Structure Learning and learning structure with Active Inference

Victorita Oana Neacsu

Wellcome Centre for Human Neuroimaging

UCL

A thesis submitted for the degree of Doctor of Philosophy

Supervised by:

Prof. Karl Friston Dr. Rick Adams Prof. John Ashburner

I, Victorita Oana Neacsu confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis is based on the Active Inference Framework (AIF), a theoretical model of information processing that views agents as inference machines. While a vast amount of research within the AIF focuses on inference and associative learning, Structure Learning (SL) is a newer and less established aspect of the AIF landscape. This thesis aims to clarify Structure Learning through three main lines of inquiry: defining Structure Learning, illustrating its implementations, and offering evidence for this construct's alignment with human behaviour. The thesis starts with a synthesis of structure learning in the general literature from research in humans, ethology, and in silico (Chapter 1). Chapter 2 will introduce three main levels of information processing in the AIF (Active Inference, Parametric Learning, and Structure Learning), and shows how various features of SL – from the general literature – relate to SL as implemented in the AIF. Chapter 3 will showcase computational simulations of a geocaching task using a deep AIF model. We show that synthetic agents learn the environmental structure through Active Inference and Parametric Learning, resulting in two types of foraging: goal-directed navigation and epistemically driven exploration. In Chapter 4, a deep hierarchical AIF model is employed to elucidate how SL influences concept learning. When endowed with SL, synthetic agents show improved performance during spatial foraging: they accumulate more rewards and show higher information gain. Chapter 5 illustrates the learning of a more abstract type of structure: learning about regularities in the environment in the form of abstract rules that underlie observed outcomes. The work in this chapter is the first to date to show evidence for Structure Learning (as implemented in the AIF) in a cognitive task in humans. In Chapter 6, I will briefly recapitulate the findings, discuss their implications, and suggest possible future directions.

Impact Statement

Recreating phenomena of human cognition using computational modelling allows for two-fold functionality. The first avenue leads toward implementing higher-order cognition in Artificial (General) Intelligence. The applications of developing Artificial (General) Intelligence are numerous but should be approached with ethical scrutiny. The second avenue brings us closer to generating digital twins: simplified models of physiology, behaviour, or dynamics that characterise specific phenomena in question. With digital twins, it could be possible (in the future) to test hypotheses non-invasively, model potential interventions, or generate new predictions. For example, in future psychopharmacology research using digital twins, one could start by modelling and improving interventions in silico, as opposed to starting with invasive, time-consuming, and costly procedures. Another example could be the phenotyping of learning to facilitate educational programmes, or modelling how individuals change their beliefs to help combat extremist beliefs or aid clinical and therapeutic interventions. The current thesis foreshadows these putative developments. The aim was to provide evidence for a novel type of learning that goes beyond associative (Hebbian) learning. This is important because associative learning does not explain the type of learning observed as a result of neurodevelopment, sleep, introspection, or rest, nor does it account for the rapid learning that involves components not initially included in the original contingencies. The contribution of this work lies in advancing the understanding of what mechanisms characterise human cognition (and action), with the purpose of eventually recreating it in silico.

Acknowledgments

I extend my sincere gratitude to my supervisors, friends, colleagues, and family who have supported me throughout my doctoral journey. To name a few:

I am deeply thankful to Prof. Karl Friston for his invaluable supervision and mentorship, and to Dr. Rick Adams for his guidance and unique insights. Many thanks also to Prof. John Ashburner for his support during the upgrade, and our engaging interactions in the context of the SSCC. My gratitude is also extended to colleagues who have become cherished friends, and people who have directly or indirectly contributed to the work in this thesis: Berk Mirza, Lance da Costa, Amy Nelson, Laura Convertino, Ryan Smith, Mikael Brudfors, Axel Constant, Berk Mirza, Tim Tierney, Avital Hahami, Peter Zeidman, Noor Sajid, Edda Bilek, Anjali Bhat, Sungwoo Lee, and Berk Mirza. No, it is not a mistake, I am thanking Berk thrice.

Special thanks to Prof. Cathy Price for not making my name public when I accidentally set off the fire alarm at the FIL at approximately 6:30pm on one not-so-sunny not-so-warm autumn day, resulting in the evacuation of the entire building. Oops. Sorry again.

Heartfelt thanks to administrative and IT staff at the FIL, including, but not limited to: Kamlyn Ramkissoon, David Bradbury, Maddy Scott, Monica Bumbury, Alphonso Reid, Mohammed Mazid, Liam Reilly, Ric Davis, Chris Freemantle, Cassandra Hugill, and Jorje Diaz. My appreciation also goes to Tracy Skinner at the ION for her support and guidance.

Finally, I am grateful to my treasured close friends, Ann, Claudia, Cristina, Danaja, Daniel, Gabrielle, Josh (names are in alphabetical order to avoid accusations of favouritism), and my wonderful partner Sebastian, for their companionship throughout this journey, for providing entertainment and support, and for their general excellence as human beings.

UCL Research Paper Declaration Form

referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

1. **For a research manuscript that has already been published** (if not yet published, please skip to section 2)

   a) **What is the title of the manuscript?**
   Synthetic spatial foraging with Active Inference in a geocaching task

   b) **Please include a link to or doi for the work**
   https://doi.org/10.3389/fnins.2022.802396

   c) **Where was the work published?**
   Frontiers in Neuroscience

   d) **Who published the work?** (e.g. OUP)
   Frontiers Media SA

   e) **When was the work published?**
   08 February 2022

   f) **List the manuscript's authors in the order they appear on the publication**
   Victorita Neacsu, Laura Convertino, Karl Friston

   g) **Was the work peer reviewed?**
   Yes

   h) **Have you retained the copyright?**
   Published under Creative Commons Attribution 4.0 International (CC BY) license which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

   i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)
   No

   If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

   ☒

   *I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

   a) **What is the current title of the manuscript?**
   Click or tap here to enter text.

   b) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

Click or tap here to enter text.

c) **Where is the work intended to be published?** (e.g. journal names)
Click or tap here to enter text.

d) **List the manuscript's authors in the intended authorship order**
Click or tap here to enter text.

e) **Stage of publication** (e.g. in submission)
Click or tap here to enter text.

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)
VN: conceptualisation, formal analysis, visualisation. VN, LC, KF: methodology and writing – review and editing. VN and LC: writing.
4. **In which chapter(s) of your thesis can this material be found?**
Chapters 2 and 3
5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Victorita Oana Neacsu

Date:

01.04.2024

Supervisor/ Senior Author (where appropriate)

Prof. Karl J. Friston

Date

02.04.2024

UCL Research Paper Declaration Form
referencing the doctoral candidate's own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

6. For a research manuscript that has already been published (if not yet published, please skip to section 2)

j) **What is the title of the manuscript?**

Structure learning enhances concept formation in synthetic Active Inference agents

k) **Please include a link to or doi for the work**

https://doi.org/10.1371/journal.pone.0277199

l) **Where was the work published?**

PLOS ONE

m) **Who published the work?** (e.g. OUP)

PLOS (Public Library of Science)

n) **When was the work published?**

14 November 2022

o) **List the manuscript's authors in the order they appear on the publication**

Victorita Neacsu, M Berk Mirza, Rick Adams, Karl Friston

p) **Was the work peer reviewed?**

Yes

q) **Have you retained the copyright?**

Published under Creative Commons Attribution 4.0 International (CC BY) license which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

r) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

No

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☒

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

7. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

   f) **What is the current title of the manuscript?**

   Click or tap here to enter text.

   g) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

   Click or tap here to enter text.

   h) **Where is the work intended to be published?** (e.g. journal names)

   Click or tap here to enter text.

   i) **List the manuscript's authors in the intended authorship order**

   Click or tap here to enter text.

   j) **Stage of publication** (e.g. in submission)

   Click or tap here to enter text.

8. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)
   VN: conceptualisation, formal analysis, writing, and illustration. VN, BM, RA, KF: methodology, revisions.
9. **In which chapter(s) of your thesis can this material be found?**

   Chapters 2 and 4

10. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

    Candidate

    Victorita Oana Neacsu

    Date:

    01.04.2024

    Supervisor/ Senior Author (where appropriate)

    Prof. Karl J. Friston

    Date

    02.04.2024

Table of Contents

General Introduction

How the brain models its environment is a main topic of investigation in cognitive and computational neurosciences, as well as in philosophy, theoretical, and artificial intelligence research. Whereas more than one century ago, the metaphor of how the brain works portrayed the brain as a hydraulic system – where the nervous system was represented by pipes and its activity as pressure exchange (Weidman 1994) – the current metaphor of brain functioning is *brain as machine* or *brain as computer* (Albus 2008, Lake, Ullman et al. 2017, Seth and Tsakiris 2018, Matassi and Martinez 2023). While it has its critics, this metaphor has produced a variety of discoveries, from neurons playing Pong (Kagan, Kitchen et al. 2022), to neural tuning that embodies prior expectations in the visual system (Harrison, Bays et al. 2023), and *C. elegans* circuits showing the computation and integration of temporal derivatives of sensory input with behavioural states to generate adaptive behaviours (Lockery 2011), as well as general theoretical frameworks of brain and behaviour (Tenenbaum, Griffiths et al. 2006, Friston 2010, Kemp, Tenenbaum et al. 2010, Tenenbaum, Kemp et al. 2011, Wiering and Van Otterlo 2012, Clark 2015, Friston, FitzGerald et al. 2017, Gupta, Mendonca et al. 2018, Friston, Moran et al. 2021, Ellis, Wong et al. 2023).

The work presented in this thesis has at its basis one such theoretical account of information processing, called the Active Inference Framework (AIF). A vast amount of work in the landscape of AIF concerns inference and associative (Hebbian) learning (Friston, Schwartenbeck et al. 2013, Friston, Rigoli et al. 2015, Friston and Buzsáki 2016, Mirza, Adams et al. 2016, Friston, FitzGerald et al. 2017, Parr and Friston 2017, Kaplan and Friston 2018, Parr and Friston 2018, Constant, Ramstead et al. 2019, Mirza, Adams et al. 2019, Da Costa, Parr et al. 2020, Hesp, Smith et al. 2021, Smith, Friston et al. 2022). Structure Learning, on the other hand, is a novel, but less established component of the AIF landscape (Friston, Lin et al.

2017, Smith, Schwartenbeck et al. 2020, Friston, Da Costa et al. 2023), and its potential underlying mechanisms are of ongoing interest more generally (Hu, Ma et al. 2021, Stoianov, Maisto et al. 2022, Ellis, Wong et al. 2023, Ghilardi, Meyer et al. 2023, Kitson, Constantinou et al. 2023, Kurth-Nelson, Behrens et al. 2023). These aspects, i.e., lack of consensus of the mechanisms involved, and its status as an emergent topic, result in an ambiguous construct of structure learning as perceived in the general literature.

This thesis centres on addressing the construct's ambiguity with three primary lines of inquiry: "*What precisely is Structure Learning?"*, *"How is Structure Learning operationalised?"*, and *"To what extent does its mechanisms align with observed human behaviour?".* The aim is to provide an integrative examination of structure learning, clarify its underlying mechanisms, and show evidence for Structure Learning as implemented in the Active Inference Framework.

The thesis will start by providing a synthesis of structure learning (and learning structure) in the general literature from research with humans, ethology research, and in silico experiments (Chapter 1). Chapter 2 will introduce the three main levels (i.e., types) of information processing in the AIF landscape (Active Inference, Parametric Learning, and Structure Learning), and present how various features of SL from the general literature (introduced in Chapter 1) relate to Structure Learning as implemented in the AIF. Please note that in this thesis, I will use *structure learning* (lowercase) in relation to the general literature, which may or may or may not coincide with *Structure Learning* (uppercase) as defined later on in the context of the AIF. Chapter 3 will illustrate computational simulations of a geocaching task using a deep (temporal) AIF model and numerical studies set. Here, synthetic agents come to learn the structure of their environment using two (of the three) levels of optimisation presented in Chapter 2: (Active) Inference and Parametric Learning. Chapter 4 provides evidence for Structure Learning (implemented with Bayesian Model Reduction) in the context

of concept formation using numerical experiments. In this chapter, a deep hierarchical AIF model of goal directed behaviour is employed, to elucidate how SL influences concept learning in a spatial foraging task, where agents explore various rooms and attempt to collect rewards. The work in this chapter was the first to feature a comparison of information gain between online (i.e., Active Inference, Parametric Learning) and offline (i.e., Structure Learning) processes, and the first to incorporate action (in addition to an observation model) in the service of Structure Learning.

In Chapters 3 and 4, the learning of structure involves physical elements, i.e., elements that can be observed directly, such as the colour of a room, or the location of a reward. Chapter 5 on the other hand, involves the learning of a more abstract type of structure: learning about regularities or observed patterns in the environment in the form of abstract rules that underlie observed outcomes. This chapter will report findings from experiments using an abstract rule-learning task. Three experiments were carried out (two empirical, and one involving computational simulations). The work in this chapter is the first work to date to show evidence for Structure Learning as implemented with Bayesian Model Selection in a cognitive task in humans. Whereas the models involved in Chapters 3, 4, and 5 concern spatial foraging, concept learning, and abstract rule learning, the three main AIF mechanisms I will present are generalisable to multiple phenomena. In Chapter 6, I will briefly recapitulate the findings and discuss their implications, as well as suggest possible future directions.

Chapter 1

Structure learning and learning structure - a synopsis

1.1 Structure learning and learning structure in human cognition

Humans are very competent at abstracting relationships between various elements in the environment, constructing knowledge that represents these relationships, as well as performing sophisticated sets of behaviours, often with limited experience or exposure. This adaptability stems from our ability to learn flexibly. What follows is a synthesis of studies on structure learning in humans, its intersection with other learning mechanisms, and the implications of this research for what we later on come to operationally define as Structure Learning.

1.1.1 Structure learning in development

Structure learning — in the sense of learning causal structure — has been reported in infants as young as 6 months old (Emberson, Richards et al. 2015). In this study, researchers used an audio-visual omission paradigm and recorded responses using functional near-infrared spectroscopy (fNIRS). Infants were exposed to auditory stimuli that were followed by visual stimuli 80% of the time. This created a temporal association, where a sound predicted an upcoming visual stimulus. When exposed to visual omissions (i.e., when playing the sound did not result in the expected visual stimulus), the infants' sensory expectations were violated, manifested as a cortical response in the occipital lobe. Furthermore, this cortical response was not present when the auditory stimulus did not predict (i.e., was not associated with) a visual event. This study (Emberson, Richards et al. 2015) essentially presents evidence supporting the idea of expectation-based feedback, suggesting that even very young infants display competence in learning the structure of their environment. Young infants have been shown to exhibit this competence even in more complex scenarios (Monroy, Gerson et al. 2019), where authors show that infants aged 8-11 months display sensitivity to structure. In the learning phase, infants observed probabilistic action sequences comprising 7 steps, where an actor

interacted with 6 unique objects. In the testing phase, when infants were exposed to a sequence that included deviant pairs, they detected violations to complex regularities (i.e., structure) learnt previously.

Whereas this research (Emberson, Richards et al. 2015, Monroy, Gerson et al. 2019) and others (Karuza, Emberson et al. 2017, Emberson, Misyak et al. 2019) focused on passive responses during the learning experience, young infants have been shown to demonstrate an awareness of structural relationships also during active engagement with the environment (Stahl and Feigenson 2015, Baek, Jaffe-Dax et al. 2020). For example, at approximately 6 months of age, when infants develop the ability to reach and manipulate objects, they are able to use their motor skills to select and explore specific objects in a way that mimics saccades (i.e., visual attention) (Needham, Barrett et al. 2002). During this type of exploration, infants preferentially grasp objects that are expectation-inconsistent (e.g., toys that violate expectations). Furthermore, they engage in hypothesis testing that reflects a specific violation: for example, infants who had previously observed an object that appeared to float in midair, drop it multiple times; on the other hand, observing an object that appeared to pass through the wall, resulted in repeatedly striking the object against the wall (Stahl and Feigenson 2015).

Testing hypotheses about the physical properties of objects in the environment aligns with the broader concept of sensorimotor learning, where infants learn sensorimotor control through both embodied trial-and-error and action observation. Learning through trial-and-error and action observation together has been referred to as predictive motor activation (Monroy, Meyer et al. 2019, Ghilardi, Meyer et al. 2023). The idea here is that knowledge gained through observational experience during infancy generates action-prediction signals in the motor system. Such motor predictions can stem from statistical regularities that were learnt through observation, with statistical learning as a prime candidate mechanism (Ghilardi, Meyer et al. 2023). Stated differently, statistical learning is required to learn regularities (i.e., structure) in

the environment, and after these structures have been internalised, they generate sensorimotor predictions. The sensorimotor predictions then result in sensorimotor control. In this paradigm, statistical learning has been referred to as the ability to identify regularities between sensory input and motor output, whether these regularities are internal (e.g., physical properties of the human body) or external (i.e., physical and temporal properties of the environment). For example, consider a scenario where an infant watches their parent pick up a cup, an action with various potential outcomes (e.g., affordances). If the parent consistently follows this action by switching on the kettle, retrieving milk from the fridge, and then adding tea and sugar, the infant, after repeated instances, will come to anticipate the subsequent steps when their parent next picks up a cup. The predictive motor activation model is assumed to be a major contributing factor to the inception of motor control in adulthood (Monroy, Meyer et al. 2019). In other words, learning (i.e., internalising) the structure of the environment allows infants to develop their models of the world, gradually constructing the understanding of sensorimotor control in self and others.

1.1.2. Structure learning as adaptive generalisation

Generalisation in cognitive science refers to the process by which individuals apply knowledge or skills learnt in one context to novel, similar situations (Tenenbaum and Griffiths 2001, Donchin, Francis et al. 2003, Braun, Aertsen et al. 2009, Mulavara, Cohen et al. 2009). That is, it involves recognising commonalities between different instances and applying previously learnt information to novel scenarios. Although generalisation enables individuals to make predictions, draw conclusions, and navigate their environment effectively, it can also lead to errors or biases when applied inappropriately. One useful strategy in enhancing adaptive generalisation is that of practicing with a multitude of examples (or tasks) that are related to

the task (Mulavara, Cohen et al. 2009). In this example study, neurotypical adults engaged in a task involving avoiding obstacles while walking on a treadmill. During the training phase, they were equipped with either three distinct sets of visual distortion lenses, a single pair of visual distortion lenses, or sham lenses. Subsequent testing involved a different set of visual distortion lenses. It was found that participants who underwent training with multiple lenses demonstrated superior adaptation to the novel lenses, compared to those who trained with only one set of lenses or those in the sham condition. This suggests that the most effective form of structure learning via generalisation necessitates two components: task or goal similarity; and slight variations within the task category. This dual-component form of generalisation is akin to multi-task learning in computer science (Caruana 1997), where generalisation in artificial neural networks is enhanced by inductive transfer, with tasks learnt in parallel, but using a shared representation (here, a shared hidden layer).

Furthermore, generalisation via practicing with multiple examples is not limited to implementing these (motor) actions physically. There is now substantial evidence that motor imagery (i.e., mental simulation of an action) can enhance motor learning (Doyon and Benali 2005, Di Rienzo, Debarnot et al. 2016, Toth, McNeill et al. 2020). Motor learning is defined as an improvement in motor performance and can occur both online and offline. In the first case, ‘online’ means that the performance was assessed before and immediately after practicing via mental imagery, indicating that the learning has occurred as a direct result of practicing, such as in (Mizuguchi and Kanosue 2017). However, offline learning involves an indirect result: where practice, followed by a latent period of at least approximately 6 hours in the absence of additional practice (Doyon and Benali 2005) improves performance, or where (motor) performance is improved subsequently to sleep (Blischke and Erlacher 2007, Hill, Tononi et al. 2008, Schmid, Erlacher et al. 2020).

Structure learning via generalisation can therefore involve physical structure that is both (or either) extrapersonal, such as in the study with visual distortion lenses (Mulavara, Cohen et al. 2009) and intrapersonal, such as in motor learning studies (Toth, McNeill et al. 2020). Furthermore, learning can occur online, where a direct effect is observed immediately after practicing; or offline, where indirect effects are observed as a result of consolidation of memory traces for example during sleep (Di Rienzo, Debarnot et al. 2016), or just as a result of a latent period of time in the absence of additional practice (Doyon and Benali 2005).

1.1.3 Structure learning as 'learning to learn' (meta-learning)

Imagine you are a race engineer at a Formula 1 team, trying to establish the car set-up in different track conditions (for different track circuits). The car is equipped with numerous adjustments, such as suspension settings, aerodynamic configurations, transmission, etc.; let us assume there are approximately 50 types of possible adjustments. The car is performing well at one of the track circuits, and you are aiming to adjust the set-up for the next track circuit. How do you go about adjusting the car set-up from one track to another? One approach could be to employ optimisation techniques that adjust each setting (i.e., parameter) individually, and explore the entire multidimensional set of possibilities. However, with experience, you might discover that certain track circuits exhibit consistent (possibly non-linear) relationships among the settings (i.e., parameters). This insight would allow you to create a new meta-setting, that adjusts (and constrains) several settings concomitantly. Consequently, when faced with deciding the set-up for the next track circuit, the search for optimal settings is narrowed down to a small subset of the full parameter space, speeding up the 'learning' process.

This example illustrates structure learning as a form of meta-learning or learning to learn (Braun, Mehring et al. 2010). In this theoretical paradigm, structure learning corresponds

to constructing the meta-setting, and parametric learning corresponds to adjusting the meta-setting (following structure learning). In essence, structure learning as defined in this paradigm involves reducing the dimensionality of 'search-space' that the subject would have to explore in order to adapt to new circumstances (Needham, Bradford et al. 2007). Reducing this dimensionality can be achieved by extracting invariants between different input-output mappings (Braun, Mehring et al. 2010).

Adapting to new circumstances via the learning of structure has been emphasised in adaptive control theory. In a similar manner to meta-learning, adaptive control theory involves two levels of adaptation: structural control and parametric control (Braun, Mehring et al. 2010). In structural adaptive control, the form of the task or environment is itself unknown, requiring the development of task-relevant representations, encoded by an internal model. In contrast, parametric adaptive control assumes a structural form, requiring only the estimation of the current (latent) parameters in play. Training in various dynamical tasks, such as grasping and moving objects with varying properties, has been shown to lead to the formation of these (task structure) representations (Shadmehr and Mussa-Ivaldi 1994, Braun, Aertsen et al. 2009), and their associated adaptive strategies (Braun, Aertsen et al. 2009). For instance, when presented with unpredictable visuo-motor rotations, participants learn to adapt as they progress through the trials (Braun, Aertsen et al. 2009), in a way specific to the task structure, and only when input-output mappings are perturbed; when perturbations were outside the scope of the input-output mappings, such as when stimuli 'jumped' instead of being rotated, the parametric adaptation observed previously (i.e., for task relevant perturbation) did not materialise.

1.1.4 Structure learning as causal reasoning

In the realm of causal reasoning, structure learning, defined as the challenge of inferring causal connections between latent (hidden) and observable variables, is a recurring theme (Steyvers, Tenenbaum et al. 2003, Gopnik, Glymour et al. 2004, Kemp and Tenenbaum 2009). A classic example from psychology illustrates this aspect of learning vividly. Consider a scenario where observations are made on three variables: temperature, ice-cream sales, and the frequency of shark attacks. In this example, there is a strong correlation between these variables, such that one could potentially assume that selling more ice-creams will cause sharks to attack more often. However, in reality, this type of statistical covariance would not imply a causal relationship between shark attacks and ice-cream. A more plausible explanation would be that as temperatures rise, more people swim in the sea and buy ice-cream, but it is improbable that the temperature change directly triggers shark attacks.

Humans, however, find it difficult to infer causal structure when given correlational observations only (such as in the example given above), primarily due to the abundance of potential conditional dependencies (Pearl 2000, Steyvers, Tenenbaum et al. 2003, Meder, Hagmayer et al. 2009). Conditional dependencies can be thought of as the result of controlling for confounding factors or covariates (Novick and Cheng 2004). Identifying conditional independencies is therefore hard, since it involves tracing of concomitant change in multiple variables. Furthermore, as the number of variables increases, the combinatorics likewise increase, but exponentially. Tracing multiple variables in time gives rise to a combinatorial explosion that quickly becomes computationally intractable. Faced with these computational limitations in inferring causal structure, humans resort to, and have been shown to rely on shortcuts and supplementary cues such as temporal structure (i.e., cause precedes effect) (Goldvarg and Johnson-Laird 2001, Lagnado and Sloman 2006), counterfactuals (Lagnado,

Gerstenberg et al. 2013, Halpern 2016), and active interventions (Pearl 2000, Lagnado and Sloman 2004, Waldmann and Hagmayer 2005).

1.1.5 Structure learning in abstract thinking, problem-solving, and insight

The distinction between two principal approaches to problem-solving has been researched for almost a century, beginning with Poincare first suggesting in 1913 that solutions to problems manifest into consciousness only after being deemed acceptable (Poincaré 2022). One of these two approaches, the analytic approach, involves applying past experiences to tackle problems gradually, where a solution is reached after taking incremental steps from the initial problem specification and criteria (Jung-Beeman, Bowden et al. 2004, Fleck and Weisberg 2013, Weisberg 2013, Webb, Little et al. 2016). The other approach, the insight-based approach, is triggered by moments of impasse, and involves a sudden breakthrough that draws from elements outside the initial problem formulation. For example, in one study, authors employ an insight task where subjects solve verbal reasoning problems; the results provide evidence for a differential in brain activity between the two processes (Jung-Beeman, Bowden et al. 2004). Using two neuroimaging methods, the authors show increased activity in the right anterior temporal area for problems solved using insight, as compared to non-insight (analytic) solutions; this area has been associated with making long-range connections during comprehension.

The experience of insight is defined by its four essential features: the impasse, the reorganisation of existing knowledge, the ‘aha’ moment, and a subjective feeling of certainty (Weisberg 2013). Problems devised to produce insight generally revolve around cognitive restructuring, i.e., a sudden change in the way some entity is perceived; whereas non-insight problems do not involve cognitive restructuring (Klein and Jarosz 2011). In a more recent study

(Webb, Little et al. 2016), the authors propose additional differentiations, beyond the two learning mechanisms outlined earlier, with an emphasis on the nature of the task. In this study, participants engaged in several insight and non-insight tasks. This research found that for problems devised to produce insight, correct solutions were associated with a greater percentage of reported insight, compared to non-insight solutions. Additionally, correct solutions elicited stronger feelings of insight as compared to incorrect solutions. The certainty element observed in insight has significant consequences for exploratory behaviour: high confidence in new knowledge will decrease the need to solicit new information from the environment.

1.1.6 Structure learning and concept learning

Learning latent structure is fundamental to human cognition. Humans do not experience sensory information as a flow of collections of brightness, colours, textures, size; they make use of concepts in making sense of their surroundings. The ability to extract similarities and identify dissimilarities across sets of experienced (sensorial or autobiographical) events is a crucial element of (structured) knowledge building (Zeithamova, Mack et al. 2019). This type of relational thinking is known as concept learning, first proposed in the book 'A study of thinking' (Bruner, Goodnow et al. 1956). More specifically, concepts are mental representations that allow humans to compare and contrast collections or sets of events and their respective elements. Various researchers have since developed and expanded on the concept of concept learning. For instance, the prototype theory of concept learning (Geeraerts 2006) suggests that biological agents possess a central example, a 'common representation' of a particular set, and then judge how (semantically) close or far new experiences are in relation to the prototype. Another instance is the exemplar theory of concept learning, involving

abstraction of features, whereby concepts are characterised as a set of rules, and agents assess new experiences (of events or objects, etc.) based solely on their respective properties, and whether they fit the definitions or not (Rouder and Ratcliff 2004, Tenenbaum, Griffiths et al. 2006). Recent work suggests that concept learning involves both prototype (e.g., generalised) and exemplar (e.g., specific) representations (Bowman and Zeithamova 2020).

One notable feature of concept learning involves the ability to quickly grasp new concepts, and effortlessly apply them to unfamiliar situations. Recent neuroimaging research (Mack, Preston et al. 2020) suggests that the ventromedial prefrontal cortex (vmPFC) monitors the efficient mappings between stimuli and categories, emphasising information that matters and down-weighting irrelevant characteristics (a.k.a., filtering). The component of concept learning that entails linking information is thought to arise through memory integration as an interplay between the vmPFC and the hippocampus (Zeithamova, Mack et al. 2019). Furthermore, category representation plays an important role in concept formation and the ability to generalise concepts. This feat has been attributed to the function of rostrolateral prefrontal cortex (rlPFC), believed to integrate decisional information and stimulus novelty, to determine the optimal time for employing inferential processes of category learning (O'Bryan, Worthy et al. 2018, Zeithamova, Mack et al. 2019).

1.1.7 Structure learning and replay

Replay in the brain was first observed in the context of animal research on spatial navigation with rodents (Skaggs and McNaughton 1996, Nádasdy, Hirase et al. 1999). During spatial exploration, hippocampal neurons encode the animal's concurrent location. On the other hand, during periods of rest (e.g., sleep), these same neurons occasionally exhibit a spontaneous sequence of firing patterns. The spontaneous firing patterns recapitulate paths that the rodents

recently explored, but these patterns were temporally compressed. This phenomenon is referred to as ‘replay’. Replay occurs across a range of states such as rest, sleep (Deuker, Olligs et al. 2013, Gruber, Ritchey et al. 2016), and wakeful pauses from active engagement (Tambini and Davachi 2019); and is considered to be a significant aspect of hippocampal function, constituting a substantial portion of neural activity during rest periods (Buzsáki 2015). Initially it was proposed that the hippocampus rapidly stores new experiences, and replay serves as a mechanism for transferring this knowledge into a more stable form in the cortex, a process known as consolidation (Wilson and McNaughton 1994). Since then, views on replay have evolved from the replay of sequences as rehashing previous experience, to replay as a form of compositional computation that synthesises information into (relational) structures to derive new knowledge (Wittkuhn, Chien et al. 2021, Kurth-Nelson, Behrens et al. 2023).

The ’replay as compositional computation’ view (Kurth-Nelson, Behrens et al. 2023) proposes that replay essentially sequences hippocampal representations of role-bound entities, and chains them into structures. In this framework, replay can involve any (re)arrangement of sequences (i.e., it goes beyond rehashing previous experiences), but implies that each element in this arrangement is bound to a representation of its role in the arrangement, altogether allowing for the construction of quite complex structures. This ability to reorganise knowledge is suggested to underlie the flexibility of human creativity and imagination, as well as to give rise to generalisation. We will come back to replay in terms of ethology research and in silico (here, using reinforcement learning models and neural networks) in the next section, and section 1.3 respectively.

1.2 Structure learning and learning structure in ethology

Picture a scenario where a crow is attempting to eat a walnut. The crow could, for example, try to crack the nut open using its beak, or it could drop it from a height in the hope that it will shatter. Crows in both Japan and New Caledonia have come up with another, quite remarkable solution: they strategically drop the walnut at the traffic lights, waiting for a car to effectively crack the nut open, allowing the crow to retrieve and eat the nut (Cristol and Switzer 1999, Nihei and Higuchi 2001).

The capacity to abstract structured interrelationships underlying the experienced world is widely accepted in human cognition. In animal research however, this capacity is commonly disputed. More specifically, the debate concerns the limits of animals' abilities regarding the abstraction of contingencies in the environment: are animals truly capable of abstraction, or do they simply display complex behaviours when interacting with the environment without actually *understanding* (Shettleworth 2010)? One prominent perspective that supports the capacity to *understand* in animals postulates that at the core of the equivalent human ability to abstract, lies a set of more primitive domain-specific cognitive systems, from which a more complex, domain-general set of proficiencies evolved (Pinker and Jackendoff 2005, Dehaene, Meyniel et al. 2015). In other words, this perspective suggests that humans and other animals are likewise capable of abstracting interrelationships, but humans are more proficient because they possess a more elaborate cognitive milieu.

Going back to the crow and walnut example, in order to display this type of behaviour, the animal must have representations of the objects (i.e., walnut, car, road, etc.), representations of the relationships between these objects, and some model of causality between its beliefs about behaviours, the behaviours of objects, and the outcomes of those behaviours. Furthermore, these representations have to be dynamic. Structure learning entails precisely

these cognitive capacities. In biological organisms, what appears to be learnt is a collection of interdependencies, a general belief system of rules that govern a collection of tasks, and when to apply them, essentially reducing the dimensionality of space of possible hypotheses that the organism has to examine in order to adapt to novel tasks, problems, or environments (Jordan 1998, Vapnik 1999, Needham, Bradford et al. 2007, Pearl 2014)

One classic account of animal cognition in support of this view shows that animals are not only able to learn the structure of particular tasks, but also capable of generalising between tasks with similar structures, demonstrating transfer of knowledge (Harlow 1949). In Harlow's experiments, monkeys had to choose one of two objects, of which only one was rewarding. For a number of trials, the animal had to select between these objects, followed by changing the object type; if the animal chooses the rewarding object from the first step, the optimal strategy is to continue choosing that object; if the rewarding object is not selected, the strategy is to swap the object choice. This process was carried out for several blocks and Harlow observed that the monkeys were able to reach peak performance during the second trial of each new block (Harlow 1949). It must be the case that as time progressed, the animals succeeded in internalising a representation of the rules defining this particular task structure, allowing for structural learning, or learning to learn (Braun, Mehring et al. 2010).

Harlow argued that when encountering novel tasks, animals learn gradually using trial-and-error, and then generalise to new tasks (in a similar class) only when they have been exposed to several examples of comparable tasks (Harlow 1949). Through this process, animals build what he coined as a 'learning set'. This notion was brought forward to connect opposing concepts in Gestalt psychology (Schrier 1984, Reznikova 2007, Van Merriënboer and De Bruin 2014), where the 'whole is more than the sum of its parts' (e.g., *understanding*), and the (reductionist) behaviouristic approaches that strongly supported reinforcement strategies, such as trial-and-error or stimulus-and-response (Van Merriënboer and De Bruin 2014). The

learning set theory merges these two perspectives, suggesting that animals are able to learn rules from individual experiences, and then apply these rules to work out solutions to novel (problem) sets.

Harlow was not the first ethology researcher to propose that animals internalise models that encode relationships between elements in their surroundings. Latent learning, one of the first accounts of animal (structure) learning, was put forward by Tolman as early as 1930 (Tolman and Honzik 1930, Tolman 1948). Tolman and Honzik remarked that rats exhibit complex and flexible behaviours, such as taking shortcuts to obtain rewards (Tolman and Honzik 1930), or finding new routes when the old ones were obstructed (Tolman, Ritchie et al. 1946). In latent learning, the type of learning that occurs without reinforcement was explained by engaging the notion of an *intervening variable*. 'Cognition' was thought to be this intervening variable, since it *intervenes* between both stimulus and response. Cognitive maps were formed as a result of learnt behaviours (Tolman 1948), a concept that directly relates to *learning sets* in Harlow's work: organisms construct systematic maps to represent their physical environment by means of complex cognitive processes, not just simple conditioning processes. The most central examples in animal literature to reveal these *cognitive maps* or *learning sets* arise in the spatial navigation literature, involving the hippocampal-entorhinal system. Hippocampal place cell activity is restricted to particular locations in space (O'Keefe and Dostrovsky 1971), whereas entorhinal grid cells are active for multiple spatial fields that are equally spaced on a triangular or hexagonal grid (Hafting, Fyhn et al. 2005), allowing for vector and distance relationships to be encoded.

Animals rely on a diverse set of internal (generative) models that require structural knowledge of the external world, as evidenced by a plethora of experimental paradigms, such as the experiments described previously, as well as experimental research showing tool usage in anthropoids (Whiten, Horner et al. 2005, Seed, Call et al. 2009, Nieder 2013), causal

inference and spatial mapping in rats (Tse, Langston et al. 2007, McKenzie, Frank et al. 2014, Laurent and Balleine 2015), or complex social cooperation in corvids (Clayton and Emery 2007). In a striking example, cockatoos were capable of picking locks (Auersperg, Kacelnik et al. 2013), a five-step process requiring the birds to remove a pin, unscrew a screw, withdraw a bolt, rotate a wheel, and pull out a lever. When the five steps were reconfigured, the animals correctly identified this change, exhibiting transfer.

Recent work on replay further supports the idea that animals are proficient at internalising and applying models of the environment in a dynamic and flexible manner, i.e., learning environmental structure (Widloski and Foster 2022). In this experiment, rats engaged in a spatial navigation task, and were able to learn the locations of liquid chocolate in a (square) environment with randomly changing barriers. Furthermore, results from monitoring (hippocampal) replays consistently demonstrated new goal-directed trajectories around each altered barrier set-up. These adaptive replays were quickly learnt and did not depend on a remapping of place cells. The results from this experiment suggest a clear distinction between stable responses of place fields, which remained tied to sensory cues, and the flexible adjustments (i.e., compositional reorganisation) seen in replays, which adapted to reflect the learnt conditions in the environment.

1.3 Structure learning and learning structure *in silico*: synthetic cognition and agent-based modelling

Any resolve to comprehend the cognitive and neurophysiological implementation of (structure) learning necessitates the identification and specification of underlying mechanisms. While some cognitive mechanisms of learning – such as (sensory) evidence integration in decision making (Waskom and Kiani 2018) – are more conspicuous, other more complex (e.g., hierarchical, dynamic, and temporal) learning mechanisms are less apparent. A multitude of theoretical approaches has been developed to characterise these underlying computational processes (Gopnik, Glymour et al. 2004, Love, Medin et al. 2004, Gopnik 2011, Gershman 2015, Friston, FitzGerald et al. 2016, Friston, FitzGerald et al. 2017, Behrens, Muller et al. 2018, Niv 2019). The core assumptions of these approaches frame cognitive-behavioural mechanisms through the lens of inferential, statistical, and probabilistic processes, with a central focus on organising knowledge. On one hand, reverse-engineering learning models enhance our understanding of human cognition. On the other hand, building these frameworks feeds back into in-silico research, allowing for advancements in Artificial Intelligence, computational neuroscience, and computational biology.

1.3.1 Structure learning and graphical models in Machine Learning

Structure learning is a fundamental task in machine learning and statistics and involves uncovering the underlying dependencies and relationships between variables in a dataset. Causal (structured) relationships are often characterised by graphical models (Pearl 2000, Gopnik 2011, Pearl 2014) such as Bayesian networks (a.k.a. Bayes nets). By accurately learning the structure of a graphical model, one can better understand the underlying causal or associative relationships in the data, leading to more effective predictions and interventions.

Graphical models are generally directed (cyclic or acyclic) graphs (Drton and Maathuis 2017) that describe specific structured latent relationships. Nodes usually represent variables (e.g., number of shark attacks, ice-cream sales), and the edges (or lack thereof) represent relationships between these variables (e.g., causal influence or causal independence). Agents, whether biological or synthetic, are assumed to maintain and update probability distributions over potential structures that explain what is being observed. Both the causal structure itself (i.e., structure learning) and the strength of causal relationships (i.e., Parametric Learning) can be learnt. The combinatorial explosion of search spaces is a challenging aspect of causal inference for both humans and machines. In the machine learning version of structure learning, reducing the dimensionality of a search space can be thought of as an abstraction of (structural) invariances between different mappings (e.g., in the Bayes nets) encoding contingencies in the world. This process facilitates generalisation and therefore optimises the efficacy of any learning algorithm (Vapnik 1999) by increasing the learning rate for problems and tasks with similar structures.

Bayes nets are suitable for representing joint distributions over sets of random (and latent) variables (Larranaga, Karshenas et al. 2013). For example, a Bayes net (i.e., model) with a particular network structure *S,* comprised of *N* random latent variables representing the input $I_1, I_2, \ldots, I_N$; *M* control variables representing the output $O_1, O_2, \ldots, O_M$ and model parameters μ can be characterised as a joint probability distribution $P(I,O,\mu,S)$, that can be decomposed into a product of conditional probabilities: $\Pi_{i=1}^{N+M} P(O_i \mid I_i, \mu_S, S)$. The structure itself governs the dependencies between the variables, and the probabilities indicating the 'strengths' of dependencies represent the parameters of the structure. In this case, structure learning involves learning the topology of the network itself, and parametric learning consists of estimating the strength of the causal connectivity given the structure in play. As is the case with many networks, latent variables and their associated relationships alike have to be inferred,

exacerbating the problem of structure learning. Efficiently computing the joint probability distribution therefore involves the estimation of latent variables. This inferential process generally involves two stages. The first stage consists of estimating the posterior probability of a specific model *S* given observations (i.e., data). In the second stage, the posterior probability of parameter μ_S is estimated using the observations (i.e., data) and structural model (i.e., network) *S.* Temporal dependencies between the variables ($\vec{I}$,$\vec{O}$) can be formalised by extending the network to include temporal sequences, such as Dynamic Bayes Nets (Dean and Kanazawa 1989, Ghahramani 1997, Zweig and Russell 1998, Mihajlovic and Petkovic 2001).

So far, we have only considered exact inference in Bayes nets, however, uncertainty is a ubiquitous part of information processing (Pouget, Beck et al. 2013). Probabilistic models entertain representing this uncertainty by use of specifying circuitry as encoding probability distributions, and message-passing (or belief update) as encoding probabilistic inference or probabilistic computations more generally (Pouget, Beck et al. 2013). Probabilistic inference (e.g., Bayesian inference) entails the estimation of expected values or probabilistic densities from a probabilistic model (e.g., Bayes net). Since exact inference is generally computationally intractable, it requires approximation methods (Finley and Joachims 2008). One typical solution is the usage of approximation techniques such as Monte Carlo sampling (Shapiro 2003), where independent samples are drawn from the probability distribution multiple times (in order to approximate the quantity of interest), although this sampling method becomes intractable as the number of dimensions (i.e., variables) increases. Markov Chain Monte Carlo (MCMC) solves this high-dimensionality problem by drawing samples dependent on the current sample and narrowing in on the quantity of interest (Andrieu, De Freitas et al. 2003). With a discrete model space for each level of the structured set of relationships (such as in hierarchical Bayesian models), this sampling method mimics the emerging perspective that

humans (and animals in general) are only capable of evaluating a limited number of hypotheses at any point in time (Vul, Goodman et al. 2014).

When formulating the structure learning problem under probabilistic models, such as non-parametric hierarchical Bayesian models, one advantage is the capacity to simultaneously consider several (alternative) hypotheses of model structure (Tenenbaum, Kemp et al. 2011). This class of models is used in machine learning to discover the form of structured interrelations (given sets of observations) and predict how to generalise novel properties. That is, they have been used to induce transfer learning (Wilson, Fern et al. 2012). Non-parametric hierarchical models have been shown to reproduce various features of neural circuitry, such as structural hierarchies in the cortex (Wang, Kong et al. 2019). Furthermore, they have been shown to attain human level performance in a variety of cognitive tasks, such as statistical learning (Griffiths and Tenenbaum 2006, Griffiths, Sanborn et al. 2011), concept learning and formation (Kemp, Tenenbaum et al. 2010, Lake, Salakhutdinov et al. 2015, Lake, Ullman et al. 2017), or action recognition from videos (Tu, Huynh-The et al. 2019).

