-right position. I can then mathematically represent the stimuli presented at each trial to the network as a set:

$$\mathcal{P} = \{p_1, p_2\} \tag{1.4}$$

One of these choice options is the correct one (the rewarded option), denoted as O_c, and the other one is wrong[1], denoted as O_w, with $O_c, O_w \in \mathcal{P}$. The correct option is the one that comes earlier in the hierarchy $\mathcal{H}$, i.e. $O_c = h_i$, and the wrong comes later, i.e. $O_w = h_{i+1}$.

For example, in an AB trial the presented stimuli are $\mathcal{P} = \{A, B\}$, and correct and wrong choices are $O_c = A$ and $O_w = B$. Whereas in a BC trial, $\mathcal{P} = \{B, C\}$, $O_c = B$ and $O_w = C$ (figure 1.1B). The network makes a decision of an action $a \in \mathcal{P} = \{p_1, p_2\}$, calculated from unit activities following the Luce choice rule (Luce, 1959):

$$P(a = p_1) = \frac{e^{y^R_{p_1}/\beta}}{e^{y^R_{p_1}/\beta} + e^{y^R_{p_2}/\beta}} \tag{1.5}$$

In which β is the equivocation parameter that controls how deterministic the decision is ($\beta = 0$ being completely deterministic). In my simulations β was set to $\beta = 0.3$. In an AB trial, for instance, if $y^R_A > y^R_B$, the network is more likely to choose action $a = A$.

***

Having successfully reproduced network dynamics results from Kumaran and McClelland (2012) (Figure 1.2) I will explore how the connections of the REMERGE architecture can be formed over the course of the task training phase in Chapter 2. Before that, I address the current understanding of generalisation both from within CLS theory and REMERGE model and also from other perspectives.

[1] In the Transitive Inference task as performed in Ryan et al. (2009), choosing the wrong option in the results in the display of an angry face and the text "Wrong!" in the screen for a human participant.

1.5 Current knowledge of generalisation, opportunities for improvement and my approach

Besides REMERGE, other models have been proposed to solve generalisation tasks. For transitive inference, for instance, one study used a feedforward neural network to learn the task through backpropagation (De Lillo et al., 2001). Nevertheless, it is still not clear if the brain can implement backpropagation (Whittington and Bogacz, 2019). Another study proposed an algorithm called "betasort" that associates each stimulus to a position in a unit scale (Jensen et al., 2015). Betasort solves the task (whereas standard Q-learning does not), but its implementation in a neural network has not been proposed.

In sum, these approaches are cut short in biological realism on the implementation level; there needs to be developed a theory about how the brain performs generalisation.

The REMERGE model is also simplified in terms of biological realism. It has a small number of neurons, fixed connections and minimal connectivity that do not reflect the complexity of medial temporal lobe circuits. Nevertheless, it makes progress in biological plausibility by using its units and connections as representations of neural circuits in the hippocampus.

In sum, REMERGE offers an algorithmic explanation of how to solve transitive inference in a neural network inspired in the hippocampus, but it is not clear how the architecture could be constructed by a learning rule. Here, I develop a biologically plausible non-linear Hebbian plasticity rule to build REMERGE's connectivity starting from an all-to-all architecture.

I separate the problem of learning the complete architecture into two parts: 1 - how to learn feature-to-conjunctive connections and 2 - how to learn conjunctive-to-response connections. This is possible because REMERGE's architecture has no feedback connections from the response layer to the feature or the conjunctive layers, making feature-to-conjunctive weight learning independent of the conjunctive-to-response weights. In chapter 2, I cover the first part, feature-to-conjunctive learning, and, in chapter 3 the second part, conjunctive-to-response learning.

2

Learning to associate

Contents

2.1 Methods 14
2.1.1 Model inputs during learning 14
2.1.2 Plasticity 14
2.1.3 Initial connectivity 15
2.1.4 Measuring learning performance 16
2.2 Results 19
2.3 Discussion 27
2.4 Supplementary figures 31

Several animals are able to learn a transitive inference task after appropriate training. The REMERGE model is a candidate neural network whose dynamics solves the task, but we lack a learning rule that explains how the architecture could be constructed in the brain.

To study this learning process, I developed a biologically plausible, non-linear Hebbian plasticity rule to build REMERGE's connectivity and, consequently, solve the task of transitive inference. To this end, I constructed a network with identical number of layers and units as the REMERGE model. Starting with "naive" (all-to-all) connections, I searched for learning rules that can mould connections during the training phase of the task to learn REMERGE's connectivity.

In this chapter, I first discuss rules that govern feature-to-conjunctive connections. To produce the desired architecture, I develop a learning rule that strengthens the connections from coincidentally stimulated pairs of feature units to different conjunctive units. The conjunctive-to-response connections are kept fixed, at appropriate values. In Chapter 3, I will cover plasticity in conjunctive-to-response connections..

2.1 Methods

To study the learning process involved in the acquisition of the ability to do transitive inference, I first set out to construct only the feature-to-conjunctive connections from the REMERGE model. I subject the network to the transitive inference protocol and find a non-linear Hebbian learning rule capable of building REMERGE's conjunctive architecture. In the following, I describe the details of the stimulation setting, the model and analytical methodology.

2.1.1 Model inputs during learning

In most learning simulations (except for the one in Figure 2.1), the stimulation protocol is slightly altered from the original REMERGE model (Kumaran and McClelland, 2012, equation 1.1). Instead, the inputs are clamped to the feature units activities (Figure 2.2); during an AB trial, for example, feature units A and B are clamped to high activity ($y_A^F = y_B^F = 1$) and C and D to zero activity ($y_C^F = y_D^F = 0$). The clamp is maintained for 2.5s after which there is an inter-trial interval of 0.5s when all feature units are clamped to baseline activity $\beta^F = 0.5$ (Figure 2.3).

2.1.2 Plasticity

When plasticity is included in the model, the connections between feature and conjunctive layer units evolve according to a mathematical rule in which the variation of the weights is a function of pre- and post-synaptic units' activities.

The standard Hebbian rule used for the simulation in Figure 2.1 that showed the failure to remain in the desired connectivity was as follows:

$$\frac{dw_{cf}^{CF}}{dt} = \frac{dw_{fc}^{FC}}{dt} = (y_c^C - \beta^C)(y_f^F - \beta^F) \tag{2.1}$$

in which the parameters were set to $\beta^C = 0.33$ and $\beta^F = 0.5$.

I then adapted the Hebbian rule above by applying a Rectified Linear Function (i.e. $max(\bullet, 0)$) to the conjunctive term in the plasticity rule:

$$\tau_p \frac{dw_{cf}^{CF}}{dt} = \tau_p \frac{dw_{fc}^{FC}}{dt} = (y_f^F - \beta^F)[y_c^C - \beta^C]_+ \tag{2.2}$$

which I denoted as RELU-Hebbian rule (RELU stands for Rectified Linear Unit). In the equation, w_{cf} is the connection weight between feature unit f (pre-synaptic) and conjunctive unit c (post-synaptic), and vice-versa for w_{fc}. In other words, feedforward and feedback connections have the same plasticity rule, for tractability. $[\bullet]_+$ is a Rectified Linear Function (i.e. $max(\bullet, 0)$). y_c^C is the firing rate of conjunctive unit c and y_f^F is the firing rate of the feature unit f. β^F and β^C are the baseline activities of feature and conjunctive units, set to $\beta^F = 0.5$ and $\beta^C = 0.33$ (the activity elicited when each conjunctive unit receives input from just two feature units). τ_p is the time constant of the learning rule, modulating how fast the synaptic weights change. The weights are hard-bounded between 0 and 1.

There were no self-connections, nor connections between different feature units. The connections between conjunctive units were not modelled explicitly, these units interact only via the hedged softmax function (equation 1.3), whose parameters were maintained constant during the whole simulation.

2.1.3 Initial connectivity

For each new episode (a set of trials; defined in section 1.4.1), the weights were randomly initialised. Three different weight initialisations were used:

- Flat all-to-all initialisation: $w_{fc} = w_{cf} = 0.5, \forall(f, c)$ except for $(f, c) = (B, C_2)$ (to break the symmetry in the architecture and thus allow competition and learning to differentiate the connections);

- Random initialisation: weights were initialised by sampling their initial values from a continuous uniform distribution $\mathcal{U}(0, 1)$;

- Random normalised initialisation: weights were also initialised randomly, but constrained such that the sum of the connections to each conjunctive unit was a fixed quantity. That quantity was 2, because in REMERGE's architecture each conjunctive unit receives two connections of strength $w = 1$ and the other two are zero. Mathematically, the constraint is $\sum_f w_{cf} = 2, \forall c$;

2.1.4 Measuring learning performance

Using 4 feature units, there are 6 possible connectivities that are equivalent to REMERGE (Figure 2.4A). The first pair of feature units (AB) can connect with any of the 3 conjunctive units; the second pair (BC), can connect to one of the remaining two conjunctive units; and the last pair (CD) has to connect with the last remaining conjunctive unit. Thus, the number of possible REMERGE architectures is indeed $3! = 6$. In general, with $N_{conjunctive}$ conjunctive units and $N_{feature}$ feature units, such that $N_{feature} = N_{conjunctive} + 1$, the number of equivalent REMERGE connectivities is equal to $N_{conjunctive}!$.

At each trial n, to evaluate the performance of my learning rule, I computed the distance $D_a(n)$ from the weight matrix to architecture $a = 1, 2, ..., 6$ (each of the 6 equivalent REMERGE architectures):

$$D_a(n) = \frac{\sum_{i,j}(w_{ij}^n - w_{ij}^{REMERGE_a})^2}{N_W} \tag{2.3}$$

In which, I normalised the squared distance (the numerator) by the number of connections $N_W = n_{conjunctive} \times n_{feature}$. Plotting $D_a(n)$ across the n trials of an episode one gets the episode's distance curve (Figure 2.4B).

In each episode, because of variations due to randomness in weight initialisations and stimulus presentation sequences, the network learns a different architecture from the 6 possibilities. The distance curves described above quantify the distance to each of the possible architectures. To compare the learning process across

episodes that converge to different REMERGE architectures, I needed a single error curve for each episode.

To have such a metric, I chose only one of the possible architectures as the target in that episode. The target a_{target} is considered to be the architecture at minimum distance to the network connectivity on the last trial n^f (e.g. a_{target} identifies the orange curve in Figure 2.4B among the others):

$$a_{target} = \underset{a}{\operatorname{argmin}}\, D_a(n^f) \tag{2.4}$$

Hence, the error curve $E(n)$ is the distance curve to the target architecture in that episode:

$$E(n) = D_{a_{target}}(n) \tag{2.5}$$

The median and interquartile range of $E(n)$ across several episodes depict the cross-episode collective performance of the learning algorithm, which should average out variations due to randomness.

Finally, to evaluate whether the learning rule managed to effectively learn REMERGE's architecture, I classified episodes as having converged to REMERGE architecture, or failed to learn, based on their error in the last trial, $E(n^f)$, being smaller than a convergence threshold Θ. If $E(n^f) < \Theta$, the episode is considered to have converged to a REMERGE architecture. Convergence bars were plotted to visualise the proportion of episodes that converged (e.g. Figure 2.4). The thresholds were set to $\Theta^{FC} = 0.09$ for feature-to-conjunctive layer, $\Theta^{CR} = 0.15$ for conjunctive-to-response layer and $\Theta^{full} = 0.09$ for the whole architecture. These values were set such that if the network ever reached an error lower than the threshold during the episode, it would be almost certain that it would reaching errors very near zero.

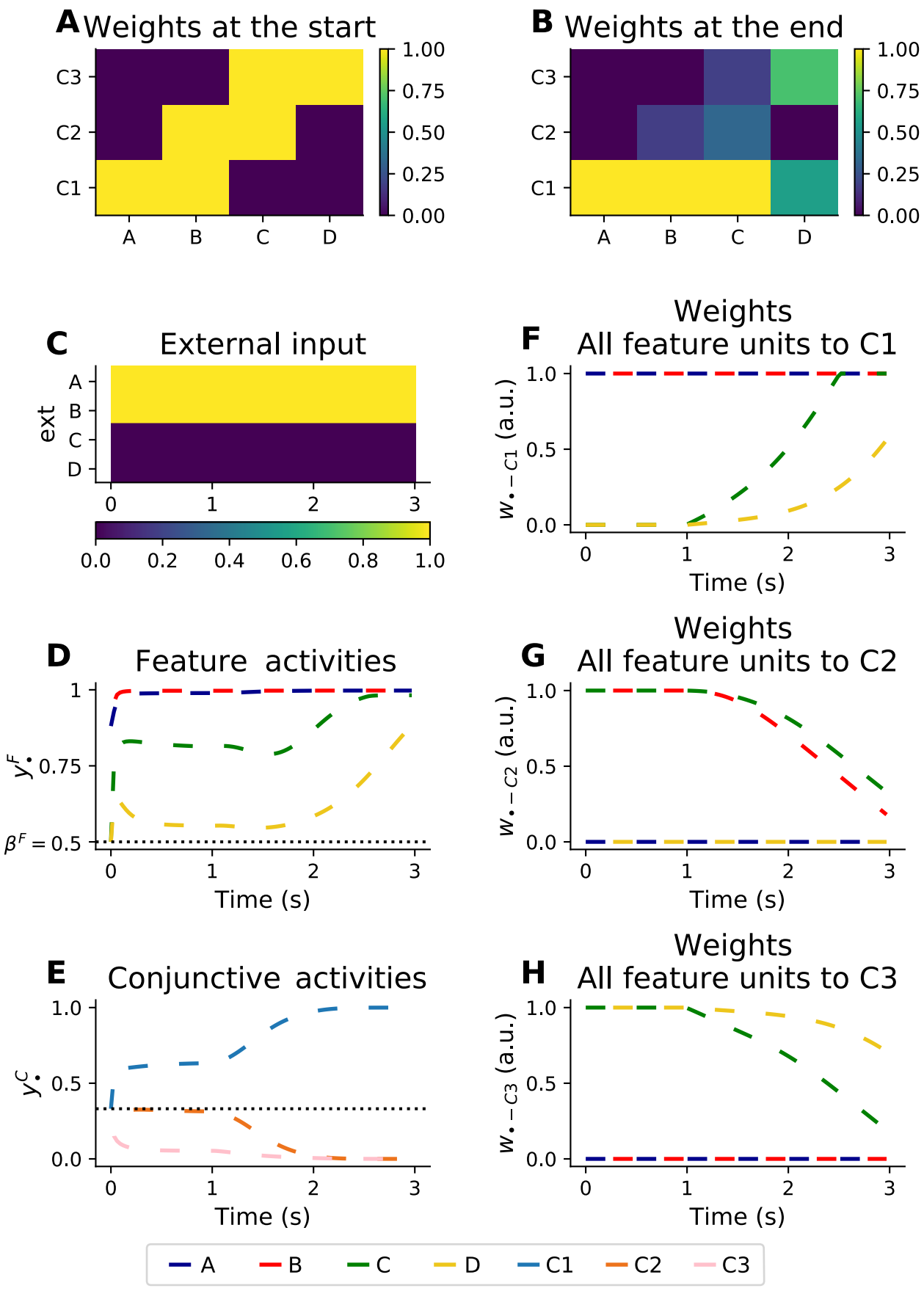

A Weights at the start
C3
C2
C1
A
B
C
D
1.00
0.75
0.50
0.25
0.00
B Weights at the end
C External input
ext
A
B
C
D
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
D Feature activities
$y^F_\bullet$
$\beta^F = 0.5$
Time (s)
E Conjunctive activities
$y^C_\bullet$
Time (s)
F Weights
All feature units to C1
$w_{\bullet - C1}$ (a.u.)
Time (s)
G Weights
All feature units to C2
$w_{\bullet - C2}$ (a.u.)
Time (s)
H Weights
All feature units to C3
$w_{\bullet - C3}$ (a.u.)
Time (s)
A
B
C
D
C1
C2
C3

Figure 2.1: ***Failure to learn****. Standard Hebbian learning rule disrupts REMERGE's conjunctive architecture because of recurrent activation. The learning rule connects all features to a single conjunctive unit and removes other connections.* **A.** *Weight matrix of the network initialised with REMERGE connectivity.* **B.** *After 3s of stimulation of features A and B, the Hebbian learning rule is under way of connecting all features to conjunctive unit C1 and depotentiating all other connections to other conjunctive units (tendency clearly illustrated in panels F, G, H).* **C.** *External input shown across time. Features A and B receive inputs, while C and D do not.* **D and E.** *Activity of feature and conjunctive units over 3s of stimulation. Plasticity is enabled after 1s. Dashed lines represent baseline activities for each layer* $\beta^F = 0.5$ *and* $\beta^C = 0.33$. **F-H.** *Weight variation over time. All weights to conjunctive unit C1 increase until they would reach the boundary value of 1, whereas all weights to C2 and C3 are depotentiated until they would reach zero if the simulation lasted more.* $\tau_p = 300$ *ms in this simulation.*

Only when the final squared error was lower than 0.09 per connection, the connectivity was considered to have converged to one of the target architectures. The proportion of episodes in which there was convergence and how fast the connectivity converged varied for different initial conditions and values of parameters τ_p, θ, and κ.

Besides measuring learning through a representational equivalence with REMERGE, one can also measure how often the network choose the correct actions, i.e. measure its accuracy. I will explore that in the next chapter.

2.2 Results

Learning how to perform transitive inferences is an important cognitive ability that involves generalisation of previous experiences. The REMERGE model performs those inferences, but learning to construct its feature-to-conjunctive connectivity with biologically plausible learning rules remains elusive.

Due to recurrency in the circuit, standard Hebbian plasticity can lead to formation of undesired connections (Figure 2.1). In fact, even when the connectivity is initialised at REMERGE values (Figure 2.1A), the plasticity rule strengthens connections between the stimulated and all the recurrently activated feature units to the conjunctive unit that wins the competition in its layer (Figure 2.1B). For instance, if stimuli A and B are presented to the network by stimulating feature units A and B (Figure 2.1C), the recurrent connectivity leads to activation of

feature units C and D (Figure 2.1D and E). The simultaneous activation of feature units C and D, and of conjunctive unit C1 spur strengthening their connection through the Hebbian learning rule (Figure 2.1F). Moreover, because conjunctions C2 and C3 are below baseline, the Hebbian rule also weakens their connections to all feature units (Figure 2.1G and H).

To avoid the weakening of connections to conjunctive units whose activity is below baseline (C2 and C3 in the previous example, due to lost competition over C1), I adapted the Hebbian plasticity rule by applying a rectified linear (RELU) function to the conjunctive term of the learning rule (Equation 2.2, Methods). And to circumvent connecting all features to the winning conjunction, I clamped the inputs to feature units during learning, removing the recurrent influence from conjunctive units (Figure 2.2). These adaptations enable the rule to successfully build the desired conjunctive connectivity. The successful learning rule, starting from random all-to-all feature-to-conjunctive connections, leads the connectivity pattern closer to REMERGE's conjunctive structure (Figure 2.3).

During a single trial, a conjunction between a contiguous pair of feature units and a single unit in the conjunctive layer gradually forms. Trial after trial, connections are gradually altered so that the REMERGE architecture is built under the influence of the learning rule and of pre- and post-synaptic units' activities (Figure 2.3A).

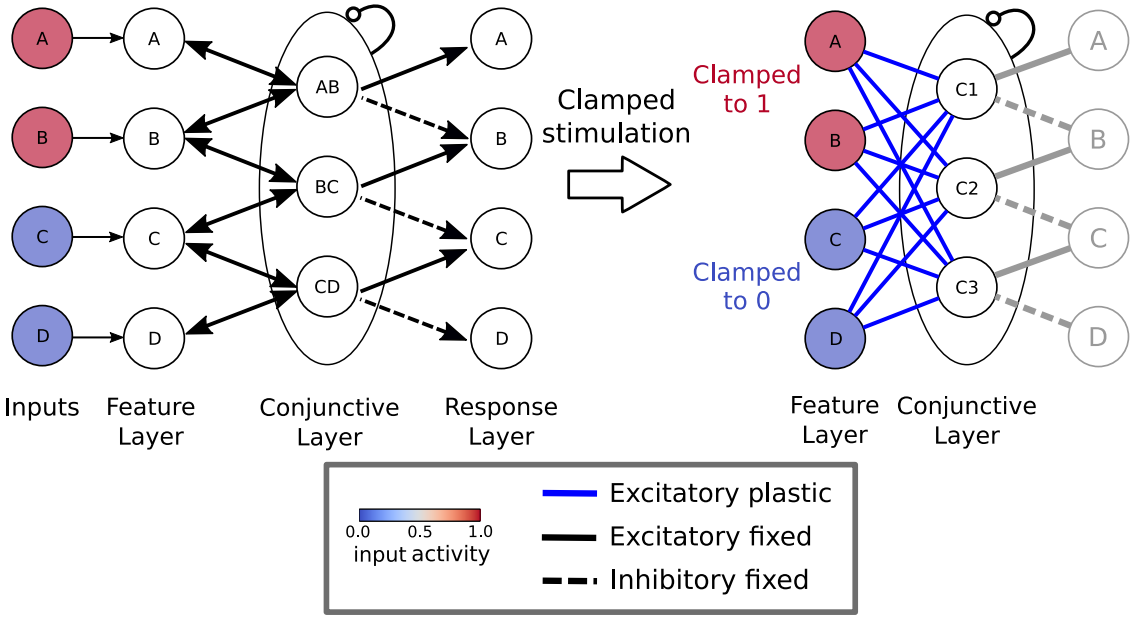
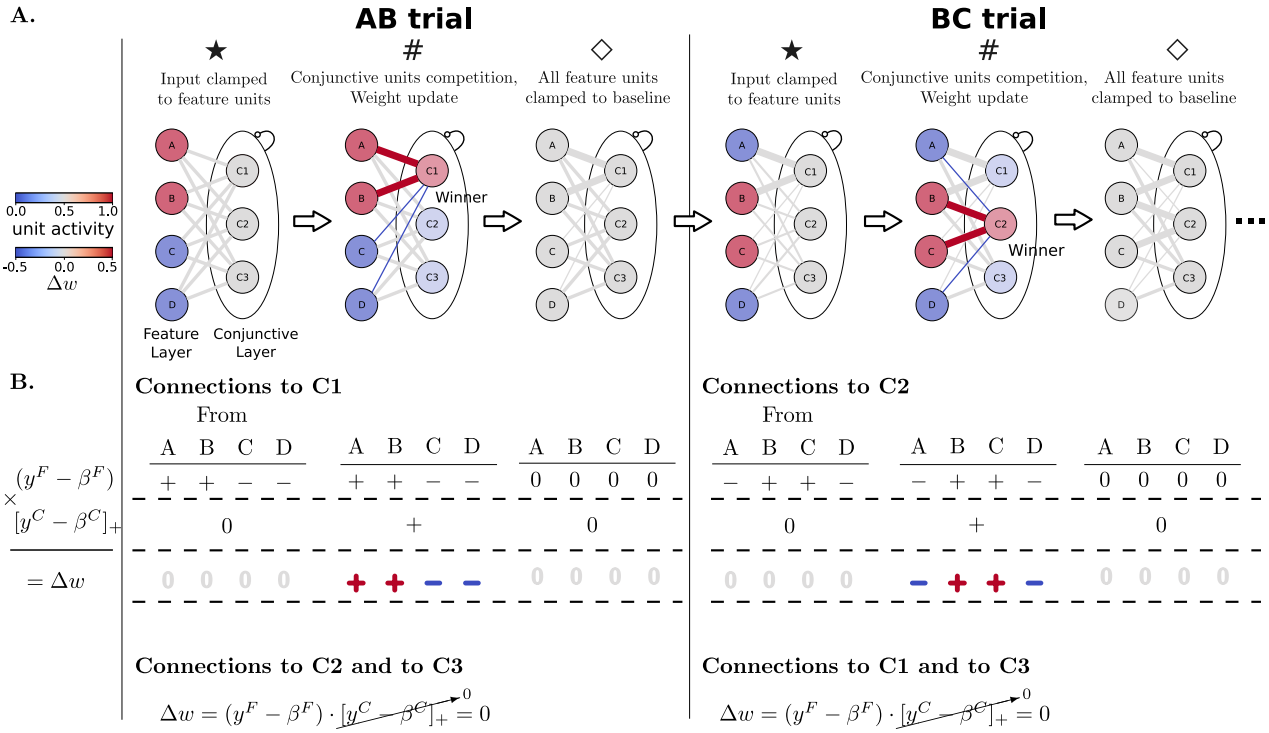


Figure 2.2: *Synaptic weights between feature and conjunctive layers were made plastic and model stimulation is simplified by clamping inputs to feature units (see text for discussion of simplification consequences). On the left the REMERGE model with all its connections and inputs in an AB trial. On the right the response layer is drawn in grey because only the plasticity between feature and conjunctive layer is analysed; connections to the response layer are kept fixed at the REMERGE model values. The network starts with all-to-all connections (drawn with blue lines without arrow heads representing bidirectional connections) that are changed throughout the task protocol as a result of pre- and post-synaptic activity. Conjunctive units are now denoted C1, C2 and C3, because their connections to feature pairs cannot be determined a priori.*


A.
AB trial
BC trial
★
#
◇
Input clamped to feature units
Conjunctive units competition, Weight update
All feature units clamped to baseline
Winner
Feature Layer
Conjunctive Layer
unit activity
Δw
B.
Connections to C1
Connections to C2
From
$(y^F - \beta^F)$
$\times$
$[y^C - \beta^C]_+$
$= \Delta w$
Connections to C2 and to C3
Connections to C1 and to C3
$\Delta w = (y^F - \beta^F) \cdot [y^C - \beta^C]_+ = 0$