Another example approach to structure learning using graphical models involves scoring models based on some predefined metric using genetic algorithms (Larranaga, Kuijpers et al. 1996, Larranaga, Poza et al. 1996, Ji, Wei et al. 2013, Kitson, Constantinou et al. 2023). Genetic algorithms take inspiration from research in biology, with techniques such as genetic mixing, genetic mutation, swarming, etc.; for example, in genetic mixing, edges are taken from two different graphs to create a new graph; genetic mutation is typically implemented by making random changes to graph edges. One example study involves combining genetic algorithm techniques with ordering-based search techniques to create a *memetic* algorithm called MINOBS, i.e., memetic insert neighbourhood ordering-based search (Lee and van Beek 2017). This quite complex algorithm can be described as follows: initially, hill-climbing search is applied to an initial set (i.e., population) of orderings. Hill-climbing involves making small,

iterative adjustments to reach a locally optimal solution. Once the initial set has been optimised through hill-climbing, techniques such as mutation and pruning are applied. Pruning here refers to selecting the best individuals (i.e., models) from the population, which produces a new population (here, maintaining the original size of the population); hill-climbing then resumes on the new population, and the process repeats until some termination condition is met.

To summarise, structure learning in Machine Learning plays a crucial role in uncovering complex dependencies and relationships within datasets, particularly using graphical models. By employing a diversity of statistical techniques, scoring metrics, and optimisation algorithms, one can infer the underlying structure from data, leading to improved understanding and predictions.

1.3.2 Structure learning and Reinforcement Learning

In Reinforcement Learning (RL), the problem of representing the structure of the environment is reproduced in terms of finding policies (i.e., actions or sets of actions) that will maximise cumulative reward (Kaelbling, Littman et al. 1996, Dayan and Watkins 2002, Behrens, Muller et al. 2018). Agent-environment interactions are typically modelled as (discrete state space) Markov decision processes, comprised of the following components: the environment (defined as a set of states), a set of possible actions, a function for state transitions (i.e., a model that specifies the probabilistic transitions from the current to the next state after taking an action), and a reward function). The Markovian aspect entails that states and rewards at time t+1 depend only on the state and action of the current time point. Research shows that deep RL agents have been demonstrated to equal and even surpass human performance in specialised task domains, such as Go (Silver, Huang et al. 2016) or Atari (Mnih, Kavukcuoglu et al. 2015). In deep RL, knowledge is encoded as high-dimensional vectors, acquired through extensive training with

large datasets. Learning (e.g., of structures) corresponds to processing these representations through mathematical operations within the deep neural network.

In a representative study, researchers employ a Stratified Rule-Aware Network (SRAN) to solve Raven's Progressive Matrices using deep learning (Hu, Ma et al. 2021). In Raven's Progressive Matrices, agents are presented with a 3 x 3 grid of simple visual stimuli (i.e., images) with the final section left empty. The objective is to select one image (from a set of 8 options) that would complete the grid in a way that adheres to the implicit rules governing the arrangement of the other images. This is achieved by examining the patterns in the first two rows and/or columns and deducing the overarching rules that guide the attributes of the images. These rules are then used to determine the suitable image for the empty section. The proposed framework (i.e., Stratified Rule-Aware Network) for solving this task is designed to extract rule embeddings at various levels of granularity (i.e., cell-wise, individual-wise, and ecological), and then fuse this information across levels of granularity to create a new set of rule embeddings. Furthermore, the framework involves a rule-similarity metric, that estimates the similarity between rule embeddings. For any given (Raven's Progressive Matrices) question, the process is as follows: initially, the first two rows/columns of the original grid are input into the framework to deduce the rule (using extraction and fusion) with its associated rule embedding (called the dominant rule); next, the framework is applied to each of the 8 matrices (i.e., the original grid + each of the 8 candidate answers) to generate rule embeddings based on the completed matrix. The correct answer is then chosen by selecting the candidate answer with the highest similarity score to dominant rule embedding. In other words, the model derives a rule embedding for the original 2 x 3 or 3 x 2 matrix, and a rule embedding for each of the eight 3 x 3 matrices (original matrix + each answer), and compares the similarity between the rule embeddings to figure out the correct answer (Hu, Ma et al. 2021).

Accounts using deep learning, however, are less representative of human cognition, given the high volumes of training data necessary to achieve this level of performance, as well as an inability to generalise to variations in task conditions. In contrast to deep learning, in deep meta-reinforcement learning, agents quickly adapt to new tasks, by leveraging structural knowledge attained via exposure to a set of similar tasks (Wang, Kurth-Nelson et al. 2016). In this study, agents undertake a series of bandit (i.e., decision making) tasks whose parametrisation varies. The agent receives as inputs the consequences of the action taken in the previous step, and its associated reward value. In parallel, the weights of a recurrent (i.e., Long Short-Term Memory) network are tuned according to the reward value, by way of employing a memory cell to maintain long-term dependencies in the (sequential) data. This procedure allows agents to become attuned to the shared structure of the training tasks, a feature that evokes the type of structure learning (as generalisation) observed in animals – i.e., learning sets and cognitive maps (Tolman 1948, Harlow 1949) and humans – i.e., transfer learning and generalisation (Braun, Aertsen et al. 2009, Mulavara, Cohen et al. 2009). Although this architecture (Wang, Kurth-Nelson et al. 2016) allows for better structure learning (i.e., as generalisation), performance can degrade in settings where temporal dependencies cover a longer horizon (Huisman, Van Rijn et al. 2021).

1.3.3 Structure learning and replay in computational modelling

One focus in the field of replay involves constructing synthetic (i.e., artificial) agents capable of learning generative models from experience. These models serve various purposes, such as inferring new connections based on latent structural rules (Evans and Burgess 2019), deducing the appropriate context for new data (Stoianov, Maisto et al. 2022), image classification based on continual learning problems (Van de Ven, Siegelmann et al. 2020), and transferring

knowledge to novel tasks to mitigate catastrophic forgetting (Shin, Lee et al. 2017). For example, in the framework proposed by one recent study (Stoianov, Maisto et al. 2022), agents learn generative models from trajectories across a maze. These models can then generate new trajectories consistent with the current maze configuration during offline periods (i.e., during replay). As the agents encounter new mazes, they continue generating novel trajectories offline, drawing from their experience with various maze types to prevent loss of information about any single maze. The hierarchical organisation of this model essentially implements inductive biases to identify individual elements of experience (first hierarchical level), organise them into sequences (second level), and cluster them into maps (third level). This structure leads to trajectories being grouped into specific maze contexts, enabling inference (of maze categories) when encountering new data (i.e., observations). In other words, agents endowed with replay in the context of a hierarchical generative model, show improved continual learning of multiple mazes, situating replay as a promising candidate mechanism for learning about (transition) structure.

Other approaches to replay in computational modelling involve extracting graph properties from observed transition structures (Eysenbach, Salakhutdinov et al. 2019). In this work, the challenge of reaching a distal goal is broken down into a series of simpler tasks, each focused on achieving a specific subgoal. Using a planning algorithm, abstractions of the environment are represented as a graph of nodes and edges. The graph was constructed using reinforcement learning, where a goal-oriented value function assigned weights to edges, and nodes were derived from previously observed states stored in the replay buffer. Graph search was then applied on the replay buffer, which automatically generated the sequence of subgoals. Essentially, this model allows agents to learn representations using replay to infer graph structures that can be used for planning. Agents are thus enabled to generalise and solve tasks with sparse rewards over prolonged periods of time (here, 100 steps).

1.3.4 Structure learning as network discovery

The above review reflects the various psychological perspectives on structure learning. One might argue that this review is overinclusive and that some of the examples above speak more to the learning of associations and contingencies, as opposed to the structure or architecture of the models used for associative learning. Having said this, installing causal or statistical structure into the architecture of the brain lies at the heart of structure learning: c.f.,., the good regulator theorem (Conant and Ashby 1970, Seth 2014).

In the pragmatic world of complex system modelling and data analysis, structure learning is usually understood in a particular and specific sense: namely, the selection or discovery of network architectures apt for providing an accurate and efficient account of some data. On this view, structure learning reduces to Bayesian Model Selection, defined as the selection of the model that maximises model evidence — or renders the data most likely under the models considered. Technically, this means that the structure learning problem is the problem of discovering a model that maximises the marginal likelihood of the data explained by the model (the marginal likelihood marginalises over the unknown parameters of any given model). This view of structure learning rests on the ability to score, or compare, models in terms of their evidence. Typically, various estimates of evidence are used: for example, cross validation accuracy, sampling techniques, information criteria, and approximate (i.e., variational) Bayesian inference. Of these, the most efficient rests on variational approximations — i.e., evidence bounds (Winn and Bishop 2005) — on model evidence that can be evaluated quickly and efficiently, because one assumes a functional form for the requisite probability distributions. Practically, this kind of Bayesian Model Selection is used routinely in network discovery in complex systems; ranging from neuronal networks themselves (Friedman and Koller 2003, Penny, Stephan et al. 2004, Penny 2012, Seghier and Friston 2013), through to epidemiological modelling (Friston, Flandin et al. 2022). In the next chapter, we will see that

this narrow definition of structure learning is leveraged in the Active Inference Framework; enabling the process of Bayesian Model Selection to be linked to the developmental and psychological formulations above.

Associating structure learning with Bayesian Model Selection raises important questions about the processes that identify candidate models. Normally, most procedures used in practice can be regarded as a form of greedy search; namely, accepting or rejecting a new model based upon whether its evidence increases or decreases (given the same data). In this setting, the exploration of model space is, in and of itself, structured — in the sense that new models are obtained from old models by various operations such as deletion, mutation and insertion – c.f., genetic algorithms (Kitson, Constantinou et al. 2023). For example, one can consider a model with an extra component or element and ask whether the increase in model complexity is warranted by the increase in model accuracy — by comparing the evidence for the old and new models. In some situations, one can place priors over adding a new component, in the spirit of nonparametric Bayes (Gershman and Blei 2012, Collins and Frank 2013) or species discovery (Efron and Thisted 1976, Friston, Da Costa et al. 2023). In the next chapter, we will pursue a biomimetic perspective and consider Structure Learning as an aspect of maximising the evidence for generative world models, under the Active Inference Framework.

1.4 Summary

Structure learning and learning structure appear to be universal features of information processing in humans and animals, and are of ongoing interest for theoretical (e.g., agent based) models of biological computation. Currently, scholars from various fields employ the term *structure learning* to describe a range of phenomena and frameworks: *statistical learning* (Vapnik 1999, Karuza, Emberson et al. 2017, Monroy, Meyer et al. 2017, Emberson, Misyak et al. 2019, Monroy, Meyer et al. 2019), *adaptive generalisation* (Donchin, Francis et al. 2003, Braun, Aertsen et al. 2009, Mulavara, Cohen et al. 2009), *learning to learn* (Braun, Mehring et al. 2010, Wang, Kurth-Nelson et al. 2016, Behrens, Muller et al. 2018), *concept learning* (Smith, Schwartenbeck et al. 2020), and *probabilistic learning* (Griffiths and Tenenbaum 2006, Kemp, Tenenbaum et al. 2010, Lake, Salakhutdinov et al. 2015). Conversely, other fields employ *structure learning* mechanisms and ideas without directly referring to the process as such, for researching topics such as *predictive motor activation* (Ghilardi, Meyer et al. 2023), *motor imagery* (Doyon and Benali 2005, Di Rienzo, Debarnot et al. 2016), *causal reasoning* (Goldvarg and Johnson-Laird 2001, Steyvers, Tenenbaum et al. 2003, Gopnik, Glymour et al. 2004, Kemp and Tenenbaum 2009, Meder, Hagmayer et al. 2009, Lagnado, Gerstenberg et al. 2013), *abstract reasoning* (Jung-Beeman, Bowden et al. 2004), *replay* (Deuker, Olligs et al. 2013, Eysenbach, Salakhutdinov et al. 2019, Tambini and Davachi 2019, Wittkuhn, Chien et al. 2021, Stoianov, Maisto et al. 2022, Widloski and Foster 2022, Kurth-Nelson, Behrens et al. 2023), and *deep (meta) reinforcement learning* (Mnih, Kavukcuoglu et al. 2015, Silver, Huang et al. 2016, Wang, Kurth-Nelson et al. 2016, Hu, Ma et al. 2021).

Altogether, these frameworks provide accounts of this phenomenon from different perspectives, with different uses and goals, which themselves provide different predictions. That is, properties and mechanisms of structure learning vary widely depending upon the

specific framework in question. However, it is possible to distil these accounts into a simplified and unified definition that takes into consideration the various accounts of structure learning from the literature discussed so far. Structure learning can be thought of as the ability to internalise a model of contingencies (i.e., conditional [in]dependencies) among different elements in space-time, that can be generalised and leveraged with ease.

This definition, while inclusive of the perspectives discussed thus far, presents as exceedingly generic and falls short of providing a (unified) mechanistic account. In what follows, we will come to see that under the Active Inference Framework, Structure Learning has an operationally defined computational form (Friston, Parr et al. 2018). We will provide an extensive and specific definition of Structure Learning at the end of Chapter 2, which foregrounds and embeds Structure Learning within the Active Inference Framework (AIF), and touches upon the various accounts of structure learning in the general literature in light of the AIF.

Chapter 2

The three main levels of optimisation in the Active Inference Framework

Partially based on:

Structure learning enhances concept formation in synthetic Active Inference agents (Neacsu, Mirza et al. 2022)

Synthetic spatial foraging with Active Inference in a geocaching task (Neacsu, Convertino et al. 2022)

In this chapter, I introduce the Active Inference Framework (AIF) and establish the notion of Structure Learning under Bayesian Model Selection (BMS), and more specifically, as implemented by Bayesian Model Reduction (BMR). The basic idea underlying the Active Inference Framework is that agents are inference machines that minimise (variational) free energy, or equivalently, maximise model evidence. This can be interpreted as self-evidencing (Hohwy 2016); namely, minimising uncertainty about the environment (Friston, FitzGerald et al. 2017). Implicit in the AIF formulation is a (generative or world) model of the environment in the form of beliefs or joint probability distributions that encode contingencies in the world. The environment can be thought of as being either extrapersonal (Kaplan and Friston 2018) or the body (Seth and Friston 2016), or both (Hesp, Smith et al. 2021).

There are at least three very distinct kinds of optimisation in the AIF landscape: *(Active) Inference* (AI)*, Parametric Learning* (PL)*,* and *Structure Learning* (SL) (Friston, Moran et al. 2021)*.* At the fastest timescale, we have the (Active) Inference of hidden or latent causes (i.e., states and/or policies); we can think of this process as that of performing approximate Bayes (Sunnåker, Busetto et al. 2013). At a slower timescale, we have Parametric Learning. This type of learning has been associated with evidence accumulation via a mathematical process that mimics Hebbian learning (Friston, FitzGerald et al. 2016, Friston, FitzGerald et al. 2017, Friston, Lin et al. 2017). At the slowest timescale, we have Structure Learning. In the AIF, this type of optimisation has been associated with synaptic homeostasis and synaptic pruning (Kiebel and Friston 2011, Hobson and Friston 2012, Friston, Lin et al. 2017, Hobson, Gott et al. 2021).

Briefly speaking, the AIF agent uses sensory data to update its beliefs about latent states and the most likely policies (i.e., actions, sets of actions, or plans) it should pursue. This is known as (planning as) inference. Since perception and action are optimised in tandem, AIF agents hold and optimise beliefs about their own behaviour. They select actions from the

posterior beliefs about policies (e.g., plans), whereby a new observation is solicited, in line with the goal of fulfilling prior preferences and resolving uncertainty. The (variational) inference process in AIF can therefore be thought of as optimising posterior beliefs about the causes of sensorial experience for past, present, and future (hidden) states, based on observations, and depending upon the pursuit of specific policies – a.k.a. *(Active) Inference* (Friston, FitzGerald et al. 2017).

The process known as *Parametric Learning* (PL) involves the optimisation of beliefs about relationships implicit in the interaction between different (latent) variables in the environment, where actions are chosen to resolve uncertainty about the parameters of a generative model. These parameters can encode beliefs (usually — in discrete state space models — as concentration parameters) about likelihood (of outcomes given states), transitions (among states), preferences (for outcomes), initial states, and policies. In AIF literature, these model parameters are usually called **A**, **B**, **C**, **D**, and **E** arrays or tensors, respectively. Please see Table 2.1 for a glossary of terms.

Table 2.1 Glossary of terms

Notation/Term	Meaning
$o_\tau \in \{0,1\}$ $\mathbf{o}_\tau \in [0,1]$ $\widehat{\mathbf{o}}_\tau = \ln \mathbf{o}_\tau$	Outcomes, their posterior expectations and logarithms
$\tilde{o} = (o_1, ...o_t)$	Sequences of outcomes until the current time point
$s_\tau \in \{0,1\}$ $\mathbf{s}_\tau^\pi \in [0,1]$ $\widehat{\mathbf{s}}_\tau^\pi = \ln \mathbf{s}_\tau^\pi$	Hidden states and their posterior expectations and logarithms, conditioned on each policy
$\tilde{s} = (s_1, ..., s_T)$	Sequences of hidden states until the end of the current trial

$\pi = (\pi_1, \ldots, \pi_K) : \pi \in \{0,1\}$ $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K) : \boldsymbol{\pi} \in [0,1]$ $\widehat{\boldsymbol{\pi}} = \ln \boldsymbol{\pi}$	Policies specifying action sequences, their posterior expectations and logarithms
$u = \pi(t)$	Action or control variables for each factor
$\gamma, \boldsymbol{\gamma} = 1/\boldsymbol{\beta}$	The precision (inverse temperature) of beliefs about policies and its posterior expectation
β	Prior expectation of temperature (inverse precision) of beliefs about policies
$\mathbf{A} \in [0,1]$ $\widehat{\mathbf{A}} = \psi(\mathbf{a}) - \psi(\mathbf{a}_0)$	Likelihood matrix mapping from hidden states to outcomes and its expected logarithm
$a^m \in \mathbb{R}$ $\mathbf{a}^m \in \mathbb{R}$ $\mathbf{a}'^m \in \mathbb{R}$ $a'^m \in \mathbb{R}$	Prior concentration parameters of the likelihood Posterior concentration parameters of likelihood Reduced posterior of the likelihood Prior concentration parameters for the reduced model (of the likelihood)
$\mathbf{B}_\tau^\pi = \mathbf{B}(u = \pi(\tau)) \in [0,1]$ $\widehat{\mathbf{B}}_\tau^\pi = \ln \mathbf{B}_\tau^\pi$	Transition probability for hidden states under each action prescribed by a policy at a particular time, and their logarithms
$\mathbf{C} := \widehat{\mathbf{B}}_\tau^0 \in [0,1]$ $\widehat{\mathbf{C}} = \ln \mathbf{C}$	Transition probability for hidden states under a habit and their logarithm
$\mathbf{U}_\tau = \ln P(o_\tau)$	Logarithm of prior preference or utility over outcomes
$\mathbf{D} \in [0,1]$	Prior expectation of each state at the beginning of each trial
$\mathbf{E} \in [0,1]$	Prior expectation of each policy at the beginning of each trial
Q	Approximate posterior distribution over the latent causes of the generative model – e.g. $\mathbf{s}$, A, π
$\mathbf{F} : \mathbf{F}_\pi = F(\pi) = \sum_\tau F(\pi, \tau) \in \mathbb{R}$	Variational free energy for each policy
$\mathbf{G} : \mathbf{G}_\pi = G(\pi) = \sum_\tau G(\pi, \tau) \in \mathbb{R}$	Expected free energy for each policy
$\mathbf{H} = -diag(\breve{\mathbf{A}} \cdot \widehat{\mathbf{A}})$	The vector encoding the entropy or ambiguity over outcomes for each hidden state
$\mathbf{s}_t = \sum_\pi \boldsymbol{\pi}_\pi \cdot \mathbf{s}_t^\pi$	Bayesian model average of hidden states over policies
$Cat(\mathrm{A})$ $Dir(a)$	Categorical and Dirichlet distributions, defined in terms of their sufficient statistics (probabilities and concentration parameters)
$\sigma(-\mathbf{G})_\pi = \frac{\exp(-\mathbf{G}_\pi)}{\sum_\pi \exp(-\mathbf{G}_\pi)}$	Softmax function, returning a vector that can be treated as a proper probability distribution

$\widehat{\mathbf{A}} = E_Q[\ln \mathbf{A}] = \psi(\mathbf{a}) - \psi(\mathbf{a}_0)$ $\breve{\mathbf{A}} = E_Q[\mathbf{A}_{ij}] = \mathbf{a} \times \mathbf{a}_0^{-1}$ $\mathbf{a}_{0ij} = \sum_i \mathbf{a}_{ij}$	Expected outcome probabilities for each hidden states and their expected logarithms
Bayesian surprise	A measure of salience based on the (Kullback-Leibler) divergence between the recognition and prior densities. It measures the information in the data that can be recognised.
Conditional density/posterior density	The probability distribution of causes or model parameters, given some data; i.e., a probabilistic mapping from observed data (consequences) to causes.
(Kullback-Leibler) Divergence	Information divergence, information gain or relative entropy is a non-commutative measure of the difference between two probability distributions.
Empirical prior	Priors that are induced by hierarchical models; they provide constraints on the recognition density is the usual way but depend on the data.
Entropy	The average surprise of outcomes sampled from a probability distribution or density. A density with low entropy means, on average, the outcome is relatively predictable (certain).
Generative model	A probabilistic mapping from causes to observed consequences (data). It is usually specified in terms of the likelihood of getting some data given their causes (parameters of a model) and priors on the parameters
Gradient descent	An optimisation scheme that finds a minimum of a function by changing its arguments in proportion to the negative of the gradient of the function at the current value.
Precision	The inverse variance or dispersion of a random variable. The precision matrix of several variables is also called a concentration matrix. It quantifies the degree of certainty about the variables
Prior	The probability distribution or density on the causes of data that encode beliefs about those causes prior to observing the data.

Surprise	Surprisal or self-information is the negative log-probability of an outcome. An improbable outcome is therefore surprising.
Uncertainty	A measure of unpredictability or expected surprise (c.f., entropy). The uncertainly about a variable is often quantified with its variance (inverse precision).

Further to minimising uncertainty about hidden states and parameters, agents also minimise uncertainty about their generative models per se, also known as Structure Learning (SL). Generative models are essentially alternative hypotheses about the potential causes that generate the agent's observations. With Structure Learning, one considers competing hypotheses about these causes. Agents can therefore minimise uncertainty about their model based on (Bayesian) model comparison, where the winning model becomes the hypothesis under which observed outcomes are the least surprising – i.e., the most likely hypothesis (having reduced all other types of uncertainty). In order to optimise the other types of uncertainty (i.e., about latent states or parameters), agents need sensorial (or factual) information, meaning that experience is needed. On the other hand, Bayesian Model Selection (e.g., Bayesian Model Reduction, Bayesian Model Expansion), operates in the absence of further sensory experience, since it proceeds by best explaining the experiences accumulated up until that point in time.

In the Active Inference Framework, extrinsic (i.e., pragmatic) value and intrinsic (epistemic) value (i.e., novelty and salience) are optimised simultaneously. This is the case because policy selection is based on Expected Free Energy, which itself implies a dual pursuit: that of utility maximisation and maximising information gain (Friston, Lin et al. 2017). These complementary imperatives are combined into a single objective function (Expected Free

Energy), such that the pragmatic and epistemic imperatives contextualise each other to provide the right balance of 'exploit' and 'explore' behaviour. Typically, in a novel setting, the behaviour of AIF agents is dominated by exploration until the epistemic values of available policies fall as uncertainty is resolved, at which point extrinsic value starts to dominate, manifesting as 'exploit' type behaviour. The degree to which agents explore therefore depends on the precision of their prior preferences that underwrite goals. Note that because these (extrinsic and intrinsic) values are (log) probabilities, their combination in Expected Free Energy corresponds to a multiplication of probabilities. This means that a policy only has intrinsic value (i.e., information-seeking value) provided that it has a non-trivial extrinsic value (i.e., goal-seeking value).

Optimality – in the current Bayesian context – includes the joint principles of optimal Bayesian decision-making under uncertainty, and the principles of optimal Bayesian design. This is most clearly seen in terms of the two parts of the Expected Free Energy objective function that can be described in terms of intrinsic motivation (value) – that scores the exploratory aspect of optimal behaviour – and the second part, which is the extrinsic motivation (value) that can be read as minimising the expected loss or maximising expected reward. Optimality entails both maximising expected (or extrinsic) rewards, and minimising uncertainty (or maximising information gain). By definition and construction, our AIF agents are optimal in this sense. For a more extensive account of AIF and associated tenets, please see (Friston, FitzGerald et al. 2016, Friston, FitzGerald et al. 2017, Friston, Parr et al. 2017, Friston, Parr et al. 2018, Smith, Friston et al. 2022).

In what follows, I briefly review inference, learning (i.e., Parametric Learning) and model selection (i.e., Structure Learning) in terms of belief updating that minimises Variational and Expected Free Energies.

2.1 (Active) Inference

The first equation below describes the process of *inference* as the minimisation of Variational Free Energy – also known as an evidence bound (Winn, Bishop et al. 2005) – with regards to the sufficient statistics of an approximate posterior distribution over the hidden causes x (representing hidden states s, and policies π):

$$Q(x) = \arg\min_{Q(x)} F \approx P(x \mid \tilde{o}) \qquad \textit{Variational free energy} \quad (1)$$

$$F = E_Q[\ln Q(x) - \ln P(\tilde{o} \mid x) - \ln P(x)] \qquad (1.1)$$

$$= E_Q[\ln Q(x) - \ln P(x \mid \tilde{o}) - \ln P(\tilde{o})] \qquad (1.2)$$

$$= \underbrace{D_{KL}[Q(x) \,||\, P(x \mid \tilde{o})]}_{\textit{relative entropy}} - \underbrace{\ln P(\tilde{o})}_{\textit{log evidence}} \qquad (1.3)$$

$$= \underbrace{D_{KL}[Q(x) \,||\, P(x)]}_{\textit{complexity}} - \underbrace{E_Q[\ln P(\tilde{o} \mid x)]}_{\textit{accuracy}} \qquad (1.4)$$

Where $\tilde{o} = (o_1, ..., o_t)$ designates observed outcomes up until the current time. These equations can be regarded as specifying the process of perception. They show that minimising Variational Free Energy brings the Bayesian beliefs closer to the true posterior beliefs by minimising the relative entropy term (that is never less than zero). This is equivalent to forming beliefs about hidden states of affairs that provide an accurate and parsimonious – complexity minimising – explanation of observed outcomes. Complexity here is simply the difference between posterior

and prior beliefs, i.e., the degree to which one 'changes one's mind' when updating prior to posterior beliefs.

Action and planning are usually formulated as selecting the action (from the most plausible policy) that has the least Expected Free Energy:

$$\pi^* = \arg\min_\pi = \sum_\tau G(\pi, \tau) \qquad \textit{Expected free energy} \ (2)$$

$$G(\pi,\tau) = E_{\tilde{Q}}[\ln Q(\mathbf{A}, s_\tau \mid \pi) - \ln P(\mathbf{A}, s_\tau, o_\tau \mid \tilde{o}, \pi)] \quad (2.1)$$

$$= \underbrace{E_{\tilde{Q}}[\ln Q(\mathbf{A}) - \ln Q(\mathbf{A} \mid s_\tau, o_\tau, \pi)]}_{\text{(Negative) novelty}} + \underbrace{E_{\tilde{Q}}[\ln Q(o_\tau \mid \pi) - \ln Q(o_\tau \mid s_\tau, \pi)]}_{\text{(Negative) salience}} - \underbrace{E_{\tilde{Q}}[\ln P(o_\tau)]}_{\text{Extrinsic value}} \quad (2.2)$$

Where $\tilde{Q} = Q(o_\tau, s_\tau \mid \pi) = P(o_\tau \mid s_\tau) Q(s_\tau \mid \pi)$. This set of equations identifies the best policy and accompanying action at the next time step. Notice that this kind of planning – based upon Expected Free Energy – involves averaging the free energy expected following a policy under the predicted outcomes. This means that the expected accuracy becomes extrinsic value, namely, the extent to which outcomes conform to prior preferences. In economics, this term is known as utility (Fishburn 1970), and in behavioural psychology, it corresponds to reward (Barto, Mirolli et al. 2013, Cox and Witten 2019). Similarly, the expected relative entropy becomes an information gain pertaining to unknown model parameters (labelled novelty) and unknown hidden states (labelled salience). These are sometimes referred to as intrinsic or epistemic values, and form the basis of artificial curiosity (Schmidhuber 2006, Ngo, Luciw et al. 2012, Schillaci, Pico Villalpando et al. 2020). They quantify the value of the evidence accumulated if agents were to pursue a particular plan. Maximising these intrinsic values can

be seen as a form of optimal information gain or active learning (MacKay 1992, Oudeyer and Baranes 2008, Baranes and Oudeyer 2013), where curiosity resolves uncertainty about states of the world and their contingencies – in accord with the principles of optimum Bayesian experimental design (Lindley 1956).

Whereas salience is associated with beliefs about the current state of affairs in the world, and how they will unfold in the future, novelty is the reducible (epistemic) uncertainty about the probabilistic contingencies themselves, and the causal structure they entail (i.e., the causal structure of the environment). In other words, novelty affords the opportunity to resolve uncertainty about what would happen if agents engaged in a specific course of action (i.e., 'what would happen if I did this*?*'). An alternative way of decomposing the Expected Free Energy is into expected (in)accuracy and complexity – that can be understood as ambiguity and risk; namely, the uncertainty that pertains to ambiguous outcomes*,* and the risk that actions will bring about outcomes that diverge from prior preferences.

2.2 Parametric Learning

Parametric Learning can be thought of as resolving uncertainty about (generative) model parameters. AIF agents have implicit priors (e.g., **A**) and hyper-priors (i.e., concentration parameters, e.g., *a*) encoding beliefs about model parameters (Friston, FitzGerald et al. 2016). Since parametric beliefs (e.g., **A**) are represented as categorical distributions, a suitable hyper-prior encoding the mapping between relevant couplings (e.g., state-outcome) is specified in terms of Dirichlet concentration parameters. Given a state *s*, the belief about the probability of an outcome *o* is:

$$P(o \mid s, \mathbf{A}) = Cat(\mathbf{A}) \tag{3}$$

$$P(\mathbf{A} \mid a) = Dir(a) \Rightarrow \begin{cases} E_{P(\mathbf{A}|a)}\left[\mathbf{A}_{ij}\right] = \dfrac{a_{ij}}{\sum_k a_{kj}} \\ E_{P(\mathbf{A}|a)}\left[\ln \mathbf{A}_{ij}\right] = \psi(a_{ij}) - \psi\left(\sum_k a_{kj}\right) \end{cases} \tag{4}$$

Where ψ represents the digamma (logarithmic derivative of gamma) function – please see (Friston, FitzGerald et al. 2016) and (Smith, Friston et al. 2022) for further details. Agents accumulate Dirichlet parameters as they are exposed to new observations, allowing them to learn. The updates over these (e.g., likelihood) parameters involve accumulating the Dirichlet parameters that represent the mapping from hidden states to the observed outcome. For example, updates to the concentration parameters of the likelihood mappings are defined as:

$$\mathbf{a} = a + \eta \times \sum_\tau \mathbf{s}_\tau \otimes o_\tau \tag{5}$$

Where a and $\mathbf{a}$ represent prior and posterior concentration parameters respectively, $\mathbf{s}_\tau$ corresponds to the posterior expectations about the hidden states, and η corresponds to the learning rate. The $\otimes$ sign indicates an outer product. In other words, learning involves accumulating Dirichlet counts, modulated by the learning rate. What is being counted here are co-occurring instances between states and observations: for each co-occurrence of a given state-outcome instance, a count is added to the concentration parameters.

Since accumulating (Dirichlet) concentration parameters (e.g., over the likelihood) is equivalent to the type of change observed in synaptic (Hebbian) plasticity (Brown, Zhao et al. 2009, Friston, FitzGerald et al. 2017), this specific type of update can be thought of as a

synaptic strengthening, every time neurons encoding states and observations (coupled by that synapse) are active simultaneously. Parametric Learning as implemented in the Active Inference Framework is then a mathematical description of Hebbian associative plasticity. Note that in this particular example, noisy mappings (of the likelihood mappings) would correspond to an imprecise likelihood array, where AIF agents would make inferences under observational uncertainty, such as being in a dimly lit room.

2.3 Structure Learning and Bayesian Model Reduction

The previous two sections summarised the computational processes entailed by online and active engagement with the environment. We now turn to a different type of learning: learning the structure of a model between periods of active engagement with the environment – with the purpose of optimising models of the world – using a quantity called marginal likelihood (a.k.a. model evidence). This type of learning is operationalised as Structure Learning (SL). Structure learning therefore ensues in the absence of additional (sensorial) evidence. The mentioned periods of active engagement can vary between having just one instance of evidence accumulation using, for example (Active) Inference and Parametric Learning; to having prolonged periods of interactions with the environment followed by Structure Learning.

Generally speaking, in the AIF landscape, Structure Learning comes in two flavours. One approach to Structure Learning involves reinverting models using fictive data (i.e., data simulated as a result of applying a generative model to a sequence of behavioural observations and actions) – such as in Dynamic Causal Modelling, or fitting (Friston and Penny 2011), and then scoring these models in terms of their model evidence or the free energy bound on the model evidence (more details about this approach to follow in section 5.5). The other approach

involves applying Bayesian Model Reduction (BMR) or Bayesian Model Expansion (BME); these types of SL offer a more expedited method by circumventing the necessity of reinversion (of models) due to a priori knowledge of their functional forms. In other words, after experiencing a sequence of events, one approach involves post-hoc re-simulation of the entire sequence of observations and actions under different alternative models and accumulating evidence for different models based on individual events; whilst the other approach involves comparing the (posterior) concentration parameters (i.e., model) with alternative explanations (i.e., models) in terms of their model evidence.

In light of the AIF, Structure Learning is therefore synonymous with Bayesian Model (comparison and) Selection, and benefits from its computational formulations, which includes the two approaches mentioned above (Friston and Penny 2011, Friston, Lin et al. 2017, Friston, Parr et al. 2018, Smith, Schwartenbeck et al. 2020, Neacsu, Mirza et al. 2022). Bayesian Model Selection entails the comparison and selection of models with the greatest model evidence (i.e., least free energy, greatest marginal likelihood) (Friston and Penny 2011). As a result, Structure Learning can be thought of as a form of model selection, where agents (synthetic or biological) compare and assess alternative hypotheses (i.e., models) defined by different (prior) configurations of their generative models (Friston, Parr et al. 2018), evaluating them against a single objective function. Since in the AIF these (e.g., likelihood) mappings implicitly encode connection strengths, the reorganisation may entail the removal of existent 'synaptic' connections, or coupling of otherwise non-existent 'synaptic' connections. Although the capacity for such connections exists in the architecture of the generative model itself, the connections themselves are not hard-coded. Synthetic agents may perform an exhaustive search over the hypothesis space, by considering all the associations found in the realm of possible combinatorics (of the specific elements or features involved). To illustrate, consider the following example. An agent is looking at two distinct faces over several trials, and there are

two possible emotions being conveyed (e.g., happiness, sadness). The agent starts with uniform beliefs. After a few trials, the agent 'believes' that face 1 is 'happy' and face 2 is 'sad'. If it engages in Bayesian Model Selection with the current posterior beliefs, it can compare the current hypothesis against the hypothesis that face 1 is 'sad' and face 2 is 'happy'. If retrospectively, there is more evidence for the second hypothesis, then its associated connectivity changes and the trials resume with this hypothesis (i.e., model) instead. This means that although the capacity for this specific belief structure was there, there was no connectivity between face 1 and 'sad' just before BMS. In this sense, Bayesian Model Selection (e.g., reduction, expansion) involves reorganisation.

Although the work in this thesis largely concerns Structure Learning as implemented by BMR, we will see in Chapter 5 that Structure Learning will be implemented in terms of fitting behavioural responses (i.e., BMS more generally, using reinversion and re-simulation of the observed events) with models that include or disallow Structure Learning (here BMS implemented with BMR). For now, we can expand on the formal definition of Structure Learning as implemented by BMR. Bayesian Model Reduction involves applying Bayes' rule to full (i.e., original) and reduced (i.e., alternative) models, and evaluating the change in free energy (i.e., log Bayes factor or model evidence ratio) for each model. It is a post-hoc optimisation, and it is applied to posterior beliefs (i.e., after all the data have been 'seen'). Essentially, BMR refines the agents' current beliefs based on comparing alternative models defined in terms of their priors. Applying BMR reduces model complexity by changing the probability mass of the accumulated beliefs such that the precision of valid (i.e., informative, likely) contingencies is increased, and the precision of any redundant parameters is decreased or eliminated altogether.

The relative evidence for a full (i.e., original, prior) and alternative (i.e., reduced) model with priors *a'* can be derived by applying Bayes' rule to all models. For example, with two models, original prior vs. alternative, this is illustrated as follows:

$$\frac{P(\mathrm{A} \mid \tilde{o}, m_{alt})}{P(\mathrm{A} \mid \tilde{o}, m_{full})} = \frac{P(\mathrm{A} \mid m_{alt})P(\tilde{o} \mid m_{full})}{P(\mathrm{A} \mid m_{full})P(\tilde{o} \mid m_{alt})}; \tag{6.1}$$

$$\frac{P(\tilde{o} \mid m_{full})}{P(\tilde{o} \mid m_{alt})} = \int P(\mathrm{A} \mid \tilde{o}, m_{full}) \frac{P(\mathrm{A} \mid m_{alt})}{P(\mathrm{A} \mid m_{full})} dA \approx \int Q(\mathrm{A}) \frac{P(\mathrm{A} \mid a')}{P(\mathrm{A} \mid a)} dA \tag{6.2}$$

$$= \frac{\mathrm{B}(a)\mathrm{B}(\mathbf{a} + a' - a)}{\mathrm{B}(\mathbf{a})\mathrm{B}(a')} \tag{6.3}$$

$$P(\mathrm{A} \mid a) = Dir(a) = \mathrm{B}(a) \prod_i \mathrm{A}_i^{a-1} \tag{6.4}$$

For a full derivation, please see (Friston, Parr et al. 2018). In equations 6.3 and 6.4, $\mathrm{B}(\cdot)$ denotes the multivariate beta function. The evidence ratio in equation 6.2 may now be expressed as the change (i.e., increase or decrease) in free energy as following:

$$\Delta F = \ln P(\tilde{o} \mid m_{full}) - \ln P(\tilde{o} \mid m_{alt}) \tag{7.1}$$

$$= \ln \mathrm{B}(\mathbf{a}) + \ln \mathrm{B}(a') - \ln \mathrm{B}(a) - \ln \mathrm{B}(\mathbf{a}') \tag{7.2}$$

And

$$\mathbf{a}' = \mathbf{a} + a' - a \tag{7.3}$$

Where **a'** is the reduced posterior, **a** represents the posterior concentration parameters, *a'* represents the prior concentration parameters defining an alternative (e.g., reduced) model, and

a represents the prior concentration parameters defining the full (i.e., original) model. Note the simplicity of these (local) update rules and their implicit biological plausibility (Friston, Parr et al. 2018). The equalities in equations 7.2 and 7.3 allow ΔF to be computed in a biologically plausible way that underwrites synaptic regression or pruning. In other words, the change in free energy that would have been observed under alternative hypotheses (i.e., alternative/reduced models) can be used to remove or retain certain connections depending upon whether the free energy bound on model evidence increases or decreases. This measure is therefore used to either accept or reject alternative hypotheses (as defined by their concentration parameters). Usually, redundant parameters are removed when $\Delta F \leq -3$, corresponding to a Bayes factor approximately equivalent to 0.05, meaning that the selected (alternative) model is 20 times more likely than the original (full) model (Friston, Parr et al. 2018).

The alternative (i.e., reduced) posteriors that emerge from the equations above – if any alternative model is accepted – are defined as follows:

$$Q(\mathrm{A} \mid m_{alt}) = \mathrm{B}(\mathbf{a}')^{-1} \prod_i \mathrm{A}_i^{\mathbf{a}'-1} = Dir(\mathbf{a}') \tag{8}$$

Bayesian Model Reduction is an efficient (and analytic) off-the-shelf procedure that scores alternative models – in terms of model evidence and accompanying posterior – given the priors and the posterior under a parent (i.e., original, full) model. For more technical details, please see (Friston, Parr et al. 2018). The above equations (illustrating the optimisation of likelihood matrices) describe BMR for Dirichlet processes that are apt for the models used in

this thesis. Please see table 1 in (Friston, Parr et al. 2018) for equations corresponding to other kinds of (probabilistic) distributions.

2.4 Features of structure learning from the general literature in relation to the AIF

In this section, I will discuss how various features of structure learning as observed in the literature at large relate to the Active Inference Framework of information processing and belief update, with its three implicit levels of optimisation.

2.4.1 AIF and Bayesian surprise

In sub-section 1.1.1. we saw evidence of structure learning in infants as young as 6 months old (Emberson, Richards et al. 2015, Friston, FitzGerald et al. 2016, Monroy, Gerson et al. 2019). Commonly, this research is carried out using tasks that include expectation-violation features; that is, structure learning is shown via evidence of learning of (statistical) structures of the experienced world, which entails internalising (a model of) expectations that are then violated through omission, or through the addition of elements that were not present in the first instance. In the AIF landscape, this violation of expectations has been defined in terms of Bayesian surprise or surprisal (Friston, Rigoli et al. 2015, Friston, Lin et al. 2017) and does not automatically invoke BMS, but its form does involve the implicit capacity for Parametric Learning. Surprise in the AIF landscape is defined as the (negative) log probability of (sensory) observations/outcomes (e.g., improbability of sensory inputs), and the impetus is for it to be minimised (Friston, Rigoli et al. 2015, Isomura 2022). This measure is the same measure

defined earlier in equation 2.2 as information gain (decomposed in terms of novelty and salience), and it has also been referred to as epistemic value or mutual information (Friston, Rigoli et al. 2015, Mirza, Adams et al. 2019). In Chapter 4, we will see that it is possible to compare this metric (information gain) not only between trial-to-trial instances of Parametric Learning, but also between Parametric and Structure Learning.

Furthermore, a byproduct of statistical learning – i.e., the learning of structure as interpreted by (Monroy, Gerson et al. 2019, Monroy, Meyer et al. 2019) for example – was the predictive motor activation hypothesis, where the knowledge built through action observation during infancy generates action-prediction signals in the motor system. Interestingly, this hypothesis is supported in part by work in silico using the AIF (Friston, Mattout et al. 2011), which suggests that a common (generative) model underlies the perception (e.g., prediction) of both self and other's actions. The differentiating factor between the two comes in the form of precision over proprioceptive (i.e., sensory) signals. Here again, Structure Learning is not a prerequisite for the learning of structures, although it can be applied to the scheme by implementing concentration parameters over the relevant contingencies.

2.4.2 AIF, habit formation, and motor learning

In sub-section 1.1.2, research on motor imagery suggested that motor imagery enhances motor learning (Doyon and Benali 2005, Di Rienzo, Debarnot et al. 2016, Toth, McNeill et al. 2020). Motor imagery here was defined as the (mental) simulation of actions, which can occur both online (Mizuguchi and Kanosue 2017) and offline (Doyon and Benali 2005, Blischke and Erlacher 2007, Hill, Tononi et al. 2008). In terms of the AIF landscape, online motor imagery can be associated with habit formation, where learning occurs over the vector that encodes a probability distribution over the set of available policies, **E** (Friston, FitzGerald et al. 2016,

Maisto, Friston et al. 2019, Smith, Ramstead et al. 2022). For example, (Friston, FitzGerald et al. 2016), where agents are equipped with a hyperprior over **E**, illustrates how habits develop spontaneously through sequential policy optimisation. Numerical experiments using the AIF are belief-based, in that they entail *beliefs* about state-action mappings rather than just state-action mappings observed for example in Reinforcement Learning settings (Dayan and Watkins 2002). This allows habits to essentially be learnt through the observation of 'one's own goal directed behaviour' (Friston, FitzGerald et al. 2016).

In light of these in silico experiments with the AIF (Friston, FitzGerald et al. 2016, Maisto, Friston et al. 2019), learning as a result of online motor imagery (i.e., habit formation) can therefore be accounted for without the need for Structure Learning (as BMS). Recall that Parametric Learning occurs online and necessitates continual sensory evidence to learn - by allowing concentration parameters (e.g., over **E**) to accumulate over trials; the habit is acquired when the (relative) posterior probability of the habit has risen sufficiently so that the behaviour is routinely executed. However, offline motor learning, which occurs in the absence of further practice does require that a mechanism – such as the one implied by BMS – is available. Explaining phenomena such as motor learning therefore requires both Parametric Learning (in the case of online motor learning), as well as something beyond associative (Parametric) learning (for offline motor learning). Since some offline motor learning involves evaluating the learning post-sleep (Blischke and Erlacher 2007, Hill, Tononi et al. 2008), and Structure Learning has been linked to consolidation during sleep (Hobson and Friston 2012, Friston, Lin et al. 2017), this makes SL (as BMS) a good future candidate for explaining the processes involved in offline motor learning.

2.4.3 AIF and the dual nature of learning

Multiple other sub-sections discussed the dual nature of learning (of contingencies). In meta-learning or learning to learn (sub-sections 1.1.3, 1.2, and 1.3.1) these learning processes have been explicitly referred to as Parametric and Structure (or structural) Learning (Braun, Mehring et al. 2010). The two types of learning have direct equivalents and similar definitions to their counterparts in the AIF landscape. In terms of sensorimotor control for example, the two types of learning reflect the difference between adapting representations by steady evidence accumulation and its implicit belief-updating (i.e., parametric adaptation, parametric learning); and forming new belief structures by associating the existing elements in novel ways (i.e., structural learning). Other accounts (sub-section 1.1.5) consider this dual aspect more generally, differentiating between the analytic, incremental approach to problem-solving, and the insight-based approach (Jung-Beeman, Bowden et al. 2004, Fleck and Weisberg 2013, Webb, Little et al. 2016), which are in line with the phenomenological manifestation of (gradual) Parametric Learning and (sudden) Structure Learning, respectively – in the AIF.

This dual aspect of learning is also found within some examples of deep learning, such as in the study where in silico agents solved Raven's Progressive Matrices – sub-section 1.3.2 (Hu, Ma et al. 2021). In this work, one part involved learning the rule embeddings using a neural network called SRAN for each of the 9 matrices (8 completed matrices, and the (partial) original matrix). The other part involved selecting the correct answer by comparing the (partial) original matrix with each of the completed matrices and selecting the answer where the rule embedding was most similar to the rule embedding of the (partial) original matrix – that is, selecting the rule embedding with maximal probability under the prior embedding of the (partial) original matrix. This (latter) process is reminiscent of the model selection (implemented post-hoc using BMS) found in the AIF: the similarity score is equivalent to the

model evidence estimate when comparing alternative models (i.e., hypotheses) with BMR for example.

2.4.4 AIF and causal reasoning

In causal reasoning (a.k.a. causal learning – sub-sections 1.1.4 and 1.3.1), the learning of structures is often formulated in terms of inference of causal connections between latent and observable variables (Pearl 2000, Steyvers, Tenenbaum et al. 2003, Gopnik, Glymour et al. 2004, Gopnik, Schulz et al. 2007), and is generally implemented with graphical models such as Bayesian Networks. Once more, the distinction between belief-based and non-belief-based aspects proves useful in this context. In the AIF landscape, modelling assumptions entail a separation between observed samples and the underlying (causal) structure that generates these observations; that is, AIF models are belief-based (Friston and Buzsáki 2016, Friston, FitzGerald et al. 2016, Friston, FitzGerald et al. 2017, Smith, Parr et al. 2019, Da Costa, Parr et al. 2020). Observations (i.e., cue samples) are used to make probabilistic inferences about the generative (causal) structure. Being belief-based, these models allow a monitoring of concomitant variables: the monitoring happens purely by virtue of having a probability distribution over contingencies, which changes and updates as a result of experience. Furthermore, the supplementary cues suggested in sub-section 1.1.4 – that are employed by humans to help solve the (difficult) problem of inferring causal structure, are integral parts of the AIF. Counterfactuals and affordances can be represented both by policies (i.e., sets of potential courses of actions), each with their respective outcomes (Corcoran, Pezzulo et al. 2020); or by alternative models (i.e., alternative hypotheses using BMS). Active interventions (e.g., 'What would happen if I did that?') can be thought of as the *active* part of the Active

Inference, where the goal is to select policies that reduce uncertainty about contingencies, while simultaneously maximising extrinsic value (Mirza, Adams et al. 2016).

2.4.5 AIF and concept formation

In sub-section 1.1.6, concept learning was defined as the ability to extract similarities and identify dissimilarities across sets of experienced events, involving aspects of rapid learning, and generalisation (Zeithamova, Mack et al. 2019). These aspects are illustrated in one set of numerical experiments using the AIF (Smith, Schwartenbeck et al. 2020), where concept formation was shown to emerge naturally in agents equipped with hyper-parameters over the likelihood array encoding contingencies in the environment. This study involved employing a subtype of BMS called Bayesian Model Expansion to learn concepts for different animals based on their individual features (such as size, colour, etc.); agents here were equipped with extra connectivity 'slots' that can be engaged when the evidence for a model with an extra 'slot' is higher than the evidence for a model without this 'slot'. In Chapter 4, we will show an extension to this work on concept formation by incorporating an active interaction with the environment – that is, agents can also solicit information further to merely observing it; and show how concept formation can be nuanced and enhanced by SL (implemented with BMR).

2.4.6 AIF and replay

Some of the research on replay (Zha, Lai et al. 2019, Liu, Mattar et al. 2021, Wise, Liu et al. 2021, Wimmer, Liu et al. 2023) comes closest to proposing computational mechanisms beyond associative learning that may underlie structure learning. However, the objective function generally used in this field (i.e., that of maximising cumulative rewards) lacks elements of epistemic gain, and self-evidencing (i.e., maximising model evidence) as seen in the AIF. One

replay RL-based study attempted to include an aspect of epistemic gain through the learning of a replay policy (Zha, Lai et al. 2019). In this (in silico) experiment, agents maximise both cumulative rewards, and a replay policy metric that provides agents with the most useful experiences (that can be read as an information gain).

In a related example using the AIF (Parr and Pezzulo 2021) - where synthetic agents engage in a T-maze task – (originally unplanned) results demonstrated a spontaneous emergence of replayed events, in that the beliefs about the agent's location in the maze during maze-solving were replayed in epochs 2 and 3 (of three epochs in total). The interpretation of replay here is that synthetic agents are replaying previous actions in the attempt to make sense of them in the context of the beliefs held. In this example, both the prior and likelihood contribute to the (spontaneous) manifestation of replay, because in the AIF, beliefs about the future are propagated forward in time, thereby uncoupling the actual actions taken from the assessment of possible actions.

In theory, however, replay events could also coincide with one or both approaches to Structure Learning found in the AIF – where experiences are re-simulated and the model evidence is assessed; or where just the posterior concentration parameters are assessed (against alternative explanations) in terms of their model evidence. That is, replay through the lens of the AIF could further involve the rehearsal of (observed) experiences, assessed via model evidence metrics through extensive reinversion and assessment of fictive data; or the hypothesis comparison between post-hoc beliefs (accumulated up until that point in time) and other alternative explanations of the observed data.

2.5 Summary

The Active Inference Framework postulates that agents are inference machines, engaged in minimising (variational) free energy or conversely, maximising model evidence. Implicit in this framework is the representation of the world — via a generative model — taking the form of beliefs, encoded as probability distributions over contingencies. The variational inference process in AIF can be thought of as estimating and optimising posterior beliefs about the causes of sensory experience for past, present, and future (latent) states, based on observations and depending upon the pursuit of specific courses of actions.

There are three main levels of optimisation in the AIF landscape: (Active) Inference, Parametric Learning, and Structure Learning. Active Inference involves the inference of hidden or latent causes and can be thought of as a form of approximate Bayes. Parametric Learning is associated with gradual evidence accumulation via a process that mimics Hebbian learning. Structure Learning involves the selection of models (i.e., BMS more generally, BMR, BME), with the purpose of optimising model evidence, and has been associated with synaptic homeostasis (Kiebel and Friston 2011, Hobson and Friston 2012, Friston, Lin et al. 2017, Hobson, Gott et al. 2021).

Having discussed and synthesised structure learning generally, and Structure Learning in the AIF in relation to the former synthesis, we can now consider a comprehensive definition of Structure Learning: Structure Learning involves the minimisation of uncertainty about models per se, using a measure called model evidence. It entails the comparison and selection of alternative models (i.e., hypotheses) with the purpose of maximising model evidence via a process called Bayesian Model Selection (BMS). There are two main types of BMS: the first approach entails re-simulating experienced events, whilst the latter involves the comparison between post-hoc beliefs and alternative models of the observed data (i.e., using BMR, or

BME). Structure Learning happens in the absence of novel evidence (i.e., it occurs offline); this could be interleaved with episodes of (Active) Inference and Parametric Learning, or it could be after a period of interaction with the environment. Structure Learning goes beyond (Hebbian, gradual) associative (Parametric) learning, granting (biological and synthetic) agents the capacity for an instantaneous, or rapid type of learning.

Chapter 3

Synthetic spatial foraging in a geocaching task

Based on: Synthetic spatial foraging with Active Inference in a geocaching task (Neacsu, Convertino et al. 2022)

3.1 Introduction

Foraging is a type of goal-directed search process whereby (biological or synthetic) agents explore a given space with the purpose of discovering resources of (sometimes) limited availability. This search process is encountered in the literature under various frameworks such as navigation (Montague, Dayan et al. 1995, Rutledge, Lazzario et al. 2009, Humphries and Prescott 2010, Pearson, Watson et al. 2014, Constantino and Daw 2015, Kaplan and Friston 2018), attention and visual salience (Itti and Koch 2000, Parkhurst, Law et al. 2002), or semantic memory (Hills, Jones et al. 2012, Todd and Hills 2020). Each of these perspectives considers different components of complex multi-network and multi-function behaviour. Successful foraging in certain animals engages the prefrontal cortex (Jung, Qin et al. 1998), decision making and reward circuits - such as the dorsal anterior cingulate cortex (Calhoun and Hayden 2015) and the basal ganglia – as well as hippocampal and para-hippocampal areas involved in spatial navigation (Seamans, Floresco et al. 1998, Kolling, Behrens et al. 2012, Barry and Burgess 2014), and planning.

Foraging is a crucial survival skill found across species, though its expression varies depending on the species. This species-specific aspect of foraging becomes apparent when examining the sub-processes involved, most of which usually attributed to the prefrontal cortex in humans and primates (Rudebeck and Izquierdo 2021). These sub-processes encompass evaluation (such as value-based decision making), prediction and action (such as learning about uncertainty, action selection, patch-leaving problems and matching), and social cognition (Rudebeck and Izquierdo 2021). In recent years, complementary work in neuroscience (especially in the field of human and primate decision making) and ethology has appealed to a more universal understanding of decision making in light of core (information) foraging processes. Simultaneously, the evolution of foraging-related structures across species becomes

essential for the development of the decision-making skills observed in humans (Mobbs, Trimmer et al. 2018). For the purpose of the current work, the focus is on one of these aspects, namely on uncertainty reduction through exploration. Uncertainty has a non-trivial role in foraging (Anselme and Güntürkün 2019), with higher levels of uncertainty leading to increased exploratory behaviour and foraging motivation in both animals and humans. This 'boost' is reflected in an increased dopaminergic response from the mid-brain, in particular the nucleus accumbens (Le Heron, Kolling et al. 2020).

The role of uncertainty in cognition has been investigated under different assumptions and frameworks (Grupe and Nitschke 2013, Hasson 2017, Peters, McEwen et al. 2017, Mukherjee, Lam et al. 2021, Walker, Navarro et al. 2021). In learning processes, uncertainty is closely linked to statistical and parametric learning, in that the latter manifests patterns of consistent associations over separate experiences, while the modulation of the former directly impacts the latter via predictive processes (Hasson 2017). In terms of action, uncertainty plays a pivotal role in driving epistemic behaviour (namely, information gathering). Within the Active Inference Framework, several forms of uncertainty exist: uncertainty about (hidden) states given a policy, uncertainty about policies in terms of expected future states, future outcomes, and model parameters, uncertainty about model parameters given a model, and uncertainty about the model per se (Friston, Lin et al. 2017). In the current chapter, all forms of uncertainty besides the latter are in play, and reduced through state estimation (minimising surprise), epistemic planning (Expected Free Energy), and epistemic learning (with respect to likelihood and transition parameters). We focus on the specific role of uncertainty in spatial foraging to elucidate, both theoretically and neurophysiologically, how goal-directed epistemic behaviour depends on the level of uncertainty about internal representations of the state of the world – and the planned exchange with that world.

This chapter will present a generative model of foraging using a geocaching task. The aim is to demonstrate how both epistemic ('explore') and reward-seeking ('exploit') behaviours arise from the same generative model of the world, focusing on one of the core aspects of foraging – uncertainty reduction – as contextualising spatial exploration and action selection. The simplified, naturalistic behaviour is reproduced using a goal-directed task. Moreover, a series of neurophysiological simulations are reported, providing evidence for the biological plausibility of the model, and the role of dopamine in foraging and uncertainty reduction, as shown in previous studies (Fiorillo, Tobler et al. 2003, Niv, Duff et al. 2005, Friston, Schwartenbeck et al. 2014, Li, Cao et al. 2016, Gershman 2017, Jo, Heymann et al. 2018, Le Heron, Kolling et al. 2020). Similarly to this proposal, (Schwartenbeck, Passecker et al. 2019) developed an account of goal-directed exploration using the AIF, providing complementary insights into the balance between explore-exploit behaviours in a T-maze task. Here, we extend on this foundation towards a generalisation of the theory in foraging behaviour in the environment, where the binary decision-making choice is substituted by multidirectional goal-directed navigation. As we will show, the same principles succeed in reproducing spatial foraging behaviour in an open environment.

For completeness, we note that the field of foraging studies has benefited from a variety of approaches and disciplines, from neuroscience of decision-making and economics (Hayden 2018, Mobbs, Trimmer et al. 2018) to computational neuroscience (Ward, Austin et al. 2000, Gheorghe, Holcombe et al. 2001, Davidson and El Hady 2019), from ethology (Stephens 2008) to social studies (Gabay and Apps 2020), with the substantial contribution of memory and spatial navigation research (Gutiérrez and Cabrera 2015, Kerster, Rhodes et al. 2016, Nauta, Khaluf et al. 2020). For a thorough perspective on the topic, please refer to relevant reviews (Hayden and Walton 2014, Hall-McMaster and Luyckx 2019, Gabay and Apps 2020).

In what follows, we describe the generative model used for numerical analyses (section 3.2). The subsequent section presents a series of illustrative simulations showcasing planning and foraging behaviour, their underlying belief updating, and prospective neurophysiological correlates. In the final section, we review the numerical experiments in light of current empirical findings in the spatial foraging literature.

3.2 The generative model

Generative models are joint probability distributions over observed outcomes, latent causes, and sequences of actions (i.e., policies), necessary to optimise beliefs and subsequent behaviour. The active side of the (Active) Inference process corresponds to inverting a generative model using observed outcomes (i.e., generating consequences from causes), and forming posterior expectations about the hidden states (i.e., recovering causes from consequences). Crucially, in the AIF, these expectations include the most likely action, hence the term Active Inference. In this section, we describe the specific generative model used to simulate purposeful behaviour and associated belief updating, and the slower accrual of evidence (i.e., associative plasticity). These distinct processes are emergent aspects of minimising the variational bound on (negative log) model evidence described in Chapter 2. These processes have a reasonable degree of biological plausibility, enabling us to simulate neuronal responses and changes in synaptic efficacy during (Active) Inference and Parametric Learning, respectively (Friston, FitzGerald et al. 2017, Friston, Parr et al. 2017).

The generative model used in the following simulations is a deep temporal model (Friston, Rosch et al. 2017) based on a discrete state space partially observable Markov decision process (POMDP). Under these sorts of models, there are generally four types of latent

causes: *hidden states* (of the world) that generate observable outcomes, *policies* (i.e., sequences of actions being pursued) that specify transitions among the hidden states, *precision* encoding confidence in beliefs about policies, and *parameters* (e.g., likelihood).

The generative model is parametrised by a set of tensors (i.e., matrices and vectors): a likelihood matrix encoding probabilistic mappings from (hidden) state factors to outcome modalities (**A**), transition probabilities among the different hidden states given particular actions (**B**), prior preferences over outcome modalities for each hidden state factor (**C**), and finally, priors over initial states (**D**). The likelihood and transition tensors are parametrised with Dirichlet (concentration) parameters that accumulate during experience: the amalgamation of a given hidden state and outcome effectively adds a concentration parameter (i.e., a count) to the appropriate element of the likelihood mapping.

Here, there are two outcome modalities: the first (*what*) registers rewarding outcomes with two levels (*reward* versus *null*). The second modality reports the current location in the space being explored (*where*). Outcomes are generated from a single hidden state factor (*location*), corresponding to locations in a 10x10 grid. Please see Figure 3.1 for a graphical depiction of the generative model. There are 5 allowable actions: up, down, left, right, and stay. These actions induce 5 transition matrices that play the role of empirical priors. The outcomes *reward: present* and *reward: null* were assigned a utility (i.e., relative log probability) of 3 and 0 respectively. With these utility values, the synthetic agent would 'prefer' (i.e., expect) a *reward: present* outcome about 20 times more than the *reward: null* outcome. The agent also prefers being in proximity of the target location (i.e., *reward: present*). In summary, we specified a minimal generative model necessary to illustrate navigation and (epistemic) foraging in which the causes of observable outcomes were locations in space. The observations available to an agent comprised two sorts. The first told it unambiguously where it was and the second described what happens at each location, in terms of preferred or non-preferred

outcomes. The agent can move around this space, taking one step at a time – knowing its location but not necessarily knowing location-specific outcomes in the reward modality.

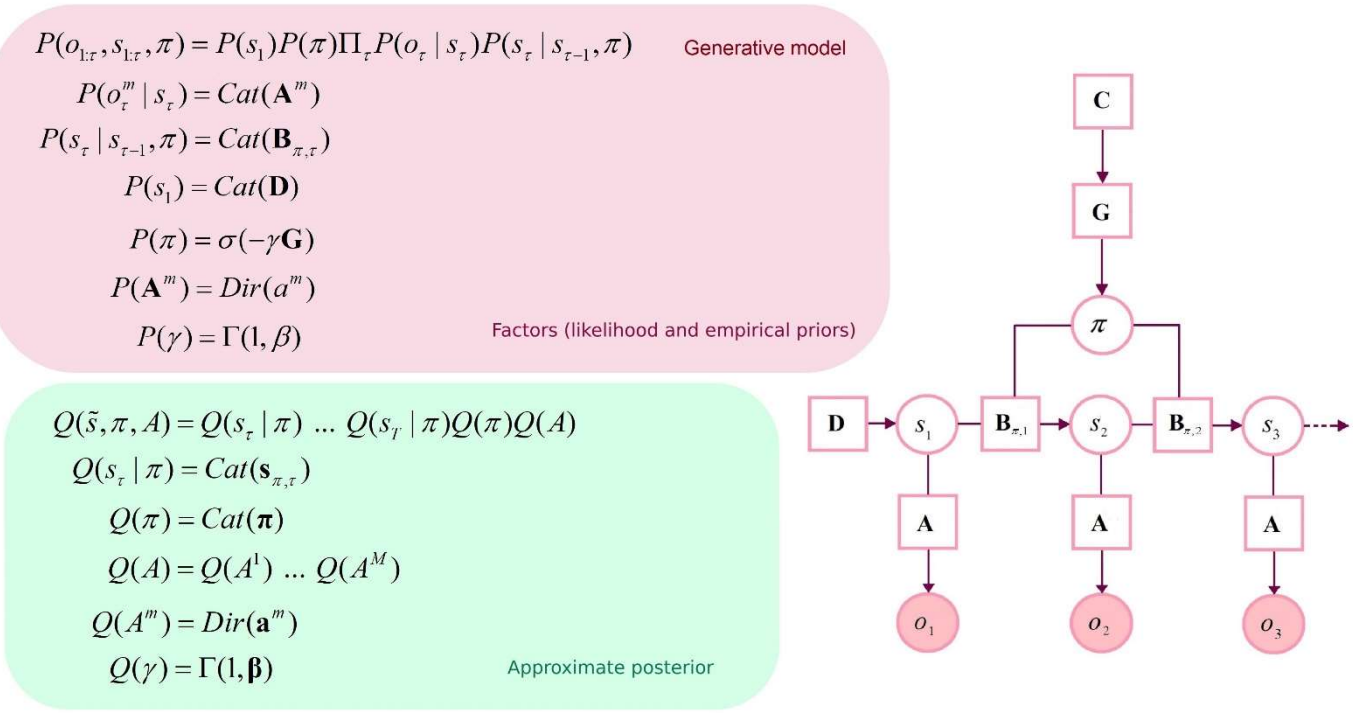


Figure 3.1 Graphical depiction of the generative model and approximate posterior. This discrete-state space temporal model has one hidden state factor: *location*. This factor generates outcomes in two outcome modalities: *where* and *what* (with two levels: reward or null). The likelihood **A** is a matrix whose elements are the probability of an outcome under every combination of hidden states. **B** represents probabilistic transitions among hidden states. Prior preferences over outcome modalities for each hidden state factor are denoted by **C**. The vector **D** specifies priors over initial states. *Cat* denotes a categorical probability distribution. *Dir* denotes a Dirichlet distribution (the conjugate prior of the *Cat* distribution). An approximate posterior distribution is needed to invert the model in variational Bayes (i.e., estimating hidden states and other variables that cause observable outcomes). This formulation uses a mean-field approximation for posterior beliefs at different time points, for different policies and parameters. Bold variables represent expectations about hidden states (in italic). Transparent circles represent random variables, and shaded circles denote observable outcomes. Squares denote model parameters and Expected Free Energy.

The **B** tensor can be thought of as an empirical prior, since it depends upon actions, which themselves are determined by policies π (i.e., it depends upon a random variable). Policies are *a priori* more probable if they minimise expected free energy **G**, which is contingent upon prior preferences about outcomes **C**, and uncertainty about outcomes under each state. Update equations (that allow agents to minimise free energy) are derived from the generative model, with consideration for neurobiological constraints – please see (Friston, FitzGerald et al. 2017, Friston, Parr et al. 2017) for a comprehensive treatment. Briefly speaking, expected hidden states are updated by means of belief propagation. In the Active Inference Framework, this is achieved using a gradient descent on (variational) free energy for each hidden variable.

Message passing (i.e., belief propagation) is implemented from representations of the past (forward message), future (backward message), and observations that update posterior beliefs over latent (hidden) states, allowing for both postdiction and prediction under each individual policy. As new outcomes emerge, more likelihood messages contribute to the belief update, which makes for more informed posteriors. This recurrent message passing can be summarised as follows: the generative process (i.e., the environment) generates outcomes that update approximate posteriors about policies (i.e., plans), which are themselves contingent upon prior preferences and intrinsic value. The policies determine the selected action, and selected actions generate new outcomes.

3.3 Simulations and results

The numerical experiments focused on navigation and local foraging respectively. In the navigation simulations, the agent observes a space comprising a 10x10 grid and navigates toward preferred target locations (specified with prior preferences over the location modality). For the foraging simulations, we zoom into a local area (also a 10x10 grid), where the agent engages in epistemic foraging to find a hidden object (i.e., rewarding location). After finding this object, the agent is given a new target location and the process repeats. The agent thus plans its trajectory towards its target location, and then explores the location to find hidden rewards. This object could be regarded as the cue that specifies the next target location. After navigating to the second target location, the agent again explores locally to find the hidden object. This process could continue ad infinitum. In this demonstration, both epistemic foraging and goal directed behaviour are evinced via the minimisation of (expected) free energy.

For the navigation phase, the agent starts at the entrance of the grid. Prior preferences prompt the agent to seek out target locations. Cues that directly inform the agent of its current location can be thought of as exteroceptive, whereas the observed outcomes (*reward* or *null*) can be thought of as interoceptive. The policy depth (i.e., planning horizon) involves four steps - that is, agents can evaluate distal (and possibly preferable) outcomes in the future, which allows them to plan and pursue the shortest trajectory towards the end goal (i.e., the first *rewarding* location). In this simulation there were ten moves in total, enough to reach the target location using the shortest available path. Synthetic agents were endowed with prior knowledge about the environment – so that they were planning their trajectory in a familiar environment.

For the local foraging simulations, the agent has additional (i.e., epistemic) incentives in the form of uncertainty about the location that contains the *rewarding* outcome. In this context, agents explore the environment, initially motivated by curiosity about the parameters

of the model (here, the likelihood matrices). In other words, their behaviour was driven by the novelty of the environment; namely, 'what would happen if I went there?'. To simulate exposure to this local novel environment, the prior Dirichlet parameters *a* of the likelihood mapping (**A**) - encoding the mapping between hidden states and '*what*' outcomes (i.e., *reward* or *null* outcomes) - were set to a small value (i.e., 1/100). As a consequence, the expected free energy **G** acquires a nontrivial novelty term (Friston, Lin et al. 2017). This phase of the simulations illustrates how agents learn about their environment by means of novelty-driven evidence accumulation. Technically, this entails the updating of Dirichlet parameters (encoding hidden state – outcome mappings) after 30 successive moves in the local environment. Once locations are visited, they lose their novelty (i.e., epistemic value), a process which endorses those policies that visit unexplored ground. Preferences for particular outcomes (i.e., *reward* and *location*) were formally the same as the prior preferences used in the navigation simulations. We also specified concentration parameters in the state transition matrix to simulate an additional type of learning – comparable to that of foraging in volatile environments – where (biological) agents have some degree of uncertainty about where exactly they will move to, based on where they have just foraged (and the actions they pursued).

Collectively, these simulations mimic the circumstances surrounding local foraging in geocaching, where agents freely explore the environment to discover a hidden object. The agent however maintains a dual imperative – to discover the environment by satisfying its curiosity, and at the same time, to realise prior preferences (i.e., of finding the object hidden in the environment). In Figure 3.2 below, we depict results of exemplar simulations for both types of simulations.

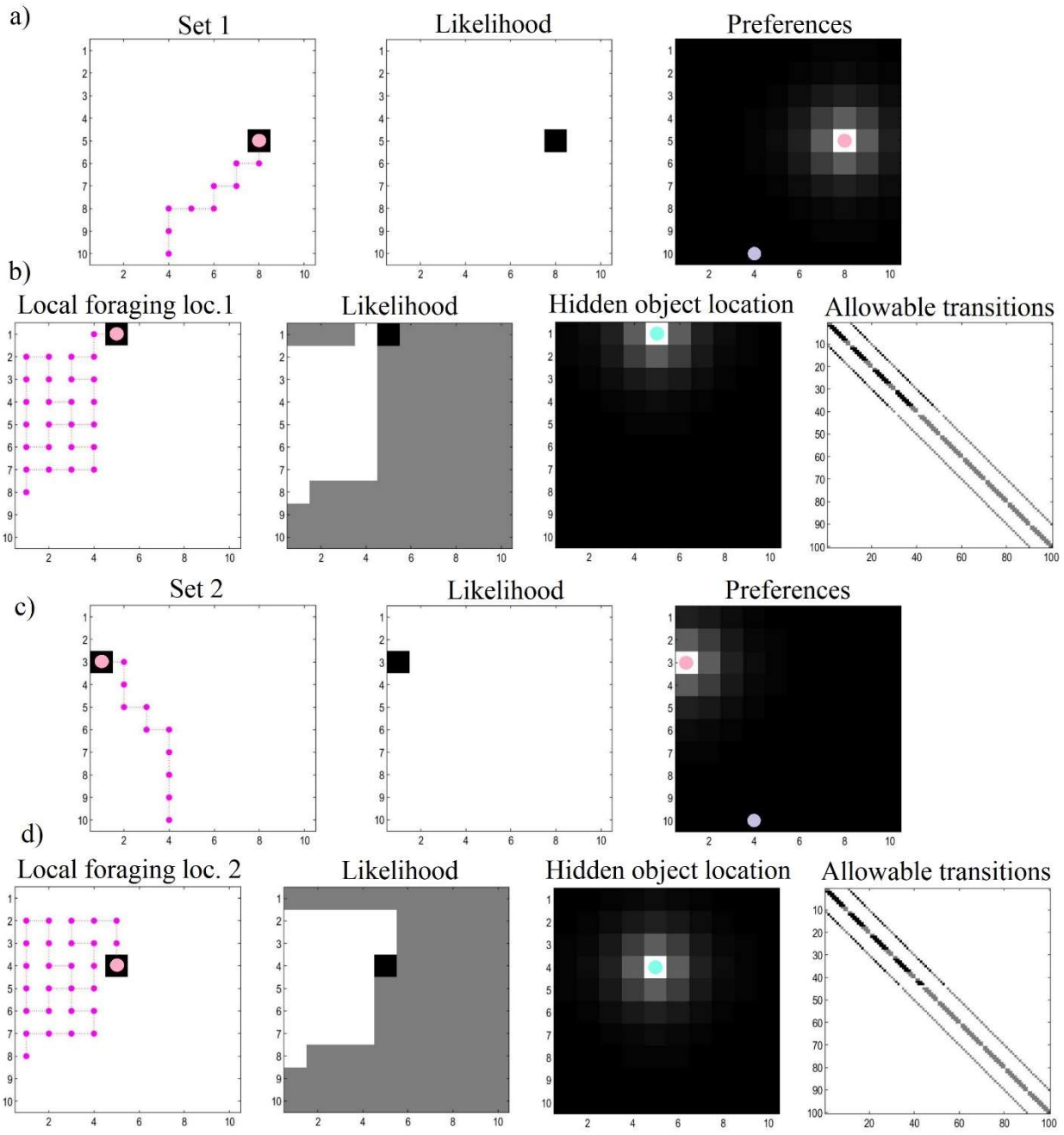


Figure 3.2 Navigation and local foraging behavioural results. a) The agent plans and executes its (shortest available) trajectory towards the first target location, driven by prior preferences. The purple dot indicates the starting location. The agent has learned the likelihood mappings, which can be interpreted as having – and making use of – a map to reach the target location. b) When the target location is reached, the agent explores the local area to find a hidden object, as it learns and discovers its environment. Here, the agent starts with a uniform distribution about the likelihood mappings, and has additional uncertainty pertaining to the transition matrix (i.e., uncertainty about where the agent finds itself given where it was previously and the action it has taken). This process involves a dual pursuit: discovering the environment and fulfilling a desire to find the hidden object. c)-d) After finding the hidden object, the agent receives a new target location and the process repeats (possibly ad infinitum).

In the Active Inference Framework, a softmax function is applied to (precision weighted) Expected Free Energy in order to optimise posterior beliefs about each policy. When new observations are available, the precision parameter is updated: the policy with the lowest (expected) free energy is more likely if the associated precision parameter is high (c.f., an inverse temperature parameter). The confidence that the inferred policy will produce preferred outcomes or resolve uncertainty about latent states is therefore represented by this precision parameter. Dopaminergic activity in the mid brain is thought to encode this type of precision (Schwartenbeck, FitzGerald et al. 2015).

Figure 3.3 illustrates representative simulated neural activity for the agent's last planning and movement sequence (i.e., ten movements) during the navigation simulations. In the current model, the phasic bursts observed in simulated dopaminergic responses (see Figure 3.3 top-right) indicate notable changes in precision at steps 1, 4, 6, and 8 (i.e., the 16^{th}, 64^{th}, 96^{th}, and 128^{th} iteration respectively, in terms of updates – since there are 16 iterations of gradient descent per time-point). Since in the AIF dopamine responses represent updates in the expected precision of the Expected Free Energy distribution over policies, these spikes can be interpreted as a change in confidence (i.e., the agent resolves uncertainty) about what policies to pursue, by eliminating other possible trajectories. In this scenario, at the first step, the agent eliminated the possibility of going right instead of up, an action that could equally have allowed it to reach the target using the minimum number of steps. At the 8^{th} movement, the agent becomes confident about fulfilling its target location, and spends steps 9 and 10 within the *rewarding* outcome. This example shows how belief updating and decision making can be unpacked in terms of uncertainty and precision.

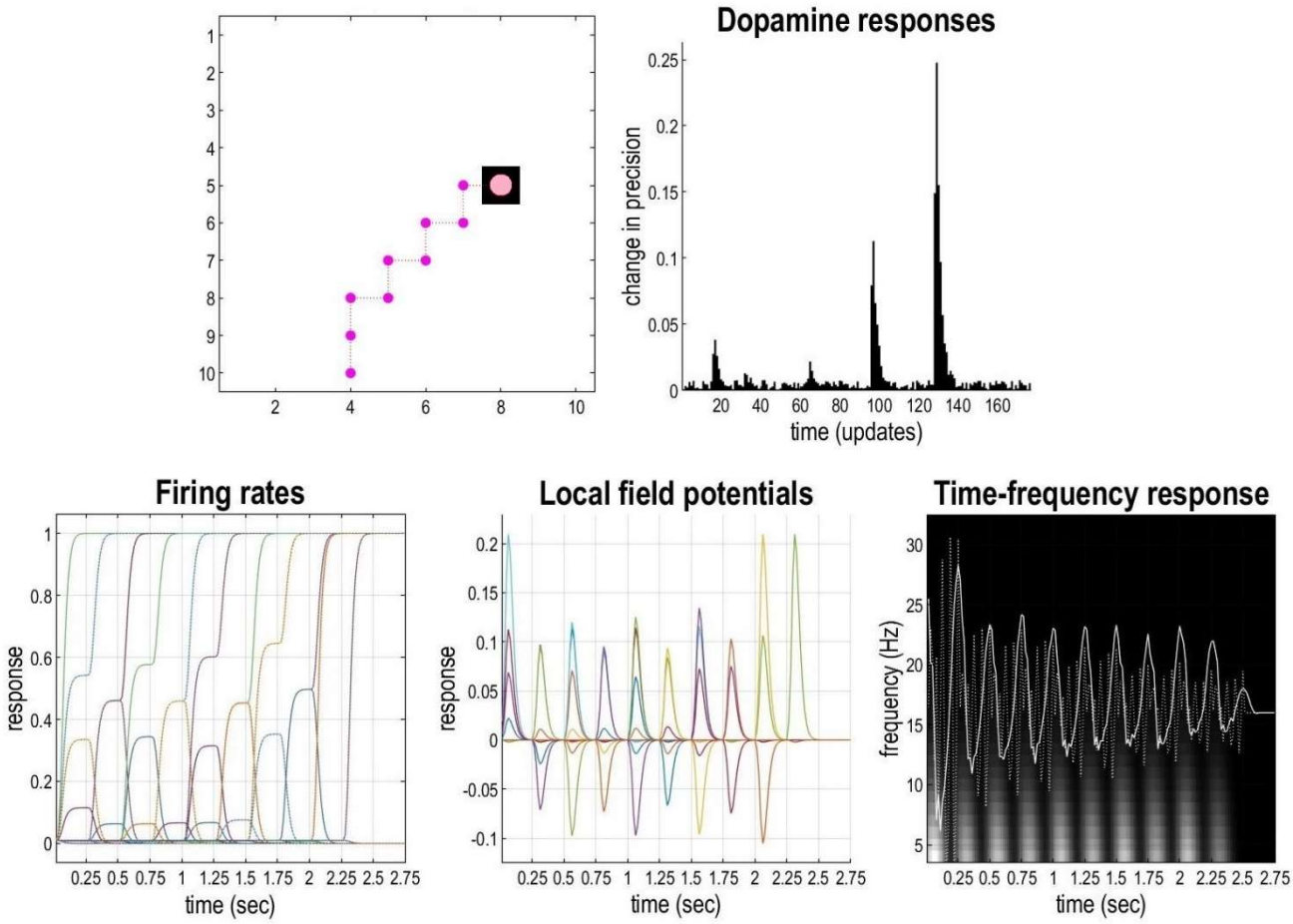


Figure 3.3 Simulated electrophysiological responses for a representative sequence of moves. The top left panel shows the agent's trajectory, followed by (synthetic) dopamine responses (top right), firing rates (bottom left), local field potentials (bottom centre) and time-frequency responses (bottom right). Please see main text for more details.

Firing rates indicate changes in beliefs over time about the *where* state factor for each time-point (Figure 3.3 bottom-left), illustrating leaps in evidence accumulation, where expectations diverge as foraging progresses. The bottom-centre panel of Figure 3.3 depicts predicted local field potentials (depolarisation), showing the rate of change in simulated firing rates for all (1100) hidden state units (coloured lines). This panel shows that visiting different locations evokes responses in different neuronal populations, and of variable degrees. Finally, the bottom-right panel displays neuronal responses associated with the *where* state beliefs before

and after filtering at 4Hz (dotted and solid line respectively). These are superimposed upon a time-frequency decomposition of the averaged local field potential (averaged over all simulated neurons). These show fluctuations in local field potentials at a theta rhythm that are phase-locked to induced responses over a wide range of frequencies (including gamma frequencies – not shown). This reproduces the characteristic theta-gamma coupling found in empirical studies of foraging and navigation in small animal studies (Bragin, Jando et al. 1995, Lisman and Redish 2009, Buzsaki and Moser 2013).

3.4 Interim discussion

Learning about the environment is fundamental for human and animal behaviour, particularly in activities like foraging, which requires the interaction of multiple processes to maintain homeostasis, on which survival depends. Recent advancements in neuroscience and ethology of foraging have underscored the necessity for a universal and ecologically valid approach to understanding foraging. Despite various disciplines contributing to the extensive study of foraging across species, the field lacks an integrative account. In the current numerical experiments, two key components of foraging were addressed: uncertainty reduction and action selection. Through computational modelling, foraging behaviour was reproduced, with due consideration given to neurophysiological correlates. The setup of this model, in its simplicity, accounts for the sequential nature of foraging; especially the gradual accumulation of knowledge during epistemic foraging.

The results reproduce two levels of foraging behaviour: goal-directed navigation (in the global environment) and epistemically-driven exploration (locally). In the first case, the goal is to follow a trajectory, given preferences for a target location. In the second this preference

seeking motivation is contextualised by explorative or epistemic imperatives. Note that because the epistemic and preference parts of Expected Free Energy are expressed as log probabilities, the policies selected can be viewed as reflecting the product of the probabilities *per se*. In other words, epistemic policies will be rejected if they have a very small probability of securing a preferred outcome. A very small probability of a preferred outcome corresponds to an aversive or surprising outcome, which means that prior preferences constrain the epistemic affordances of any behaviour (under the Active Inference Framework).

The simulations in this chapter illustrate the effect of uncertainty on behaviour and neuronal activity. This is particularly relevant in the second part of our simulations (local foraging). The degree of exploratory behaviour is modulated by the level of uncertainty about the state of the world. When uncertainty is high, action selection is built upon the exploratory imperative of reducing uncertainty. The more the agent becomes confident about its surroundings (i.e., the more uncertainty is reduced), the more action selection is guided by exploitative behaviour, when extrinsic gain is predominant, and less by exploration. Uncertainty reduction thus has a direct effect on action selection. As proposed in previous work, dopamine is suggested to encode uncertainty over policies or decisions (Schwartenbeck, FitzGerald et al. 2015). In other words, the kind of beliefs – whose precision is modulated by dopamine – are beliefs about policies (sequences of actions, resulting in action selection). At a synaptic level, the modulation of precision could be thought of as neuromodulation or synaptic gain control (Parr and Friston 2017). The firing rate of dopamine in the mid brain is reproduced in our electrophysiological simulations. As expected, the agent becomes more and more confident about its predictions, which is reflected in a progressive increase in rates of beliefs updating and reduction of uncertainty.

The current work has some clear limitations. Although it succeeds in reproducing biologically plausible and real-world oriented foraging behaviour, it does not account for

several sub-processes involved in foraging, such as aspects of spatial navigation (e.g., the navigation system of hippocampal and para-hippocampal areas), patch-leaving problems, matching and social cognition. Another limitation is the assumption that the model is given the target location as a fully formed prior preference. This could be interpreted as 'information passing' of cues between individuals of the same group. However, a more extensive account of foraging would have to address how these prior location preferences were inferred or learned.

Although restricted in its focus, the current model offers a preliminary account of foraging, both in terms of behavioural and neurophysiological responses. Future work could aim to extend this approach to include the missing elements of foraging. The Active Inference Framework is indeed equipped to account for many aspects of sentient behaviour, social behaviour included. A successful extension of the model could also reproduce and investigate the neurophysiological role of other neurotransmitters in foraging. For example, the role of norepinephrine in setting the precision of state transitions – or the role of cholinergic neurotransmission in setting the precision of sensory or likelihood mappings (Doya 2002, Doya 2008, Parr and Friston 2017). Moreover, the AIF offers a promising approach to bridge the gap not only between behaviour and neurophysiology, but also between foraging mechanisms across different species. Developmental and comparative neuroscience could benefit from in silico, evidence-informed modulation of model parameters to test different hypotheses about how foraging evolved over time – from simple living beings to more advanced primates and humans. This work offers one step towards a universal conceptual and mechanistic understanding of foraging.

Chapter 4

Structure Learning enhances concept formation in synthetic Active Inference agents

Based on: Structure learning enhances concept formation in synthetic Active Inference agents (Neacsu, Mirza et al. 2022)

4.1 Introduction

The main focus of this work is to illustrate the importance of Structure Learning as implemented by Bayesian Model Reduction. This chapter presents a computational formulation of how agents may come to form representations and concepts by foraging their environment, and how these concepts may be shaped by Structure Learning. We attempt to capture the computational mechanisms that underwrite concept formation and associated relationships (i.e., relationships between elements within contexts and relationships between contexts), as resulting from the action-perception cycles that govern agent-world interactions.