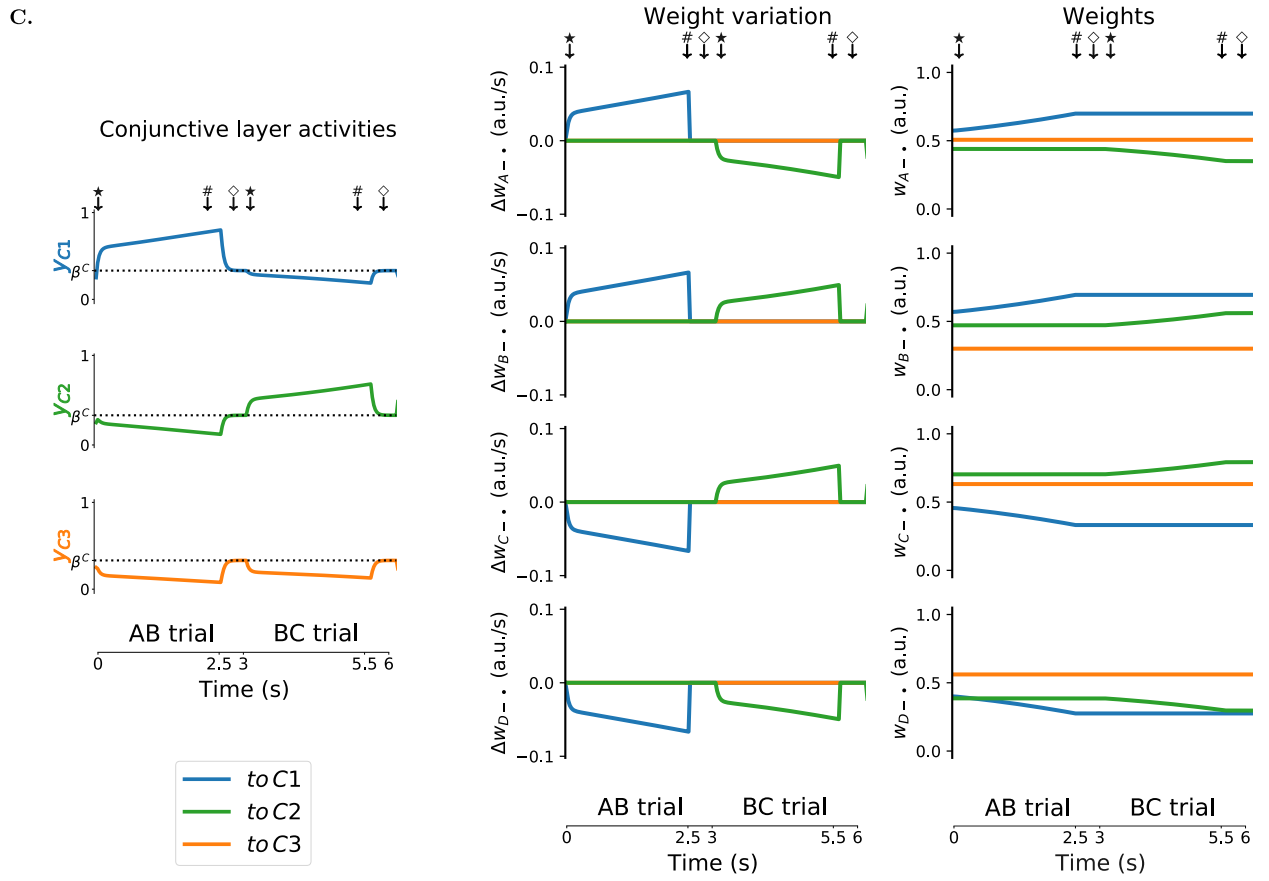

C.
Conjunctive layer activities
Weight variation
Weights
AB trial
BC trial
Time (s)
to C1
to C2
to C3

Figure 2.3: *Trial-by-trial evolution of the connections between feature and conjunctive layers under the influence of RELU-Hebbian learning rule.* **A.** *In the beginning of the experiment the connectivity of the network (top left) has all-to-all connections between feature and conjunctive units with random weights. After AB and BC trials (top right), the connectivity is closer to the REMERGE conjunctive architecture: feature pairs AB and BC strongly connected mostly to only one unit in the conjunctive layer (C1 and C2). Further trials continue to build the REMERGE architecture. Three different trial moments ($\star$, # and $\diamond$) are shown. $\star$ is the start of the trial, immediately after the input clamp is put, # is just before the inter-trial interval and $\diamond$ is halfway through the interval. The activity level of units is displayed with red for active units ($y^F = 1$), blue for inactive units ($y^F = 0$) and grey in between. Similarly, the instantaneous weight variation is depicted by the colour of the connections. Connection lines thickness represent the absolute value of weights (initially random weights are depicted all with the same thickness for clarity).* **B.** *RELU-Hebbian plasticity rule determines that weight variations depend on the multiplication of deviations from the baseline of activities in feature and conjunctive units, with a RELU function applied to the conjunctive unit's term (equation on the left). Units' activities are influenced by the inputs and the competition at the conjunctive layer. In an AB trial, at $\star$ the activity from the feature layer has not yet influenced the conjunctive layer, whose units' activities are near baseline and thus no plasticity happens. At #, because C1 wins the competition, its activity goes above baseline (represented with '+') and the connections from all feature units to C1 can be altered. Connections from A and B (above baseline β^F: '+') to C1 are strengthened (thicker red lines in panel A) and the connections from C and D (below baseline β^F: '-') to C1 are weakened (thinner blue lines). The connections to the other conjunctive units remain static because of the conjunctive activity below baseline β^C (the RELU function yields 0; connections are coloured in light grey). In the interval $\diamond$, the features are clamped to baseline, so no plasticity happens. A similar process happens in BC trial.* **C.** *Results of model simulation showing conjunctive units' activities (dotted line shows baseline activity β^C), weight variation and weight absolute values in feature-to-conjunctive connections during AB and BC trials. In this simulation $AB \rightarrow C1$ (blue trace) and $BC \rightarrow C2$ (green trace) conjunctive connections are under formation. Connections to C3 remain static.*

Two feature units are stimulated (clamped to $y^F = 1$) and two are blocked (clamped to $y^F = 0$). The conjunctive layer units compete with each other through the hedged softmax function mechanism so that usually only one unit per trial becomes highly active compared to the others and reaches activity above baseline β^F (Figure 2.3C left). Due to the RELU function in the plasticity rule, the winner unit is the only one that triggers plasticity in that trial. The RELU reduces to zero the weight variation for units with activity below baseline because of the term $[y^C - \beta^C]_+$, which equals zero in the term of the plasticity rule for these units. The RELU-Hebbian rule potentiates the winner's connections to the stimulated feature units and depress its connections to the blocked feature units (Figure 2.3B

and 2.3C centre and right). Without the RELU, a standard Hebbian learning rule would induce potentiation in the connections between losing conjunctive units and feature units clamped to 0 (which is below baseline $\beta^F = 0.5$); that is because both $(y^F - \beta^F)$ and $(y^C - \beta^C)$ in the learning rule (see equation 2.1) would be negative, yielding a positive multiplication $(y^F - \beta^F)(y^C - \beta^C)$.

After a sequence of trials, the proposed plasticity rule often leads the feature-to-conjunctive connectivity to one of the 6 possible target REMERGE architectures (Figure 2.4A). Converging to one of these target connectivities zeroes the distance to that connectivity, but not to the others (Figure 2.4B; see also supplementary Figure 2.6). The error is taken to be the distance to the nearest target (Methods). Across episodes, the connectivity errors converge to zero in most of the simulations (the connectivity reaches one of the target REMERGE architectures), but some fail to do so (Figure 2.4C).

To analyse the robustness of the learning rule to variation of parameters, I performed parameter sweeps varying several factors that influence the dynamics of the model. I varied initial connectivity, competition in the conjunctive layer by varying the temperature parameter (θ) and the hedge (κ) in the hedged softmax activation function, and speed of plasticity by varying plasticity time constant (τ_p) (Figure 2.4).

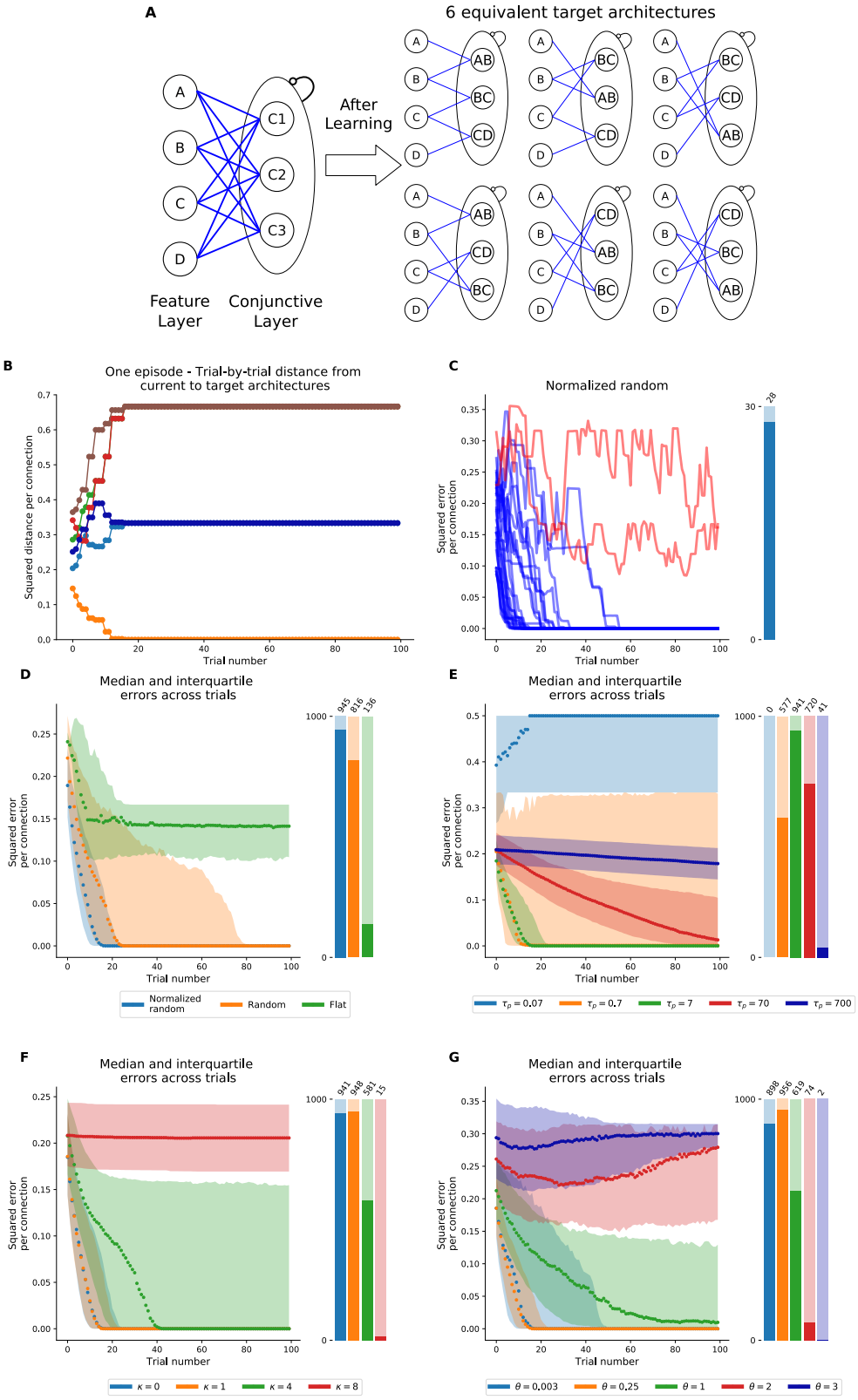

A
6 equivalent target architectures
A
B
C
D
C1
C2
C3
Feature Layer
Conjunctive Layer
After Learning
AB
BC
CD
B
One episode - Trial-by-trial distance from current to target architectures
Squared distance per connection
Trial number
C
Normalized random
Squared error per connection
Trial number
D
Median and interquartile errors across trials
Squared error per connection
Trial number
Normalized random
Random
Flat
E
Median and interquartile errors across trials
Squared error per connection
Trial number
$\tau_p = 0.07$
$\tau_p = 0.7$
$\tau_p = 7$
$\tau_p = 70$
$\tau_p = 700$
F
Median and interquartile errors across trials
Squared error per connection
Trial number
$\kappa = 0$
$\kappa = 1$
$\kappa = 4$
$\kappa = 8$
G
Median and interquartile errors across trials
Squared error per connection
Trial number
$\theta = 0.003$
$\theta = 0.25$
$\theta = 1$
$\theta = 2$
$\theta = 3$

Figure 2.4: *Learning performance and parameters robustness analysis via parameter sweep.* **A.** *The learning procedure can result in 6 different, but equivalent architectures that connect the feature pairs to conjunctive units in different permutations.* **B.** *The squared distance per connection (cf. equation 2.3) between the network's connectivity and each of the 6 possible target REMERGE architectures changes through trials in the episode. In a successful episode, the weights converge to one of the target architectures (the orange trace in this example reaches zero distance).* **C.** *Error curves show that not all episodes lead to successful REMERGE architecture learning. Error curves quantify distances to the target architecture (closest architecture in the end of the episode; cf. equation 2.5, Methods) across trials. In some episodes the learning rule successfully moulds the connectivity to close to one of the target architectures (blue error curves). Nevertheless, there are episodes in which the connectivity does not converge to one of the targets (red error curves). Randomness in the connectivity initialisation and in stimulus presentation sequence lead to variable results. Bar on the right of the plot show the number of episodes that converged (dark coloured bar on the base) and the number that did not converge (light coloured bar on the top) out of 30 randomly chosen episodes from 1000 simulations.* **D-G.** *Learning performance depends on weight matrix initialisations (***D***), plasticity time constant values τ_p (***E***), hedged-softmax parameter κ (***F***) and temperature parameter θ (***G***). Dots show median error at each trial and shaded region represent interquartile range. Convergence bars are as described for panel C above, but for 1000 simulations at each condition.*

The learning rule performance is sensitive to the initial connectivity. The network learns the desired architecture successfully when weights are initialised with normalised connections (see Methods section 2.1.3). This did not happen when the network was initialised with random weights with no normalisation, or with all the weights initialised to the same value (flat all-to-all initial connectivity; Figure 2.4D).

Normalisation affected learning because the activity of a conjunctive unit depends on the dot product between the firing rates in the feature layer $\boldsymbol{y^F} = (y^F_A, y^F_B, y^F_C, y^F_D)$ and the vector of incoming weights $\boldsymbol{w_c} = (w_{cA}, w_{cB}, w_{cC}, w_{cD})$ from all pre-synaptic feature units to conjunctive unit c:

$$\boldsymbol{w_c} \cdot \boldsymbol{y^F} \tag{2.6}$$

If, by chance of the random initialisation, the weight vector to one conjunctive unit is too small, then that unit always loses the competition and no plasticity happens so to associate it with feature unit pairs to form a conjunction.

The learning rule time constant τ_p also affects learning. Optimal τ_p is around 7s, both increasing and decreasing this value harms performance (Figure 2.4E). Hedge

κ parameter lower than 1 yields good learning performance, but it deteriorates as κ is increased further (Figure 2.4F). Temperature θ had best performance when set around 0.25. Decreasing or increasing it away from 0.25 impoverishes learning performance (Figure 2.4G).