In this work, agents possess and update an internal model that entertains temporally and physically structured processes when interpreting these orderly interrelationships. Whereas processes such as (Active) Inference and Parametric Learning have been discussed extensively in the literature (Friston, Schwartenbeck et al. 2013, Friston, Schwartenbeck et al. 2014, Schwartenbeck, FitzGerald et al. 2015, Friston, FitzGerald et al. 2016, Mirza, Adams et al. 2016, Seth and Friston 2016, Friston, FitzGerald et al. 2017, Friston, Parr et al. 2017, Friston, Rosch et al. 2017, Parr and Friston 2017, Kaplan and Friston 2018, Parr and Friston 2018, Mirza, Adams et al. 2019, Da Costa, Parr et al. 2020, Hesp, Smith et al. 2021, Neacsu, Convertino et al. 2022), little attention has been given to the type of off-line learning we define operationally as Structure Learning (Friston, Parr et al. 2018, Smith, Schwartenbeck et al. 2020), with the implicit computational form within the Active Inference Framework. This is a nascent field of inquiry raising important questions about what it means to process information off-line.

In this chapter, we take the first steps toward a comprehensive process theory of Structure Learning, grounded in a single objective function: that of maximising model evidence. In the work by Smith et al, (Smith, Schwartenbeck et al. 2020), the Structure

Learning of concepts proceeds passively (i.e., there is only an 'observation model'). Here, we build on this work by incorporating the 'active' part of the perception-action cycle, making this the first attempt to connect action and perception in the context of Structure Learning, and thereby moving toward a more ecologically valid account of model selection. Furthermore, this work was the first - to my knowledge - to feature a comparison of information gain between online (i.e., Active Inference, Parametric Learning) and off-line (Structure) Learning.

The capacity to mine for similarities and detect dissimilarities across sets of experienced (sensorial or autobiographical) events is a crucial facet of structured knowledge-building. This type of relational thinking is known as *concept learning*, first proposed by Bruner, Goodnow and Austin in 1956 in 'A study of thinking' (Bruner, Goodnow et al. 1956). More specifically, concepts are mental representations that allow biological agents to compare and contrast collections or sets, and their respective elements. Various researchers have since developed and expanded on concept learning. For instance, the prototype theory of concept learning suggests that biological agents possess a central example, a 'common representation' of a particular set, and then judge how (semantically) far (or close) new experiences are in relation to the prototype (Geeraerts 2006). Another example is that of abstraction of rules, whereby concepts are characterised as a set of rules, and agents assess new experiences (of objects, events, etc.) based solely on their respective properties and whether they fit the definitions or not (Rouder and Ratcliff 2004, Goodman, Tenenbaum et al. 2008). A further instance is that of ad hoc categories in goal-directed behaviour (Barsalou 1983), which describes a temporary and spontaneous type of concept formation, such as 'things to bring on a trip'. In this scenario, knowledge from different domains is combined to form a novel temporary structure, specific to the context in play.

There is a tight relationship between context and content: if I know the context (for example, living room, beach, street, etc.) then I can call on a conditional probability distribution

over the things I expect to see there (i.e., the content): sofa, TV, coffee table; sand, water, floaties; buildings, cars, traffic signs. And if I know what I am seeing (i.e., the content), then I can infer the contexts I may plausibly be in. This bidirectional relationship is implicit in the current computational model: during the inferential and learning processes, both these probability distributions are optimised simultaneously. That is, in order to find out which context is in play – and to fulfil desired outcomes – agents use information acquired in previous time-steps to infer the context in which they are operating. At the same time, this context places constraints on outcomes in the future, given the actions they take (e.g., things I expect to see if I look over there).

Concept learning spans several areas of inquiry relevant to both neuropsychology and computational neuroscience: the way (biological) agents form *concepts*, how they interpret *context* and *content*, what it means to *represent* concepts and *relationships* between different elements within a context or between contexts, what *similarity* means, how humans *categorise* environments, objects, and their elements into distinct entities, what role *memory* plays, what counts as *relevant* information, and so on. A growing body of work in concept formation and structure learning employs computational frameworks, such as non-parametric Bayesian models (Blei, Griffiths et al. 2003, Griffiths, Sanborn et al. 2011, Gershman and Blei 2012, Collins and Frank 2013, Stoianov, Genovesio et al. 2016), where generative models are equipped with an extendable space. The focus in this instance is on whether to incorporate additional components to the generative model, and at what point. For example, in (Collins and Frank 2013), concept learning is presented as inferring a hidden structure, and deciding whether the current structure should be reused, or a new structure should be created. This representative example is relevant to our current model, where we disentangle three processes: (Active) Inference (about latent causes), Parametric Learning (learning associations), and Structure Learning (deciding on the best generative model). Whereas non-parametric Bayesian methods

furnish one way of growing models in a principled way, our work considers Bayesian Model Reduction, which starts with an over-complete, overly expressive model, and then removes redundant components or model features, in order to minimise complexity. We use the Active Inference Framework as the most generic formulation of Bayes-optimal behaviour that is necessary to identify the best model or structure in terms of Bayesian model evidence.

Other related approaches to concept and structure learning from the machine learning literature involve model-based clustering algorithms such as Gaussian mixture models (McNicholas 2016) or hierarchical deep models (Salakhutdinov, Tenenbaum et al. 2012). Determining the optimal number of clusters in these approaches ranges from fitting all the models in a family and selecting the best using a Bayes information criterion (McNicholas 2016), to augmenting deep Boltzmann machines with hierarchical Dirichlet process priors (Salakhutdinov, Tenenbaum et al. 2012). Many of these approaches, however, require large amounts of training data, unlike their human counterparts (Mnih, Kavukcuoglu et al. 2015) and are generally difficult to evaluate in terms of Bayesian model evidence (Penny 2012, Fourment, Magee et al. 2020).

The neurobiological literature concerning concept learning is vast (McClelland, McNaughton et al. 1995, Love, Medin et al. 2004, Love and Gureckis 2007, Mack, Love et al. 2016, Bowman and Zeithamova 2018, Hutter and Wilson 2018, Mack, Love et al. 2018, Mok and Love 2019, Zeithamova, Mack et al. 2019, Barron, Auksztulewicz et al. 2020, Mack, Preston et al. 2020). Many models focus on hippocampal-neocortical interactions, and more specifically on the interaction between the hippocampus and the prefrontal cortex (PFC), given their wide involvement in generalised knowledge-building (Eichenbaum 2017, Rubin, Schwarb et al. 2017, Gruber, Hsieh et al. 2018). In one relevant study Mack, Love, and Preston (Mack, Love et al. 2016) address the swiftness and flexibility of incorporating new knowledge and sensory information into existing models of the world. In their study, they employ

neuroimaging and a computational model called SUSTAIN (Love, Medin et al. 2004) to ascertain the neural mechanisms underlying this aspect of concept learning (i.e., integrating new information with pre-existing concepts). Subjects viewed and categorised complex visual objects (insects) into groups by attending to either one or two features (width of legs or antennae and pincers, respectively). Although the objects presented remained constant, they belonged to different categories based on the number of features attended and their specific combinations. Subjects therefore had to integrate new and old representations of the objects in line with this foundational structure. Neuroimaging results confirmed the computational model prediction that objects encoded by similar representations should also evoke similar neural activity patterns. Further, in the hippocampus, these conceptual representations were shown to evolve and reorganise as a result of assimilating new information (in this case, new object features).

The work in this chapter implies the use of a generative model (i.e., beliefs encoding probability distributions over observed outcomes and hidden causes). As we will see below, the reorganisation of knowledge entails restructuring these beliefs as a result of Bayesian Model Reduction (BMR). Belief adaptation corresponds to Parametric Learning, where beliefs change gradually, based on observing data. Whereas adaptation is in relation to the environment (i.e., agents 'adapting to' the environment as they gather sensorial information), reorganisation is in relation to the generative model per se (i.e., agents minimising complexity, maximising model evidence). In other words, adaptation is due to, and a result of moment-to-moment interactions with the environment, whereas reorganisation entails off-line (model) optimisation in the absence of evidence. This reorganisation may or may not be adaptive (to the environment), based on whether or how the environment changes.

Here, we combine the three primary AIF mechanisms of information processing (Active Inference, Parametric Learning, and Structure Learning), to examine whether we can

reproduce cardinal aspects of concept learning, context learning, and representation, which naturally emerge from self-evidencing (Hohwy 2016). Practically, we used simulations of agents situated in a novel environment. This environment comprised several rooms, two pairs of which had an identical form. Within each room, a particular location afforded a reward. The agents had two hierarchical levels of action: they could forage within each room at the lower level, or move between rooms at the higher level. In ethology, this could be construed as a simple patch foraging paradigm (Constantino and Daw 2015). To begin with, agents had an imprecise representation of the possible types of rooms they would encounter, but more importantly, they did not know *a priori* which context (i.e., room) they were in, or the unique affordances of the different rooms. By simply optimising the evidence for their model of the environment, we hypothesised that the agents would come to learn and remember the number of rooms and reward locations, thereby forming a representation of their active engagement with the environment. Beliefs over hidden states, parameters, and the structure of their (actively explored) environment are encoded by – and underlie – these representations. The computational principles underlying the Active Inference Framework have been demonstrated in other contexts such as saccadic eye movements (Mirza, Adams et al. 2016, Parr and Friston 2018), or at a more abstract level, prosocial behaviours (Constant, Ramstead et al. 2019), and emotional constructs (Smith, Parr et al. 2019). Here, we adopt a minimal model of spatial foraging and Structure Learning in order to clarify underlying processes and demonstrate key ideas. This is the first paper, to our knowledge, to apply the principles of Structure Learning to spatial foraging during an active engagement with the environment.

In what follows, Structure Learning will refer to Bayesian Model Selection, and in particular, Bayesian Model Reduction, to find the best model of an active engagement with the environment. In virtue of the fact that these models are based upon discrete state-space models (namely, Markov decision processes), different models are distinguished by the presence or

absence of a particular mapping among discrete states. In the context of the likelihood mappings, this will be between latent states of affairs in the world and observable outcomes. This means that Structure Learning can be cast as exploring a space of mappings among discrete states. We will demonstrate concept learning by applying Parametric and Structure Learning to the likelihood mappings, reading 'concepts' as the latent causes that generate observable outcomes. Whereas Parametric and Structure Learning are mechanistic processes, concept learning is a teleological description of what these processes look like, from a psychological or constructivist perspective.

The remainder of this chapter comprises three sections. In the first, we provide a specific description of the generative model used to unpack these ideas and demonstrate the learning of likelihoods under Structure Learning and Parametric Learning. The second section presents a series of simulations (i.e., numerical analyses) showcasing characteristic behaviours and their associated belief updating. The key hypotheses were a) as agents forage their environment, they come to form representations – that is, precise (probabilistic) beliefs encoding the structure of the environment; b) Structure Learning in the form of Bayesian Model Reduction (BMR) assists concept formation and performance. With ongoing exposure to the environment, we hoped to see the emergence of concept learning and improved performance both as a function of gradual (Parametric) Learning and BMR. That is, agents will come to learn that there is a limited number of rooms, with a particular topology, find the reward more often, and gather more reward overall. The final section reviews the numerical experiments in light of existing empirical findings in cognition and neurophysiology.

4.2 The generative model

This section provides a specific description of the generative model used to unpack and demonstrate the learning of likelihoods, simulating behaviour in terms of moment-to-moment belief updating, slower accumulation of evidence under Parametric Learning – in the form of associative plasticity – and Structure Learning, in the form of Bayesian Model Reduction. These distinct processes are emergent aspects of minimising the variational bound on (negative log) model evidence. The generative model used in the following simulations is a deep or hierarchical temporal model (George and Hawkins 2009, Friston, Rosch et al. 2017) based on discrete states in a partially observable Markov decision process (POMDP). It comprises two Markov decision processes, where the outputs of the higher level generate the initial (hidden) states of the lower level. These types of models have been used previously to model reading and language processing (Friston, Rosch et al. 2017). Here, we use it to model spatial foraging within and between different contexts. Each level of the generative model is parametrised by a set of matrices and vectors (more generally, arrays): a likelihood matrix encoding probabilistic mappings from states to outcomes (**A**), transition probabilities among the different hidden states (**B**), prior preferences over outcomes (**C**), and finally, priors over initial states (**D**). The likelihood matrices are parametrised with Dirichlet (concentration) parameters that are accumulated with experience: the combination of a given hidden state and outcome effectively adds a concentration parameter (i.e., a count) to the appropriate element of the likelihood mapping.

The model used in this work generates three outcome modalities (at the lower level): the *location* within a room, a *reward* outcome, and a room or *context* specific cue (i.e., the room colour). The *location* modality has 16 levels corresponding to locations in a 4 x 4 grid. The *reward* modality has two levels: present or absent. The *context* modality has 16 levels

corresponding to 16 possible rooms. The hidden state factors generating these outcomes comprise two factors: *location* (inside a specific room) and *context* (room identity). The *location* factor has 16 levels corresponding to sensed locations (i.e., 4 x 4 grid), while the *context* factor has 16 levels corresponding to the room identity (i.e., contextual cue). In other words, we have two hidden state factors (*location* and *context*) generating three outcome modalities (*location, reward,* and *context*) at the lower level. The link to the higher level is via the *context* hidden state factor. The content of the higher level therefore becomes the context for the lower level via this hidden state factor - please see Figure 4.1 for an illustration of the generative model. As an analogy, being in a specific *building* (i.e. higher level) entails a specific set of available *rooms* (lower level). For example, a school has laboratories and classrooms, each with their own configurations, types of furniture, etc. The generative model acts as a simplification of the contingencies entailed by a specific set of buildings, rooms, and their properties. Note that we deliberately reproduce the (4 x 4) structure of the lower level at the higher level (i.e., 16 rooms with 16 locations). The implication here is that the generative model can be extended hierarchically to furnish very deep inference and learning in a multiscale environment, with an implicit coarse graining over successive scales (i.e., 16 buildings with 16 rooms with 16 locations).

Policies entailed four moves, where each move could be in one of four directions (up, down, left, right). This means that there are $4^4 = 256$ policies (i.e., plans) that could change the location state (but not the context state). Agents could reach any of the locations in a room from the starting location within this specified number of steps given the appropriate policy (or policies).

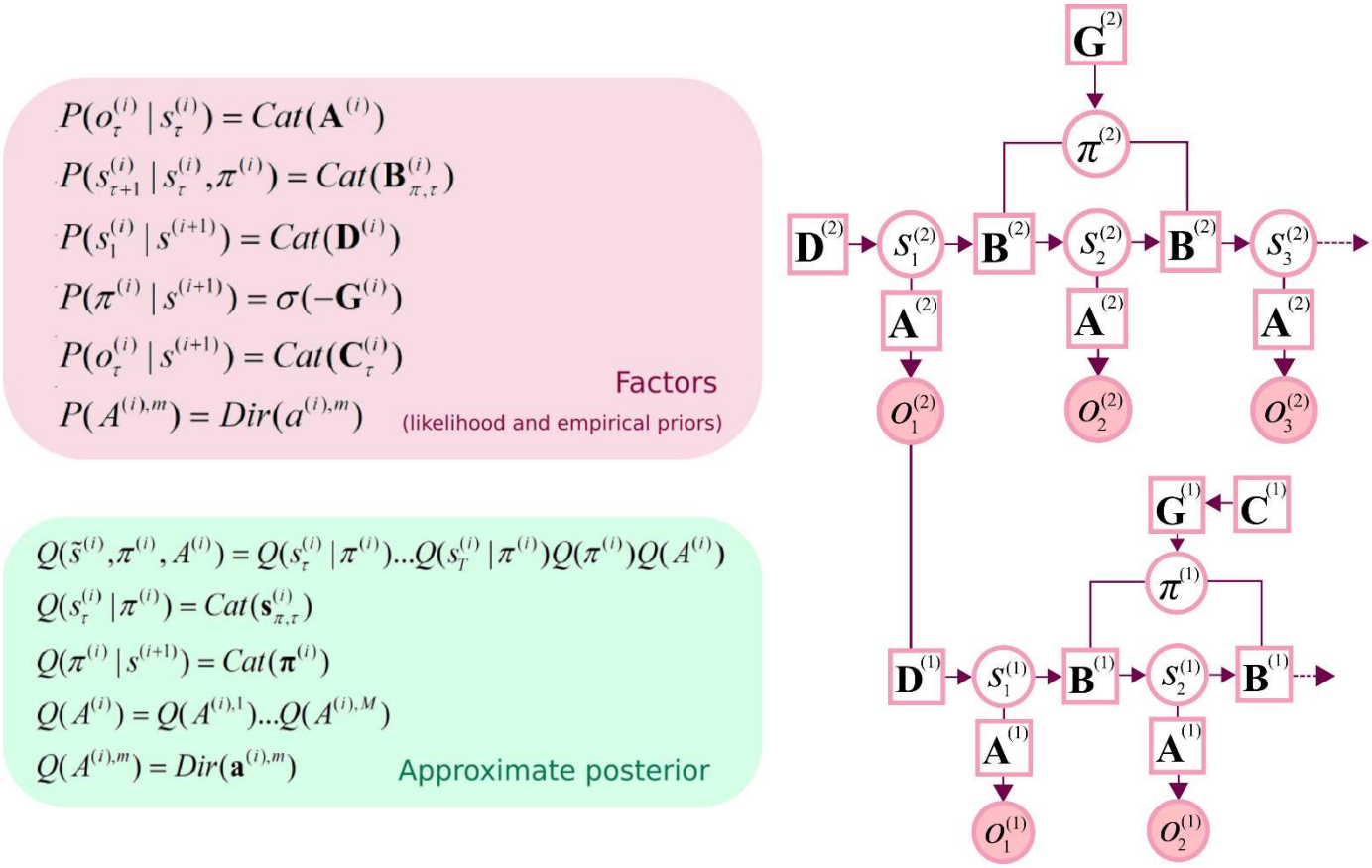


Figure 4.1 Graphical depiction of the generative model. This deep (temporal) generative model has two hierarchical levels. At the lower level there are two hidden state factors: *location* and *context*. These generate outcomes in three outcome modalities: *location*, *reward,* and *context* (i.e., room cue). At the higher level, there is one hidden state factor and outcome modality: *context* (room identity); the link between the higher and lower level is via the *context* factor. Latent states at the higher level generate initial states for the lower level, which themselves unfold to generate a sequence of outcomes. Lower levels cycle for a sequence of 5 time-steps for each transition of the higher level, and there are 5 epochs in the higher level for every iteration. This scheduling endows the generative model with a deep temporal structure. The likelihood **A** is a matrix whose elements are the probability of an outcome under every combination of hidden states. **B** represents probabilistic transitions between hidden states, which depend on actions determined by policies π. **C** specifies prior preferences and **D** specifies priors over initial states. *Cat* denotes a categorical probability distribution. *Dir* denotes a Dirichlet distribution (the conjugate prior of the *Cat* distribution). Please see Table 2.1 for a glossary of terms.

In each room, one location provided a preferred outcome or *reward: present* and there was a null outcome or *reward: absent* everywhere else. The outcomes *reward: present* and *reward: absent* were assigned a relative log probability (or utility) of 3 and 0, respectively. With these utilities, the agent would expect (or 'prefer') a *reward: present* outcome ≈ 20 times more than

the *reward: absent* outcome. At the lower level, the starting location was always the same, namely location 7 (i.e., to the left and below the centre of the room). Please see Figure 4.2a for three example illustrations of simulated behaviour in one of the 16 rooms.

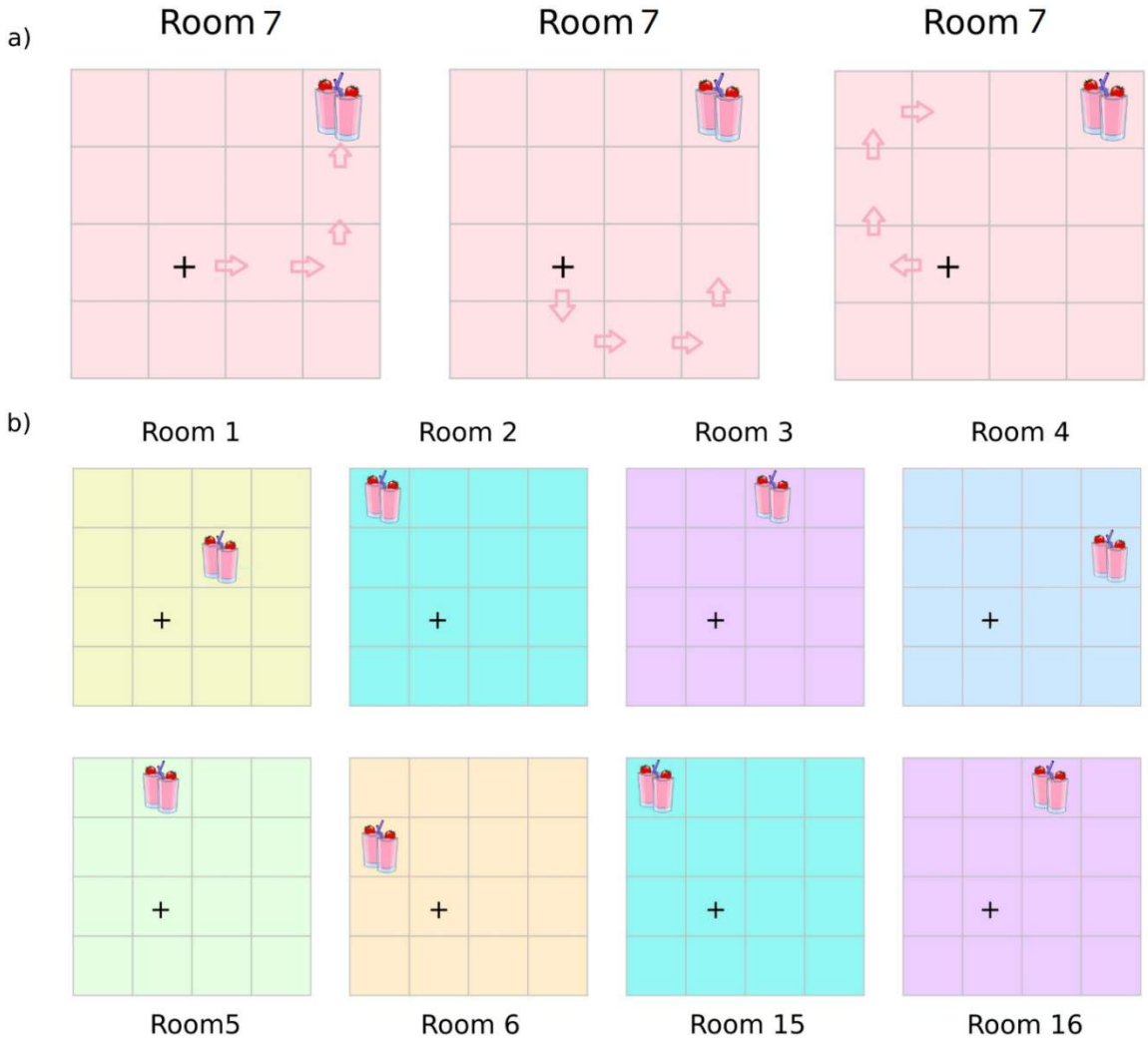


Figure 4.2 Example paths and room types. a) Examples of simulated paths or policies that agents could choose in one of the 16 possible rooms. The agents are allowed to make 4 moves, regardless of whether they find the reward or not. b) Sample rooms (out of 16 possible rooms). Rooms 2 and 15, as well as rooms 3 and 16 share the same contextual cue (colour) and reward location (locations 1 and 9 respectively). Every higher-level block involves foraging five rooms (whose identity is unknown) and exploring each of them in order, for four time-steps in the lower level.

The structure of the second level process is similar, however, the second level *room identity* states generate the initial state of the *context* (i.e., room) factor at the lower level. More specifically, the room identity (i.e., the content from the point of view of the second level) becomes the context from the point of view of the first level.

A crucial aspect of this (deep) generative model is that the higher level generates the initial states at the lower level. This means that for every transition at the higher level, there is a succession of transitions at the lower level. For every room the agent visits, there are five time-points at the lower level. The agent had control over only the *location* state through actions at the lower level, while changes in the *context* (i.e., room) depend on the state transitions at the higher level. This diachronic construction means that the *context* state cannot change in the course of a trial at the lower level. The ensuing state transitions relax the Markovian constraints on belief updating – and accompanying behaviour – given the implicit separation of temporal scales (Friston, Rosch et al. 2017). Note that the current model features similar contexts (i.e., rooms). This is important: although agents can only forage within a specific room, they can generalise the concept of that room to other similar rooms: 'This is a living room' (as opposed to a bedroom). Or 'This is an apartment', as opposed to 'This is an office space'.

At the lower level, agents initially have uniform beliefs about the context they find themselves in (i.e., the room identities), and they are not equipped with any preferred trajectory or sequential passage through the rooms. Furthermore, the agents have an imprecise mapping or knowledge of reward locations. This means that upon entering one of the 16 possible rooms, the agents were initially unaware of the identity of the context in which they were foraging, and believed they could be in any of the 16 rooms. Conceptually, this can be thought of as having an imprecise set of beliefs about what a room can contain: 'I know that I am entering Room 2, but I do not know what colour or reward location it entails, nor the relationship between the room colour and reward location'.

The Dirichlet parameters encoding the confidence or precision about these various beliefs (i.e., the likelihood mapping) were set to low values, such that accumulated experience would have a substantive effect on the corresponding posterior expectations about probabilistic contingencies. Importantly, although the process generating outcomes comprises 16 rooms, there were only 14 unique rooms, as described by the contextual cue (colour) and reward location: rooms 2 and 15 are identical both in terms of the colour and reward location, and so are rooms 3 and 16 (please see Figure 4.2b for an illustrative sample of rooms). This means that the agents have to learn there are only 14 unique context-specific reward locations. This presents a learning problem for the agents at multiple levels. First, they have to explore each room optimally, to resolve their uncertainty about whether there is a reward or not at each location. Furthermore, they have to explore all the rooms to resolve uncertainty about which colours, and reward locations would be elicited in the different rooms they explore. The agents can therefore leverage the information they know about the rooms (i.e., configuration) in order to pursue the reward. Notice that, by construction, this hierarchical model can be extended to arbitrary depth. That is, we could have rooms of rooms of rooms, navigated at progressively slower timescales. For example, one could forage within rooms, apartments, buildings, boroughs, cities and so on, whereby agents take several steps within a room during which the apartment, building, etc., does not change.

The generative model generates outcomes by first evaluating the Expected Free Energy for each policy (at the higher level), and selecting the most likely policy (Figures 4.1 and 4.3). Latent states are generated based on the transition probabilities specified for this policy. Latent states then generate outcomes in one modality (for this model, *context*), and the process repeats for the lower level, whereby the outcomes are generated in three modalities: *location, reward,* and *context.* Perception (i.e., inference about latent states) is equivalent to inversion of this generative model (given a sequence of outcomes). Learning corresponds to parametric updates.

Figure 4.3 summarises the associated belief updates about hidden states, policies and ensuing action selection using the free energy minimising solutions presented in Chapter 2.

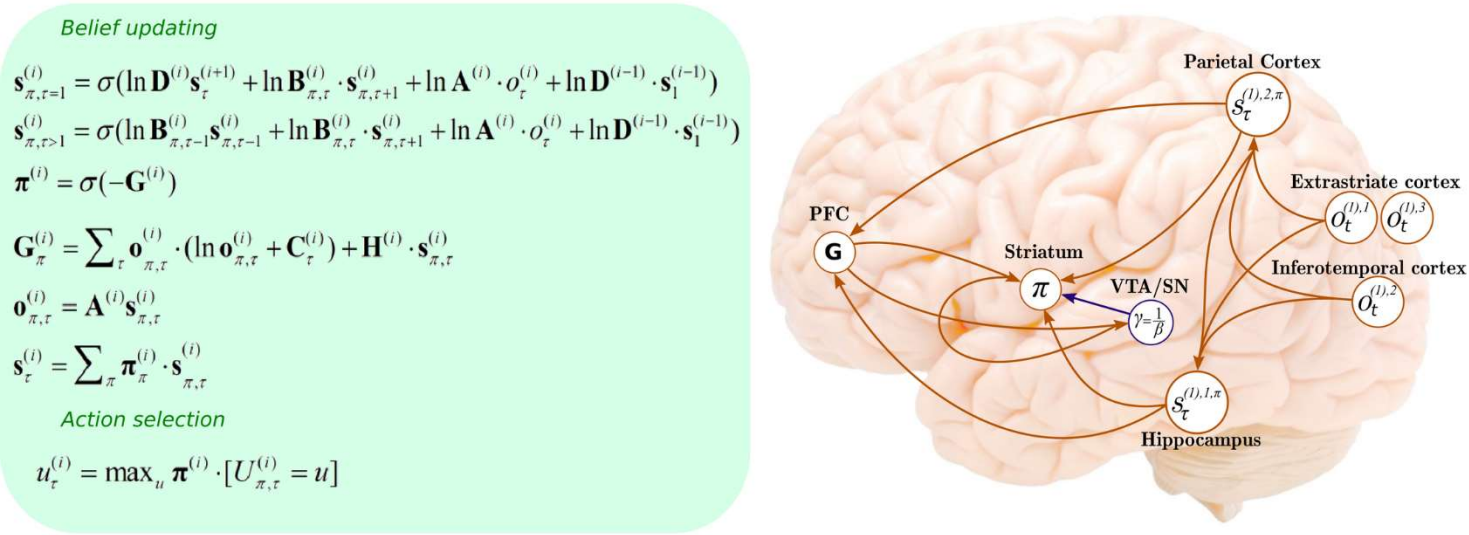


Figure 4.3 Schematic overview of belief updating. Left panel: belief updates defining Active Inference: state-estimation, policy evaluation and action selection. These belief updates are expressed in terms of expectations, which play the role of sufficient statistics for these categorical variables. Right panel: here, the expectations that are updated are assigned to various brain areas. This depiction is purely schematic, and its purpose is to illustrate a rudimentary functional anatomy implied by the functional form of the belief updating. Here, we have assigned observed outcomes to the occipital cortex, given its involvement in visual processing of spatial location (Haxby, Grady et al. 1991, Haxby, Horwitz et al. 1994), whereas *reward* outcomes are assigned to the inferotemporal cortex given its contributions to forming stimulus-reward associations (Spiegler and Mishkin 1981). Hidden states encoding the context have been associated with the hippocampal formation (Rudy, Barrientos et al. 2002, Miller, Neufang et al. 2013), and the remaining states encoding sampling location have been assigned to the parietal cortex, given its role in the encoding of multiple action-based spatial representations (Andersen 1995, Colby and Duhamel 1996, Silver and Kastner 2009). The evaluation of policies, in terms of their expected free energy, has been placed in the ventral prefrontal cortex. Expectations about policies *per se* and the precision of these beliefs have been associated with striatal and ventral tegmental areas, respectively, to indicate a putative role for dopamine in encoding precision (Friston, FitzGerald et al. 2017). The arrows denote message passing among the sufficient statistics of each factor or marginal. First and second digits in the superscript (e.g., $o^{(1),1}$) indicate the hierarchical level and modality, respectively. Please see glossary in Table 2.1 and (Friston, FitzGerald et al. 2017) for a detailed explanation of the equations and notation.

This model entails online planning – in the sense that, at each point in time, the agent evaluates future trajectories in terms of Expected Free Energy, and action is sampled from beliefs about those policies. Briefly speaking, agents form expectations about future states by projecting their posterior beliefs to the future epochs, under each policy (Friston, FitzGerald et al. 2016). Policies are then evaluated under these beliefs in terms of their Expected Free Energy, which involves goal-fulfilling and uncertainty-resolving components: c.f., expected value and information gain, respectively. This renders policy selection (implicitly) contingent upon expectations of future states under each policy. It is this aspect that lends synthetic agents the ability to plan (and explore). The sampled action is more likely to originate from the policy with a lower Expected Free Energy. The selected action generates a new observation, and the perception-action cycle continues.

Please note that for the lower level in these simulations, the entire set of policies (4^4 = 256) is in play. The set comprises a combination of every possible move (i.e., up, down, left, right) over 4 time-points (also called deep policies). However, in a given trial, agents can eliminate unlikely policies based on their evidence using an Occam's window (i.e., if the difference in log probability between a policy and the most likely policy is smaller than -3). This means that the agent computes combinatorics over actions up until the very last time point. The policies at the higher level refer to the potential rooms the agent can visit (i.e., 1-16); here, agents consider policies over just one time step into the future.

4.3 Simulations and results

The main focus of the work in this section is to illustrate the importance of Structure Learning as implemented by Bayesian Model Reduction. This will be demonstrated in the context of the Active Inference Framework by showing how it enables agents to form concepts and improve their performance, as scored by information gain, and the total reward gathered.

The kind of behaviour we hoped to elicit with this generative model can be described as follows: on repeated exposure to the rooms, the agents would explore optimally, defined by a trajectory that avoids previously visited (i.e., uninformative) locations, thereby enabling the agents to learn efficiently and remember which locations are rewarding and which locations are not. Note that this learning is context-specific, in virtue of including a context factor in the generative model. That is, each room has a location with reward and contextual cue (i.e., colour) to which the agent has access, regardless of the sampled location. Preferred outcomes are time sensitive, in that the agents are only permitted to explore for up to 4 steps in each room they forage. If the reward is not found within those 4 steps, foraging in that particular room ceases, and they move to a different room. This process repeats for a given number of blocks at the higher level, namely 2, 10, 20, 30, 40, and 50 blocks (of 5 rooms each). For the lower-level process, this means that rooms were respectively sampled for 10, 50, 100, 150, 200, and 250 trials, with five time-steps each.

In one group, Bayesian Model Reduction was applied to reassign Dirichlet parameters after each set of training blocks. This can be thought of as optimising the model structure in the absence of any further sensory information. In this instance, Bayesian Model Reduction is used to assess the evidence for reduced models that describe the structure of the environment. For example, one hypothesis (i.e., alternative model) describes an environment where each possible room has its own unique identity (despite the contextual cues and reward locations –

see Figure 4.4a). A second example hypothesis depicts the identical pairs of rooms as having a 50% probability of mapping on to identical contextual cues (Figure 4.4b). Another hypothesis expresses the rooms with similar reward locations and contextual cues as sharing a representation, therefore specifying the existence of only 14 rooms (Figure 4.4c). In a fourth example, rooms 15 and 16 have a uniform distribution over all the potential rooms, that is, these rooms are equally likely to have any of the other possible identities (Figure 4.4d). These exemplar hypotheses describe potential model spaces depicting likelihood mappings for the *context* factor. Values along a column must add to 1 since they represent a probability distribution across the different possible rooms – that is, each column represents the *context* states (i.e., room identity), and each row represents *context* outcomes (i.e., room colour). Since the combinatorics of potential model spaces are extremely high, we restrained the number of potential alternative hypotheses such that they reflect concentration parameters for the likelihood *context* matrix that were observed as a result of the training trials (i.e., such that they reflect the learnt state-outcome associations).

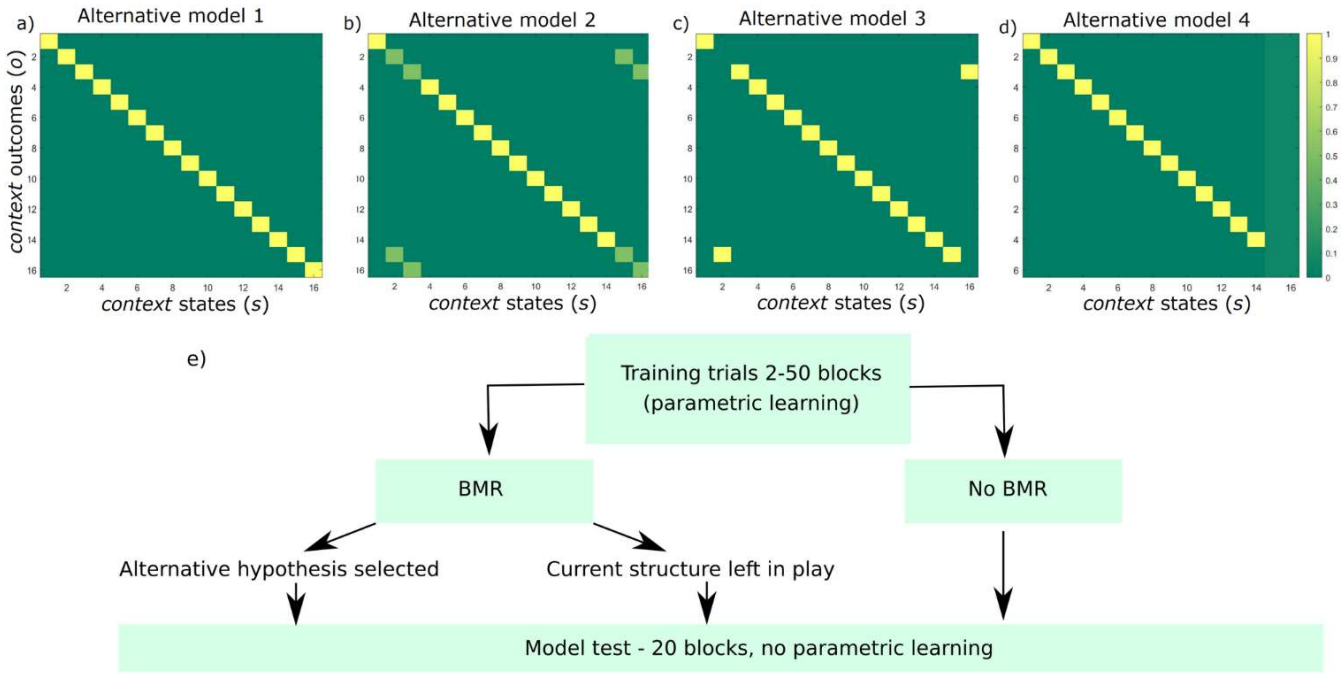


Figure 4.4 Example alternative models and flow chart depicting simulations. a)-d) Example alternative models (i.e., hypotheses). The generative process (also Alternative model 1) and alternative hypotheses were subject to Bayesian Model Reduction, focusing on the likelihood mappings encoding the *context* modality. Matrices represent a mapping from *context* states (columns) to the *context* outcomes (rows) – this can be thought of as *room identity* (*s*) to *room colour* (*o*). a) Note the identity matrix defined in the generative process is also used as an alternative hypothesis for model comparison. b) The second hypothesis depicts the identical pairs of rooms as having a 50% probability. c) The third hypothesis represents rooms 2 & 15 as being Room 15 and rooms 3 & 16 as Room 3. d) In the fourth hypothesis, rooms 15 and 16 do not exist, having a uniform distribution over all the other potential rooms – that is, rooms 15 and 16 are equally likely to have any other possible identities. e) Flow chart depicting the core simulations for all 120 agents – 60 undergoing the 'BMR' group, and 60 in the 'No BMR' group.

In the first half of the simulations (i.e., training trials), agents accumulated concentration parameters to learn the mappings from *context* and *location* states to the *reward* and *context* outcomes. For the testing phase, we ran simulations in both groups (i.e., models that did and did not undergo Bayesian Model Reduction) for a further 20 blocks at the higher level (i.e., 100 more trials involving the 16 potential rooms – please see Figure 4.4e for a graphical illustration of the simulation setup). During these 20 test blocks, agents in both groups (i.e., BMR versus No BMR) were precluded from accruing further concentration parameters, such that the performance with those specific posterior distributions accumulated up to that point could be assessed (Figure 4.4e).

In what follows, we first show how agents learn associations and form concepts about the identities and configurations of the rooms as a result of (Active) Inference, Parametric Learning, and model selection (i.e., Structure Learning - Bayesian Model Reduction). Next, we show the performance benefits of Bayesian Model Reduction, in terms of information gain, and the amount of reward accumulated. The next set of simulation results addresses the capacity of agents to learn and infer that certain rooms are identical, as defined by reward location and contextual cue (i.e., colour). Finally, we show one source of individual differences

in concept formation, namely how a stronger preference for obtaining reward impacts concept acquisition and performance.

4.3.1 Agents form concepts by inferring and learning the structure of their environment

For the training part of these simulations, agents started with uniform beliefs about which context (i.e., room) they find themselves in. As far as they were concerned, upon entering a (randomly) chosen room, they could be in any of the 16 possible room types. This is specified in the structural prior of the Dirichlet concentration parameters of the *context* likelihood matrix, initialised as a uniform 10^{-1}. The *reward* concentration parameters were initialised as 10^{-1} (i.e., imprecise priors) for the most plausible associations, and 0 otherwise. This means that agents had a set of initial beliefs about the configuration of rooms (in terms of potential reward locations), but were unable to make use of them without being able to discern the identity of the rooms. Initial concentration parameters over these modalities can be thought of as an *a priori* set of associations (i.e., synaptic connectivity) about contingencies in the world that sets the scene for subsequent inference and learning. Initialising the *context* likelihood mapping with uniform concentration parameters - and the *reward* likelihood mapping with small concentration parameters for the most plausible associations - is based on the following considerations: initialising with uniform concentration parameters would cause the agent to attribute any context and reward outcomes equally to all room identities (i.e., a uniform posterior distribution over the room identity states), essentially preventing the agent from learning distinct associations between states and outcomes. We could have initialised likelihood mappings with random concentration parameters, but this would have destroyed the relationship between true and learnt room labels (e.g., room 1 is learned as room 5). For a

straightforward interpretation of the reduced models (used in the BMR analysis), we therefore used the reward modality to resolve ambiguity about room identity.

As trials progress, agents update their beliefs about room identity (i.e., context) reflected in both the configuration of their associations (**a** matrix), and the increased probability of finding the reward. Figure 4.5a shows the averaged performance for agents foraging 50, 100, 150, 200, and 250 times (i.e., 10, 20, 30, 40, and 50 blocks respectively at the higher level), in terms of reward accumulated. Performance per block increases for all the agents, with an initial concavity, suggesting a preference for exploratory behaviour. In Figure 4.5b we show how the beliefs about context change through time for one agent, and accordingly becoming increasingly precise. Updates to the likelihood concentration parameters proceed as described above. A simple interpretation of this (Parametric) Learning is a change in connectivity between observations and the specific context, quantified by the number of times they are inferred to co-occur (Friston, FitzGerald et al. 2017). For this model, this corresponds to the number of times the reward location and contextual cue are associated with a particular room. In Figure 4.5b, rows represent the *context* outcome (i.e., room colour) and, and columns represent context state (i.e., room identity). This exemplifies the notion of concept acquisition: agents start with uniform beliefs about state-outcome associations and as a result of inference and learning, they acquire an explicit (and reasonably precise) representation of environmental contingencies.