It is necessary to conduct detailed study of the network and plasticity dynamics for each case, but solely from these results we can hypothesise the following causes for the decrease in learning performance when parameters are varied:

- When plasticity is too fast (small τ_p), the network probably keeps unlearning previously learned conjunctions. With each trial overriding learning of previous ones;

- Increasing the hedge κ diminishes activity in the conjunctive layer. This potentially thwarts conjunctive units from reaching activity above plasticity baseline. Below baseline of plasticity, the RELU non-linearity would impede plasticity and hamper learning the target conjunctive architecture;

- Decreasing the degree of competition (by increasing the magnitude of θ) may lead to more than one conjunctive unit reaching activities above plasticity baseline and thus disrupting the formation of a conjunction with only one conjunctive unit;

2.3 Discussion

The model results indicate that non-linear unsupervised learning in feature-to-conjunctive connections is sufficient to learn the conjunctive connections present in REMERGE, and thus support learning of generalisation in the hippocampal system. We can consider RELU-Hebbian a candidate plasticity rule the brain may use to wire up its connections while learning the task. Specifically, this learning rule associates jointly presented stimuli by connecting its responsive feature units to a single conjunctive unit. Another learning rule will be explored in the conjunctive-to-response layers to learn how to use the formed conjunctive representations to guide choices such that the attained reward is maximised.

An important difference in the stimulation protocol I used compared to the protocol used by Kumaran and McClelland (2012) is that the stimuli are clamped to the feature units. The clamp is used because REMERGE's recurrent connections allow activity to spread to feature units that were not initially stimulated, making standard Hebbian learning rules without clamp problematic. Activity in non-stimulated feature units results in strengthening undesired connections from these non-stimulated feature units to the active conjunctive unit that wins competition in the conjunctive layer. The network grows connections that are not present in the REMERGE model (Figure 2.1) and, thus, may perform worse in the task.

I thus analyse learning in this scenario in which the clamp is used to avoid this problem of strengthening connections to "non-stimulated-but-recurrently-activated" features. The clamp removes any recurrent influence the activity of conjunctive units could have in the feature units. Future research will need to investigate learning rules that are robust in the face of recurrent activity. Nevertheless, the brain may indeed present different processing modes. During active waking, the presence of Acetylcholine in the hippocampus reduces the connectivity between the hippocampus and cortex (as illustrated in Figure 1 of Hasselmo, 1999; see also discussions in Gais and Born, 2004; Lewis and Durrant, 2011), remaining with a circuit that would work very similarly to what is shown in the right side of Figure 2.2. On the other hand, when confronted with a test trial of a pair of stimuli never presented together before, it could well be the case that the animal's surprise would initiate a momentary switch to brain a processing mode commonly seen during quiet waking (and also during slow wave sleep). In this mode, there is low Acetylcholine and the recurrence of the circuit is re-established as in the right side of Figure 2.2. With recurrence, REMERGE's recurrent similarity computation is performed, enabling the generalisation capacity of the circuit.

Moreover, it should be noted that the use of a clamp during encoding trials opens way to ask if in generalisation trials the features should be clamped too. They should not. To study this, I will consider, for instance, a network with 6 feature units and 5 conjunctive units, similar to the configuration used in the REMERGE model study

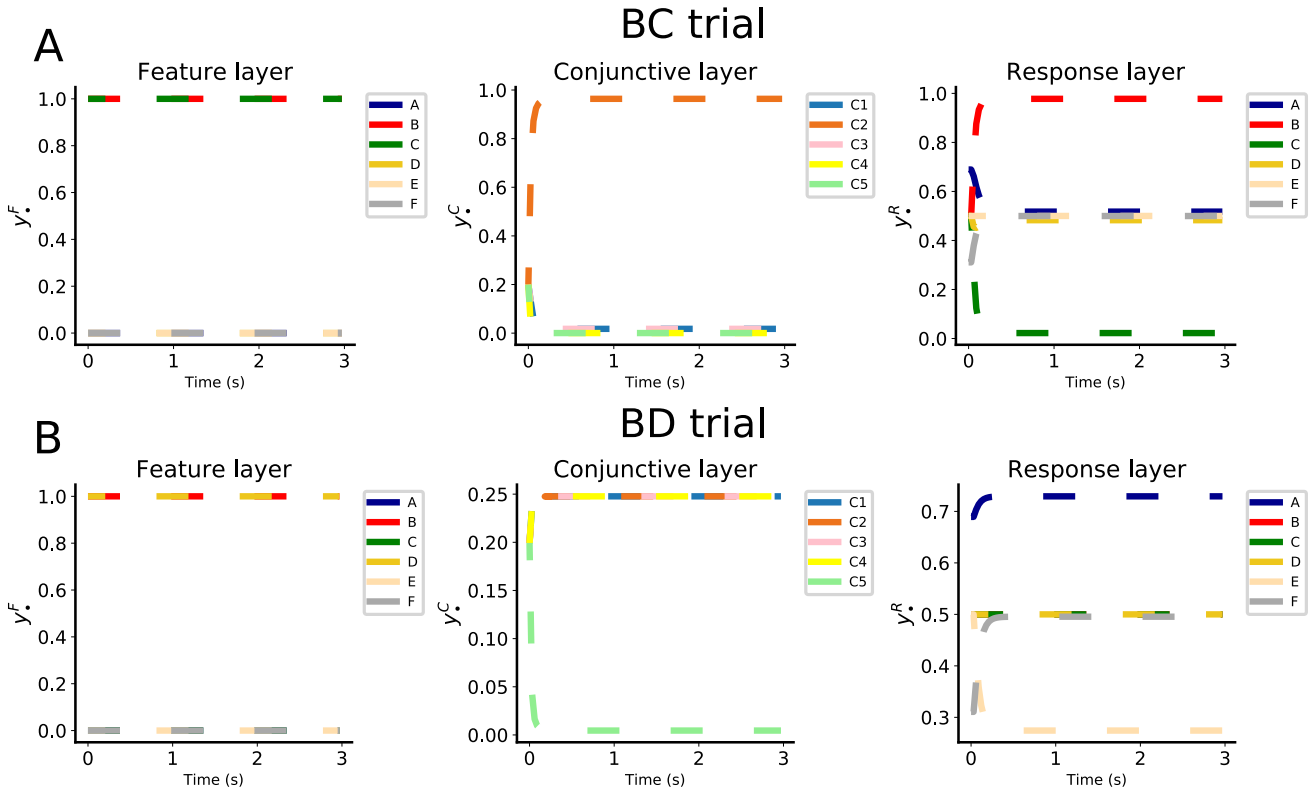


Figure 2.5: *Clamping stimuli to the feature layer impairs generalisation ability.* **A.** *Clamping features B and C to 1 and A and D to 0 does change the activity of response units compared to the external stimulation (as in Figure 1.2), which would affect the proportion of times the network would choose B over C. Nevertheless, B is still preferred over C.* **B.** *In a generalisation trial BD, clamping does harm accurate behaviour. Response activities of B and D are of the same level. Thus, the network does not demonstrate the generalisation bias towards B, as it should.*

(Kumaran and McClelland, 2012). Simulating a BD generalisation trial by clamping inputs to the feature units, instead of using inputs as external stimulation, the network shows no preference towards B nor D, failing to demonstrate generalisation capacity (Figure 2.5). The network does not make correct generalisation when the features are clamped, it needs the recurrent feature activation and the resulting graded activation in the conjunctive layer in order to bias responses towards the responses according to the ordered linear hierarchy introduced in section 1.4.3.

Perhaps during training trials upstream brain areas exert a stronger control over hippocampal inputs (represented by the feature units in my network), whereas in generalisation trials this neuronal activity control is relaxed.

Although the biological plausibility of this model is still limited, compared

to others in the literature it innovates by using biologically plausible plasticity and a neural network architecture. One previous study used backpropagation to train feedforward networks to solve transitive inference (De Lillo et al., 2001). But implementation of backpropagation in the brain is still a matter of ongoing research (Whittington and Bogacz, 2019). Another study explicitly associated each stimulus of the task with a position on a unit scale (from 0 to 1) and updated the position using task rewards (Jensen et al., 2015). How to implement such an algorithm in a neural network is unclear. Thus, although the model presented here entails a small neural network to represent the complex circuitry of hippocampal system, it is innovative for making use of neural networks and biologically plausible learning. A similar model applied to a working memory task has been able to reproduce a wide range of behavioural results (Manohar et al., 2019); further research with my model applied in other generalisation tasks may also yield interesting explanations and predictions.

One way this model can be tested experimentally is through representation similarity analysis (Kriegeskorte et al., 2008) during training trials. The comparison between the model and fMRI signals in the brain of human subjects can inform if the neural data show evidence of gradual formation of feature selectivity in Dentate Gyrus and CA3 regions, which is predicted from the formation of the conjunctive connections from feature units to conjunctive layer units.

Furthermore, in order to analyse the importance of the proposed learning rule to transitive inference, going beyond the formation of a REMERGE-like architecture for a small network, the number of stimuli used must be increased to at least 5 (cf. section 1.4). Nevertheless, the model architecture with 5 feature units is very similar to the one with 4. I evaluated the capacity for learning in this simpler architecture as a first step. Further research is necessary to investigate if the learning rule can also learn the architecture with more units (for a brief exploration of this model with 6 stimuli, see supplementary section 3.4).

A possible computational level interpretation of the feature-to-conjunctive learning rule is that it controls the formation of a synaptic representation that can later be used for recognition memory. A detailed comparison would be necessary, but

it could be an implementation of cognitive-level differentiation models of recognition memory (Criss and McClelland, 2006), which become generalisation-apt through recurrent activation of linked episodes.

Lastly, to validate the model, further research is needed to compare the results of simulations to neural data from experiments. Specially experiments that record neural activity during the learning phase of the task.

2.4 Supplementary figures

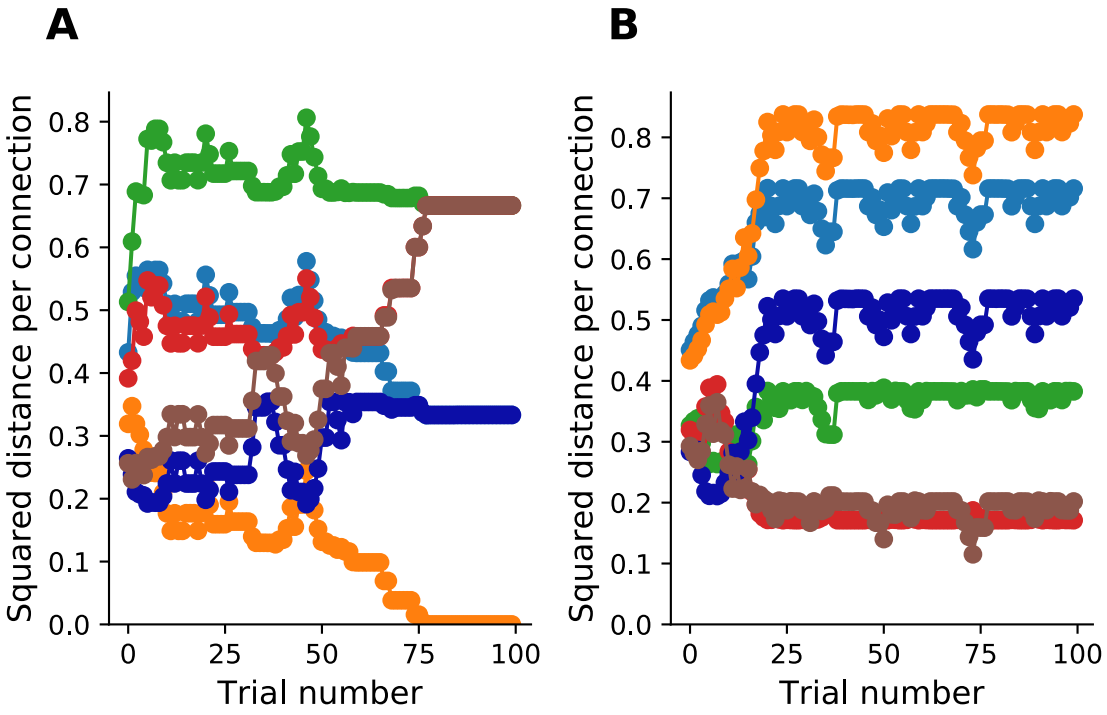


Figure 2.6: *Example episodes in which learning took longer or failed.* **A.** *Only after more than 70 trials the network finally converged to one of the 6 REMERGE architecture alternatives.* **B.** *The network did not converge to any of the REMERGE architectures in 100 trials.*

3

Learning to respond

Contents

3.1 Methods 33
3.1.1 Reward signal 34
3.1.2 Model-based reward complementarity through *switch* signal 34
3.1.3 Modulatory signal 35
3.1.4 Eligibility trace 35
3.1.5 Reward modulated learning rule 36
3.1.6 Integrating learning in all layers 36
3.1.7 Measuring behavioural accuracy 36
3.2 Results 38
3.2.1 Simpler model without switch signal 42
3.2.2 All plastic network 44
3.3 Discussion 47
3.4 Supplementary section: Learning rule performance in a network for 6 stimuli 48

In the previous chapter, I showed one approach to learn feature-to-conjunctive connectivity so that each conjunctive unit becomes responsive to a different pair of stimulated feature units. In this chapter, I address the second half of the learning process, learning connections that determine the behavioural responses: conjunctive-to-response weights.

After learning feature-to-conjunctive weights, conjunctive units code for the type of trial (AB, BC or CD) the network is exposed to when stimulated. Conjunctive-to-

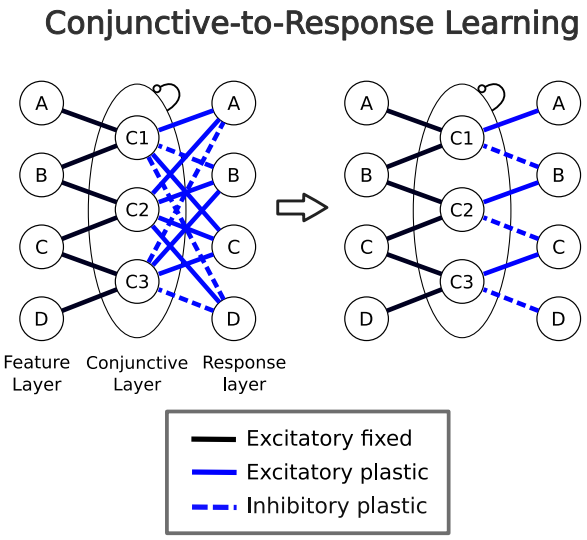


Figure 3.1: *State of the network in the beginning (left) and in the end (right)*

response weight learning will then associate each conjunctive unit, and its associated trial type, with the contextually-appropriate response.

In order to learn the connections to response units, it is necessary to use reinforcement signals from the task: the presence or lack of reward after each choice. These signals inform what are the correct choices. The learning rule then uses that information to change the weights to favour rewarded choices in the future.

Such a reward-modulated scheme is not implausible in the hippocampal circuits modelled by my network. Reinforcement signals are believed to be delivered to the hippocampus via dopaminergic projections from the ventral tegmental area and from the locus coeruleus (McNamara and Dupret, 2017).

Here, first I kept feature-to-conjunctive connections weights fixed at REMERGE's conjunctive architecture and learn conjunctive-to-response connections (Figure 3.1). Afterwards, I integrated the learning rule proposed here with the one proposed in chapter 2 to learn the full architecture simultaneously.

3.1 Methods

In order to learn how to make correct choices when presented with different pairs of stimuli at each trial, I built a plasticity rule that takes into account reward signals. These signals inform the weight change direction that favours rewarding choices.

The learning rule is designed such that the network learns to make the choices that lead to positive reward. Learning takes the network from an initial architecture in which conjunctive units have random (sampled from a uniform distribution $\mathcal{U}(-1, 1)$) all-to-all connections to response units, to a final state in which the pattern of connections is similar to the connections between these layers in the REMERGE model (Figure 3.1). Nevertheless, albeit similar to REMERGE, there might remain small differences in the strength of connections. These could remain due to continuous decay of connections (modelling forgetting) and if these alterations do not impact decisions, thus not affecting the reinforcement that would alter those connections.

As described in the introductory chapter 1, actions a are taken at each trial choosing stimuli among the ones presented in that trial, $\mathcal{P} = \{O_c, O_w\}$. O_c being the correct stimulus, after which reward is delivered, and O_w the wrong one. I will now describe the reinforcement process through the delivery of reward, and how the learning rule uses reinforcement to mould the connectivity.

3.1.1 Reward signal

The reward signal in a trial is modelled as positive one if the correct action is taken. Otherwise, it is modelled as negative one:

$$rwd = \begin{cases} 1, \text{if } a = O_c \\ -1, \text{otherwise (i.e. } a = O_w) \end{cases} \tag{3.1}$$

3.1.2 Model-based reward complementarity through *switch* signal

As previously noted, at each task trial, one choice is correct (O_c) and one is wrong (O_w). Generally an animal would learn to account for a mistake by not only counting against the chosen stimulus, but also in favour of the alternative. I.e. in an AB trial in which the network mistakenly chooses B ($a = B$) and received a negative reward ($rwd = -1$), the response unit B will be negatively affected and the response unit A will be positively affected. That is modelled by

introducing a *switch* signal to make sure that the response unit correspondent to the action taken will be positively influenced by the reward and that the response unit corresponding to the alternative response will be oppositely influenced by the reward. The switch is calculated as follows:

$$switch_r = \begin{cases} 1, \text{if } r = a \\ -1, \text{otherwise (i.e. } r \neq a) \end{cases} \tag{3.2}$$

Thus, we can define an effective reward:

$$\widehat{rwd}_r = rwd \times switch_r \tag{3.3}$$

In the real task, this complementarity is gradually learned across trials as the subject learns the structure of the task. I perform simulations both assuming it has already been learned from the start, and using a simpler model that include solely the influence of the reward to the response related with response action, with no switch signal.

3.1.3 Modulatory signal

To modulate the learning of the connection to the response units, I use a **reward-modulated learning rule**. It makes use of a modulatory signal defined by:

$$m_r(t) = \begin{cases} \widehat{rwd}_r, \text{if } r \in \mathcal{P} \\ 0, \text{otherwise} \end{cases} \tag{3.4}$$

or, when not using the switch signal, the modulation is non-zero only in the synapses onto the response unit r correspondent to the action taken a:

$$m_r(t) = \begin{cases} rwd, \text{if } r = a \\ 0, \text{otherwise} \end{cases} \tag{3.5}$$

3.1.4 Eligibility trace

I use an eligibility trace which modulates the plasticity rule using a Hebbian mechanism. Only units that are sufficiently co-active undergo plasticity (Gerstner et al., 2018):

$$\tau_{et} \frac{de_{rc}}{dt} = -e_{rc} + H(y_r^R, y_c^C) = -e_{rc} + \alpha_{et}\, y_r^R \left[y_c^C - \beta^C\right]_+ \tag{3.6}$$

The time constant of the eligibility trace $\tau_{et} = 400$ ms, bigger than the time constant of the neuron $\tau_n = 70$ ms (equation 1.1), so that the trace does not decay before the reward is received. $\alpha_{et} = 5$ is a parameter that regulates the size of the eligibility trace (it was set at $\alpha_{et} = 15$ for the simulation without switch signals).

3.1.5 Reward modulated learning rule

The resulting plasticity rule for conjunctive-to-response units obeys the following dynamics:

$$\tau_p \frac{dw_{rc}}{dt} = m_r(t)\, e_{rc}(t) - \frac{\tau_p}{\tau_d} w_{rc} \tag{3.7}$$

w_{rc} is the connection weight between conjunctive unit c and response unit r. The term $-\frac{\tau_p}{\tau_d} w_{rc}$ on the right implements a slow weight decay. $\tau_d = 300$ s.

3.1.6 Integrating learning in all layers

I simulate the system simultaneously enabling the unsupervised learning rule developed in chapter 2 for feature-to-conjunctive connections and the reward-modulated learning rule developed in this chapter for conjunctive-to-response connections.

Both sets of weights start with all-to-all configurations. Feature-to-conjunctive connections are bounded in $[0, 1]$ and conjunctive-to-response in $[-1, 1]$ (Figure 3.7A).

3.1.7 Measuring behavioural accuracy

The accuracy of the choices made by the network was measured by counting how often it made the correct choice ($a = O_C$). A running average of the proportion of correct decisions taken was calculated using a Gaussian filter (SD=10) to smooth the decisions across trials (see, for example, Figure 3.2C).

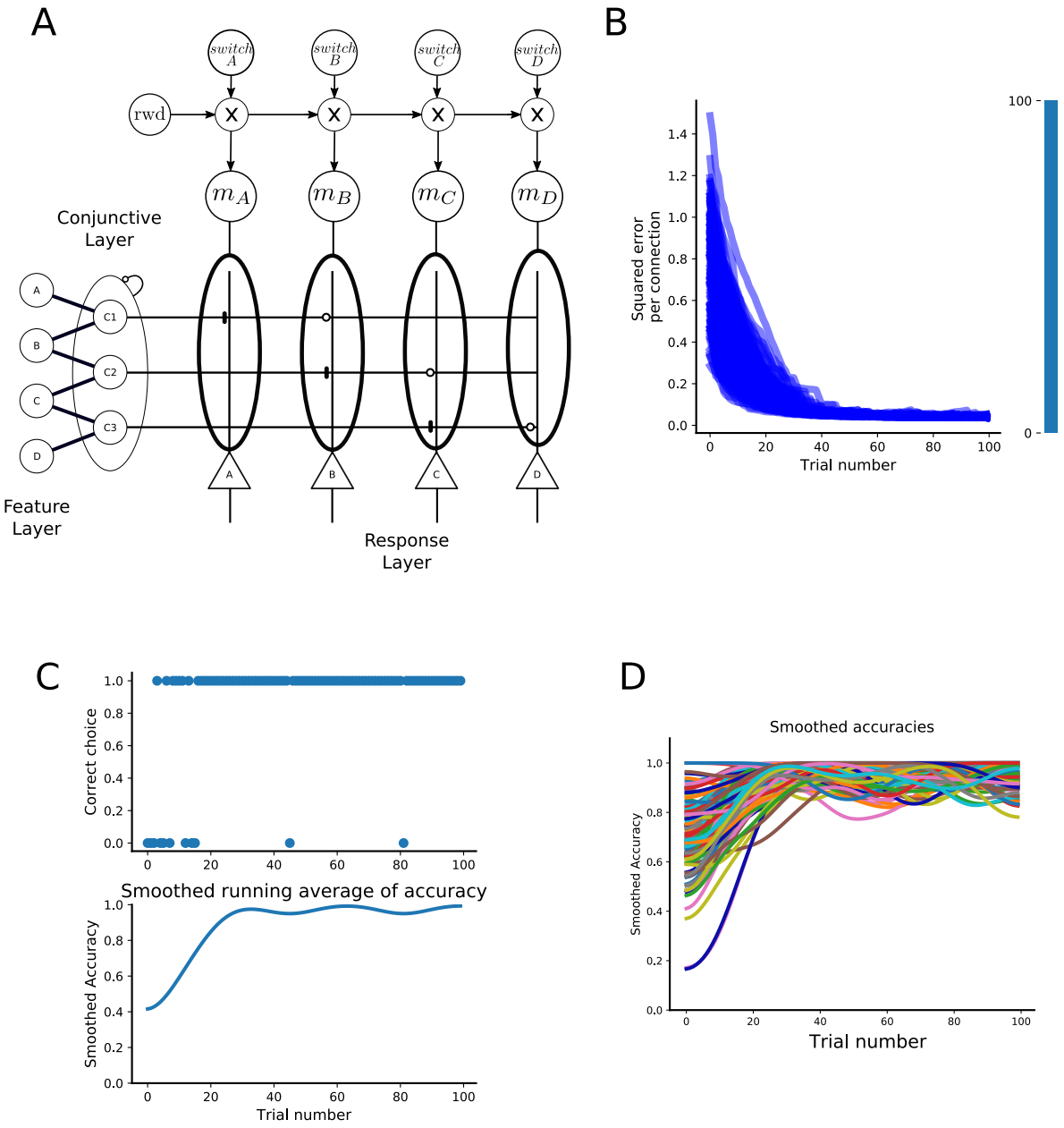


Figure 3.2: Model-based mechanism of reward modulated learning rule **A.** *Different representation for the REMERGE circuit shows synaptic connections as line crossings. The circuit now includes modulatory units that bias learning taking into account task reward. The modulatory signals are calculated by multiplying a global reward signal by a switch signal (Methods section 3.1.2).* **B.** *Error curves show that the learning rule successfully builds the desired conjunctive-to-response connectivity in all simulated episodes out of 100 ones.* **C.** *Quality of network choices (correct, 1, incorrect, 0) across trials (top) and smoothed average of rewards (accuracy curve; bottom). The network learns the desired architecture and eventual mistakes happen (e.g. around trial number 80) because of the probabilistic way the choices are made (Luce choice rule; Methods).* **D.** *Accuracy curves for all episodes show that the plasticity rule manages to achieve good accuracy in the task.*

3.2 Results

In order to learn how to make the correct choices, the network must make use of the choice-contingent reward signals it receives. The learning rule developed here constructs conjunctive-to-response weights through reward modulation of a Hebbian rule (Methods).

Contrary to the flexibility of binding between feature and conjunctive units, where 6 possible equivalent architectures could be learned (Figure 2.4A), conjunctive-to-response weights have to converge to a single target connectivity pattern adequate to the learned feature-to-conjunctive weights. The target is such that the conjunction that binds features A and B must have an excitatory connection to response unit A and an inhibitory connection with B, and analogously for the other two conjunctions (Figure 3.1; see also Figure 3.7 for all possibilities for different feature-to-conjunctive connectivities).

The learning rule for conjunctive-to-response connections makes use of neurons that individually modulate the connections (Figure 3.2A and Methods section 3.1.3; note eligibility trace e_{rc}'s dependence on r and c). Furthermore, it has a weight decay to ensure that connections that are not affected by rewards are forgotten (Figure 3.4 and Methods section 3.1.5).