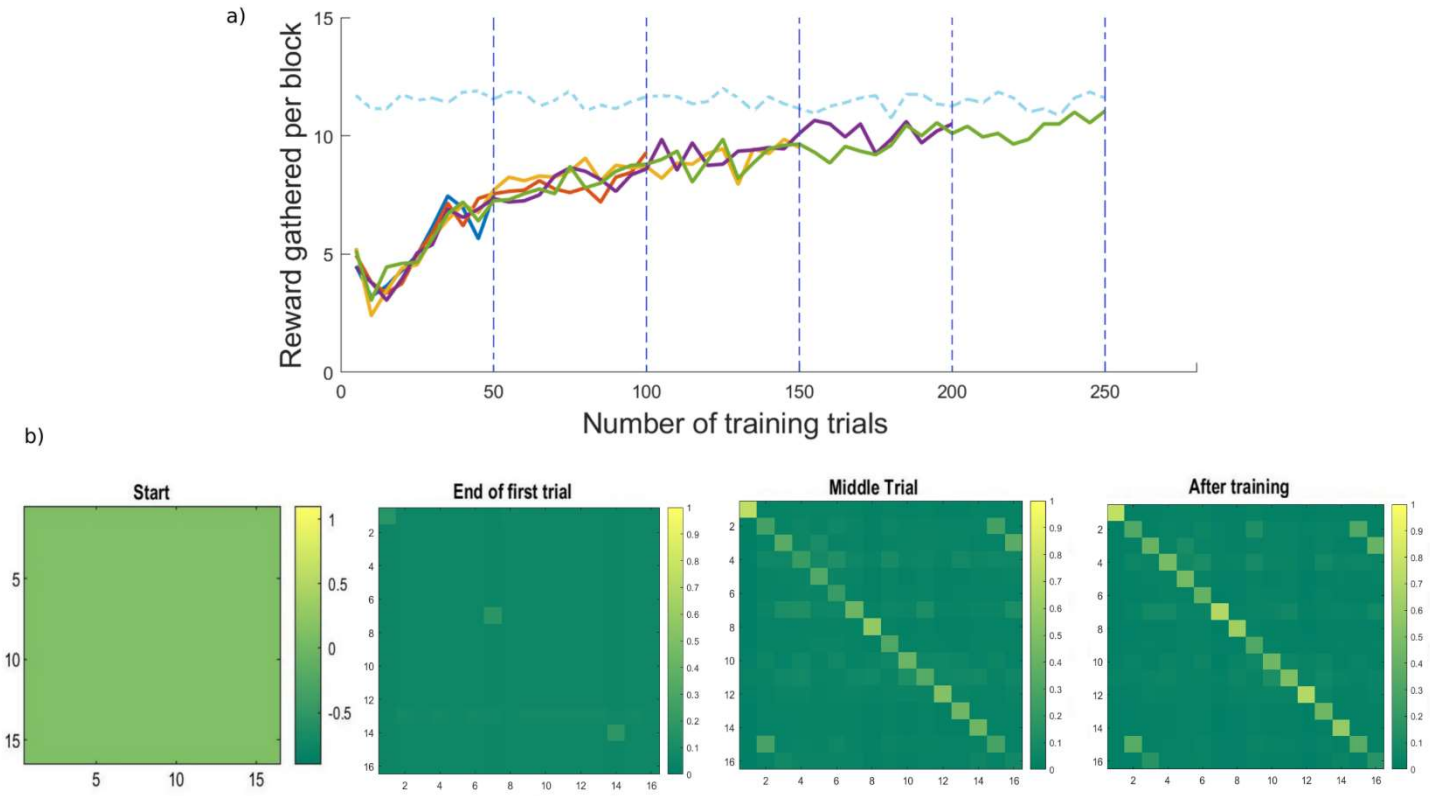


Figure 4.5 Average performance with learning. a) Progressive increase in performance scored by the amount of reward gained per block. For each higher-level block, five rooms at the lower level are explored. The performance is averaged over 20 simulated agents for each of the training settings: 10, 20, 30, 40, and 50 blocks (i.e., 50, 100, 150, 200, and 250 rooms respectively). Please note that the dashed blue line illustrates a cap in performance, represented by total reward gathered per block, averaged over 20 fully knowledgeable agents foraging for N=50 blocks (i.e., agents that start with fully precise likelihood matrices). As agents progress through the simulations, they accumulate more reward per trial. The concavity at the beginning of training reflects exploratory behaviour; i.e., intrinsic value predominated over the extrinsic value of rewards. b) Learning: the progressive updates to the concentration parameters over state-outcome associations from a uniform distribution to a more precise one, representing concept formation. The agent forages for N=50 blocks at the higher level (i.e., 250 lower-level trials). Middle trial represents the end of block number 25 (at the higher level).

The purpose of this sub-section was not only to show that synthetic agents are able to form concepts (resulting in better performance), but also to validate the generative model in its current implementation and set the scene for illustrating benefits of BMR in later sub-sections. Although Figure 4.5b depicts one particular agent's learning trajectory as it forages its

environment - there are 6 training conditions: we stop the training after 2, 10, 20, 30, 40, or 50 blocks. For each of the 6 conditions, there are 20 agents. This numerical experiment is used later to illustrate how state-outcome contingencies (and therefore performance) change with or without BMR, and whether these effects vary with the amount of training.

4.3.2 Performance benefits of employing BMR

Along with a gradual learning of contingencies about the external world, concept formation can also be a result of, and enhanced by, Bayesian Model Reduction – a faster and saltatory type of learning. In Figure 4.6 we show results for three of the (120) simulated agents. We can see in Figure 4.6a (left panel) the generative process generating the data; that is, the 'environment' that agents forage. It is useful to consider here the distinction between the generative process (i.e., environment) and the generative model (i.e., the agent). Active Inference Framework agents do not have direct access to knowledge about (hidden) states of the environment and must infer them based on observable outcomes. Although the structure of environment that generates data is an identity mapping (mapping from hidden states to observable outcomes) in the current work, this does not preclude the agents from acquiring a different mapping; so long as it helps the agents recognise the environment in a useful way that allows them to minimise uncertainty and gather rewards. In other words, the representations (i.e., concepts) that agents form, need not be identical to the actual form of the environment (and seldom are).

As previously mentioned, all agents start their foraging with an imprecise uniform distribution over their representations (of room identities). This means that they are not aware which rooms they are foraging in (Figure 4.6a, right panel). Figure 4.6b shows the posterior concentration parameters for three different agents after training for 2 (4.6b, left), 20 (4.6b,

centre), and 50 (4.6b, right) blocks (at the higher-level) - i.e., after 10, 100, and 250 training trials respectively). Training blocks consist of (Active) inferential and Parametric Learning processes described in Chapter 2. During this training period, therefore, agents learn gradually as they accumulate evidence about contingencies in the environment. After these training blocks, half of the agents undergo BMR. As a result of BMR, redundant parameters may be pruned, and agents form more precise representations about contingencies in the environment (encoded as state-outcome associations – Figure 4.6c). Because BMR maximises the evidence for a particular model or hypothesis, it does not just find the simplest possible model, but discovers the best balance between accuracy and complexity, thereby precluding over-pruning. Figure 4.6d quantifies the associated information gain using the Kullback-Leibler (KL) divergence between the posteriors and priors over likelihood parameters. The KL divergence is measured in nats: units of information based on natural logarithms. It can be seen in Figures 4.6d and 4.6e that Bayesian Model Reduction greatly enhances information gain. Furthermore, engaging BMR after 2 training trials provides a very marked change in concentration parameters relative to the two other conditions (Figures 4.6b and 4.6c below). We can also see in Figure 4.6c that agents who experienced fewer blocks ended up selecting different alternative hypotheses as compared to the agents training for more blocks after undergoing BMR. We expand on these results in the next sub-section.

Figure 4.6e shows a comparison of information gain (in nats) before and after BMR. This is averaged for all agents in the BMR group, in each of the 6 conditions (i.e., N = 2, 10, 20, 30, 40, 50). As expected, information gain shows an upward trend with the amount of training blocks both before and after BMR. The trends observed in Figure 4.6d for the three different agents hold for the entire set of agents: there is a marked difference in information

gain when comparing between before (light green bars) and after (light blue bars) undergoing BMR.

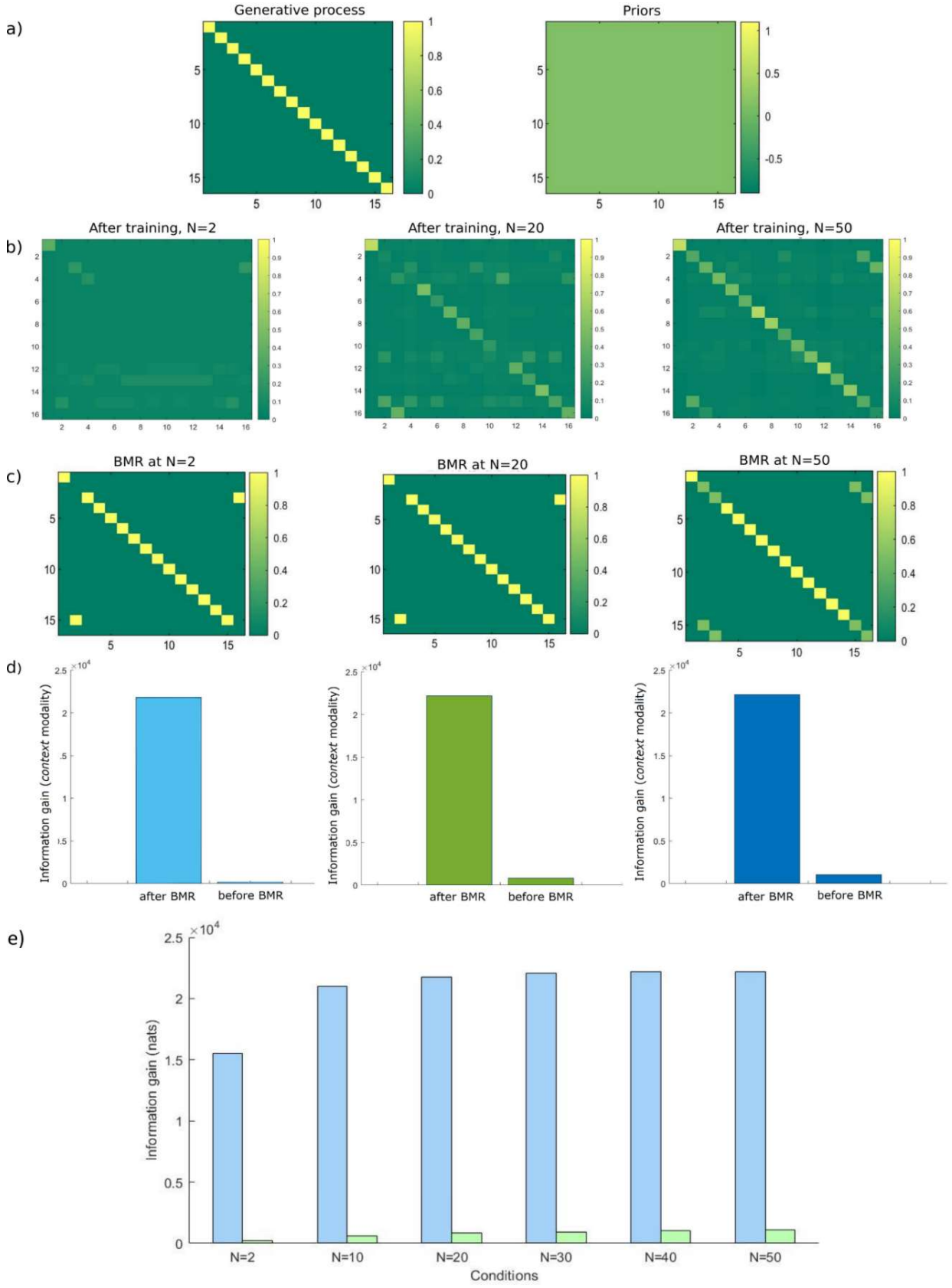


Figure 4.6 Likelihood mappings from hidden states to *context* outcomes, before and after BMR, and how these learned mappings affect concept formation for three different agents. Matrices represent the likelihood mapping from *context* states (i.e., columns) to *context* outcomes (i.e., rows). a) The process generating the actual state-outcome mappings (left) and the uniform concentration parameters that agents start with (right). b) Likelihood matrices for the three agents (averaged over all locations) at the end of 2, 20, and 50 training blocks at the higher level of foraging (from left to right). c) Likelihood matrices after BMR, showing the reduced set of state-outcome associations (i.e., likelihood) for the *context* factor. d) Information gain for the *context* modality before and after BMR for each of the three agents. e) Comparison of information gain before and after BMR, averaged over agents for each condition; light blue bars denote information gain after BMR whereas light green bars denote information gain before BMR (i.e., after N training blocks).

Next, we turn to the benefits of Bayesian Model Reduction for goal-directed behaviour in terms of performance, defined as the time spent with reward – and how often the reward was found. As described in the introduction, we can regard BMR as off-line hypothesis testing in the absence of further information. Figure 4.7 shows the comparison between the two groups in question: BMR versus No BMR. In a training phase, agents foraged the environment for either 10, 50, 100, 150, 200, or 250 trials (i.e., 2, 10, 20, 30, 40, and 50 blocks at the higher level). In one group, the agents were subject to Bayesian Model Reduction (i.e., 'BMR' group). If the free energy (i.e., negative model evidence) is lower for any of the potential hypotheses, the mixture of Dirichlet parameters is accepted and redundant 'synaptic' connections are effectively pruned. Conversely, if the free energy is higher, the original structure is left in play. Following BMR, the agents continued foraging the rooms for a further 20 blocks (at the higher level). In the other group, agents continued foraging the rooms with the posteriors accumulated during the training trials, without BMR (i.e., 'No BMR' group).

Choosing to stop training after different numbers of blocks (i.e., 2, 10, 20, 30, 40, 50 blocks) allows for a comparison between different stages in the learning trajectory and its effects on performance: we can see in Figure 4.7 that in the 'No BMR' group there is a sharp

jump in performance from 2 to 10 to 20 training blocks, which levels out with an increased number of training blocks, but remains below the performance for agents in the 'BMR' group.

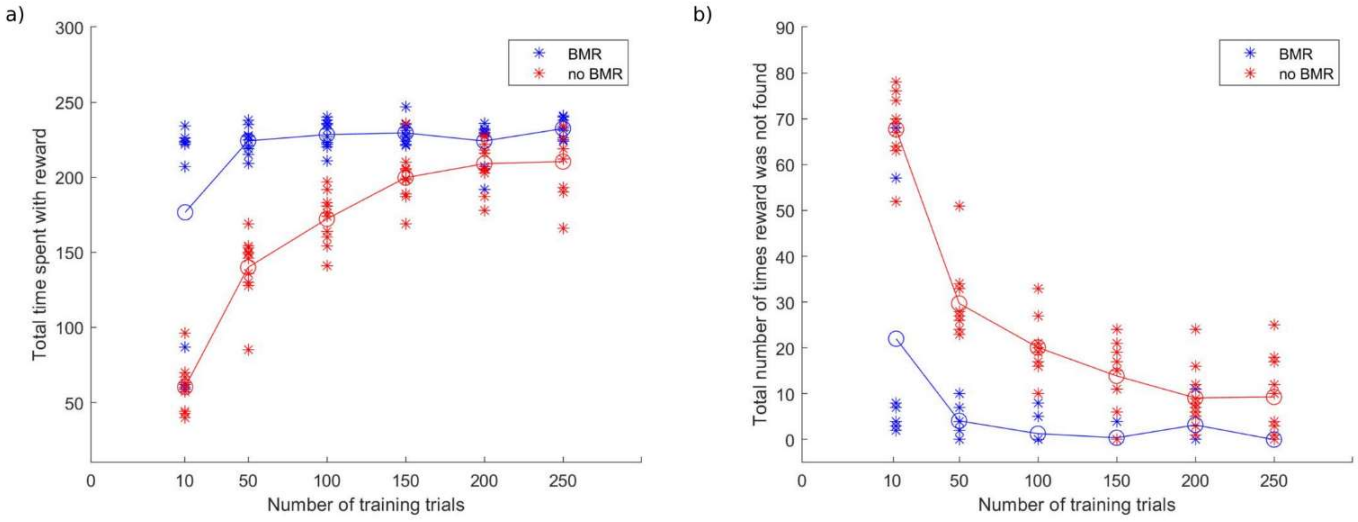


Figure 4.7 Performance comparison between agents undergoing BMR versus continuing with the posteriors accumulated after a specified number of training trials (at the lower level). Each asterisk represents an agent; circles represent performance averaged over agents at a specified number of training trials and their respective group (BMR vs No BMR). Agents with *n* training trials are assigned to one of the two groups, and then continue to forage for another 100 (lower level) trials. a) Total reward gained – agents undergoing BMR perform almost at peak, even before foraging through all of the 16 rooms. b) The number of times reward was (not) found – agents in the 'no BMR' group spend more time foraging without finding the reward. Performance improves for agents in both groups as they undergo more training trials.

Agents in the BMR group appear to perform almost at peak even before foraging in all of the 16 room types. This is because although the agents are not exposed to new sensory information, the beliefs encoding the *context* factor become very precise, precluding further epistemic foraging, and therefore emphasising extrinsic value (i.e., agents become relatively more exploitative). The performance for the agents in both groups improves gradually, levelling off after training for approximately 50 higher level blocks (i.e., 250 lower-level trials).

Furthermore, the agents in the 'No BMR' group spend more time foraging the rooms without finding any reward, a performance characteristic that does improve with more training.

4.3.3 Agents can learn that rooms with similar configurations are identical

One aspect of concept formation concerns the ability to represent invariance and symmetries. In this sub-section, we show that agents learn to associate the rooms with identical configurations (i.e., identical colour and reward location) to form associations that encode similarity, defined as the state-outcome connectivity of the *context* factor. This is a result of learning; however, this aspect is evinced more clearly following Bayesian Model Reduction. Most importantly, none of the simulated (sixty) agents undergoing BMR settled on the hypothesis specifying an identity mapping for the *context* factor. This means that despite having a process generating observed outcomes (i.e., generative process) with an identity mapping (i.e., each room has an individual identity), none of the agents judged this mapping (of state-outcome associations) as being the most parsimonious (i.e., explaining the observations accurately, in as simple a way as possible). The representations (i.e., concepts) formed by synthetic agents can – but do not have to - reflect the actual form of the environment, as long as they aid synthetic agents in interpreting the environment in a useful way (i.e., allows them to minimise uncertainty and gather rewards).

The agents come to recognise the rooms as being different only when their configurations could be disambiguated (i.e., as a result learning, rooms with different reward locations and contextual cue were not confused with each other). Figure 4.8a shows the final encoding of environmental structure (i.e., *context* mappings) for three different agents as examples. The agent on the left in Figure 4.8a believes that rooms 2&15 are room 15, and rooms 3&16 are room 16. The middle agent believes that rooms 2&15 are room 2, and rooms

3&16 are room 16. The agent on the right, however, believes that there is a 50% probability of being in either of the rooms with identical configurations. For example, when this agent is in room 15, it believes that it could be in either room 15 or room 2, with equal probability.

There is diversity in terms of these learnt mappings, based on variations in foraging the rooms as a result of the order of observations. As noted above, when implementing BMR for Dirichlet hyperparameters (in this case the *context* likelihood mappings), agents compute a relative log evidence (i.e., free energy) for each model, and compare this score to the evidence of the parent model. Subsequently, agents select the model with the greatest evidence (Figures 4.8b and 4.8c). The most frequently chosen alternative hypothesis (i.e., alternative model) is the one where there is a 50% probability of being in either of the rooms represented by identical configurations (i.e., model/hypothesis 7). Figure 4.8b also shows the percentage of time the alternative hypothesis (i.e., alternative model) with a 50-50 probability for the rooms with identical configurations was chosen when applying BMR for the entire set of agents (i.e., when applying BMR to all the agents after various training blocks), consisting of 120 synthetic agents.

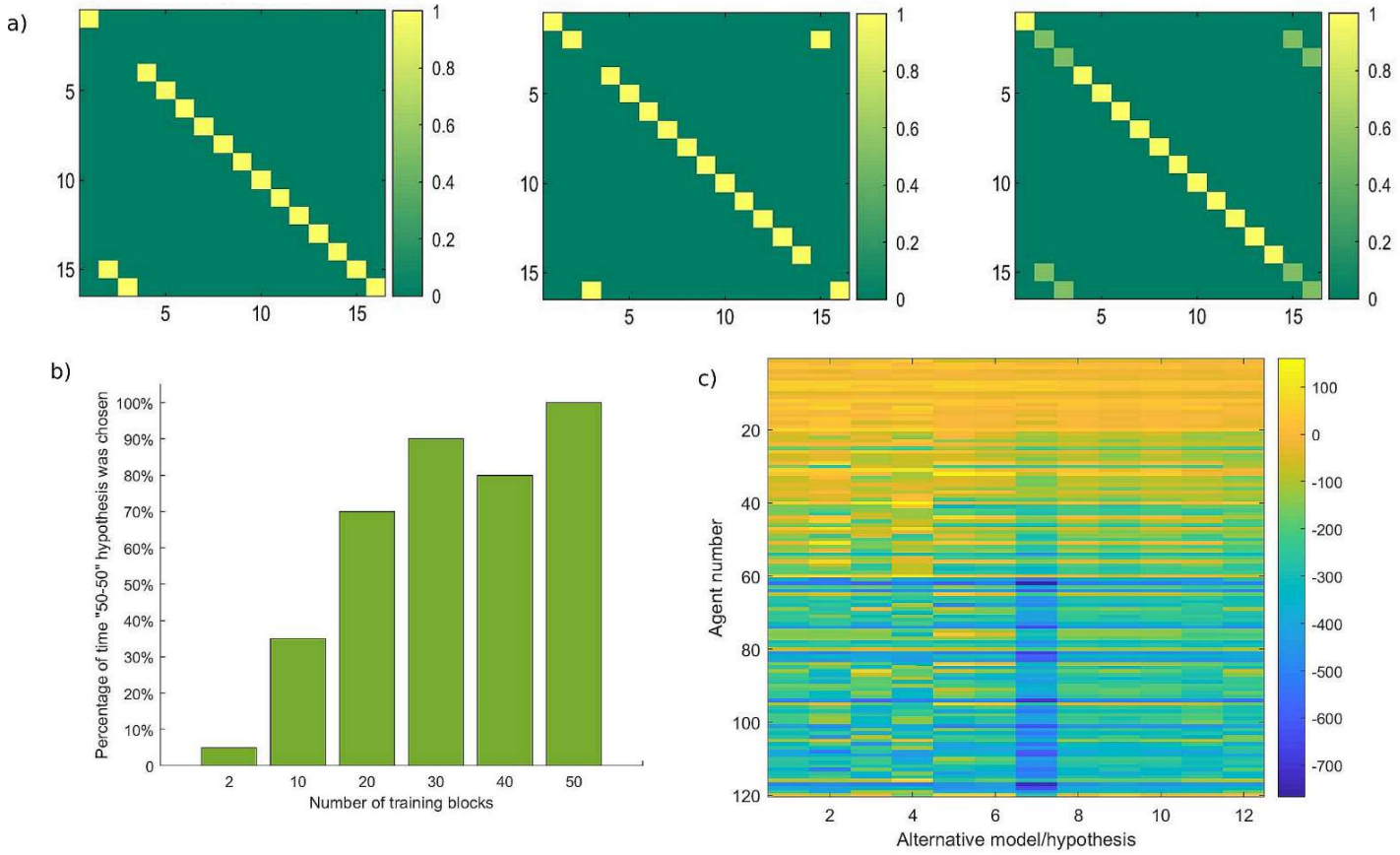


Figure 4.8. Possible representations of *similarity* between the rooms for three different agents after BMR and the most frequently chosen hypothesis during BMR. a) Likelihood matrices representing the reduced posterior concentration parameters. The matrices represent the *context* state-outcome mappings with rows representing the context state, and the columns representing the context outcome. The likelihood mapping for the first agent shows that rooms 2&15 as having the identity of context 15, and rooms 3&16 as being context 16. The second agent's beliefs show that rooms 3&16 have the identity of context (room) 16, and rooms 2&15 as having the identity of context 2. The third agent believes that there is an equal probability for the rooms that are identical in terms of their configuration: 2&15 can be either context 2 or 15 and rooms 3&16 are equally likely to be either context (room) 3 or 16. b) The percentage of time the hypothesis with an equal ('50-50') probability for the rooms with identical configurations was chosen by the agents, for different numbers of training blocks. At N=50 (i.e., after 50 higher level training blocks) this hypothesis is chosen 100% of time – that is, all 20 agents training for 50 higher level blocks, selected this hypothesis as being the most parsimonious, explaining the observations with the least model complexity. c) The negative log evidence for the twelve alternative hypotheses/models (x axis) for the entire set of agents (y axis, 120 agents). Model 7 appears to consistently have the greatest evidence (i.e., least free energy).

Interestingly, in addition to learning associations that encode similarity between rooms, in some cases agents also showed a similarity in simulated neural activity as characterised by (simulated) local field potentials, firing rates and dopaminergic responses. We illustrate the electrophysiological responses, associated with belief updating, for *one agent* foraging two rooms with identical configurations (rooms 2 and 15), during the training blocks (N=50) (Figure 4.9). The agent follows a similar trajectory in these two rooms, gathering reward for the last two time-steps. The top-left panel of Figure 4.9 shows average local field potentials over all the units encoding the *context* factor before (dotted line) and after (solid line) bandpass filtering at 4Hz, juxtaposed with its time frequency decomposition. The lower-left panel illustrates evidence accumulation for these units. The top-right panel shows the rate of change of neuronal firing. Finally, the lower-right panel illustrates simulated dopaminergic responses defined as an amalgamation of precision and its rate of change. This is reminiscent of demonstrations (using Representational Similarity Analysis) that neural activity patterns to repeated presentations of identical or related stimuli are likewise very similar (Mack, Love et al. 2016). Please see the Appendix for a contrast in electrophysiological activity that ensues as a result of *the same agent foraging the same room, with the same trajectory* (i.e., Room 15) at two different trials, and of *foraging different rooms* (i.e., Rooms 4 and 12) *with a similar trajectory*.

Posterior beliefs about policies are obtained by applying a softmax function to precision weighted (negative) Expected Free Energy of each policy. The precision parameter is estimated as new observations become available, and it plays the role of an inverse temperature, meaning that the policy with the least Expected Free Energy becomes more likely to be selected if the precision parameter is high. In other words, this precision encodes the confidence that the inferred policies will lead to preferred outcomes or will resolve uncertainty about the hidden states. Previous work (Schwartenbeck, FitzGerald et al. 2015) suggests that the dopaminergic

activity in the mid-brain might encode this kind of precision. In our paradigm, the phasic bursts we see in simulated dopaminergic responses indicate that at step 2 (i.e., the 32nd iteration in terms of updates – Figures 4.9a and 4.9b, bottom right panels) the agent becomes more confident (i.e., resolves uncertainty) about which policies to pursue, having eliminated the possibility that the room it is foraging is room 14, given that it did not discover a reward at location 6, which is the rewarding location for room 14. During the second spike at step 4 (i.e., the 64th iteration) the agent eliminates further possible policies, having become more confident that the room it is foraging is neither room 3 nor 16 (with reward at location 9), since it found a reward at location 1. This example illustrates how one can unpack belief updating and decision-making, while encoding uncertainty and precision.

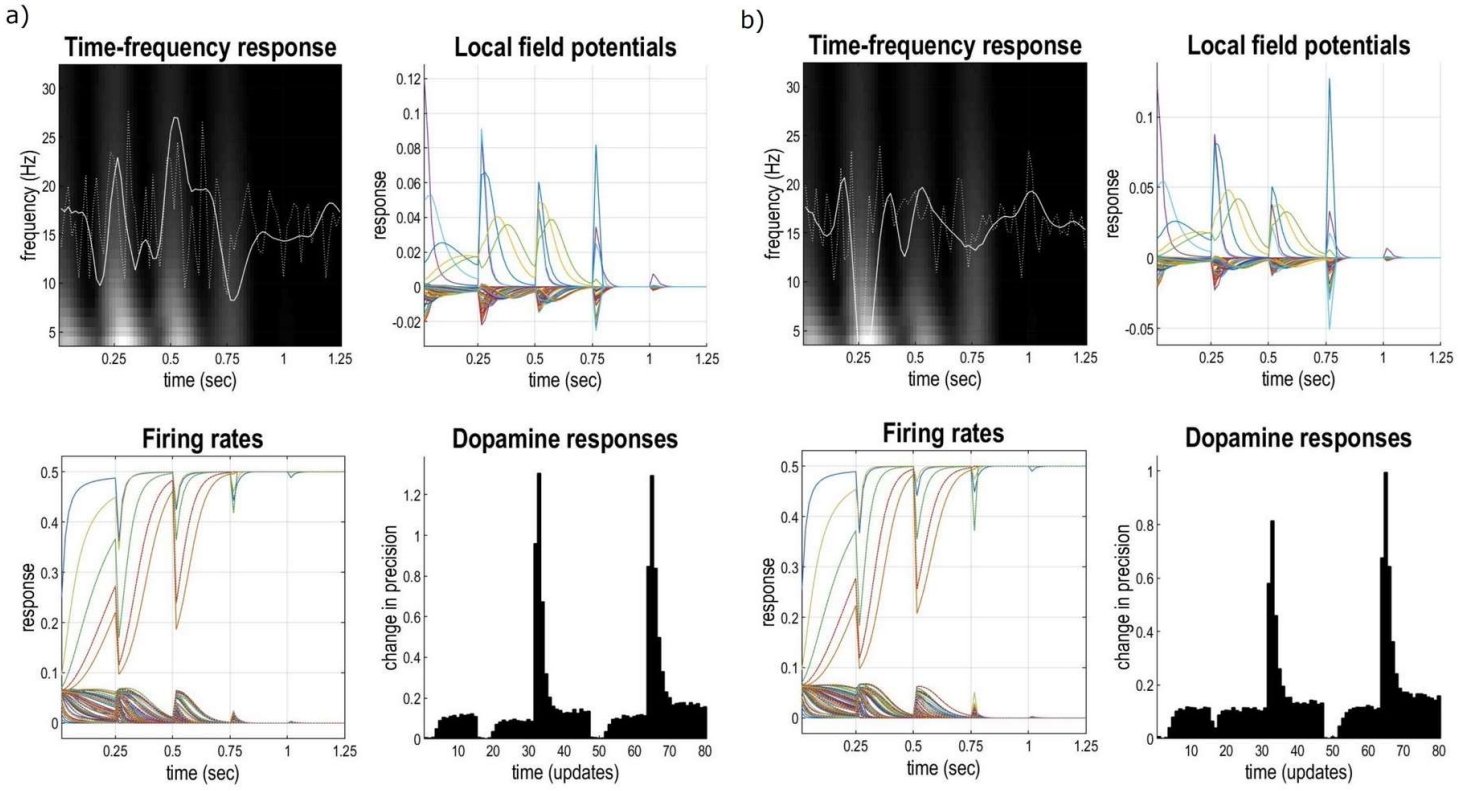


Figure 4.9 Neural activity for a synthetic agent in two rooms with identical configurations. In these epochs, the agent forages the two rooms in the same manner – that is, it follows the same trajectory of locations. During the last two steps, the agent encounters the reward and stays with the reward for one more step. Please see main text above for more details. a) Room 2 b) Room 15.

4.3.4 The strength of prior preferences impacts concept formation

One source of individual differences in concept formation reflects the preference for some outcomes over others. We asked whether prior preferences (i.e., regarding reward) influence learning and subsequent performance, by simulating three agents who experienced the same number of training blocks at the higher level (N=20). For one agent, we reduced the precision of prior preference over outcomes (reward) to 0.5 and 0 elsewhere (as compared to the default used in all other simulations of 3 and 0 elsewhere). This means that the agent has a weaker reward preference, compared to its conspecifics. Figure 4.10a shows the learned *context* likelihood matrices for the agents that have different degrees of preferences for reward. The third agent (Figure 4.10a, right) starts with a fully precise set of likelihood matrices. We use this agent as a baseline to help illustrate the performance comparison between agents with higher and lower precision in prior preferences, providing a cap on the total amount of rewards that agents can gather. The agent with weaker preferences does not accumulate as much reward (Figure 4.10b) but learns more (Figure 4.10c). Here, the degree of learning was assessed with the information gain or KL divergence between posterior and prior Dirichlet concentration parameters for the *context* likelihood matrix. This quantifies how much the agent has learned about the state-outcome associations from the start of the simulations. This example shows that the agent with a weaker preference for rewards is more sensitive to epistemic incentives, and subsequently, learns more efficiently.

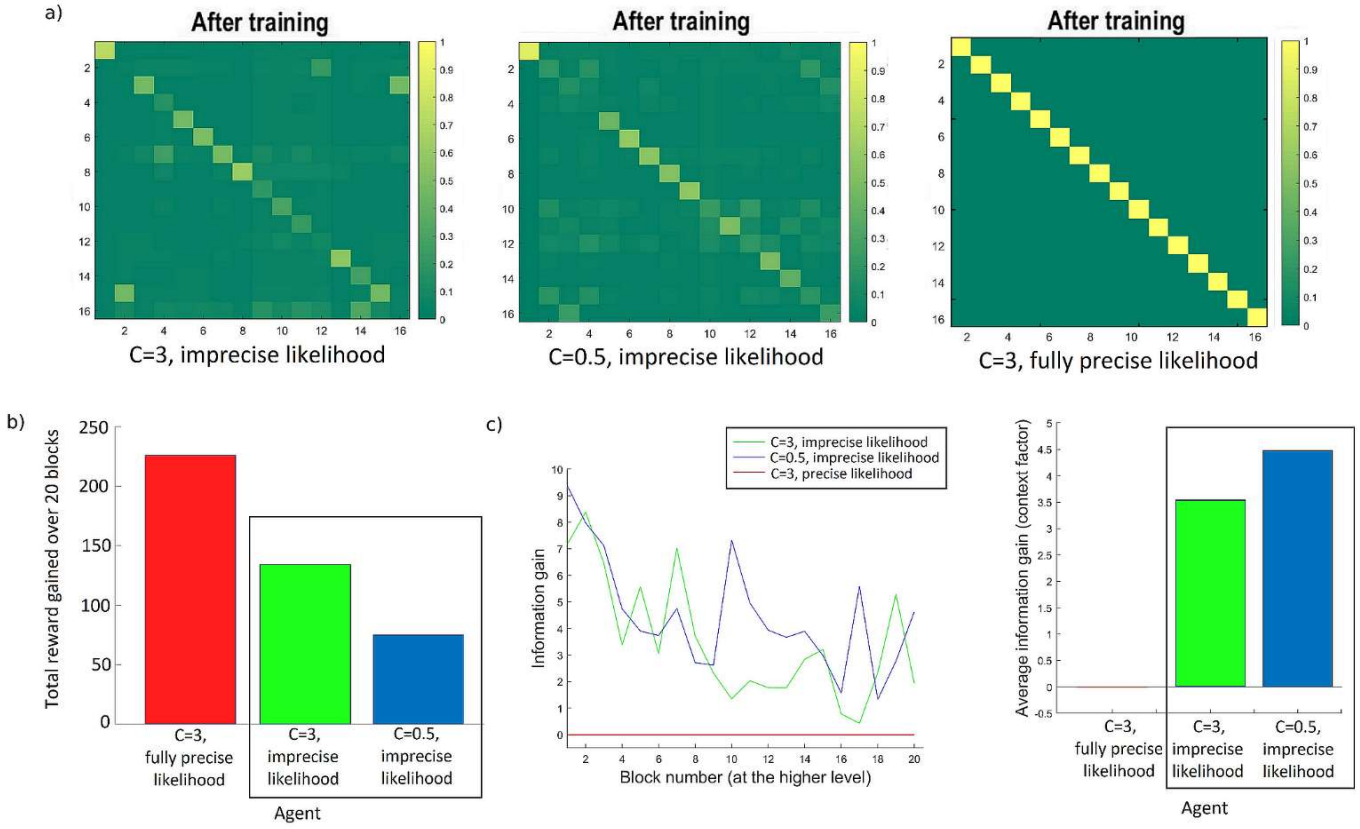


Figure 4.10 Performance comparison between an agent with a strong preference for reward (C=3) versus an agent with a weaker preference (C=0.5). a) Likelihood mappings after 20 training blocks, including a fully knowledgeable agent (right). b) Total reward accumulated over 20 (higher level) blocks. c) Comparison between the two agents, in terms of the information gain associated with the *context* modality. The agent with weaker preferences does not accumulate as much reward (Figure 4.10b) but learns more (Figure 4.10c).

4.4 Interim discussion

This work focused on the importance of Structure Learning as implemented by Bayesian Model Reduction. A series of simulations was presented: we first established that agents form concepts and gather rewards as a result of interacting with their environment. Consequently, we demonstrated the benefits of Bayesian Model Reduction, defined as an improvement in performance. Synthetic agents foraged a novel environment defined as a set of (16) rooms (contexts), each having 16 available locations. These agents came to learn the identities of the

rooms (via updates to concentration parameters) and form precise beliefs about the structure of the environment they were foraging i.e., they formed concepts. By learning their environment, agents gradually start performing better – they develop the ability to find the reward more often and gather more reward overall.

Bayesian Model Reduction enhanced the implicit concept acquisition and formation, endowing (synthetic) agents with representations that are more precise. In sub-sections 4.3.1 and 4.3.2 we saw that the parametric beliefs encoding the likelihood mappings of the *context* factor change (i.e., diverge from the initial uniform beliefs) depending upon the amount of training and whether or not BMR has been applied. There was a marked difference in information gain after BMR relative to before BMR. Undergoing BMR entails higher information gain with regards to the parameters encoding the (state-outcome) representations. To my knowledge, this was the first attempt to compare information gain in AIF agents using BMR – by comparing between information gain following BMR, and information gain as a result of Parametric Learning. Future work could use this metric to ascertain how and when it would be most useful for both machines and humans to engage in Structure Learning.

Interestingly, during the learning (i.e., training) process, agents occasionally 'mislabel' rooms, i.e., appear to be confusing room identities. This is most likely due to the way these agents forage the environment: for example, if we have one agent foraging in a room, and the foraged locations do not contain the reward, the agent is likely to label this room as any of the other rooms whose corresponding locations do not contain a reward, everything else being equal.

Further to the effect of increased precision in concept formation, we have illustrated the performance benefits of BMR (sub-section 4.3.2) in the context of goal-directed behaviour. In these numerical experiments, alternative hypotheses about the structure of the likelihood matrix

(i.e., the agent's beliefs about the set of rooms and their identities) were entertained. Selecting the most parsimonious hypothesis (i.e., the one with the greatest evidence) allowed the agents to minimise their uncertainty about their environment, and to use this knowledge (i.e., room identity as defined by its cue) to secure rewards. For example, after BMR, agents become more confident that the room they are foraging is pastel orange (i.e., room 6), and they can use this information to head to the reward (left and up on the first move), rather than responding to epistemic affordances or novelty. Agents undergoing BMR performed consistently better than agents not undergoing BMR. Furthermore, agents in the BMR group performed well, even before foraging all the rooms. This is important because it speaks to the ability of generative models with higher evidence to be generalised to new data and contexts (MacKay 2003). This reflects the implicit idea behind (Love and Gureckis 2007, Mack, Love et al. 2016, Mack, Love et al. 2018) whereby the reorganisation and restructuring of information changes the form of concepts, regardless of whether more external information is assimilated or not.

This offline aspect of reorganisation and restructuring was also observed in the problem-solving literature, where no further sensorial (or factual) information is necessary for insight (Kounios, Frymiare et al. 2006, Kounios, Fleck et al. 2008, Weisberg 2013, Shen, Tong et al. 2018, Tik, Sladky et al. 2018). Furthermore, Bayesian Model Reduction has been associated with physiological processes such as the regression of synaptic connections or pruning, observed during periods of sleep (Tononi and Cirelli 2006). In this setting, the structure of generative models is learned by minimising model complexity in the absence of sensory data, when accuracy does not contribute to log evidence (Hobson and Friston 2012, Pezzulo, Zorzi et al. 2021). For example, in sleep, endogenous activity – that resembles neural message passing in wakefulness – has been interpreted as the generation of fictive data to evaluate model evidence: c.f., (Hinton, Dayan et al. 1995). That is, fictive episodes are

'replayed', in the absence of (precise) sensory information, in order to optimise generative models (with the implication that this kind of model reduction facilitates generalisation).

In sub-section 4.3.3 agents were shown to exhibit a pronounced inter-agent variability: a characteristic that pertains to the outcomes sampled, rather than the agents themselves. This has implications in the realm of individual differences, because it can potentially elucidate how different individuals sampling different (sensorial) observations can reach the same conclusion (i.e., alternative hypothesis defining contingencies in the outside world), as well as how similar individuals sampling similar observations can reach different conclusions, as observed for example in (Finlayson, Neacsu et al. 2020), where individuals exposed to similar sensory information diverged in terms of whether they perceived a (bistable) stimulus as a vase or face.

We have shown that as agents forage and learn about their environment, they also come to ascribe the same identity to rooms with similar configurations (i.e., colour and reward location). Furthermore, the similarity in representation was accompanied by very similar neurophysiological responses, as seen empirically in the concept learning literature by Love, Medin et al. (2004), Love and Gureckis (2007). However, it remains to be investigated whether this phenomenon holds universally. For instance, it is unclear whether similar representations at time-points far apart evoke the same effect, or whether there is a relationship describing discrepancies between items of the same class. Interestingly, in sections 4.3.1 and 4.3.3, we saw that for conditions with shorter training duration (i.e., N=2, N=10), Bayesian Model Reduction appears to promote generalisation, with agents perceiving the rooms with identical configurations as being one room. After more experience however (i.e., for the conditions with longer training duration), the agents' beliefs seem to diverge again, 'perceiving' the rooms with similar configurations as having a 50-50% probability of being each of the two possible rooms. That is, agents are retaining both representations in an attempt to maintain a 'flexible' set of beliefs, regardless of having evidence to the contrary (i.e., that they do not need two separate

concepts). A future research direction could identify this potential computational benefit of developing and retaining a 'flexible' set of beliefs about contingencies in the lived world, in light of the Active Inference Framework. That is, what are the useful measures when deciding whether to retain a more flexible but less precise set of beliefs, versus a more rigid but also more precise set of beliefs?

Finally, we considered one source of individual differences, namely the strength of prior preferences for reward. Preferences affected concept acquisition and therefore the way agents formed representations. An agent with an imprecise preference for reward explored its environment more, and diverged more from its prior beliefs encoding state-outcome associations. Making more exploratory choices in this case hindered performance, in terms of reward gained, as well as the number of times the reward was found. These results are reminiscent of work by (Tschantz, Seth et al. 2020): here, the authors demonstrate that in Active Inference, uncertainties pertaining to the agents' goals and preferences are prioritised over other types of uncertainty. The Active Inference Framework thus provides a Bayes optimal and principled approach to balancing epistemic (i.e., exploratory) and instrumental (i.e., exploitative) actions. As predicted by (Tschantz, Seth et al. 2020), this balance depends on the shape of agents' beliefs; in our case underwritten by prior preferences. In light of these results (and results presented in this chapter), agents minimise uncertainty insofar as it is required for fulfilling their goals, whatever they may be defined as. When the imperative for satisfying prior preferences is diminished in relation to epistemic imperatives, the balance between exploration and exploitation shifts towards explorative behaviour, and vice versa. Future computational and empirical work may involve assessing agents with different levels of reward preferences, to see whether agents with a strong preference end up forfeiting exploratory behaviour (and, with it, predictive power) in an attempt to obtain rewards.

Chapter 5

Evidence for Structure Learning using an abstract rule-learning task in humans

5.1 Introduction

Abstract thinking and reasoning have been popular avenues of empirical research since the inception of cognitive science. Although there is a plethora of computational research attempting to model this type of higher-order cognition (Goodman, Tenenbaum et al. 2008, Barsalou 2009, Goodman, Tenenbaum et al. 2014, Blass and Forbus 2016, Lake, Ullman et al. 2017, Conway 2020, Mitchell 2021, Combs, Lu et al. 2023, Ellis, Wong et al. 2023), many unanswered questions remain on the nature of the underlying information-processing.

The focus of this chapter is to provide evidence for Structure Learning, using an abstract rule learning task in which both human and synthetic agents engage. The main feature of this task is the stark difference in performance between discovering and not discovering the underlying rules, or the necessity of reasoning (above and beyond associative learning), that is often accompanied by a moment of insight. The two conditions (i.e., *discovered* vs *undiscovered*) emerge naturally from the experimental configuration of the task. That is, human subjects either discover the rules or they do not, with no middle ground. Here, we evaluate a set of potential mechanisms for abstract thinking and rule learning — in the context of the Active Inference Framework (AIF) — using principles of (Active) Inference, Parametric Learning, and Structure Learning. Specifically, using empirical choice behaviour, and *in silico* simulations of abstract thinking, we evaluate the evidence for different belief updating mechanisms in subjects who did, and did not, discover the rule.

5.1.1 What is abstract thinking and rule learning?

Abstract rule learning is generally defined as a process of deriving and understanding general patterns or principles that are applicable across various instances or contexts (Kayser and D'Esposito 2012). This cognitive skill allows individuals and agents to flexibly and rapidly

make sense of — and respond to — new situations. The concept of abstract thinking is often referred to in the context of cognitive science, artificial intelligence, and machine learning (Lake, Salakhutdinov et al. 2015, Mitchell 2021, Combs, Lu et al. 2023). In cognitive science, concepts, analogy, and abstractions, are foundational areas of inquiry. Research on concepts has explored various theories; for instance, concepts as perceptual simulations based on accumulating evidence (Barsalou 2009), or concepts as probabilistic predictive representations (Goodman, Tenenbaum et al. 2014), suggesting that concepts (and abstractions) are essentially models of the world, employed in perception (and action) to generate predictions and counterfactuals.

5.1.2 Generative models – abstractions, representations, and concepts in the AIF

In the Active Inference Framework, (generative) models refer to abstractions, representations, and concepts in an interchangeable manner. The reason for their interchangeability inherits from an important aspect of the framework: the idea here is that these terms all refer to particular (beliefs about) conditional interdependencies, meaning that they encode probability distributions over various contingencies relevant to the model at hand (Smith, Schwartenbeck et al. 2020, Neacsu, Mirza et al. 2022). This is pertinent to the current analysis because it establishes a consistent framework upon which the three levels of belief updating or computational processing (i.e., inference, learning, and selection) unfold.

5.1.3 General hypothesis and structure of remaining sections

Abstract thinking and reasoning require a process over and above the associative (i.e., Hebbian) learning (of the parameters of any given model with a particular structure). One proposed mechanism for abstract reasoning (here, abstract rule learning) is that of Structure Learning.

The general hypothesis — explored in this chapter — is that abstract rule learning in human subjects involves the (Bayesian model) selection of a generative model whose structure is apt to explain sensory evidence: a.k.a., Structure Learning (SL). Specific to the task that will be considered, the hypothesis is that the behaviour of participants who discovered the rules will be better explained by a model that includes SL — in addition to Active Inference (AI) and Parametric Learning (PL), and the behaviour of participants who did not discover the rules will be better explained by a model without SL; i.e., (Active) inference and/or (Parametric) learning in the absence of Structure Learning via model selection.

The remainder of this chapter is structured as follows: the first part (section 5.2) describes the behavioural task used in the empirical work. This task is an abstract-rule learning task, where subjects have to discover the set of rules that generate observed patterns of (visual) cues. The next part (section 5.3) describes the computational (generative) model — i.e., the Markov decision process MDP — used to fit behavioural data under the Active Inference Framework, and subsequently simulate behaviour *in silico*. The following two sections (sections 5.4 and 5.5) present behavioural and fitting results from Experiment 1; in section 5.4, the experimental paradigm is validated, as being fit for purpose in identifying the type of sudden learning that accompanies SL (section 5.5). Sections 5.6 and 5.7 present behavioural and fitting results from Experiment 2, which incorporates an additional feature to the behavioural task (a ’70-30’ feature). The following section (section 5.8) presents results from simulated behaviour with numerical experiments (i.e., synthetic agents), using the MDP model described in section 5.3. Crucially, the ensuing behaviour mimics that observed in human subjects, and considers further inquiry into Structure Learning. The final section (section 5.9) discusses the implications and limitations of the results and suggests future directions of research.

5.2 Behavioural task

This section introduces the general task structure that was used for both behavioural experiments. The differences between Experiment 1 and Experiment 2 will be introduced in the relevant sections below. The task can be thought of as an abstract rule-learning task, with rules that are almost impossible to discover via associative learning. Skilled performance in the task rests on reasoning, with a drastic increase in performance after rules are discovered, a phenomenon usually investigated in the insight literature (Kounios, Frymiare et al. 2006).

Subjects completed the experiment via the online platform provided by Prolific (www.prolific.com). A device restriction was applied, such that subjects had to use a Desktop while carrying out the study (i.e., subjects could not use phones or tablets to perform the task).

Before starting the experiment, subjects were informed that they would be presented with visual and auditory stimuli; that they will participate in a reasoning task; and that they will be asked some simple questions about themselves and their experience of the task. They then gave informed consent, and after entering their Prolific ID, they were presented with two pages of instructions. The first page informed participants that they will be presented with a combination of three images for each trial (either face, tool, or house), that there are three hidden rules that will help them figure out the correct image on each trial, and that the top-centre image cues the rule. The second page told participants that the images will be masked, that they had to 'click-to-reveal' to unmask the images. And that they had a maximum number of reveals per trial, but that they could select an answer earlier if they felt confident. Subjects then engaged in 20 practice trials. After the practice trials, reminders for both instruction pages were presented, followed by carrying out the task.

The number of blocks (and therefore trials) differed between Experiment 1 and Experiment 2. However, the trial structure was the same across experiments. Each trial entailed

a randomly generated arrangement of three distinct images: ‘face’, ‘tool’, and ‘house’. These were generated as follows. Firstly, a list of centre images was created, with an equal distribution over the three images. Then, based on the rule associated with the centre image, an allowable image was selected and placed at the location indicated by the rule (i.e., centre image); finally, an image that was not the target or centre image was placed at the remaining location. The final step — in generating the trial stimuli — was to shuffle the trials using a random number generator. There was an exception to this process when the central image is ‘house’. In this case, the left and right images were generated randomly, such that they did not correspond to each other or ‘house’. Clicking on the images revealed the content for 1500 ms, after which the images became masked again. A smaller section at the bottom of the screen displayed the word ‘Answers’ and three images from which the subjects could choose (face, tool, or house).

In Experiment 1, there were three blocks, each with its own sets of rules that will be described below. Experiment 2, however, contains only 1 block, the equivalent of Block 1 in Experiment 1. This means that both experiments entailed Block 1, with the following rules:

- If there is a tool in the centre, the correct answer (i.e., target image) will be in the left location.
- If there is a house in the centre, the correct answer (i.e., target image) will be in the centre location (i.e., house)
- If there is a face in the centre, the correct answer (i.e., target image) will be in the right location.

Please see Figure 5.1 for an example image arrangement (Figure 5.1a), and how it appeared to participants at the beginning of each trial (Figure 5.1b). Figure 5.1 also displays an example of how subjects would have proceeded through trials (Figure 5.1 c-f).

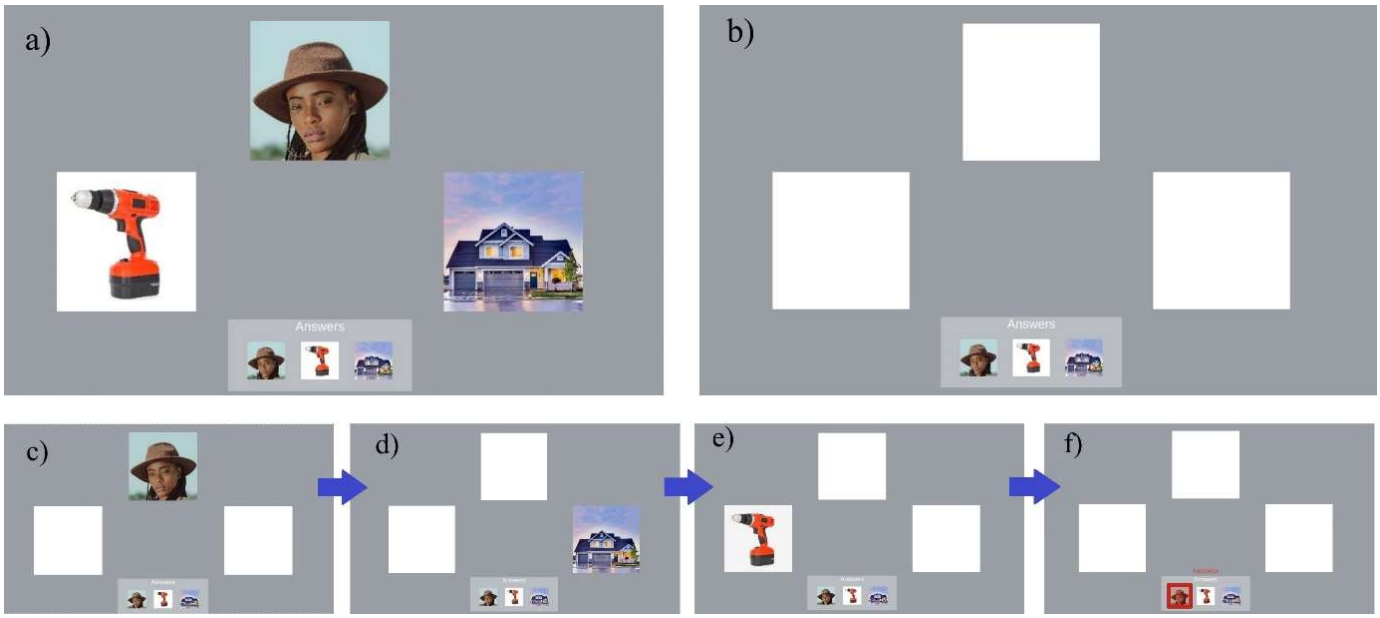


Figure 5.1 Example image arrangement and progression through trials. a) Underlying images for this specific trial. b) When subjects start each trial, the three images are masked. c)-f) An example progression through a single trial. Subject clicks on centre image and observes a face, followed by selecting the right image (revealing a house), and the left image (revealing a tool). Subject then selects 'face' and receives incorrect feedback (as the correct answer should have been house, which is the image at the right location).

An example progression through a single trial (from Figure 5.1 c-f) can be described as follows: The participant clicks on the top-centre image and observes a 'face' stimulus for 1500 ms, followed by a click on the right image (revealing a 'house'), also observed for 1500 ms; the subject then clicks on the left image (revealing the 'tool'). The subject then selects an image (here, face) from the box of Answers based on what they think the correct answer would be. If we imagine this trial as being part of Block 1, the subject will receive 'incorrect' feedback indicated by a red mask, coupled with a low-pitch tone. If, however, the subject selected 'house', they would have received 'correct' feedback, indicated by a green mask, coupled with a high-pitch tone.

The following metrics were collected during the main experiment, for each trial: the number, timestamps, and identity of the reveals; and the response (0 if correct, 1 if incorrect) and timestamp of the response.

The task was implemented using the Unity engine: version 2021.3.27 (https://unity.com/). The three visual stimuli (seen in Figure 5.1) were obtained from a public domain database called Pixabay (https://pixabay.com/) by using search terms for 'house', 'face', and 'tool'. The selected images contained no background clutter and were object-focused, to remove potential ambiguity surrounding image categories. During the task, images were displayed as squares of equal size, however, the size varied according to participants' screen size, such that each image had a ratio of image to screen of 6 to 25. Auditory stimuli for 'correct' and 'incorrect' feedback were obtained from the same website (https://pixabay.com/), using search terms for 'correct' and 'incorrect' sound effects.

5.3 The generative model

This section describes the generative model used for both fitting the behavioural data (sections 5.5 and 5.7 below) and numerical experiments (i.e., computational simulations, section 5.8 below). As noted previously, a generative model is a joint probability distribution over observed outcomes, hidden causes, and policies. The generative model used in the following simulations is a discrete-state space model, also referred to as partially observable Markov decision process (POMDP).

This model is parametrized by a set of high-dimensional arrays: the likelihood array encoding probabilistic mappings between states and outcomes (i.e., the likelihood of an

outcome given hidden states) (**A**), the transition array encoding transition probabilities among hidden states (**B**), prior preferences over outcomes (i.e., prior beliefs about future outcomes) (**C**), and priors over initial states (**D**). Here, the likelihood array is itself parametrized as a Dirichlet distribution, whose sufficient statistics are concentration parameters *a* (prior concentration parameters), and **a** (posterior concentration parameters) that accumulate with experience. Using a default parametrization for learning rate (eta = 1), this entails adding a count (i.e., a concentration parameter) to the appropriate element of the mapping, given a particular combination of a given hidden state and outcome. In other words, these concentration parameters can be interpreted as counting the number of times a specific combination (of outcomes and states) has been observed.

The paradigm involves four hidden state factors: *rule* (f1, with three levels: left, centre, right), *correct image* (f2, with three levels: tool, house, face), *location of sampling* (f3, with four levels: left, centre, right, null), and *decision* (f4, with four levels: tool, house, face, null). The purpose of the 'null' level in *location of sampling* is purely to allow for initializing the simulated trial, or to indicate the end of foraging within a trial. In the case of *decision*, the purpose of 'null' is to indicate the trials which have a decision or not – that is, any given trial would indicate 'null' at every step before and after the decision, i.e., not currently declaring a choice.

There are three outcome modalities (generated by specific combinations of these four hidden states): *what* ($\mathbf{A}^1$, with four levels: tool, house, face, null), *where* ($\mathbf{A}^2$, with four levels: left, centre, right, null), and *feedback* ($\mathbf{A}^3$, with three levels: null, correct, incorrect). Please see Figure 5.2 for a graphical depiction of the generative model.

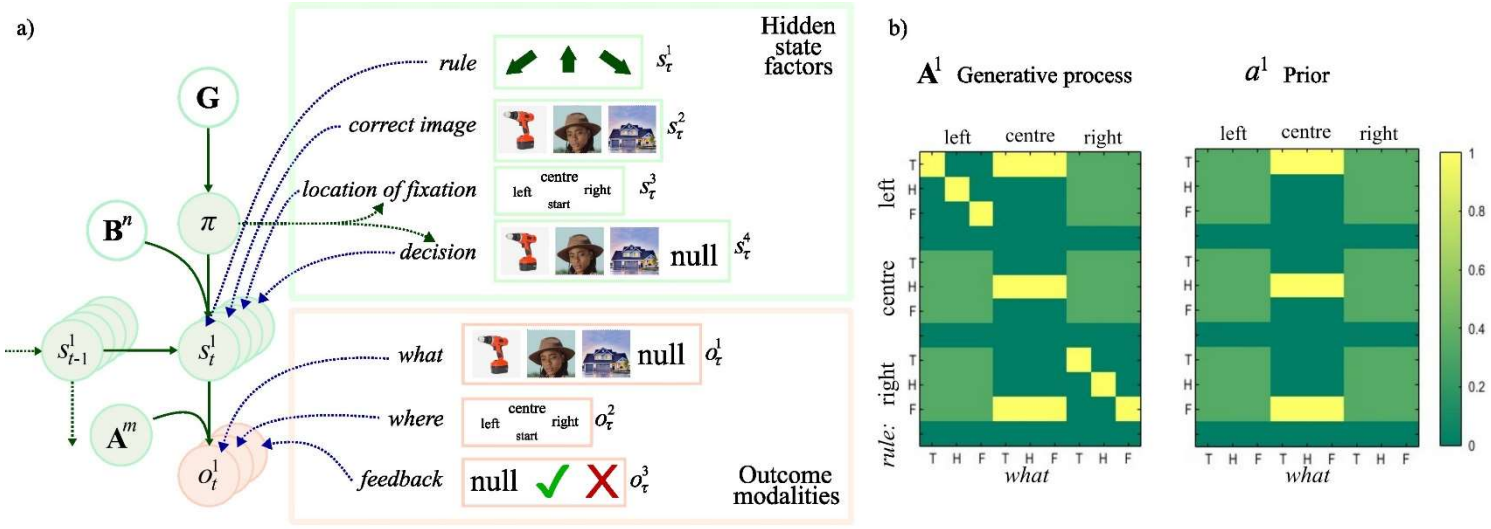


Figure 5.2 Graphical representation of the generative model. a) Left panel - the Bayesian dependency graph showing conditional dependencies. Variables in unshaded circles depict hyperpriors, and green circles indicate random variables. Outcomes (o) are generated from hidden states (s) that unfold according to probabilistic transitions (**B**), which themselves depend on policies (π). Policies are selected based on their expected free energy (**G**). Right panel - the particular hidden state factors and outcome modalities used to model abstract rule learning. The generative model has four hidden state factors: *rule, correct image, location of sampling,* and *decision*; and three outcome modalities: *what, where,* and *feedback.* b) A visual depiction of the likelihood mappings between hidden states and outcomes for the generative process ($\mathbf{A}^1$) and prior (a^1). This mapping is a five-dimensional array showing the contingencies for the modality of interest: *what* (with three levels, tool, house, and face). Each 4x3 matrix shows the mapping from the correct image to the outcomes under each *rule* (vertically) and each *sampling location* (horizontally). In other words, slices of the likelihood array have been placed side-by-side for visual inspection; the prior does not have the precise mapping of hidden states to outcomes featured by the generative process: e.g., the identity mapping from the *left* sampling location to the target image when, and only when, the rule is *left*. In this *rule* context, a tool would be seen at the centre location, in accordance with the Block 1 rule in the main text.

For example, the *feedback* modality in the case of Block 1 rules (where the rules are tool -> look left, face -> look right, house -> look centre) can be interpreted as following:

- Where *rule* is 'look centre' and *decision* is 'house', then 'correct' feedback.
- Where *rule* is 'look centre' and *decision* is not 'house', then 'incorrect' feedback.
- Where *rule* is 'look left' and *decision* = *correct image*, then 'correct' feedback.
- Where *rule* is 'look left' and *decision* ≠ *correct image*, then 'incorrect' feedback.

- Where *rule* is 'look right' and *decision* = *correct image*, then 'correct' feedback.
- Where *rule* is 'look right' and *decision* ≠ *correct image*, then 'incorrect' feedback.
- Otherwise, 'null' feedback

Overall, the type of information contained in the likelihood array (**A**) can be summarized as follows: there was 'knowledge' that there are three rules, three possible correct images, agents would 'know' where they were looking, and would be 'aware' of what image they choose (tool, house, face), and whether it is correct. Furthermore, agents would have 'knowledge' that there were three locations they could examine, and that the image at the top-centre location cued the specific rule that would enable a correct response. This description does not indicate what agents 'knew' at the beginning of the simulations or the fitting procedure, rather, this was the process generating agents' observable outcomes (i.e., the generative process). It should be noted here that 'knowledge' of these contingencies does not involve an explicit declarative type of knowledge. Rather, 'knowledge' of the contingencies is encoded in (Dirichlet) concentration parameters that can be associated with synaptic connectivity (Friston, FitzGerald et al. 2016, Da Costa, Parr et al. 2020, Da Costa, Friston et al. 2021, Neacsu, Convertino et al. 2022, Neacsu, Mirza et al. 2022).

The generative *process* can be thought of as a POMDP equipped with **A**, **B**, **C**, etc. arrays. A generative *model* can be specified in terms of concentration parameters (i.e., *a* **a**, *b* **b**, *c* **c**, etc. arrays) that may or may not recapitulate the generative process. In our example, prior knowledge supplied to the subjects recapitulated the structure of parts of the likelihood (**A**) mapping but did not include the causal structure corresponding to the rule. This meant that the generative model could, in principle, learn the rules based on outcomes supplied by the generative process. More specifically, the generative model is identical to the generative process, except for having concentration parameters that do not encode certain contingencies of the generative process, i.e., the contingencies constitute a rule. That is, the generative model

includes the contingencies to be learnt (here, the likelihood, mapping from latent states to outcomes). In other words, in the generative process, contingencies are fully known, whereas the generative model only contains priors about these contingencies with a given degree of ignorance. Here, *a*, *b*, *c*, etc., refer to prior concentration parameters (i.e., Dirichlet counts) and **a**, **b**, **c**, etc., indicate posterior concentration parameters, before and after experience dependent learning, respectively. Given this interpretation, we will see below that the simulations and the fitting procedure (i.e., fitting behavioural data to the MDP) were initialized with a specific configuration of prior concentration parameters over the likelihood array *a*. Briefly speaking, the priors used for simulations and fitting essentially indicate that some contingencies in the generative process are known, whereas others (here, the underlying rules) are not.

The set of policies entailed six possible actions: agents could sample left, sample centre, sample right, return to fixation and choose tool, return to fixation and choose house, or return to fixation and choose face. An example sequence of actions could be: sample centre, sample right, sample left, return to fixation and report face.

The transition array (**B**) involves one set of transitions for each factor: *rule*, *correct image*, *location of sampling*, and *decision*. The first two sets of hidden state factors are not controllable. That is, states for the *rule* or *correct image* do not change within a trial, making the arrays for each of these factors an identity matrix. Each trial starts with a new set of stimuli, and comprises a sequence of timesteps, where each timestep corresponds to the belief updating that follows each successive observation (e.g., saccade or 'reveals'). The third and fourth set of (probabilistic) transitions (i.e., sets of transitions for factors *location of sampling* and *decision*) depend on action, where each action changes the hidden state to where the agent chooses to sample, or the decision made (i.e., tool, face, or house).

$$\mathbf{B}^{1}_{ij}(u),\ \mathbf{B}^{2}_{ij}(u) = \begin{cases} 1,\ i = j,\ \forall u \\ 0,\ i \neq j,\ \forall u \end{cases}$$

$$\mathbf{B}^{3}_{ij}(u),\ \mathbf{B}^{4}_{ij}(u) = \begin{cases} 1,\ i = u,\ \forall j \\ 0,\ i \neq u,\ \forall j \end{cases}$$

For the prior preference over outcomes (**C**), there were no preferences for *what* and *where* outcomes, meaning that neither image was preferred *a priori*, and neither location was preferred a priori. Prior preferences over *feedback* differ slightly between the fitting procedure and the computational simulations. This is due to the variable number of reveals that participants can make in the behavioural experiments. The specifics for each will be described below. However, generally speaking, in the case of fitting, there was a preference against being incorrect, and a preference for receiving 'correct' *feedback*. In the case of computational simulations, the preference against being incorrect is likewise present. However, there is no preference for receiving 'correct' *feedback*, but there is a preference against receiving 'null' *feedback* at the final timestep, essentially forcing a decision at the end of the trial.

Prior beliefs about initial states (**D**), for factors *rule,* and *correct image* were uniform distributions. For *location of sampling* and *decision*, each trial started with a fixation cross, and a 'null' *decision*, meaning that a decision has not yet been made for the trial in question.

$$\mathbf{D}^{3}_{i},\ \mathbf{D}^{4}_{i} = \begin{cases} 1,\ i = 4 \\ 0,\ otherwise \end{cases}$$

This ensures that each trial began with states that indicate a fixation cross (for *location of sampling* factor) and null (for *decision* factor, meaning that a decision is yet to be made).

Having specified the structure of the generative process and model, let us now expand on the generative model by describing the priors used for the fitting and simulations. These

priors (i.e., *a*) concern the likelihood mapping (i.e., **A**), linking hidden states to outcomes. The priors used to initialize the fitting and simulations can be described as follows. Agents (human or simulated) had knowledge that feedback depended on choosing the correct image, as described by specifying the mappings for outcome modalities *where* and *feedback* as high for correct contingencies, and zero otherwise. Next, agents had knowledge that the three different rules are specified by the central image, but were not aware of how the rule determined outcomes. This is indicated by uniform Dirichlet counts in mappings between *correct image* and the image seen at each location, under all three rules (Figure 5.2b). The uniform beliefs regarding the link between *what* is being observed and the *correct image* are in contrast to the generative process, where the precise mappings for the 'left' *location of sampling* specify that the observed image maps to the correct image under the 'look left' rule. Please see Figure 5.2b for a depiction of the principal parts of the prior (*a*) for the likelihood array. This array corresponds to mappings from the *correct image* (i.e., tool, house, face), to the visual outcome (i.e., tool, house, face, null), for each *location of sampling* (i.e., left, centre, right, null) and for each *rule* (look left, look centre, look right), and indicates that *a priori*, subjects are not aware of the relationship between the *correct image* and *what* image they are observing. This contingency (i.e., between the *correct image* and *what* image is being observed, is the one being learnt during simulations. Similarly, for the fitting procedure, and encoded in the priors of the generative model, we assume that 'discovering the rules' entails learning the features of this specific contingency: i.e., the relationship between the *correct image* and *what* image is being observed, in a context (rule) sensitive fashion.

Although the underlying rules appear to be simple (e.g., if tool -> look left, if face -> look right, if house -> look centre), discovering a solution is not. This is because agents (simulated or human) do not have knowledge about the hidden states that underlie the outcomes being observed. For example, the observable outcomes depend on a two-way interaction

between the *correct image* and the *location of sampling*, but only when the *rule* is 'look right' or 'look left'. Agents have to learn these contingencies by accumulating evidence regarding the inferred states and the outcomes they observe, but they are unaware of which hidden states are responsible for generating these outcomes.

It is useful to note here one aspect of the likelihood array **A**, whereby some redundancy is contained. For example, in theory, if the agent were to fixate the central cue when the *correct image* is 'face', the agent would still observe a 'house' cue. However, this combination of hidden states is never instantiated because a posteriori, a centre 'house' cue means that the *correct image* is 'house', a (likelihood) mapping that is encoded in relation to the *decision* factor.

In summary, we assume that agents (human or simulated) start with (prior) beliefs that there are three distinct rules, three distinct locations, and three distinct image types; that the feedback depended on choosing the correct image; and that the rules depend on the central location; however, agents would have no concept of what the rules are, or what they mean. These priors equip agents with quite a bit of information about the problem structure, but not about the solution. One important aspect of this formulation is the possibility to encode prior beliefs using task instructions, and *vice versa*. That is, we can encode task instructions from empirical settings into a generative model, and likewise, we can implement presumed priors into task instructions to generate empirical predictions.

5.4 Behavioural Experiment 1

5.4.1 Method

5.4.1.1 Participants

Participants were recruited world-wide using the Prolific platform (www.prolific.com). The sample consisted of 32 healthy adults with normal or corrected vision, and no mental health conditions or impairments. Four participants were excluded due to failure to follow instructions correctly (i.e., made no reveals, or only revealed the central location). The final sample consisted of 28 subjects; there were 13 females (15 males), with ages ranging from 21 to 54, $M = 32.39$, $SD = 8.58$, participating in the task. In this sample, participants' countries of origin or residence included United Kingdom, South Africa, Nigeria, Canada, Ireland, Czech Republic, Zimbabwe, Poland, Sweden, Australia, and Spain. The study was approved by UCL Research Ethics Committee, and all subjects gave written informed consent, being reimbursed £7.50 for their participation.

5.4.1.2 Task specifics and procedure

In this experiment, the main task (i.e., the abstract rule-learning game) comprised 3 blocks, with 32 trials each, totalling 96 trials. The entire duration of the experiment, ranged from approximately 17 minutes to 53 minutes, with a mean time to completion of 31 minutes. The maximum number of reveals participants had for this experiment was 5 per trial. There was approximately a 33% probability for each of the three images at each of the three locations.

Each block had a set of three (hidden) rules. For the first block, these were:

- If there is a tool in the centre, the correct answer (i.e., target image) will be in the left location.

- If there is a house in the centre, the correct answer (i.e., target image) will be in the centre location (i.e., house)
- If there is a face in the centre, the correct answer (i.e., target image) will be in the right location.

Rules swapped for each block, such that in Block 2, the rules became:

- If there is a tool in the centre, the correct answer (i.e., target image) will be in the right location.
- If there is a face in the centre, the correct answer (i.e., target image) will be in the centre location (i.e., face)
- If there is a house in the centre, the correct answer (i.e., target image) will be in the left location.

And in Block 3, the rules were:

- If there is a tool in the centre, the correct answer (i.e., target image) will be in the right location.
- If there is a house in the centre, the correct answer (i.e., target image) will be in the centre location (i.e., house)
- If there is a face in the centre, the correct answer (i.e., target image) will be in the left location.

Participants were instructed before each block that rules may be swapped for the upcoming block. This message was displayed for 10 seconds, to provide a short temporal break between the blocks. Please see Figure 5.3 for a visual depiction of the rules for each block that underlie the correct response in the case of each centre image. Performance was judged by the number of correct responses per block, divided by total number of trials per block.

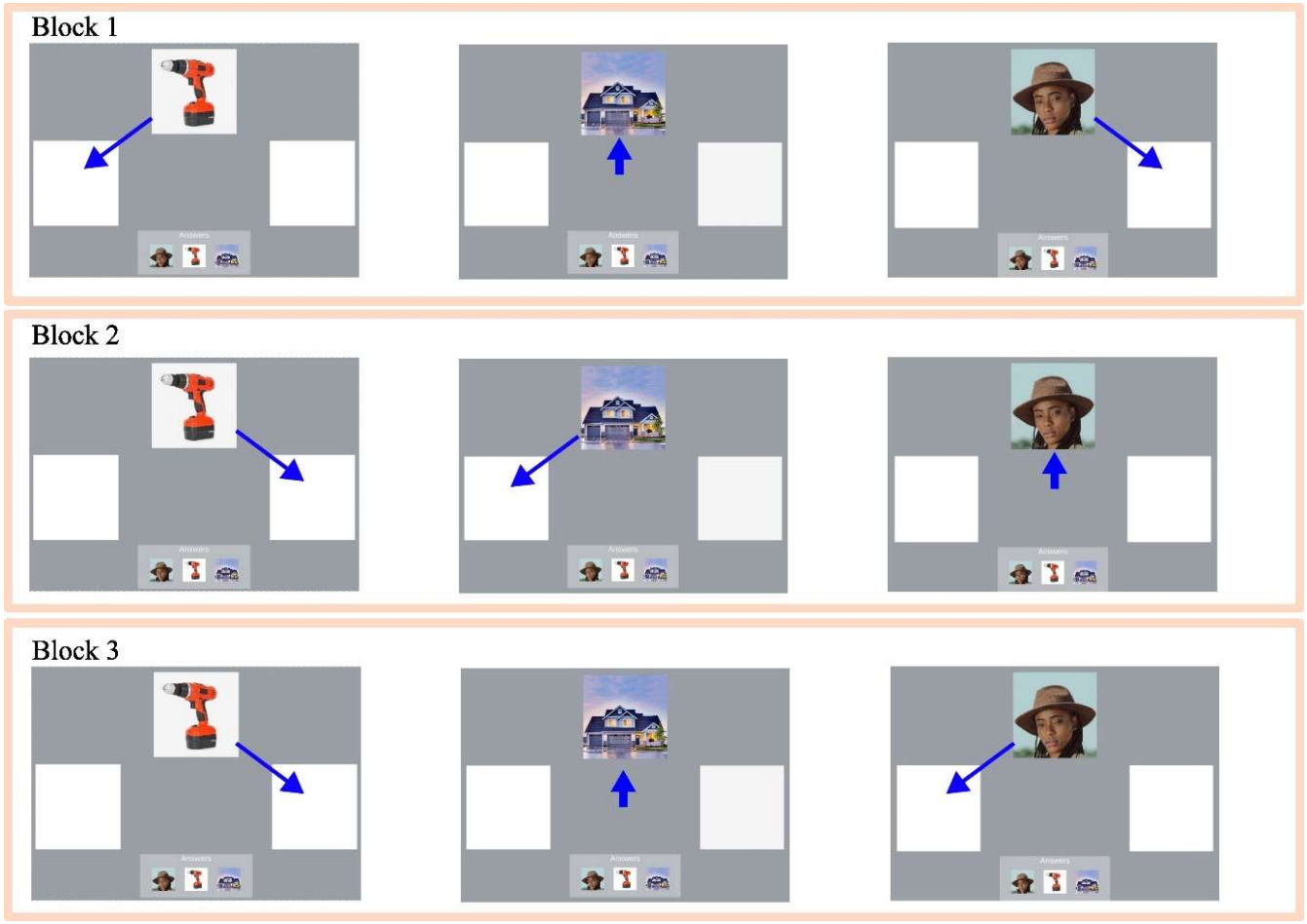


Figure 5.3 Visual schematic of the rules for each block. In Block 1, the rules can be summarized as follows: if there is a tool in the centre, the correct image will be on the left; if there is a house in the centre, the correct image will be 'house'; and if there is a face in the centre, the correct image will be on the right. For Block 2, these rules were swapped, resulting in the following set: if there is a tool in the centre, the correct image will be on the right; if there is a house in the centre, the correct image will be on the left; and if there is a face in the centre, the answer will be 'face'. For the third Block, the rules involve a correct image on the right if face is at the centre, and if house is at the centre, the correct answer will be 'house'.

5.4.1.3 Hypotheses for Experiment 1:

1. There will be rapid learning where rules are discovered, as represented by i) a significant difference in performance between subjects who discover the rules versus subjects who do not; ii) this difference in performance will be consistent across blocks;

and there will be gradual learning, as represented by iii) a significant difference between Blocks 1 and 3, regardless of whether the rules are discovered.

2. There will be a predilection for sampling novel cues, which will abate with experience, as represented by a significant decrease in the number of reveals made.

3. Similarly to hypothesis 2, there will be an overall decrease in reaction time between blocks.

5.4.2 Results

The focus of this section is to establish the behavioural paradigm's validity and illustrate a significant difference in performance between subjects who discover the rules and subjects who do not discover the rules in the context of the abstract rule-learning task described in sections 5.2 and 5.4.1. Furthermore, similarities in other performance metrics are presented, with predictions arising from the specific configuration of the task itself, as well as predictions arising from the Active Inference Framework.

To differentiate between subjects discovering the rules, and subjects who did not, we implement a threshold of maximum successive correct answers made by subjects. That is, there is a maximum number of successive correct answers that indicate whether subjects discovered the rules or not. This threshold was obtained by simulating a succession of trials with random choices. Essentially, this mimics 2000 agents playing the game for 32 trials, but with agents making choices completely at random. A histogram shows the distribution of successive correct responses (Figure 5.4). Out of 2000 agents making random choices, 779 had a maximum of 2 consecutive correct trials, and the number of agents making 3 or more consecutive correct trials decreases. There were 2 agents with 8 and 9 correct consecutive responses respectively. Following this, we define the differentiation between 'rules discovered' and 'rules not

discovered' such that (human) subjects making 10 or more consecutive correct responses were classified as 'discovered rules', whereas subjects with 9 or less consecutive correct responses were classified as 'did not discover rules'. In the empirical dataset, there were four subjects who 'did not discover rules', of which two had 9 consecutive correct responses, and two had 8 consecutive correct responses. In short, if subjects had 10 or more consecutive correct responses in any of the three blocks they were classified as 'discovered rules'.

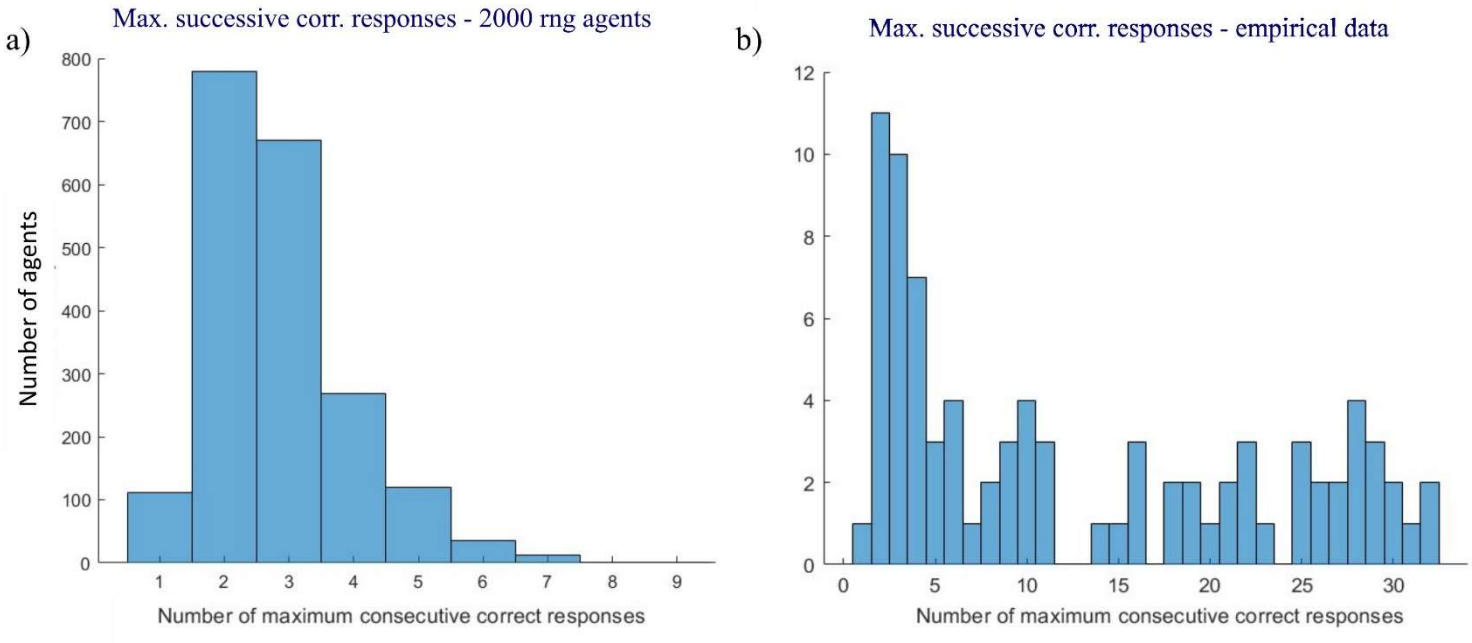


Figure 5.4 The maximum number of successive correct responses. a) Maximum number of successive correct responses for 2000 agents making random responses. Out of 2000 agents making random choices, 779 had a maximum of 2 consecutive correct trials, and the number of agents making 3 or more consecutive correct trials decreases. The maximum observed number of successive correct responses is 9. b) Maximum number of successive correct responses for the empirical data. In these illustrations, each subject provides 3 entries (1 per block).

5.4.2.1 Results from Hypothesis 1

This subsection presents analyses of performance in terms of accuracy, compared between subjects who discovered the rules, and subjects who did not. Overall, according to the criterion mentioned earlier, 16 out of 28 subjects in this dataset discovered the rules (in any of the three

blocks). Although the two groups were created based on the number of consecutive correct responses, the accuracy comparisons between the two groups and their respective results, e.g., in terms of statistical significance are nevertheless of interest, since with this paradigm it is (in theory) possible to obtain high accuracies without the need for correct consecutive responses. For example, one subject in the 'did not discover rules' group had an accuracy of 56% in Block 1. Another subject in the same group had an accuracy of 75% in Block 3 (yet without satisfying the condition of 10 or more consecutive correct responses). On the other hand, one subject in the 'discovered rules' group had an accuracy of 50% in Block 1, and another subject in this group had a 72% accuracy in Block 3 (i.e. lower than the two subjects mentioned, in the 'did not discover rules' group). Analysing performance (in terms of accuracy) is related to the criterion used, where the criterion used captures 'discovering the rules' which in turn should involve higher accuracies. In other words, the paradigm implies some form of validity of capturing the behavioural nuances of abstract rule learning, only if the criterion established also involves a significant difference in performance (e.g., higher accuracy) if the rules have been discovered.

For subjects who discovered rules, the average accuracies were M = 73.2% (SD = 18.61), M = 81.8% (SD = 15.78), and M = 87.9% (SD = 8.06) for Blocks 1, 2, and 3 respectively. The average accuracies for subjects who did not discover the rules were M = 38.5% (SD = 10.93), M = 42.2% (SD = 8.79), and M = 50.5% (SD = 13.58) for Blocks 1, 2, and 3 respectively (see Figure 5.5). Averaging across the three blocks, the accuracies are 81% for subjects who discovered the rules, vs 43.8% for subjects who did not, with a mean difference of 37.2%. In other words, discovering the rules entails that on average, subjects are 37% more accurate.

To examine performance (in terms of accuracy), a mixed ANOVA was conducted, with Block as within-subjects factor, and Group (i.e., discover vs. did not discover rules) as between-

subjects factor. There was a significant main effect of Group, F (1, 22) = 103.526, p < .001, and a significant main effect of Block, F (2,22) = 9.138, p < .001, but no significant interaction F (2, 22) = 0.316, p = .731. Simple post-hoc contrasts (for within-subject effects) reveal that mean accuracy for Block 3 was higher than for Block 1, F (2,22) = 19.820, p < .001, and Block 2, F (2,22) = 6.877, p = .014; with Bonferroni-corrected comparisons suggesting a significant difference between Blocks 3 and 1 (p < .001), and Blocks 3 and 2 (p < .05), but not Blocks 1 and 2 (p = .293). Pairwise comparisons for the main effect of Group, using a Bonferroni correction, indicate that the significant main effect reflects a significant difference between the groups for each block (p < .001).

These results suggest that the accuracy is significantly higher for subjects who discovered the rules, regardless of the Block. On the other hand, there was (gradual) learning from Block 1 to Block 3 in both groups; we can see from Figure 5.5a that there is a monotonic increase across the blocks, and the differences in accuracy between the groups remain steady, possibly suggesting two types of learning at play.