The direction of plasticity (positive or negative) is controlled by switch signals and by the reward. The switches model the complementarity of action options: at each trial, one action is correct and leads to reward ($rwd = 1$) if chosen, and the other is wrong and leads to a delay period ($rwd = -1$). The manner the switch takes the complementarity into account is by controlling if the reward is to be taken as it is, or as its opposite. Thus, it is possible to simultaneously learn (at each trial) about the option taken and its complementary, instead of having to learn about them individually in different trials. For example, suppose an AB trial in which the network chooses the wrong action ($a = B$) and receives negative reward ($rwd = -1$). In this case the connection from the winning conjunction ($c = \underset{c}{\operatorname{argmax}}\, y_c^C$) to response unit B is depressed (negative effective reward; $\widehat{rwd} = rwd \times switch_B = -1 \times 1 = -1$)

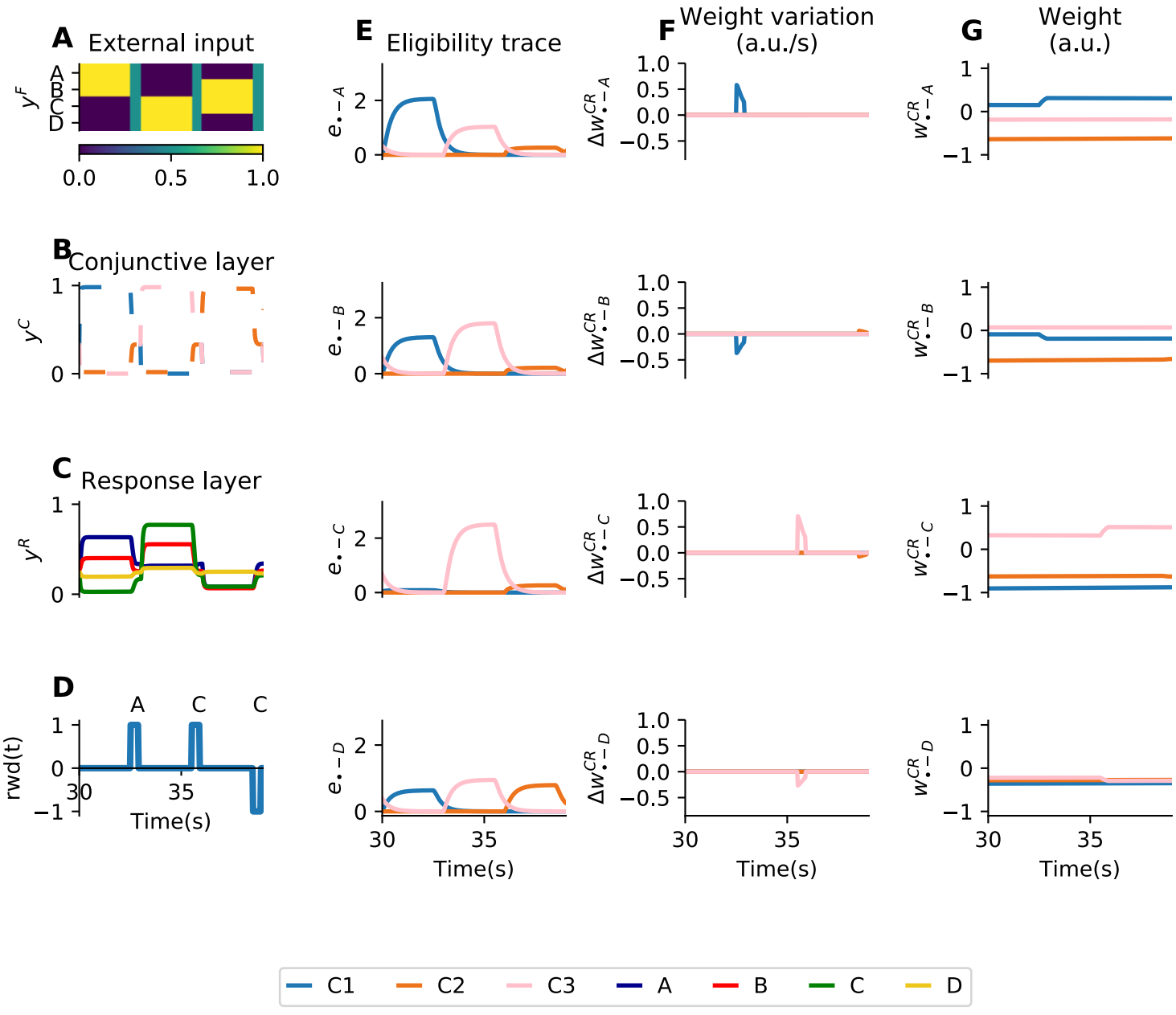


Figure 3.3: *Intra-trial dynamics of reward modulated plasticity for conjunctive-to-response learning.* **A.** *External input clamped to the feature layer.* **B.** *Conjunctive unit activities.* **C.** *Response unit activities.* **D.** *Reward signals delivered during the inter-trial intervals. The action choices made by the network are depicted with letters above the plot.* **E.** *Eligibility traces of all 12 possible conjunctive-to-response connections separated in groups by their postsynaptic targets A, B, C and D.* **F.** *Weight variation that results from the learning rule dynamics when the circuit undergoes the dynamics shown on this figure.* **G.** *Absolute values of connection weights. The small variations in weights can be seen in the curve. When weights are not affected by rewards, they slowly decay to zero.*

and the connection to response unit A is potentiated (positive effective reward; $\widehat{rwd} = rwd \times switch_A = -1 \times -1 = 1$).

Simulating the network through the TI protocol and enabling the designed reward-modulated plasticity rule, it always learns the target architecture, regardless of the initial conjunctive-to-response weights (Figure 3.2B). Across trials in an episode the proportion of correct choices increases as the network connectivity approaches REMERGE connections (Figure 3.2C and D).

To understand how the learning rule builds the desired connectivity, it is insightful to look at the activities of the network units, the reward signals and the eligibility traces, during two example trials, AB and BC. In trial AB, A and B are stimulated during the first 2.5s and then clamped to baseline for 0.5s before trial BC starts (Figure 3.3A). Because feature-to-conjunctive connections are fixed at REMERGE's values, one conjunctive unit is highly active (C1), while the others are nearly at zero activity (Figure 3.3B). At the response layer the activity of unit A is bigger than all the others, which makes the probability of choosing A high (Figure 3.3C; see Luce choice rule mentioned in section 1.4.3, Luce, 1959). In the AB trial the network chose the correct choice, A, and gets a positive reward. In the BC trial the network chose option C, which is the wrong choice, leading to the receipt of a negative reward (Figure 3.3D).

The eligibility traces determine which connections are eligible to be modified by the learning rule. On AB trial, the low activities of conjunctive units C2 and C3 make their eligibility traces low, and thus not plastic. The connections from C1 are the only ones whose eligibility traces can be higher, because of C1's high activity. Still, the eligibility trace for C1-C is low because of response unit C's low activity. Connection C1-A, on the other hand, has a especially high eligibility trace due to the high activity of both C1 and response unit A (Figure 3.3E). The weight variations are such that the excitatory connection from C1 to response unit A is strengthened, and the inhibitory connection from C1 to response unit B is made more negative (Figure 3.3F and G). The opposite direction of plasticity happens because of the switch (not shown in Figure 3.3, see Methods) that changes the

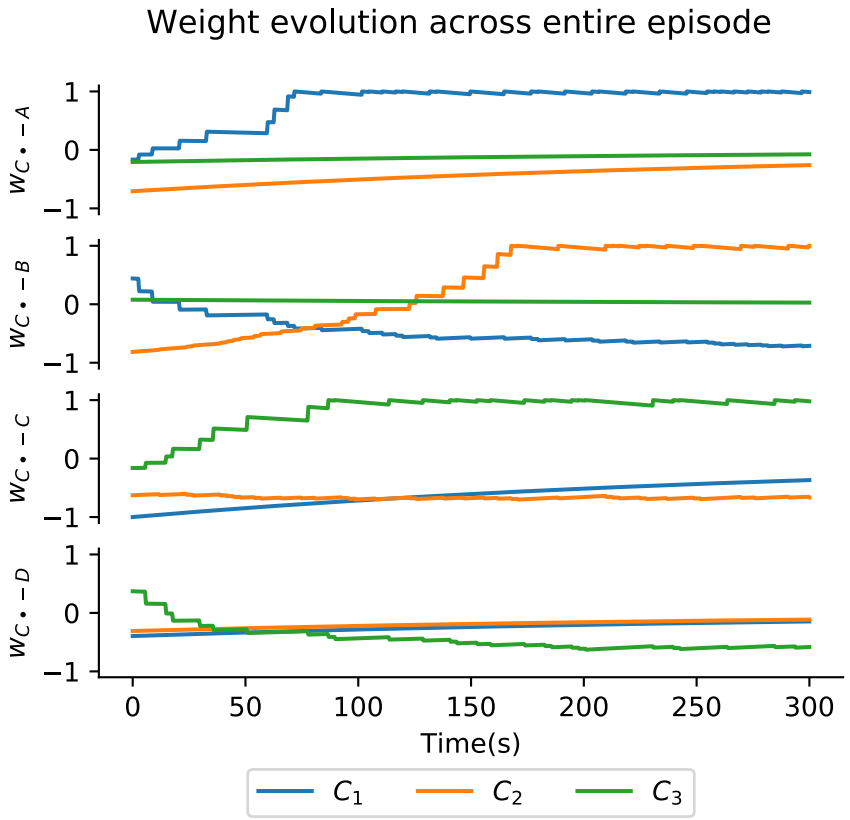


Figure 3.4: *Evolution of conjunctive-to-response weights across an episode. Reward-modulated weights are kept near +1 or -1, while weights not affected by modulation decay to zero. Note that the negative weights do not reach the boundary, differently to the positive ones. That is because of lower eligibility traces for these connections due to low activity levels of their corresponding response units.*

influence of the reward signal at each connection. Furthermore, notice that despite a similar eligibility trace to C1-B, C1-D does not elicit a change in weight because the modulation signal is specific to the connections to the response units that are correspondent to the stimulated features, and zero otherwise (Methods).

Moreover, it is worth noting that in the BC trial, despite the mistake (choosing C, instead of B), connections C2-B and C2-C are modified in the intended direction (C2-B positive variation, and C2-C negative variation; Figure 3.3F).

Finally, the weights that are not affected by modulation eventually decay to zero (Figure 3.4).

3.2.1 Simpler model without switch signal

Given that the switch mechanism used in the previous section is not a biologically plausible implementation of the model-based learning that captures complementarities of the stimulus-reward contingencies, I simulated a simpler model without using the switch. In this case, the modulatory units affect only connections to the response unit that represent the action choice taken (see Equation 3.5 in Methods).

For instance in an AB trial, the network learns about A, only if A is chosen, and not when its complementary pair B is chosen, and vice-versa; similarly for other units in other trial types (see in Figure 3.5F that only connections to one of the response units are affected in each trial). As a demonstration of this, the last two trials shown in Figure 3.5 are BC trials, and while in the first of them the network chose B and only modulated connection C2-B, in the second it chose C and only modulated connection C2-C.

When using this simpler model, the accuracy and error results are similar to what was achieved using switch signals (Figure 3.6). A noteworthy difference was that in this case the weight decay has more time to affect the connectivity. To stay close to REMERGE's conjunctive-to-response connectivity, I used a bigger α_{et} (set at 15 for the result presented) and maintained the already big τ_w (slow decay) used in previous simulations. Simulations using the switches were robust to smaller τ_w's because the frequency of modulation was much bigger as at every trial connections to two, instead of only one, response units were affected by reward modulation. The frequency of reward modulation for the model without switch is specially low for inhibitory connections, because the inhibition causes the activity of the postsynaptic response unit to be low, yielding low probability of choosing that action and consequently of modulating the associated conjunctive-to-response connection (Figure 3.5 has an example of that in the last shown trial. Despite low y_C^R the network chose action C and received a negative reward which decrease C2-C weight).

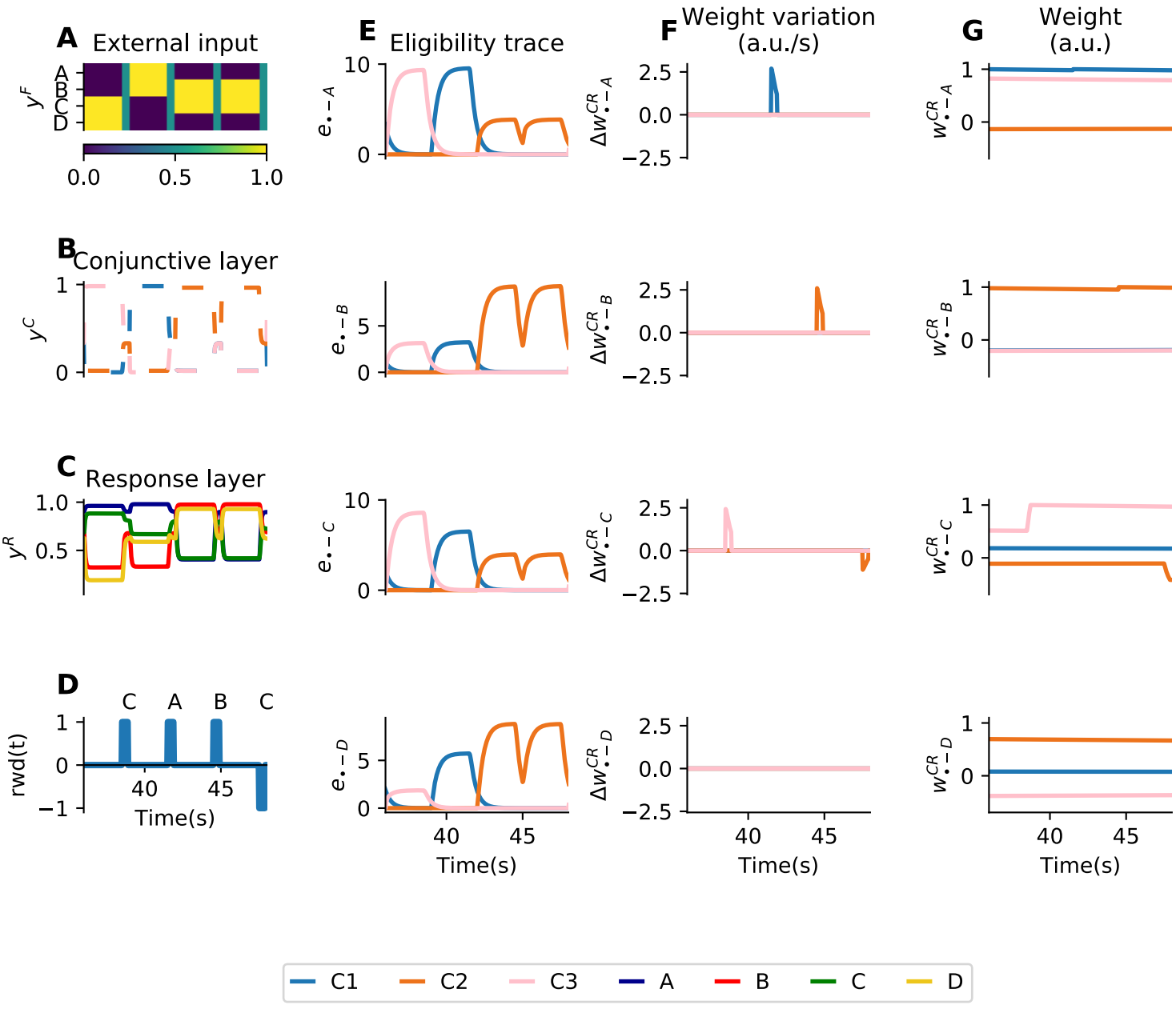


Figure 3.5: *When learning conjunctive-to-response weights without the switch signals, at each trial the network only adjust connections to one response unit. Panels* **A-G** *are analogous to the ones shown in Figure 3.3. Note that weight variations in panel* **F** *now happen only to one response unit at a time and affect only the connection weights to response units of the chosen stimulus.*