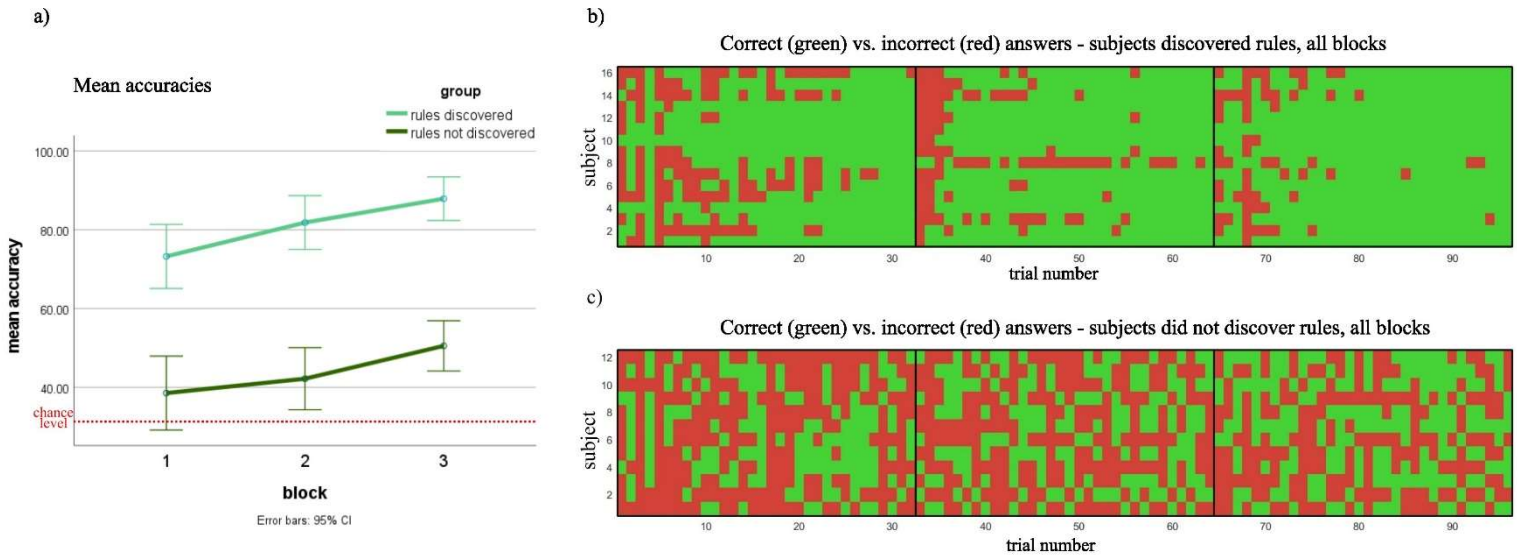


Figure 5.5 Performance comparison between subjects who discovered the rules and subjects who did not, across all 3 blocks. a) Accuracy comparison for Blocks 1, 2, and 3 respectively, showing a significantly higher performance where rules are discovered and a gradual increase in performance across blocks for both groups. b-c) Patterns of correct vs. incorrect responses between subjects who discovered the rule (b), where the number of

successive correct responses appears to be increasing, and subjects who did not (c), showing an unsystematic pattern of trial-and-error responses.

On average, performance is above chance (here, 33.3%), regardless of block number, or whether the rules were discovered or not. Furthermore, Figure 5.5(b-c) presents a visual depiction showing a matrix of correct vs incorrect responses for the two different groups across the entire experiment. There appears to be a pattern of increased successive correct responses for subjects who discovered the rules, as compared to the unsystematic pattern observed for subjects who did not discover rules. Overall, these results suggest that there is a significant difference in accuracy based on whether the rules were discovered or not, that these differences remain consistent across blocks, and that there is gradual learning as trials progress.

5.4.2.2 Results from Hypothesis 2

Next, analyses of the number of reveals are presented. It was expected that as subjects familiarize themselves with the task, there would be a decrease in the number of reveals overall. Figure 5.6 shows the mean number of reveals for each block and group. For subjects who discovered rules, the average number of reveals was $M = 2.22$ ($SD = 0.44$), $M = 1.87$ ($SD = 0.28$), and $M = 1.77$ ($SD = 0.19$) for Blocks 1, 2, and 3 respectively. The average number of reveals for subjects who did not discover the rules was $M = 1.87$ ($SD = 0.66$), $M = 1.89$ ($SD = 0.72$), and $M = 1.74$ ($SD = 0.64$) for Blocks 1, 2, and 3 respectively (see Figure 5.6).

A mixed ANOVA was carried out to test for differences between the number of reveals, with Block as within-subjects factor, and Group (i.e., discover vs. did not discover rules) as between-subjects factor. There was a significant main effect of Block, F (2, 22) = 5.477, p = .014 (using a Greenhouse-Geisser correction for sphericity), suggesting an overall decrease in the number of reveals across blocks. There was no significant main effect of Group, F (2, 22) = 0.556, p = .463, and no significant interaction F (2,22) = 2.683, p = .095. Simple post-hoc

contrasts (for within-subject effects) reveal that the mean number of reveals for Block 3 was lower than for Block 1, F (2,22) = 8.081, p = .009, and Block 2, F (2,22) = 4.807, p = .037. Interestingly, Bonferroni-corrected pairwise comparisons suggest a significant difference between Blocks 3 and 1 ($p < .01$) and Blocks 1 and 2 ($p < .05$), but not Blocks 3 and 2 (p = .633), only when rules have been discovered. For the group where rules were not discovered, pairwise comparisons retrieve no significant differences. Overall, these results suggest a decrease in the number of reveals across blocks. However, post-hoc comparisons suggest that this effect could be mainly guided by subjects who discovered the rules, although the interaction between Group and Block was non-significant.

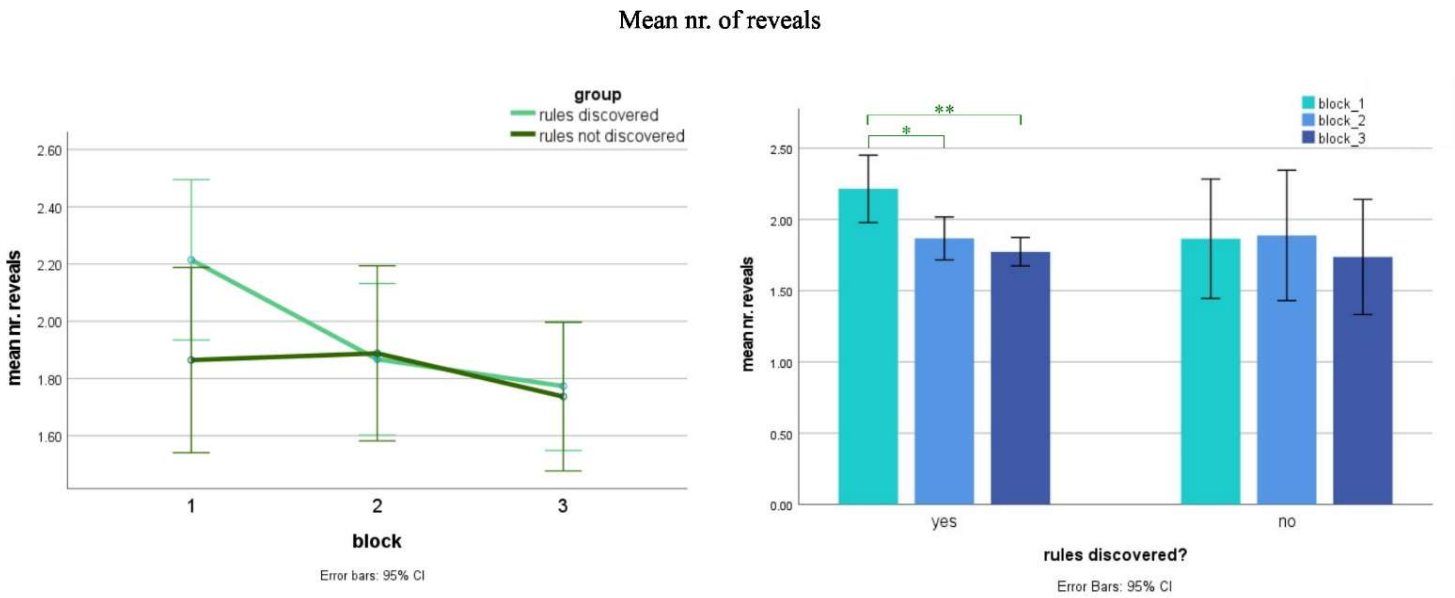


Figure 5.6 Average number of reveals across blocks and for each group. Left panel: line chart showing how the average number of reveals compares across blocks between the groups. The number of reveals decreases across blocks overall. Right panel: the average number of reveals separated by group, illustrating what could be guiding the overall decrease: subjects who discovered the rules show a decrease in the average number of reveals between Blocks 1 and 3, and Blocks 1 and 2 (with little overlap between the CIs), whereas subjects who did not discover the rules, do not show as much change (in terms of overlapping CIs).

5.4.2.3 Results from Hypothesis 3

Similarly to the results from hypothesis 2, we expected to see an overall decrease in reaction times since making fewer reveals per trial would entail a shorter trial reaction time. Figure 5.7 left panel shows mean trial reaction times for each block and group. For subjects who discovered rules, average reaction times were $M = 6.36$ ($SD = 4.25$), $M = 4.87$ ($SD = 3.49$), and $M = 4.44$ ($SD = 3.11$) for Blocks 1, 2, and 3 respectively. Average reaction times for subjects who did not discover the rules were $M = 4.25$ ($SD = 1.24$), $M = 3.99$ ($SD = 1.71$), and $M = 3.66$ ($SD = 0.82$) for Blocks 1, 2, and 3 respectively (see Figure 5.7).

A mixed ANOVA was carried out to test for differences between reaction times, with Block as within-subjects factor, and Group (i.e., discover vs. did not discover rules) as between-subjects factor. There was a significant main effect of Block, F (2, 22) = 5, $p = .024$ (using a Greenhouse-Geisser correction for sphericity), suggesting an overall decrease in reaction times across blocks. There was no significant main effect of Group, F (2,22) = 1.569, $p = .221$, and no significant interaction F (2,22) = 1.646, $p = .211$. Simple post-hoc contrasts (for within-subject effects) reveal that the mean number of reveals for Block 3 was lower than for Block 1.

Interestingly, Bonferroni-corrected pairwise comparisons suggest a significant difference between Blocks 3 and 1 ($p < .05$), but not otherwise, when rules have been discovered. For the group where rules were not discovered, pairwise comparisons retrieve no significant differences. Overall, these results suggest a decrease in reaction times across blocks, possibly guided by the decrease from Block 1 to Block 3 in subjects who discovered the rules; however, the interaction between Group and Block was non-significant.

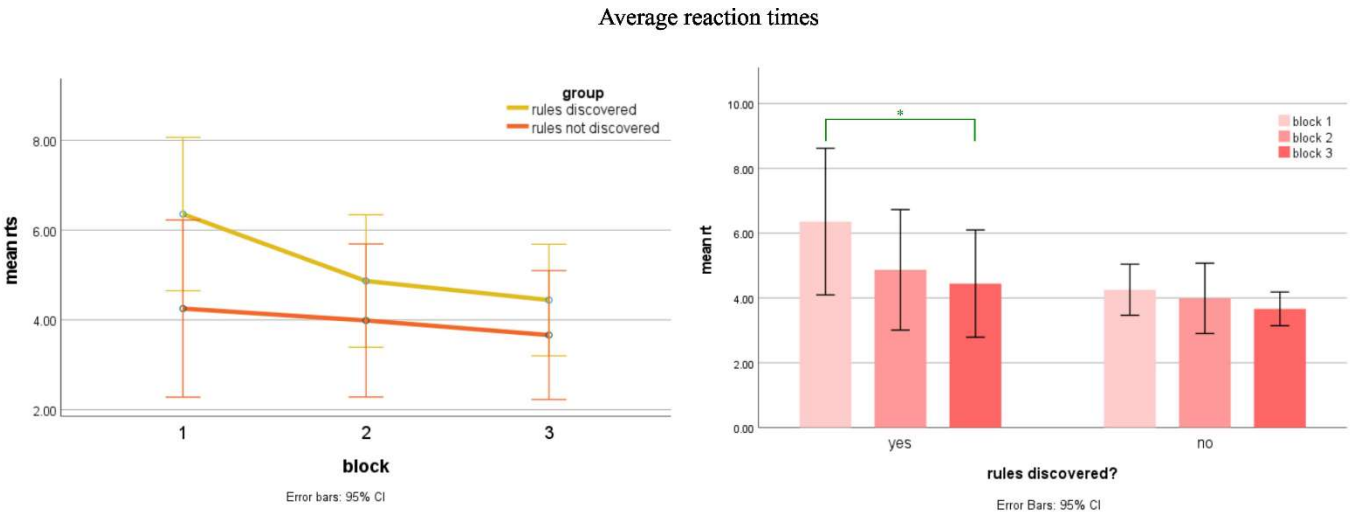


Fig 5.7 Average reaction times across blocks and for each group. Left panel: line chart showing how the average number of reveals compares across blocks between the groups. Reaction times decrease across blocks overall. Right panel: average reaction times separated by group, illustrating what could be guiding the overall decrease: subjects who discovered the rules show a decrease in the average number of reveals between Blocks 1 and 3, whereas subjects who did not discover the rules, show a smaller decrease across blocks.

In summary, the results in this section suggest that there was rapid learning where rules are discovered, indicated by a significant difference between subjects who discovered the rules as compared to subjects who did not. This difference remained consistent across the 3 different blocks. In addition to this rapid learning there was an element of gradual learning, indicated by a gradual increase in performance across the blocks, regardless of whether the rules were discovered. Furthermore, the predilection for sampling novel cues abated with experience, as shown by a decrease in number of reveals made per trial. Finally, there was an overall decrease in reaction times between blocks.

5.5 Fitting with the AIF – Experiment 1

Having now established the validity of the behavioural paradigm, we now proceed to the main hypothesis for this experiment. The hypothesis here was that SL as implemented by Bayesian Model Reduction (BMR) explains behaviour over and above Parametric Learning (i.e., associative learning) and (Active) Inference. More specifically, the behaviour of subjects who discovered the rules will be best described by a model that includes SL (i.e., BMR). And the behaviour of subjects who did not discover the rules would be best described by a model without SL (but with Parametric Learning and/or Active Inference).

The analysis for this section involves fitting participants' behavioural data under an Active Inference Framework model of choice behaviour. This was implemented using standard and modified routines (here, DEM_demo_MDP_fit_fields, spm_MDP_gen, spm_MDP_L, spm_BMS, etc.) from the (open source) SPM 12 Matlab package (available at http://www.fil.ion.ucl.ac.uk/spm/). Essentially, this procedure involves feeding trial-by-trial observations and actions made by subjects into the (PO)MDP model, and determining the probabilities of emitting each action, in relation to the actions predicted by the model. Technically, this entails calculating the marginal likelihood of empirical behaviour under different models — P (subject behaviour | model) — using Variational Bayes/Variational Laplace (Friston, Mattout et al. 2007). For a more detailed account of this procedure, please see (Stephan, Penny et al. 2009, Friston and Penny 2011, Schwartenbeck and Friston 2016, Smith, Friston et al. 2022). This allowed us to evaluate the evidence (a.k.a., marginal likelihood) for different models of the same behaviour that differed either in their structure or priors over key model parameters.

The (PO)MDP presented in section 5.3 was used for fitting the behavioural data in Experiment 1, with two additions. The first was specifying the prior preferences over outcomes

(**C**). Here, these involve a preference against being incorrect, and a preference for being correct, regardless of the timestep.

$$\mathbf{C}_\tau = \ln P(o_\tau) = \begin{cases} -4, \ o_\tau = incorrect : \forall \tau > 0 \\ 2, \ o_\tau = correct : \forall \tau > 0 \\ 0, \ otherwise \end{cases}$$

Where τ indicates the timestep or number of saccades in each trial. This is because subjects made a variable number of reveals as they progressed through the trials.

The second addition to the generative model involves specifying a prior for estimating the alpha parameter (i.e., precision in action selection, known colloquially as the 'shaky hand' parameter). The alpha parameter controls randomness in action selection under a chosen policy (higher values entail less randomness). This prior does not represent the subjects' prior beliefs, but the initial parameter values that are evaluated during fitting. The model inversion scheme assumes a Gaussian distribution over the maximum *a posteriori* estimate of the parameters (here, alpha). The prior mean for estimating alpha was set to log(1), which would be the equivalent of initialising an MDP simulation by setting alpha equal to 1. The prior variance was set to mildly informative default (1/32). In summary, the Variational Laplace scheme starts with these values (prior mean and prior variance) to evaluate the posterior that maximises the marginal likelihood of a subject's actions.

The fitting procedure was applied to models of subjective behaviour; namely, subject models where SL was included (i.e., Model 1, with SL) and when it was not. Structure Learning was instantiated with Bayesian Model Reduction (BMR) at the end of each trial. Model 1 therefore includes all three levels of processing, (Active) Inference (AI), Parametric Learning (PL) and Structure Learning (SL). BMR involves assessing the model evidence under alternative priors and selecting the model with the highest evidence. The set of hypotheses involved 20 alternative hypotheses or models; where each model corresponded to a particular

prior over the (Dirichlet concentration) parameters of the likelihood mapping (*a'*) that encodes prior beliefs about rules and contingencies. These hypotheses were chosen for their simplicity, to illustrate the phenomenon at hand under this task. Generally, BMR considers generic mappings between states and outcomes, for example, by adding or removing just one element in the likelihood mapping, such as in (Friston, Lin et al. 2017, Neacsu, Mirza et al. 2022). One can think of this process as that of increasing a particular connection strength (for example by 8), and assessing the model evidence, followed by increasing another connection strength, and again assessing the model evidence, and continuing until the available set of alternative hypotheses has been assessed. In this example, increasing a mapping by 8 would be the equivalent of having observed that particular combination of states and outcomes eight times. Please see Table 5.1 for a brief description of the alternative hypotheses used for the model where SL (i.e., BMR) is enabled, and Figure 5.8 for example visual illustrations of hypotheses 3, 6, 7, 14, and 20 (c.f. Figure 5.2b).

Table 5.1 The set of alternative hypotheses used for Bayesian Model Reduction/Bayesian Model Selection

Hypoth. nr.	Brief description
1	Generative process, knowledge only for *decision* = face
2	Generative process, knowledge only for *decision* = null
3	Generative process, information only for *rule* = look left
4	Generative process, information only for *rule* = look centre
5	Generative process
6	Random array, elements add to 1 column-wise
7	Generative process, flipped vertically
8	Generative process, flipped horizontally
9	Prior
10	Generative process, information only for *rule* = look right
11	Prior, flipped horizontally
12	Prior, flipped vertically
13	Generative process, information only for *correct image* = face
14	Generative process, information only for *correct image* = house
15	Generative process, information only for *correct image* = tool
16	Prior + Generative process (*location* = left)
17	Prior + Generative process (*location* = right)
18	Prior + Generative process (*location* = centre)
19	Generative process, *rule* = look left swapped with *rule* = look centre
20	Generative process, *rule* = look right swapped with *rule* = look centre

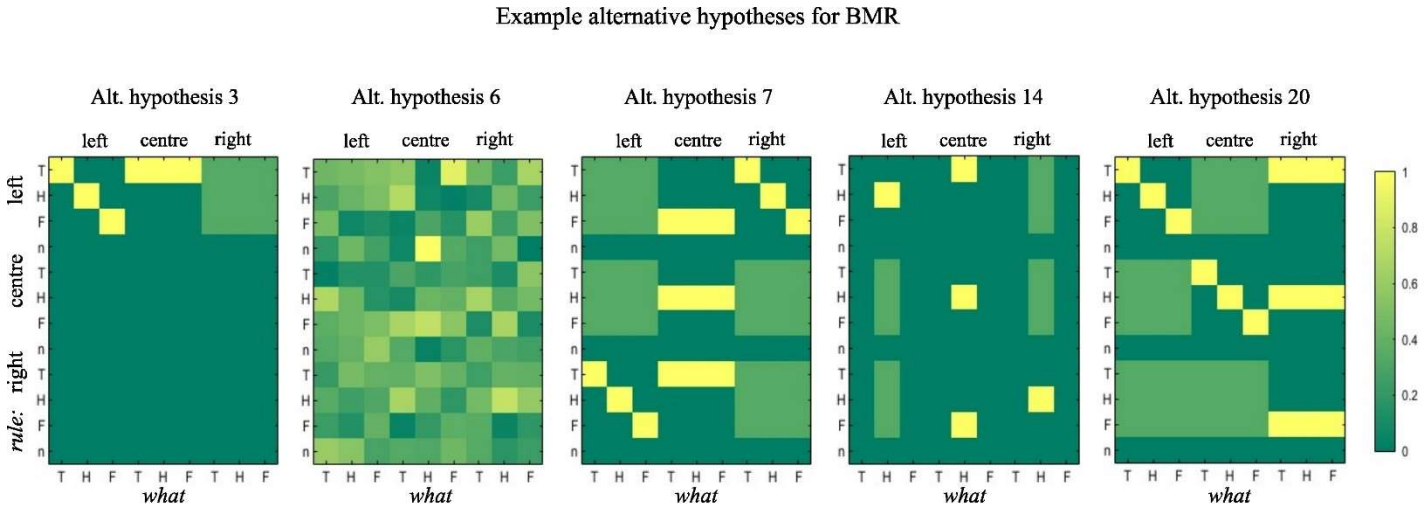


Figure 5.8 Visual illustrations of example alternative hypotheses. From left to right: alternative hypotheses 3, 6, 7, 14, and 20. Hypothesis 3 implies a partial version of the generative process (i.e., where contingencies are only known for *rule* = look left, and uniform otherwise. Alternative hypothesis 6 involves random contingencies. In alternative hypotheses 7 and 20, the generative process has been swapped, whereas alternative hypothesis 14 contains only partial information about contingencies, for *correct image* = house, which can be interpreted as discovering only one rule (out of three). C.f. Figure 5.2b, where the prior and generative process are presented.

Model 2 (without SL) involved carrying out the same fitting procedure, but without BMR. However, Parametric Learning was still present — with the accumulation of Dirichlet concentration parameters from trial to trial. Model 2 therefore included two levels of belief updating: Active Inference (AI) and Parametric Learning (PL). This subject model assumes that participants learn from trial to trial, as they accumulate evidence about the contingencies. Model 3 (just (Active) Inference) involved no update from trial to trial. This model assumes that subjects use the same generative model throughout the duration of the task, without accumulating any additional evidence for the observed contingencies. In summary, Model 1 involves a combination of all three levels of processing (AI, PL, and SL), Model 2 involves AI and PL, and Model 3 involves just Active Inference (AI).

In preparation for analysing the behavioural data, we first established the (PO)MDP's face validity. This was done by estimating alpha for the first participant and using this value (i.e., 0.9795) to simulate 32 trials. Using this simulated data, we repeated the fitting procedure, to ensure that similar values for alpha were recovered. The results of an exemplar analysis are illustrated in Figure 5.9. It can be seen that the scheme was able to recover the parameter value used to generate the data, with a reasonable degree of confidence. In this figure, the grey bars represent posterior expectations in log-space, and the pink bars represent 90% Bayesian CIs.

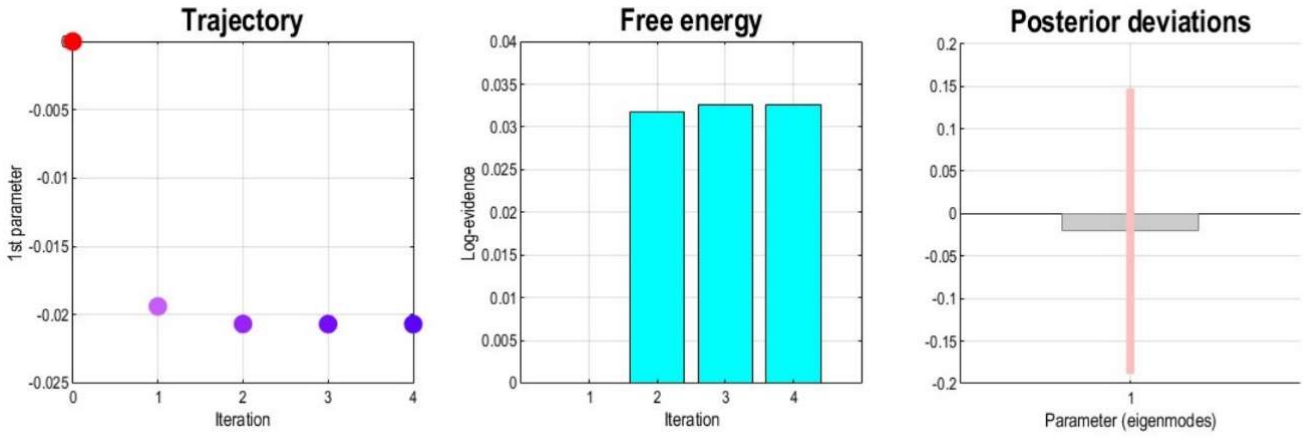


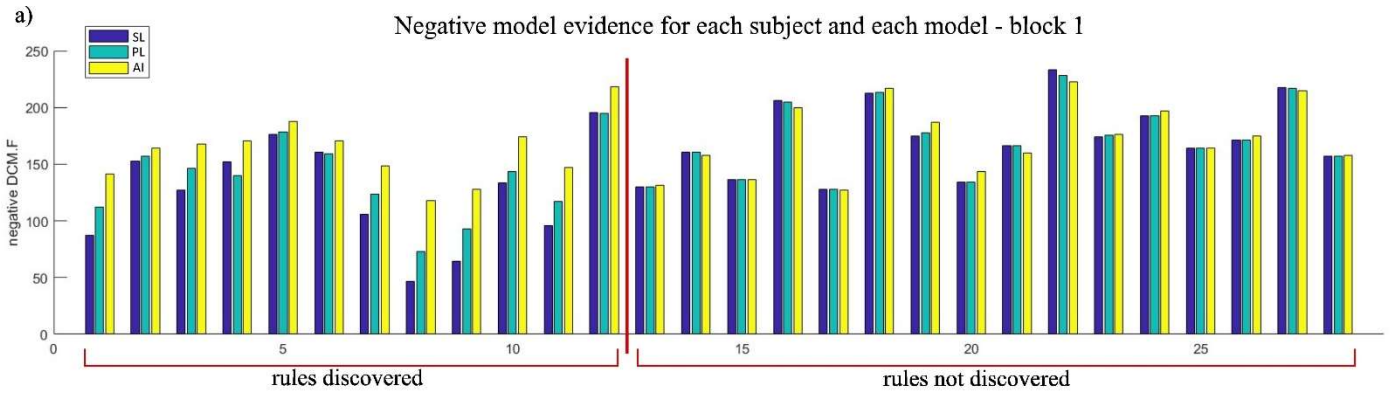
Figure 5.9 Parameter simulation and re-estimation. The simulation of 32 trials and subsequent re-fitting shows that the estimates for alpha can be reliably recovered. The parameter values are in log-space. The left panel shows how the estimated value of alpha changes with each iteration when a model is fitted to simulated data (red to blue). The middle panel shows how the free energy changes with each iteration during model inversion. The right panel shows the mean alpha value in log-space (grey bar) and the variance (pink bars) of the probability distribution over alpha (a.k.a. 90% Bayesian CIs). These results suggest that fitting can recover plausible values for alpha. Because this parameter is evaluated in log space, the posterior expectations and credible intervals can be interpreted (roughly) in terms of percent change. For example, the credible intervals lie between 15% and -18% of the value used to generate choice behaviour.

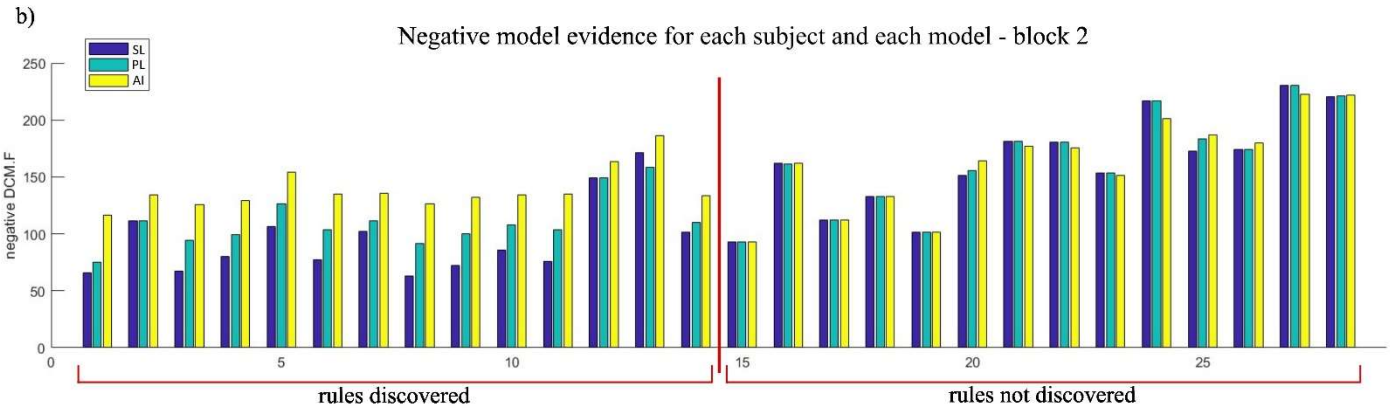
In summary, each subjects' data was fitted separately with Model 1 (SL, PL, AI), Model 2 (PL, AI) and Model 3 (AI only). The fitting scheme involves evaluating the log-probability of a

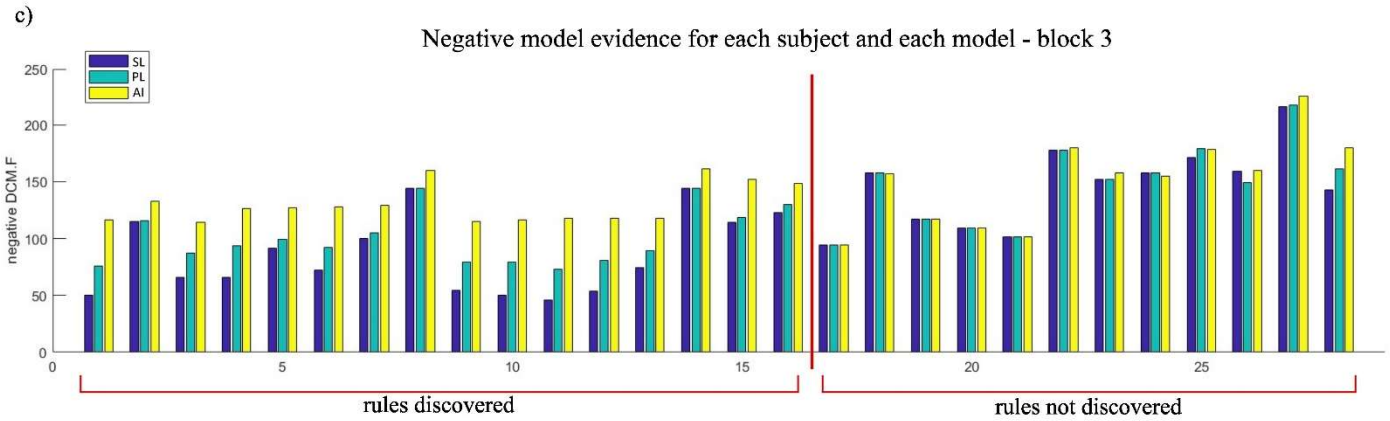
subject's actions, starting with the estimation priors on alpha, and proceeding by gradient descent in the direction of increasing likelihood, until convergence. This procedure accumulates evidence (i.e., log-likelihood - the sum of log-probabilities of selected actions under the model). After convergence, the scheme outputs DCM.F, the field of relevance for this analysis. This field, DCM.F, represents the final free energy value of the best model fit.

The (negative) free energy values for each subject are displayed in Figure 5.10(a-c), separately for each block. These values (i.e., estimates of log marginal likelihood or model evidence) effectively represent the likelihood of subjects behaving the way they did during the task, given that they used SL or not. Negative free energy is reported for ease of interpretation: The closer these values are to 0, the better the model evidence. Furthermore, since the data from each subject were conditionally independent, this means that the evidence for the three models is conditionally independent, allowing us to simply add each log-evidence (and their differences) from each subject to assess the overall evidence. Free energies were pooled (i.e., summed) over subjects for each block under each condition (i.e., discovered vs did not discover the rules) and reported in Figure 5.10d.

For all blocks, the likelihood of behavioural responses for subjects who discovered the rules was greater under Model 1 (i.e., the model that included SL) – Figure 5.10(a-c), left panels. However, this evidence varies for subjects who did not discover the rules – Figure 5.10(a-c), right panels. In Block 1, for example, behavioural responses appear to be best explained by Model 2 (i.e., PL), whereas in Block 2, the responses appear to be best explained by Model 3 (i.e., AI only). In Block 3, the model evidence appears to be highest for Model 1 (i.e., SL).


a)
Negative model evidence for each subject and each model - block 1
SL
PL
AI
negative DCM.F
rules discovered
rules not discovered


b)
Negative model evidence for each subject and each model - block 2
SL
PL
AI
negative DCM.F
rules discovered
rules not discovered


c)
Negative model evidence for each subject and each model - block 3
SL
PL
AI
negative DCM.F
rules discovered
rules not discovered

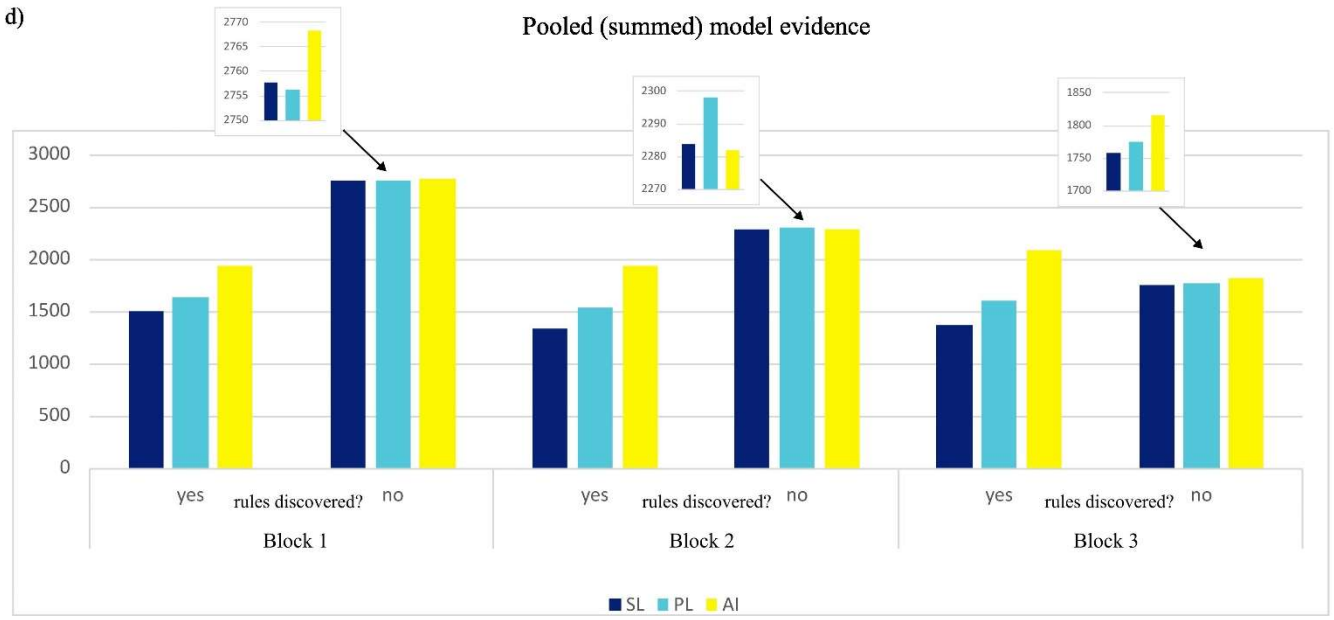
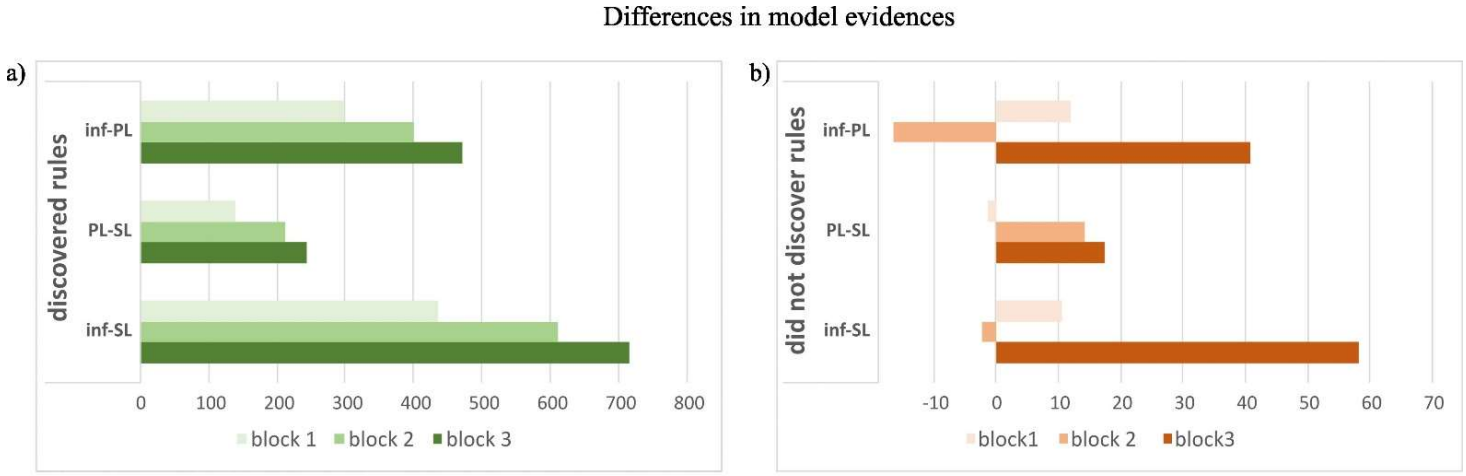

d)
Pooled (summed) model evidence
yes
rules discovered?
no
Block 1
Block 2
Block 3
SL
PL
AI

Figure 5.10 Evidence for Structure Learning. a-c) Negative free energy values (i.e., negative log-evidence) for each subject under each of the three models, for both conditions (left – rules discovered, right – rules not discovered), for each block. d) Log-evidence pooled over subjects under each model (Model1 – SL; Model2 – PL; Model3 – just-AI) and condition (discovered vs did not discover rules), for each block: Block 1 (left), Block 2 (centre), Block 3 (right). For subjects who discovered the rules, Model 1 had the highest log-evidence in each block. For subjects who did not discover the rules, the log-evidence varies across blocks (small insert boxes show a zoomed-in comparison of pooled model evidence for subjects who did not discover the rules).

Differences in model evidence for each block and each condition (i.e., discovered vs undiscovered rules) are shown in Figure 5.11. These results are pooled over participants. Here, we can see for example, that for Block 3, rules discovered (Figure 5.11 a, c), the model that incorporates SL (i.e., Model 1) had substantially more evidence. Model 1 (i.e., SL), scored approximately 715 more log-evidence as compared to Model 3 (i.e., AI only). A difference in log evidence of about 3 = log(20) corresponds to an evidence or odds ratio (a.k.a., Bayes factor) of about 20:1.

Since the differences in log-evidence are log Bayes factors, for subjects who discovered the rules the evidence for Model 1 (i.e., SL) is considered *decisive* according to Kass and Raftery (Kass and Raftery 1995). Furthermore, for subjects who discovered the rules, the differences between Model 1 (i.e., SL) and the other models increases as blocks progress (Figure 5.11a). For subjects who did not discover rules, in Block 1 we see *strong* evidence for Model 1 compared to Model 3, but the evidence is slightly *stronger* for Model 2 (compared to Model 3). In Block 3 for example, the evidence is *strong* for Model 1 (SL) compared to Model 3 (AI-only), but also *strong* for Model 2 (PL) compared to Model 3 (AI only).



c)

	discovered rules			did not discover rules		
	inf-SL	PL-SL	inf-PL	inf-SL	PL-SL	inf-PL
block 1	436.2186	138.8615	297.3571	10.5772	-1.3611	11.9383
block 2	612.3961	212.5802	399.8159	-2.1646	14.2412	-16.4058
block 3	715.1966	244.65	470.5466	58.1186	17.3509	40.7677

Figure 5.11 Differences in log-evidence. a) Differences in log-evidence between the three models for subjects who discovered the rules. The results suggest that Model 1 had substantially more evidence than Models 2 and 3. b) Differences in log-evidence between the three models for subjects who did not discover the rules. Results for this condition are inconclusive. For example, in Block 1, there is strong evidence for Model 1 (compared to Model 3), but the log-evidence is slightly stronger for Model 2 as compared to Model 3. c) Table showing exact figures for the results from a) and b).

For completeness, we computed the expected exceedance probability (at group level) using the spm_BMS function (available in SPM12). This function returns the protected exceedance probability (*PXP*). In essence, this quantifies the probability that any one model is more frequent than the others, above and beyond chance (Stephan, Penny et al. 2009, Rigoux, Stephan et al. 2014).

As expected, the *PXPs* presented in Figure 5.12 show similar trends to Figure 5.10 d-f). Here, for subjects who discovered the rules, Model 1 (i.e., the model that includes SL) has a probability of almost 1 for all three blocks. On the other hand, for subjects who did not discover the rules, *PXPs* are inconsistent, suggesting inconclusive evidence for a best model.

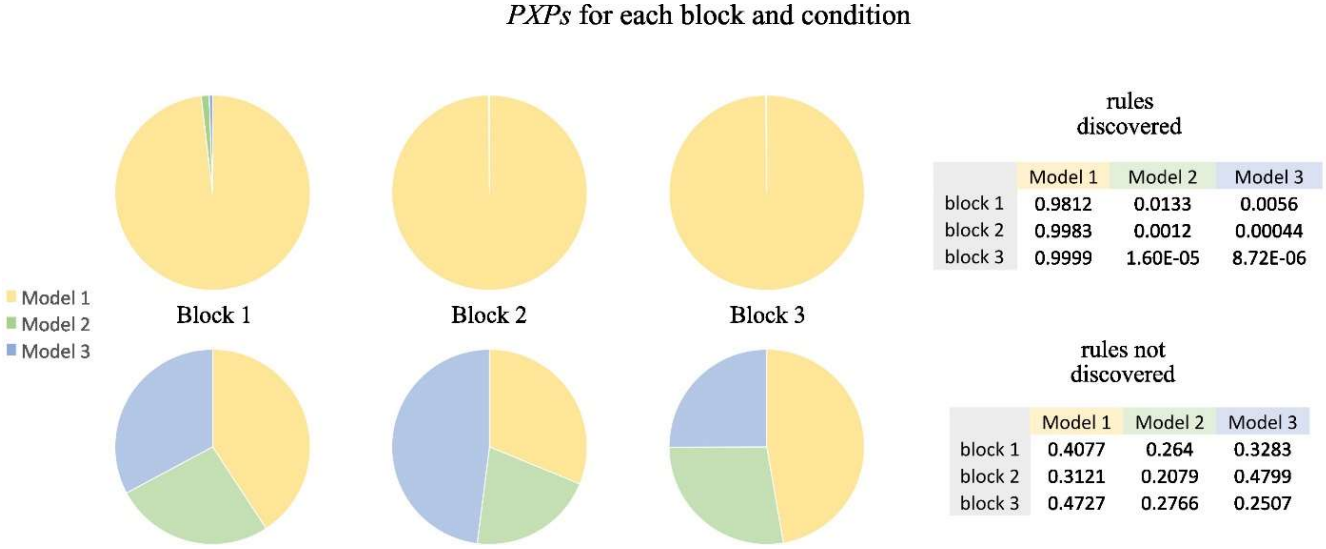


Figure 5.12 Protected exceedance probabilities. This figure shows a pie chart of the probability of any one model given other models using the *PXP* values obtained with the spm_BMS function. For subjects who discovered the rules (top half of the image), Model 1 has a *PXP* of almost 1, across all blocks. For subjects who did not discover the rules, the results are inconclusive. In Blocks 1 and 3, Model 1 is marginally most likely; in Block 2, Model 3 is marginally most likely.