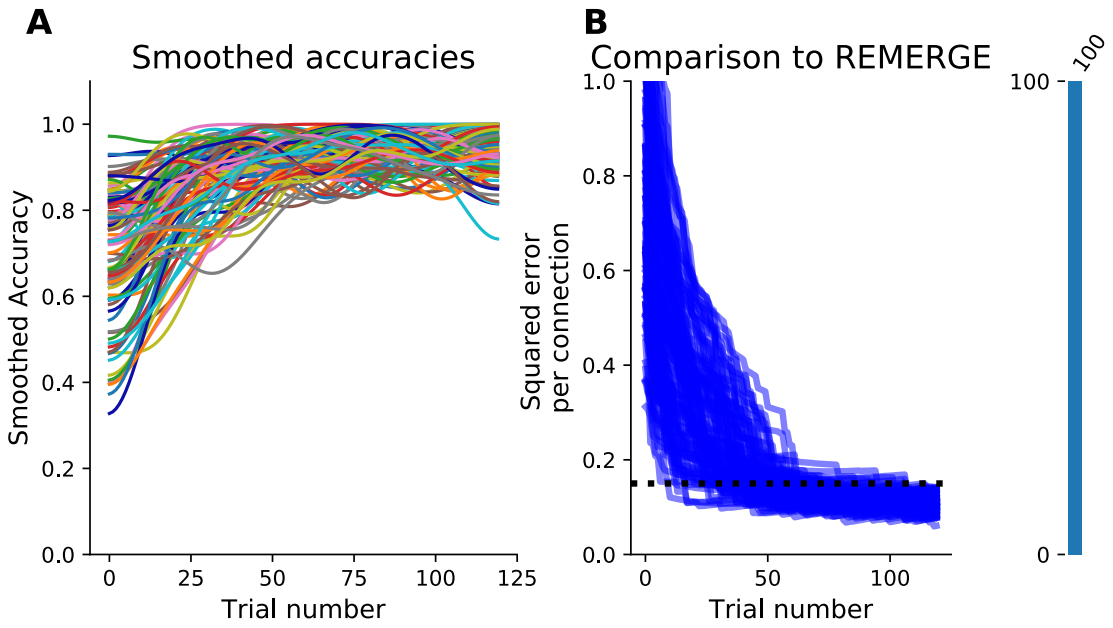


Figure 3.6: *Successful conjunctive-to-response learning without switch signal.* **A.** *Accuracy curves of 100 simulations in which the network did not use switch signals show that the accuracy reaches at least 80% correct choices.* **B.** *The connectivity of the network approaches REMERGE conjunctive-to-response connectivity. The dashed line represents the convergence threshold for conjunctive-to-response connections, which was set at 0.15.*

3.2.2 All plastic network

During the training phase of the TI experiment, there is one unsupervised learning process happening in the feature-to-conjunctive layer to bind and associate together simultaneously presented stimuli, and a supervised learning process that grasps the advantageous responses. These processes take place simultaneously, notwithstanding their mutual dependence. To emulate that, I enabled plasticity both in feature-to-conjunctive connections, using the plasticity rule discussed in Chapter 2, and in conjunctive-to-response connections using the reward-modulated plasticity rule introduced in this chapter (Figure 3.7A).

The plasticity rules are capable of acting together simultaneously in both layers to learn the REMERGE connectivity. In the majority of cases the network successfully learns the REMERGE connectivity (Figure 3.7B). It is necessary to have convergence to one of the 6 equivalent feature-to-conjunctive architectures to the determine the unique target connectivity of conjunctive-to-response connections.

Thus, when the feature-to-conjunctive connections are not properly learned, the conjunctive-to-response connections cannot be learned either (Figure 3.7C). But in cases in which feature-to-conjunctive weights are learned successfully, the network manages to achieve good behavioural performance, choosing the correct stimuli and being rewarded most of the time (Figure 3.7D).

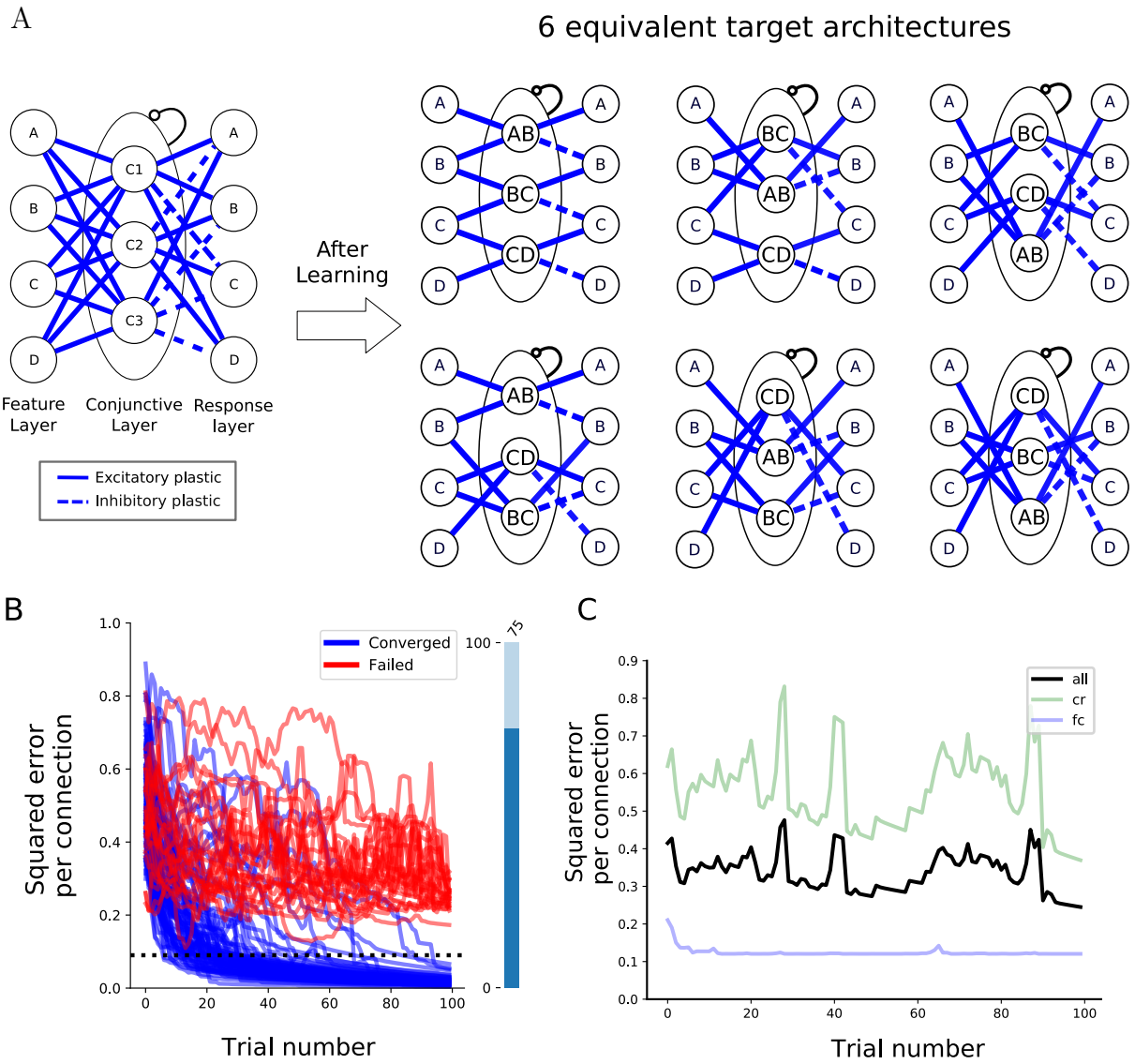

A
6 equivalent target architectures
Feature Layer
Conjunctive Layer
Response layer
After Learning
Excitatory plastic
Inhibitory plastic
B
Converged
Failed
Squared error per connection
Trial number
C
all
cr
fc
Squared error per connection
Trial number
D
Smoothed Accuracy
Trial number
Smoothed Accuracy
Trial number

Figure 3.7: *Plasticity in all layers simultaneously.* **A.** *Both feature-to-conjunctive and conjunctive-to-response connections are made plastic using the corresponding learning rules for each layer. After training the architecture reaches one of the equivalent REMERGE connectivities. For each feature-to-conjunctive connectivity pattern, there is one possible conjunctive-to-response.* **B.** *Error curves of 100 simulated episodes with full plasticity. Curves were coloured in blue for episodes in which the error ended below the convergence threshold and in red for the ones that the network failed to converge. Dashed line is the convergence threshold. Most of the error curves converge to values near zero, showing effective learning.* **C.** *Example of an episode in which the learning of feature-to-conjunctive (labelled 'fc') connections did not converge. Without convergence of the input weights to a conjunctive architecture, conjunctive-to-response ('cr') connections cannot be learned either. Error curves for feature-to-conjunctive connections are depicted in blue, conjunctive-to-responsive connections in green and the error for all connections in black ('all').* **D.** *Smoothed accuracies of behavioural choices across episodes. On the left in blue are episodes that were below a convergence threshold in panel B. On the right in red the ones that were above the threshold (did not converge).*

3.3 Discussion

In order for subjects to learn the correct actions in a transitive inference task, they rely on the rewards that accompany decisions. Rewards and reward prediction errors have a neurophysiological counterpart in the activity of some dopaminergic neurons (Schultz et al., 1997). The dopamine these neurons release can influence the plasticity of neuronal circuits so as to guide actions towards attaining rewards (Gerstner et al., 2018). Similarly, the reward-modulated learning rules designed in this chapter learned the intended neuronal architecture, and left the network making the correct choices (frequently receiving rewards).

To achieve that, a special challenge was faced to simultaneously learn feature-to-conjunctive and conjunctive-to-response weights. Enabling both plasticity rules to learn together required a flexible conjunctive-to-response learning rule; one that could adapt to the learned feature-to-conjunctive architecture. At each trial the conjunctive-to-response rule had to change connections from the highly active conjunctive unit (the layerwise competition "winner", whichever unit it were), to the pair of response units correspondent to the current trial type, A and B, B and C, or C and D (in the case with the switch; Figure 3.7), or solely to the response unit correspondent to the choice (in the case without switch; which was

not tested with plasticity enabled in both layers). The rule potentiated connections that favoured choosing correct responses, and depressed connections that favoured incorrect ones. The role of the eligibility trace in the learning rule was to make sure that the connections from the winning conjunctive unit were made eligible, i.e. plastic, and while other connections remained fixed.

The parameter values used in the integrated learning simulation were the same as the ones used when feature-to-conjunctive connections are fixed. It makes sense that the parameters used for the integrated case also work for the case with fixed feature-to-conjunctive connections, as it is a special, simpler case of when the weights are plastic but have already converged to the final feature-to-conjunctive architecture.

Moreover, throughout this chapter, I used two metrics for evaluating the plasticity rules: the error curves and the accuracy curves. The analysis of those metrics showed that, quite often, despite not being exactly at a REMERGE architecture (non-zero error), the accuracy of the model was still better than chance (50%) and comparable to values that humans attain in experiments (Kumaran and McClelland, 2012). A more extensive analysis of the relationship between different connectivity patterns and behavioural accuracy would help clarifying if a neuronal representation error metric is suitable and whether any metric different than the squared distance would be more appropriate.

In sum, the plasticity rules proposed in this chapter successfully took advantage of the reward signals present in the task to build the response associations necessary for performing the task. Furthermore, these learning rules integrate well with the unsupervised learning rules proposed in chapter 2, making it possible to learn the whole architecture simultaneously, as is probably the case in the brain.

3.4 Supplementary section: Learning rule performance in a network for 6 stimuli

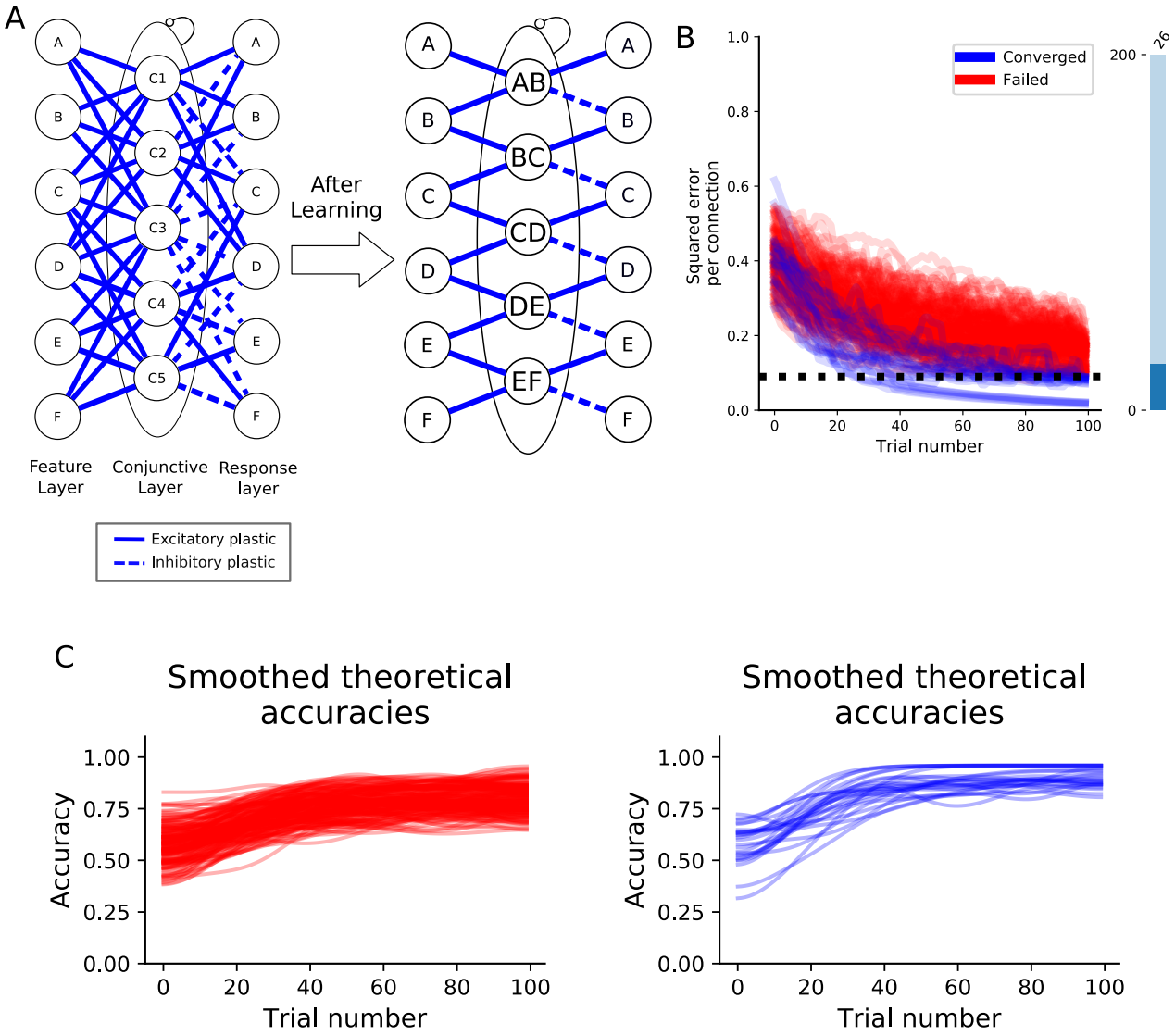


Figure 3.8: *Testing designed plasticity rules to learn 6-feature REMERGE connectivity.* **A.** *REMERGE architecture for 6 features before and after successful learning. Both feature-to-conjunctive and conjunctive-to-response connections are made plastic using the non-linear Hebbian rule for feature-to-conjunctive connections and reward-modulated switch rule for the conjunctive-to-response connections. After training the architecture reaches one of the equivalent REMERGE connectivities (in this case* $5! = 120$ *options).* **B.** *Only 26 of 200 simulated episodes reached error below convergence threshold at the end of the episode. Of these, only 6 actually reached errors that truly appeared to converge to zero. Indicating that the convergence threshold may depend on the number of features, and would need to be changed in this case. The low proportion of curves that converge might indicate that the network needs more trials to learn than with 4 stimuli, which is also expected for animals.* **C.** *Smoothed accuracies of behavioural choices across episodes separated by convergence (red - failed to converge; blue - converged). Episodes that converged show higher behavioural accuracy than the ones that did not converge, but all of them improve accuracy over the course the episode.*

4
Discussion

This work proposes biologically plausible learning rules capable of building RE-MERGE's connectivity (Kumaran and McClelland, 2012), whose dynamics is hypothesised to be responsible for generalisation in the hippocampal system. Two main plasticity rules were proposed, one unsupervised and one reward-modulated. The unsupervised rule forms associations between stimuli that are repeatedly presented together, and reward-modulated rule makes the network learn how to respond by choosing rewarding actions. Combined, they perform unsupervised learning in feature-to-conjunctive layer and reward-modulated Hebbian learning in conjunctive-to-response layer.

Experimental work is necessary to confirm the prediction that the proposed plasticity rules are the mechanism the brain uses to wire up its connections. An approach to experimentally test the proposed model is through representation similarity analysis (Kriegeskorte et al., 2008). The analysis would compare the activity of units in the conjunctive layers during training trials to, for instance, fMRI BOLD signal in the dentate gyrus hippocampal region. Across training, the change in feature-to-conjunctive connections should be reflected as change in functional connectivity between the entorhinal cortex and hippocampal DG neurons. The changes in BOLD signal in the DG should be similar to the changes in the activity of conjunctive units observed across training.

Furthermore, to check if the brain uses some kind of non-linear plasticity similar to the RELU-Hebbian proposed here, one can check for example if there is a threshold activity in the hippocampus (also probably in the DG) over which plasticity is enabled.

Some outstanding questions remain. Are these plasticity rules plausible? The architecture is strikingly simplified when compared to the actual hippocampal circuit. How could this be reasonable? How to relate such a learning rule to learning rules in a cellular level?

In a study of working memory, Manohar et al. (2019) has comparable network biological realism (few units) and plasticity rule (standard Hebbian). The plasticity yields a mechanism through which working memory can be stored both in delayed neural activity and in synaptic connections. Similarly, here I showed how my designed plasticity rules can solve the transitive inference task by building the REMERGE architecture.

In preliminary simulations that still need further analysis, the plasticity rules presented here are also successful in learning the architecture with 5 feature units. This is an important future research direction, because the number of modelled feature units needs to be increased to at least 5 so that the model is representing a task that effectively requires transitive inference capacity, rather than simple value learning. With 4 feature units the agent does not need to rely on the underlying stimulus hierarchy of solve the task, but could rather rely on the fact that the some stimuli are either never rewarded (D, as labelled here) or always rewarded (A). The reward history itself would be enough to determine the correct response.

In conclusion, I have achieved the goal of creating learning rules capable of building the connectivity proposed by Kumaran and McClelland (2012), leaving the network capable of performing the transitive inference task successfully.

Appendices

A
Parameter values

Table A.1: Simulation parameter summary of the model

Name	Value	Description
τ_n	70 ms	Time constant of rate neuron
κ	1	Hedge
θ	0.25	Temperature
β	0.3	Luce choice rule equivocation parameter
Stimulation duration	2.5 s	
Interval	0.5 s	
Trial duration	3 s	Stimulation + Interval
Reward duration	0.4 s	
τ_p	3.5 s	Plasticity time constant
β^F	0.5	Baseline in plasticity for feature layer units
β^C	0.33	Baseline in plasticity for conjunctive layer units
α_{et}(switch)	5	Size of eligibility trace
α_{et}(no switch)	15	Size of eligibility trace in simulation without switches
τ_{et}	400 ms	Time constant of eligibility trace
τ_d	300 s	Weight decay time constant
Θ^{FC}	0.09	Threshold for considering that the connectivity converged to REMERGE feature-to-conjunctive architecture
Θ^{CR}	0.15	Threshold for considering that the connectivity converged to REMERGE conjunctive-to-response architecture

Bibliography

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

Ashby F.G., and Maddox W.T. (2011). Human category learning 2.0. Annals of the New York Academy of Sciences *1224*, 147.

Bevins R.A., and Besheer J. (2006). Object recognition in rats and mice: a one-trial non-matching-to-sample learning task to study'recognition memory'. Nature protocols *1*, 1306.

Brunamonti E., Mione V., Di Bello F., Pani P., Genovesio A., and Ferraina S. (2016). Neuronal modulation in the prefrontal cortex in a transitive inference task: evidence of neuronal correlates of mental schema management. Journal of Neuroscience *36*, 1223–1236.

Buckmaster C.A., Eichenbaum H., Amaral D.G., Suzuki W.A., and Rapp P.R. (2004). Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. Journal of Neuroscience *24*, 9811–9825.

Bunsey M., and Eichenbaum H. (1996). Conservation of hippocampal memory function in rats and humans. Nature *379*, 255–257.

Cohen N.J., and Eichenbaum H. (1993). Memory, amnesia, and the hippocampal system (MIT press).

Coutureau E., Killcross A.S., Good M., Marshall V., Ward-Robinson J., and Honey R.C. (2002). Acquired equivalence and distinctiveness of cues: Ii. neural manipulations and their implications. Journal of Experimental Psychology: Animal Behavior Processes *28*, 388.

Criss A.H., and McClelland J.L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). Journal of Memory and Language *55*, 447–460.

Dayan P., and Abbott L.F. (2001). Theoretical neuroscience, vol. 806 (Cambridge, MA: MIT Press).

De Lillo C., Floreano D., and Antinucci F. (2001). Transitive choices by a simple, fully connected, backpropagation neural network: implications for the comparative study of transitive inference. Animal Cognition *4*, 61–68.

Dusek J.A., and Eichenbaum H. (1997). The hippocampus and memory for orderly stimulus relations. Proceedings of the National Academy of Sciences *94*, 7109–7114.

Eichenbaum H. (1999). The hippocampus and mechanisms of declarative memory. Behavioural brain research *103*, 123–133.

Eichenbaum H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. Neuron *44*, 109 – 120.

Eichenbaum H., Dudchenko P., Wood E., Shapiro M., and Tanila H. (1999). The hippocampus, memory, and place cells: Is it spatial memory or a memory space? Neuron *23*, 209–226.

Gais S., and Born J. (2004). Low acetylcholine during slow-wave sleep is critical for declarative memory consolidation. Proceedings of the National Academy of Sciences *101*, 2140–2144.

Gerstner W., Lehmann M., Liakoni V., Corneil D., and Brea J. (2018). Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. Frontiers in neural circuits *12*.

Hasselmo M.E. (1999). Neuromodulation: acetylcholine and memory consolidation. Trends in cognitive sciences *3*, 351–359.

Jensen G., Muñoz F., Alkan Y., Ferrera V.P., and Terrace H.S. (2015). Implicit value updating explains transitive inference performance: The betasort model. PLoS computational biology *11*, e1004523.

Koster R., Chadwick M.J., Chen Y., Berron D., Banino A., Düzel E., Hassabis D., and Kumaran D. (2018). Big-loop recurrence within the hippocampal system supports integration of information across episodes. Neuron *99*, 1342 – 1354.e6.

Kriegeskorte N., Mur M., and Bandettini P.A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. Frontiers in systems neuroscience *2*, 4.

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., and McClelland J.L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. Psychological review *119*, 573.

Lewis P.A., and Durrant S.J. (2011). Overlapping memory replay during sleep builds cognitive schemata. Trends in cognitive sciences *15*, 343–351.

Logothetis N.K., Pauls J., and Poggio T. (1995). Shape representation in the inferior temporal cortex of monkeys. Current Biology *5*, 552–563.

Luce R. (1959). Individual Choice Behavior: A theoretical analysis (New York, NY: Wiley).

Manohar S.G., Zokaei N., Fallon S.J., Vogels T., and Husain M. (2019). Neural mechanisms of attending to items in working memory. Neuroscience & Biobehavioral Reviews .

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.

McNamara C.G., and Dupret D. (2017). Two sources of dopamine for the hippocampus. Trends in neurosciences *40*, 383–384.

Moses S.N., Villate C., and Ryan J.D. (2006). An investigation of learning strategy supporting transitive inference performance in humans compared to other species. Neuropsychologia *44*, 1370–1387.

Ohl F.W., Scheich H., and Freeman W.J. (2001). Change in pattern of ongoing cortical activity with auditory category learning. Nature *412*, 733–736.

Orlov T., Yakovlev V., Hochstein S., and Zohary E. (2000). Macaque monkeys categorize images by their ordinal number. Nature *404*, 77.

Piaget J. (1928). Judgement and reasoning in the child (London: Kegan, Paul, Trench & Trubner).

Ryan J.D., Moses S.N., and Villate C. (2009). Impaired relational organization of propositions, but intact transitive inference, in aging: Implications for understanding underlying neural integrity. Neuropsychologia *47*, 338–353.

Schultz W., Dayan P., and Montague P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Shohamy D., and Wagner A.D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. Neuron *60*, 378–389.

Urcuioli P.J. (1996). 4 acquired equivalences and mediated generalization in pigeon's matching-to-sample. In Advances in Psychology (Elsevier), vol. 117, pp. 55–70.

Van Strien N.M., Cappaert N., and Witter M.P. (2009). The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network. Nature Reviews Neuroscience *10*, 272.

Wetzel W., Wagner T., Ohl F., and Scheich H. (1998). Categorical discrimination of direction in frequency-modulated tones by mongolian gerbils. Behavioural brain research *91*, 29–39.

Whittington J.C., and Bogacz R. (2019). Theories of error back-propagation in the brain. Trends in Cognitive Sciences *23*, 235 – 250.

Wixted J.T., and Squire L.R. (2010). The role of the human hippocampus in familiarity-based and recollection-based recognition memory. Behavioural brain research *215*, 197–208.

Wynne C. (1997). Pigeon transitive inference: Tests of simple accounts of a complex performance. Behavioural Processes *39*, 95–112.

Zeithamova D., and Preston A.R. (2010). Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. Journal of Neuroscience *30*, 14676–14684.

Zeithamova D., Schlichting M.L., and Preston A.R. (2012). The hippocampus and inferential reasoning: building memories to navigate future decisions. Frontiers in human neuroscience *6*, 70.