Interestingly, however, when model evidence is added together for Blocks 1, 2, and 3 for each participant, *PXP* results weakly favour Model 3 (i.e., AI only) if rules have not been discovered. Results for subjects who discovered the rules remain consistent with previous analyses. These results are shown in figure 5.13 below.

Taking these results overall, for Experiment 1 there is *decisive* evidence supporting the hypothesis that rule discovery entails Structure Learning. For subjects who did not discover the rules however, there appears to be an equally good explanation for behavioural responses using Models 1, 2, and 3, with a slight preference for Model 3. This suggests that not learning the rules could mean that subjects use the same generative model throughout the duration of the task.

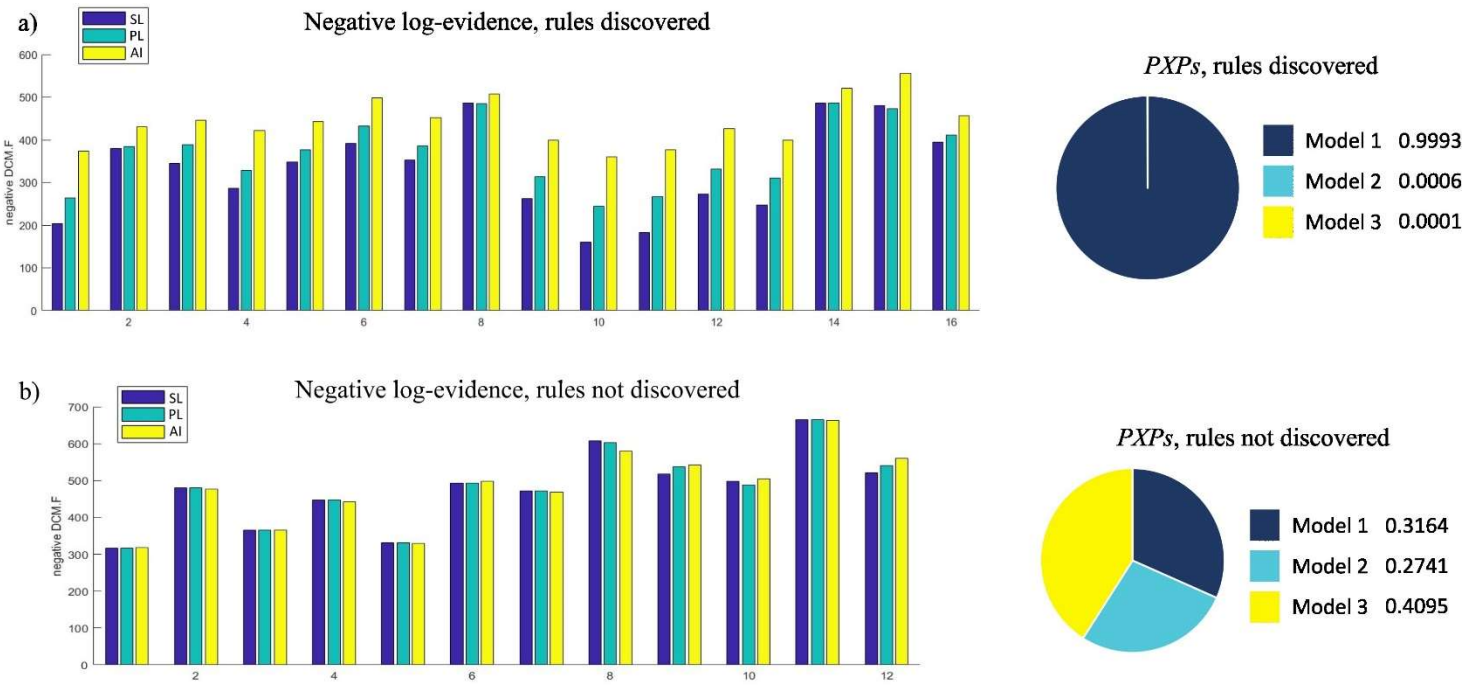


Figure 5.13 Log-evidence and *PXPs* for all subjects, pooled over blocks. Left panels - Negative free energy values (i.e., negative log-evidence) for each subject under each of the three models, for both conditions (left – rules discovered, right – rules undiscovered), pooled over blocks. Right panels – *PXPs* for each condition. For subjects who discovered the rules (top right panel), Model 1 had the highest evidence. For subjects who did not discover the rules, Model 3 is weakly favoured.

It is interesting to note that for subjects who did not discover the rules, there was a markedly lower marginal likelihood of responses under either model (Figure 5.10) in Blocks 1 and 2. This may suggest that the subjects in this group could have been using strategies that resembled the generative model less. For instance, for subject 27 (Figure 5.10), who did not discover the rules, in Blocks 1 and 2, the best model was Model 3 (i.e., AI-only). In Block 3, however, Model 1 (i.e., SL) becomes the best explanation for behavioural responses. Looking at the alternative hypotheses chosen during fitting as a result of BMR (i.e., BMS), in Block 1, hypothesis 2 was chosen at trial 6 of 32 trials (which continued for the remaining trials). In Block 2, hypothesis 14 was chosen at trial 19, and in Block 3 it was hypothesis 5 at trial 13. Hypotheses 2 and 14 are more informative than the prior, but not as informative as hypothesis 5. This means that in Block 2, the model with the best explanation pointed to a generative

model where there is only partial information on contingencies from the generative process. Similarly, in Block 2, the model with the best explanation entailed a generative model where there is only information for *correct image* = house (generative process). It is only at Block 3 that the subject started behaving in line with what would be observed for an ideal Bayesian actor (i.e., the generative process itself). It is unclear whether this subject would have discovered the rules in Block 4, if that existed, considering this information. This fit is also nuanced by this subject's accuracy, which increased modestly from 44% in Blocks 1 and 2, to 53% in Block 3; and the number of reveals made, which remains fairly constant across the 3 blocks (i.e., approximately 3 reveals per trial). One explanation, therefore, is that although the fit improved with BMR (i.e., BMS), this only got subjects closer to the generative process (i.e., to figuring out the rules), but not enough such that SL occurs. In other words, although the selected hypotheses were apt in explaining the subject's behaviour, it does not necessarily imply that SL was in play.

Contrast this with subject 6 (Figure 5.10), who discovered the rules, and whose accuracy increases drastically from 53% in Block 1, to 88% in Block 2, and 94% in Block 3. For this subject, during fitting, the hypothesis selected was always hypothesis 5 (i.e., the generative process), at trials 9 (in Blocks 1 and 2), and 10 (in Block 3). Furthermore, this subject's average number of reveals also decreases from 3 (Block 1) to 2.1 (in Block 2) to 1.9 in Block 3, meaning that altogether, this subject's behaviour is in line with an information gathering strategy that leads to the most informative hypothesis (i.e., the generative process), and subsequently, to Structure Learning.

5.6 Behavioural Experiment 2

5.6.1 Method

5.6.1.1 Participants

Participants were recruited world-wide using the Prolific platform (www.prolific.com). The sample consisted of 38 healthy adults with normal or corrected vision, and no mental health conditions or impairments. Five participants were excluded due to a failure to follow the instructions correctly (i.e., made no reveals, or only revealed the central location). The final sample consisted of 33 subjects; there were 17 females (16 males), with ages ranging from 18 to 48, M = 31.88, SD = 7.74. In this sample, participants' countries of origin or residence included United Kingdom, South Africa, Italy, United States of America, Germany, Australia, Mexico, Ireland, and Canada. The study was approved by UCL Research Ethics Committee, and all subjects gave written informed consent, being reimbursed £7.50 for their participation.

5.6.1.2 Task specifics and Procedure

In this experiment, the main task (i.e., the abstract rule-learning game) consisted of 1 block, with 32 trials. The maximum number of reveals participants had for this experiment was 4 per trial. This block had a set of three underlying hidden rules, which were identical to the rules in Experiment 1 Block 1:

- If there is a tool in the centre, the correct answer (i.e., target image) will be in the left location.
- If there is a house in the centre, the correct answer (i.e., target image) will be in the centre location (i.e., house)
- If there is a face in the centre, the correct answer (i.e., target image) will be in the right location.

Additionally, the task included a ’70-30’ feature. Given Block 1 (hidden) rules, this feature means that when ‘tool’ was at the centre location, ~ 70% of the time there would be a ‘face’ to the left, and ~ 30% of the time there would be a ‘house’ to the left, essentially coupling ‘tool’ and ‘face’ probabilistically.

Performance here indicates accuracy: i.e., the number of correct responses divided by the total number of trials. The threshold for differentiating between discovering and not discovering rules was kept from Experiment 1: if subjects had 10 or more consecutive correct responses they were classified as ‘discovered rules’.

5.6.1.3 Hypotheses for Empirical Experiment 2:

1. There will be rapid learning where rules are discovered, as represented by a significant difference in performance between subjects who discover the rules versus subjects who do not.
2. There will be a predilection for sampling novel cues which will abate with experience. In other words, the number of reveals will decrease overall as subjects familiarize themselves with the stimuli.
3. Similarly to Hypothesis 2, there will be a decrease in reaction times overall as trials progress.

5.6.2 Results

The performance in terms of accuracy was compared between subjects who discovered the rules, and those who did not. Out of 33 subjects, 16 discovered the rules. The mean accuracy for subjects who discovered the rules was M = 87.89% (SD = 11.231), whereas for subjects who did not discover the rules, the mean accuracy was M = 45.96% (SD = 14.813). Please see

Figure 5.14 (a-c). To assess whether the differences in accuracy were statistically significant, an independent-samples t-test was performed. The result was statistically significant, $t(31) = 9.119$, $p < .001$, $d = 3.176$. These results suggest that subjects who discovered the rules were, on average, approximately 42% more accurate than those who did not.

Next, trends in terms of number of reveals and reaction times (in seconds) were analysed. Figure 5.14 shows these trends across trials, separately for subjects who discovered the rules (Figure 5.14d), and subjects who did not (Figure 5.14e). Simple linear regression analyses were carried out to ascertain whether the number of reveals and reaction times decreased as trials progressed.

For subjects who discovered the rules, the results indicate that the regression model explains a significant proportion of variance in number of reveals, $F(1,30) = 23.989$, $p < .001$, $R^2 = .444$, as well as in terms of reaction times, $F(1,30) = 61.898$, $p < .001$, $R^2 = .674$. The resulting equations describing the predicted number of reveals and reaction times are as follows:

$$Predicted\ number\ of\ reveals = 2.451 - 0.031 \times trial$$
$$Predicted\ reaction\ time = 6.039 - 0.133 \times trial$$

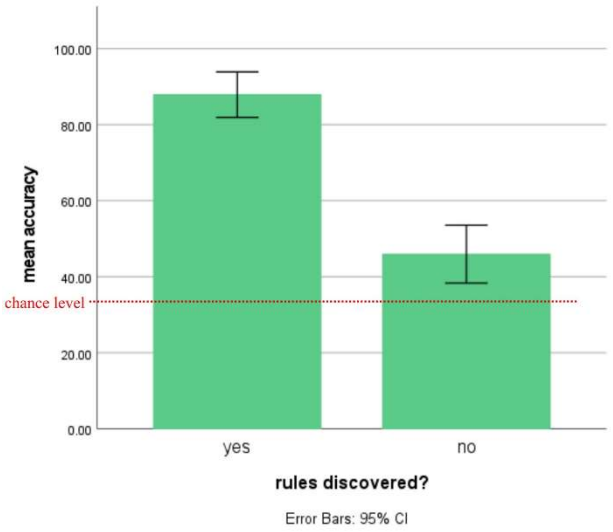
Suggesting that for every change in trial, the average number of reveals decreases by a factor of 0.031 and average reaction times decrease by a factor of 0.133.

For subjects who did not discover the rules, the results are similar, albeit slightly weaker. In terms of numbers of reveals, the coefficient of determination was weaker, $R^2 = .139$, with $F(1,30) = 4.824$, $p = .036$, and for reaction times, $R^2 = .287$, with $F(1,30) = 12.052$, $p = .002$, with the following equations:

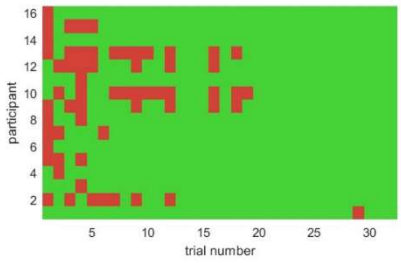
$$Predicted\ number\ of\ reveals = 2.693 - 0.006 \times trial$$
$$Predicted\ reaction\ time = 6.907 - 0.111 \times trial$$

The reason for the slope being quite small is most likely due to subjects having only 4 maximum reveals per trial. Even where rules were discovered, a correct response still required 1 or 2 reveals.

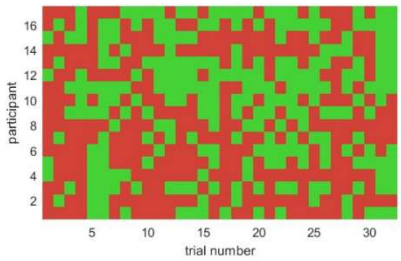
a) Accuracy comparison



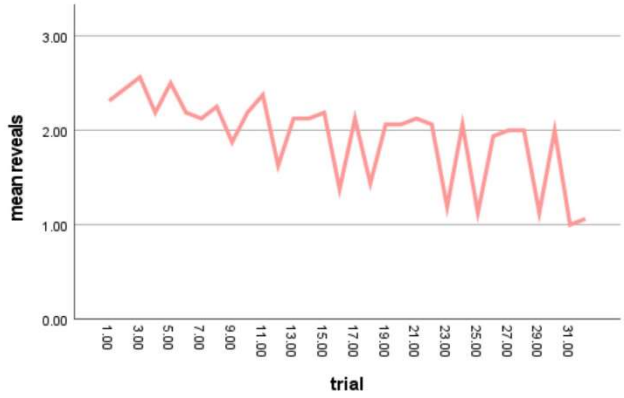
b) Correct (green) vs incorrect (red) - rules discovered

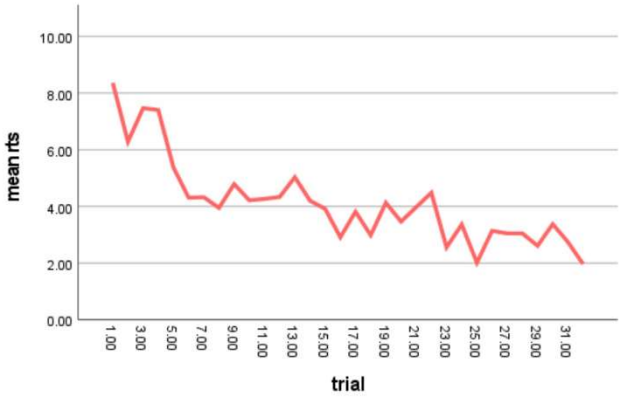


c) Correct (green) vs incorrect (red) - rules not discovered

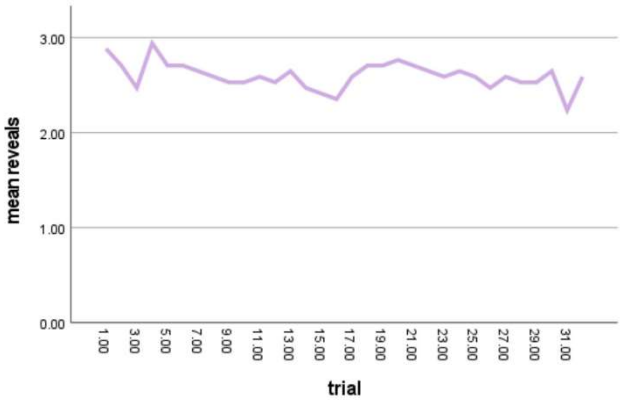


d) Rules discovered - mean nr. of reveals and mean reaction times





e) Rules not discovered - mean nr. of reveals and mean reaction times



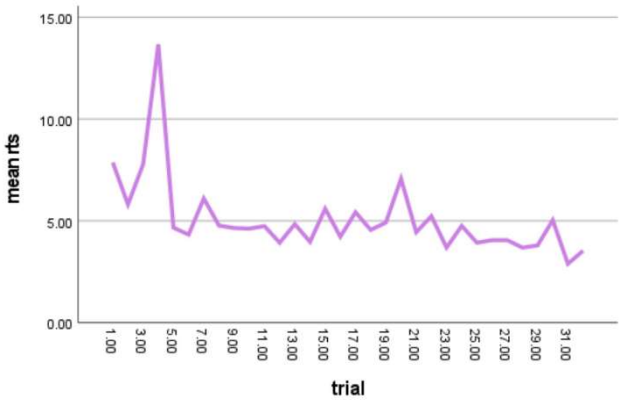


Figure 5.14 Performance comparison between subjects who discovered the rules and subjects who did not. a) Accuracy comparison showing a significantly higher performance where rules are discovered. b-c) Patterns of correct vs incorrect responses between subjects who discovered the rule (b), with subjects displaying a pattern of trial and error only until rules are discovered, and subjects who did not (c), showing an unsystematic pattern of trial-and-error responses. d-e) average number of reveals and reaction times across trials for subjects who discovered the rules (d), and subjects who did not (e), showing that as trials progress, the number of reveals and reaction times decrease.

Overall, the results in this section suggest that for Experiment 2 (similarly to Experiment 1), there is a significant difference in accuracy based on the discovery of rules. For both groups, the average performance was above chance level. The patterns of correct vs. incorrect responses seen in Figure 5.14 b) and c) are indicative of trial and error until rules are discovered, followed by a succession of correct responses only (for subjects who discovered the rules), as compared to a sustained trial and error for subjects who did not discover the rules. Furthermore, as trials progress, the number of reveals and reaction times decrease, albeit less strongly for subjects who did not discover the rules.

5.7 Fitting with the AIF – Experiment 2

As with Experiment 1, the hypothesis here is that SL as implemented by BMR explains behaviour over and above Parametric (i.e., associative) Learning (PL) and Active Inference (AI). The same insight task was used, with three differences: Experiment 2 contained only 1 block (as compared to Experiment 1, which had three blocks). The second difference is that Experiment 2 contained the 70-30 feature, where if 'tool' were at the central location, the 'face' would appear on the left ~ 70% of the time. The third aspect that differs is the number of

reveals, which was lowered from 5 (in Experiment 1) to 4 in Experiment 2. The same generative model, fitting procedure, and inversion settings were used.

Each subject's data was fitted separately with Model 1 (AI, PL, and SL), Model 2 (AI and PL), and Model 3 (AI-only). The negative free energy values for each subject are shown in Figure 5.15 a) and b), and after pooling over subjects for each condition in Figure 5.15c. The closer these values are to 0, the greater the model evidence. Figure 5.15d shows the corresponding protected exceedance probabilities (*PXPs*) for each group.

The results here show that the likelihood of behavioural responses for subjects who discovered the rules was greatest under Model 1 (i.e., model that included SL) – please see Figure 5.15c, left panel. Here, this model scored approximately 282 more log-evidence as compared to Model 2 (PL), and 803 more log-evidence as compared to Model 3 (AI-only), suggesting *decisive* evidence for Structure Learning when rules are discovered.

For subjects who did not discover the rules, the likelihood of behavioural responses is highest under Model 3 (i.e., AI only). Model 3 here scored approximately 65 more log-evidence compared to Model 2 (PL), and 68 more log-evidence as compared to Model 1 (SL) – Figure 5.15c, right panel. This result suggests *strong* evidence for the absence of Structure Learning when rules are not discovered.

These findings are further reinforced when comparing the model evidence at group level using model comparison (i.e., spm_BMS function in SPM12). For subjects who discovered rules, Model 1 (i.e., SL) has a *PXP* of almost 1, whereas for subjects who did not discover the rules, the winning model was Model 3 (AI only) with a *PXP* almost as high (i.e., *PXP* = 0.951). Overall, these results provide *decisive* evidence that discovering the rules for this task involves Structure Learning, and *strong* evidence that not discovering the rules entails that SL was not instantiated.

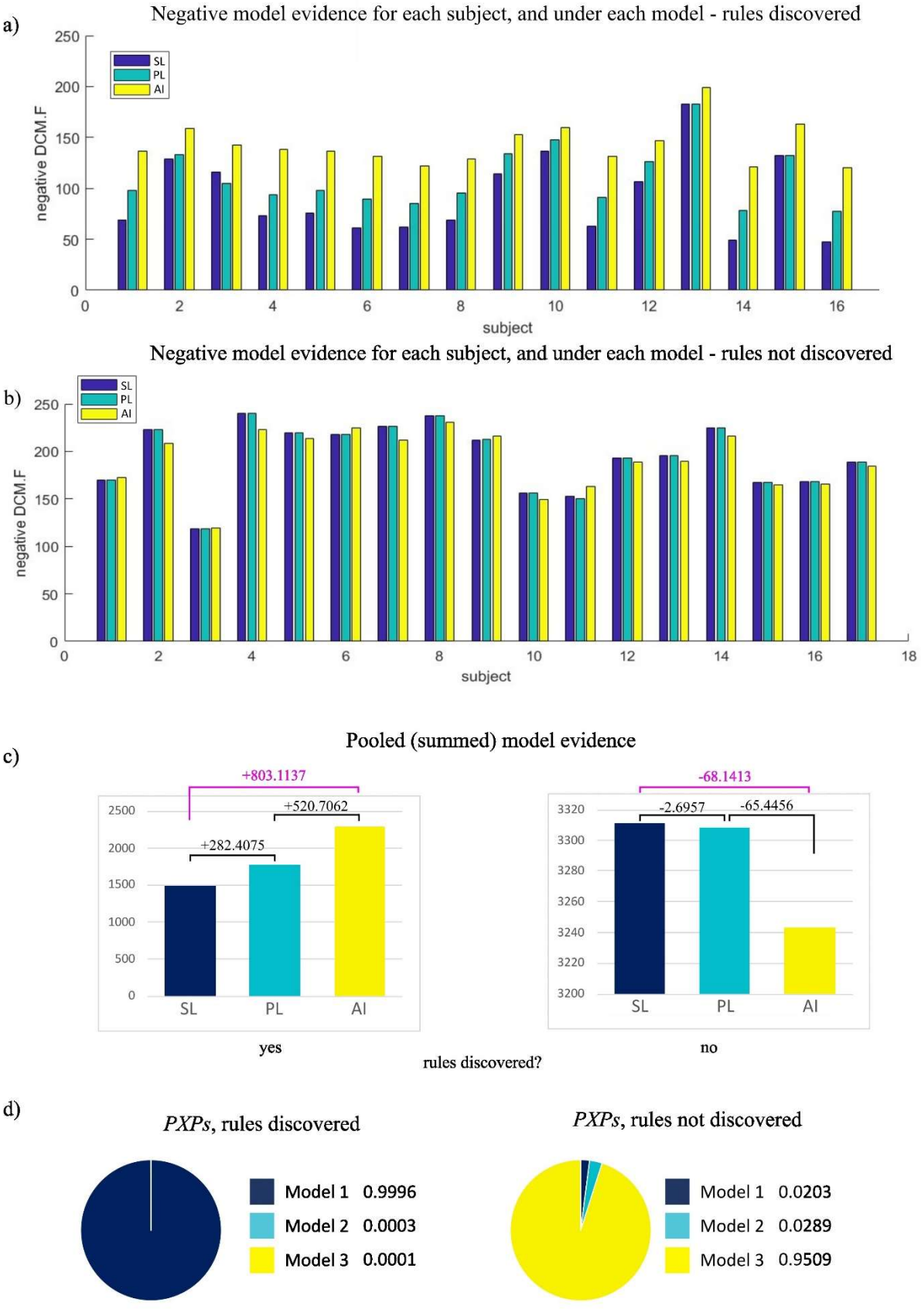


Figure 5.15 Log-evidence and *PXPs* for all subjects. a-b) Negative free energy values (i.e., negative log-evidence) for each subject under each of the three models, for both conditions - (a) discovered rules, (b) did not

discover rules. c) Log-evidence for both groups, pooled over subjects. d) *PXPs* for each group. These results suggest that for subjects who discovered the rules (a), (c-left panel) and (d-left panel), Model 1 had the highest evidence. For subjects who did not discover the rules (b), (c-right panel) and (d-right panel), the log-evidence results suggest that Model 3 provides the best explanation for behavioural responses.

In Experiment 2, we can see from Figure 5.15c that for subjects who did not discover the rules, there was a markedly lower marginal likelihood of responses under either model. This may suggest that the subjects in this condition could have been using strategies not explained by the generative model of choice behaviour. To address the differences between Experiment 1 and Experiment 2, further unplanned analyses were pursued:

Experiment 2 contained the 70-30 feature, which has a few interesting consequences. The first is that the performance for subjects who discovered the rule, is significantly higher in Experiment 2, M = 87.9% (SD = 11.231) when compared to Experiment 1 (Block 1), M = 73.2% (SD = 18.61), as evinced using an independent-samples t-test, $t(24.647) = 2.696$, p = .012, d = 0.953. On the other hand, for subjects who did not discover the rules, the performance between Experiments is not significantly different, $t(27) = 1.471$, p = .076.

Furthermore, subjects in Experiment 2 — who discovered the rules — also make significantly fewer reveals than subjects in Experiment 1, $t(30) = 2.118$, p = .043, d = 0.749, from an average of 2.21 reveals per trial in Experiment 1 to an average of 1.93 reveals per trial in Experiment 2. On the other hand, subjects who did not discover the rules make significantly more reveals in Experiment 2 as compared to Experiment 1, $t(27) = 3.386$, p = .002, d = 1.277.

In Experiment 2 – for subjects who discovered the rules — the probabilistic link between tool and house appeared to encourage the expected association: the tool-face pair had significantly higher accuracy (M = 94.6%, SD = 8.85) as compared to the tool-house pair (M =

81.3%, *SD* = 14.43), with $t(15) = 3.652$, $p = .002$, $d = 0.913$, assessed with a paired-samples t-test.

Overall, these results suggest that one explanation for the differences in behaviour between Experiment 1 and Experiment 2 could be due to the opportunity for associative learning afforded by the 70-30 feature. It could be that implicit opportunity to learn associative contingencies nudged subjects who discovered the rule to behave closer to an ideal Bayesian actor, given the additional dimension of associating tool and face. In other words, there was more information to be learnt in Experiment 2, but this information was congruent with the underlying rules that generated the observations. The resulting difference in behaviour and performance — between subjects who discovered the rules and subjects who did not — might explain why Model 3 had the most evidence for subjects who did not discover the rules; since accuracy results for subjects who did not discover the rules remain identical to Experiment 1.

This explanation is strengthened by interpreting the fitting results in terms of which hypotheses were selected after fitting Model 1: only 2 out of 17 subjects who did not discover the rules ended up having an alternative hypothesis selected during BMR; i.e., only 2 subjects showed any evidence of Structure Learning. In contrast, for subjects who discovered the rules, alternative hypotheses were selected during fitting with Model 1 for 14 out of 16 subjects.

5. 8 Numerical experiments (computational simulations)

5.8.1 Additional specifications for the generative model

This section revisits face validity in terms of the AIF model's ability to reproduce human-like behaviour in the insight (abstract rule learning) task. The ensuing simulations presented in this

section use the generative model and task described in sections 5.2 and 5.3, with a few adaptations. In brief, the emergent behaviour has notable parallels with the behaviour observed in human subjects. In what follows, we will see how abstract rule learning emerges from maximizing model evidence, and as a result of minimising expected free energy under prior beliefs that make indecisive or erroneous choices surprising.

The simulations describe agents as engaging in (Active) Inference (i.e., AI – inverting a generative model given a sequence of outcomes), PL (i.e., updating model parameters), and SL (maximizing model evidence using model selection through BMR). The addition of Structure Learning means that agents can, in principle, discover the rules that underlie the generative process. Similarly to Block 1 rules in Experiments 1 and 2, the rules for the simulations were:

- If there is a tool in the centre, the correct answer (i.e., target image) will be in the left location.
- If there is a house in the centre, the correct answer (i.e., target image) will be in the centre location (i.e., house)
- If there is a face in the centre, the correct answer (i.e., target image) will be in the right location.

As established earlier, the Active Inference Framework rests upon a generative model of observable outcomes, used to infer hidden states (i.e., the most likely causes of observed outcomes based on expected states). Since observable outcomes depend on actions, this implies the presence of expectations about outcomes under different sequences of actions (i.e., consequences under different actions). The expectations are optimized by minimising variational free energy. However, the prior probability of a specific sequence of action (i.e., a policy) depends on the expected free energy (of pursuing that policy). In turn, the expected free

energy can be decomposed into expected information gain (a.k.a., intrinsic value) and expected value (a.k.a., extrinsic value) where value corresponds to the log probability of preferred outcomes.

In brief, outcomes are generated as follows: the expected free energy for each policy is passed through a softmax function, followed by the selection of the most likely policy. Using the transition probabilities entailed by the selected policy, sequences of hidden states are generated. These sequences of hidden states then generate outcomes in one or more modalities, which restarts the implicit action-perception cycle.

In general, the behaviour of agents in an ambiguous context is dominated by exploratory behaviour (i.e., information gain), until no further uncertainty can be resolved, at which point epistemic imperatives give way to exploitative behaviour, where prior preferences dominate. One adaptation to the generative model (in section 5.3) used in the current simulations concerns the prior preferences over *feedback* outcomes ($\mathbf{C}^3$). Here, these involve a preference against being incorrect, and that the agent is likely to make a decision after the 4th timestep, even if this entailed 'incorrect' feedback.

$$\mathbf{C}_\tau = \ln P(o_\tau) = \begin{cases} -4, \ o_\tau = incorrect : \forall \tau > 0 \\ -8, \ o_\tau = null : \forall \tau > 4 \\ 0, \ otherwise \end{cases}$$

Here, τ denotes the timestep or number of samples (e.g., reveals or saccades) in each trial. Prior preferences over *feedback* differ here from the ones employed during the empirical fitting procedure. This is due to subjects in the behavioural experiments having a variable number of reveals per trial, compared to the agents in these simulations: where the length of a trial is always fixed to 5 timesteps per trial (including the initial fixation). In contrast to the fitting scheme (where there was a preference for receiving 'correct' *feedback*), for the current simulations, agents had no preference for receiving 'correct' *feedback* or no *feedback* until the

final timestep of the trial. At the final timestep, there was a preference against receiving no (i.e., null) *feedback,* essentially forcing a decision by the end of each trial. However, there was a preference against receiving 'incorrect' *feedback* at any timestep. This adaptation was chosen to emulate an instruction set that encourages uncertainty-resolving (information-seeking) behaviour until agents become sufficiently 'confident' to report a decision.

All other arrays and settings are essentially identical to the generative model described in sections 5.3 and 5.5 (the fitting to empirical choice behaviour). The focus of this section, and the numerical experiments reported, is on learning the likelihood model. The simulations were initialized as follows. The prior likelihood concentration parameters were identical to the priors used for the fitting scheme. To recapitulate, this meant that agents essentially knew the mappings for *feedback* (a^3) where the concentration parameters for correct contingencies were initialized with high precision (e.g., 128) and 0 otherwise. This can be thought of as installing a prior belief that *feedback* depends on choosing the correct image. Similarly, we used informative priors for *where* (a^2), essentially a high-precision identity matrix, regardless of other factors. This means that agents were aware that looking left entails the left *location*, regardless of what the *rule* was or *what* they were observing. The priors of interest here concern the mappings between *what* is being observed and the *correct image* (a^1). These priors entailed that agents had 'knowledge' that there were three rules as specified by the central image but were not aware of how the *rule* determined outcomes. This ignorance is instantiated by uniform concentration parameters between the *correct image* and the image seen at each location, under all three different rules for left and right locations, (and high precision concentration parameters for the central location). Altogether, these priors mean that agents would start with beliefs that there are three different rules, three different locations, and three different image types; that the feedback depended on choosing the correct image; and that the rules depend on the central location; however, agents would have no concept of what the rules are to start with.

During the simulations, each agent engaged with the environment for a total of 32 trials. In total, 100 agents were simulated. Half of these agents (i.e., 50 agents) were in a synthetic 'SL' (i.e., BMR) group, whereas the other half were in the 'AI-only' group. For the SL group, BMR was invoked at the end of each trial, employing the hypotheses (i.e., model priors) described in section 5.5. On the other hand, for the 'AI-only' condition, eta (i.e., the learning rate for model parameters) was set to a low value (e.g., 1/512), to suppress parametric learning, as observed after fitting the behavioural responses in Experiment 2.

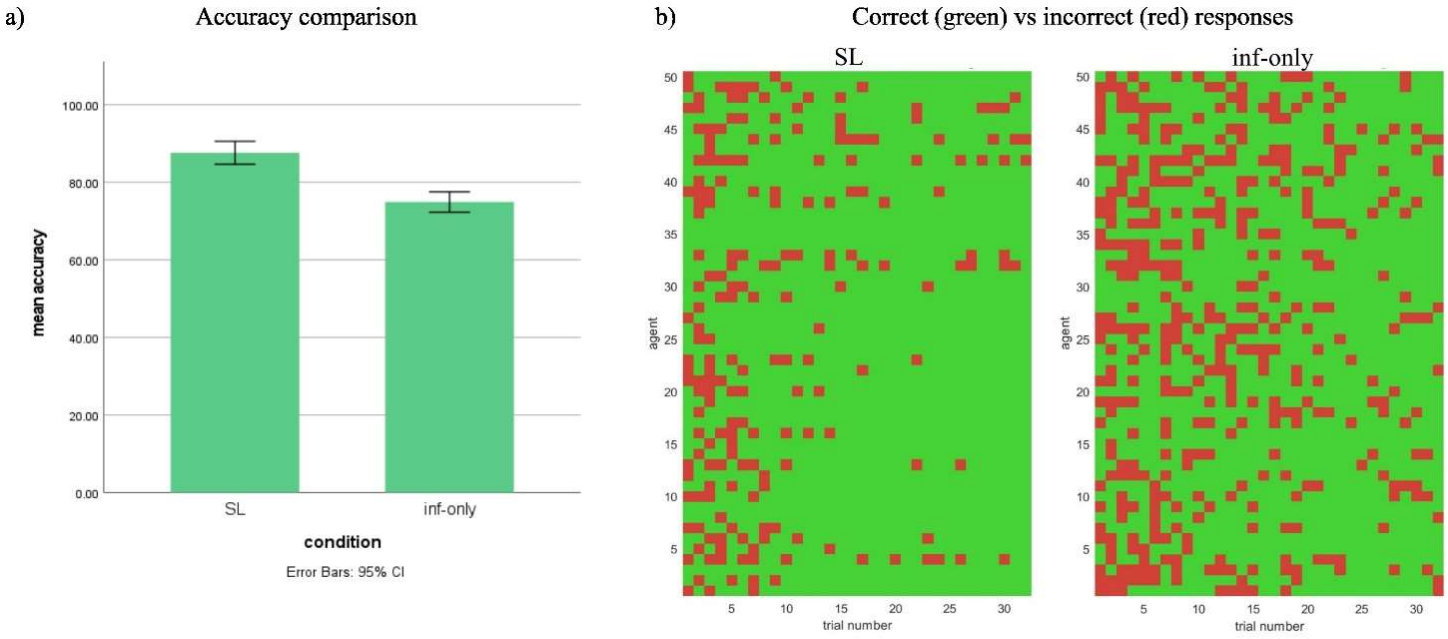
5.8.2 Hypotheses specific to numerical experiments:

1. There will be a significant difference in performance between (synthetic) agents in the SL group versus agents in the AI-only group.
2. There will be a decrease in reaction times overall, as trials progress and (synthetic) agents come to resolve uncertainty about the environment.

5.8.3 Results

The performance in terms of accuracy was compared between agents in the two groups (i.e., SL vs AI-only). The mean accuracy for agents in the SL group was $M = 87.63\%$ ($SD = 10.412$), whereas for agents in the AI-only group, the mean accuracy was $M = 74.94\%$ ($SD = 9.268$). Please see Figure 5.16 (a-b). An independent-samples t-test was conducted to assess whether the differences in accuracy are statistically significant. The results indicate that the difference in accuracy between agents in the SL condition and agents in the AI-only condition was statistically significant, $t(98) = 6.436$, $p < .001$, $d = 1.287$.

Next, we characterise behavioural trends in terms of reaction times. Simulated reaction times essentially represent the time to convergence (i.e., with Variational Bayes) for each round of message passing and action selection. Since in these simulations, each trial includes 5 timesteps, there are 5 reaction times values per trial; these were summed to provide a total reaction time per trial. Figure 5.16c shows the trends in reaction times across trials, separately for agents in the SL group (left), and AI-only group (right). Simple linear regression analyses were carried out to ascertain whether reaction times decreased as trials progressed.



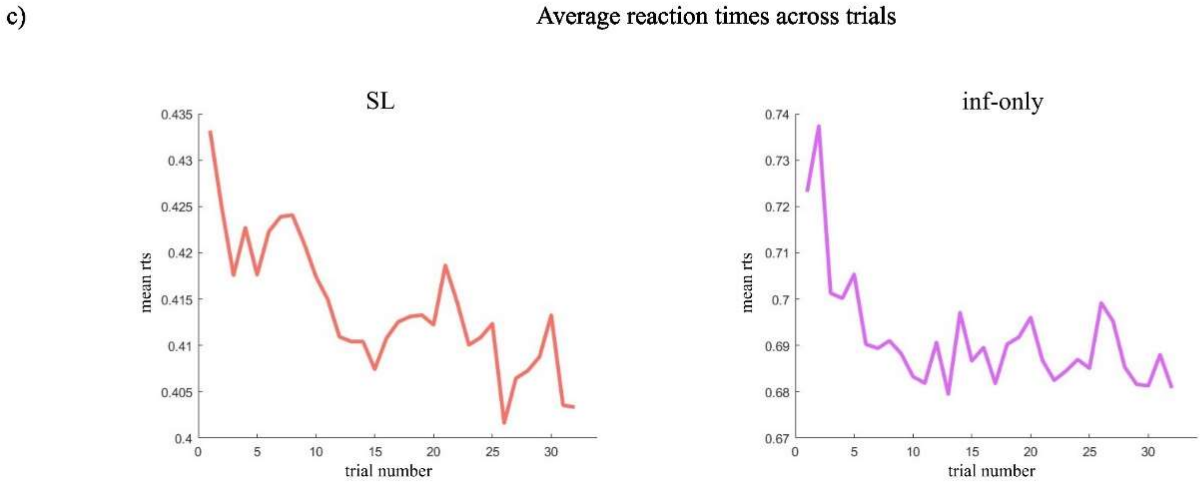


Figure 5.16 Performance comparison between (synthetic) agents in the SL vs AI-only groups. a) Accuracy comparison showing a significantly higher performance for agents in the SL group. b) Patterns of correct vs incorrect responses between agents in the SL group (left), and agents in the AI-only group (right); these patterns resemble the behavioural patterns observed Experiments 1 and 2 (c.f., Figures 5.14b-c and 5.5b-c), showing a more unsystematic pattern of trial-and-error responses in the AI-only condition (equivalent to 'rules not discovered' in Experiments 1 and 2) and trial-and-error followed by a succession of correct responses in the SL group (equivalent to 'rules discovered' in Experiments 1 and 2). c) average reaction times across trials for agents in the SL group (left) and AI-only group (right), showing that as trials progress, reaction times decrease in both groups (c.f. Figure 5.14d-e, right panels) resembling reaction time plots in Experiment 2.

For agents in the SL group, the results indicate that the regression model explains a significant proportion of variance in terms of reaction times, $F(1,30) = 55.928$, $p < .001$, $R^2 = .651$. The resulting equation describing predicted average reaction times are as follows:

$$Predicted\ reaction\ time = 0.424 - 0.001 \times trial$$

For agents in the AI-only group, the results are similar, albeit slightly weaker. The coefficient of determination for reaction times as a function of trial was $R^2 = .311$, with $F(1,30) = 13.539$, $p < .001$, and the following equation:

$$Predicted\ reaction\ time = 0.704 - 0.001 \times trial$$

These estimates suggest that although for both groups, average reaction times decrease by a factor of 0.001 per trial, the intercept is higher for agents in the AI-only condition.

So far, the results observed between the two groups (SL vs AI-only) reproduce the results observed with human subjects, where the groups were 'discovered' vs 'undiscovered'. That is, the equivalent of SL in simulated agents is that of 'discovered rules' in human subjects. Similarly to results in Experiment 2, the patterns of correct vs. incorrect responses seen in Figure 5.16b are indicative of trial and error until rules are discovered by agents, followed by a succession of correct responses. This is in contrast to a more unsystematic trial and error for

agents in the AI-only condition. Furthermore, — similar to results from Experiment 2 — as trials progress, reaction times decrease for both conditions.

There are some notable nuances in terms of Structure Learning that emerge from these computational simulations. The first is that, although BMR is applied at the end of each trial, this does not necessarily lead to selecting an alternative hypothesis (i.e., *a'*) rather than continuing with the current posterior concentration parameters (i.e., **a**). Furthermore, selecting an alternative hypothesis does not necessarily *always* lead to selecting a correct response. For the first part (i.e., selecting an alternative hypothesis), the model evidence for an alternative hypothesis has to be higher than that of the current running hypothesis (i.e., model). For the second part (i.e., for selecting a correct response), the selected hypothesis (i.e., model), although more useful than the prior, has to capture contingencies that are directly relevant to the goal of providing a correct response. In other words, the selected hypothesis has to reflect the generative process.

To illustrate this, imagine that the current simulations begin with a uniform prior over a^1. Since BMR effectively redistributes the observations previously seen, applying BMR can hinder performance if a relatively non-informative hypothesis has been selected early on, based on an insufficient number of observations. For example, if one tosses a non-biased coin, but observes tails three times, applying BMR at this stage would select an alternative hypothesis where the coin is biased (e.g., 80-20) towards tails (assuming this hypothesis exists).

Likewise, if we imagine that the prior a^1 is random, but more similar to the generative process than other random priors that are not as similar to the generative process, this will increase performance by having the appropriate alternative hypothesis selected earlier (i.e., since there is already more evidence for those specific contingencies), assuming that the set of alternative hypotheses include the true generative contingencies. Using the coin example,

imagine that a generally suspicious person is assessing whether a coin is tails-biased (e.g., 80-20) and therefore starts with a prior assumption that the coin is biased. After making three observations that result in tails, and applying BMR, the individual would correctly assume that the coin is tails biased. However, if this individual started with the assumption that it was more likely to observe heads (e.g., 20-80), applying BMR will most likely not result in selecting the 80-20 hypothesis after 3 observations of tails.

For the task used in the current experiments, quantifying when 'rules are discovered' is not clear cut. One way to address this is by looking at the trial number where an alternative hypothesis is first chosen, as well as which hypothesis was chosen. Let us now superimpose the matrix showing the time at which an alternative hypothesis was selected, over the correct and incorrect responses (see Figure 5.17). Agent 44 for example — highlighted in blue on the right panel — selects a hypothesis very early on (at trial 3), after making observations that house is at the centre (in both trials), and selecting the correct response (i.e., house) two times. The hypothesis selected however, is hypothesis 20, which is a flipped version of the generative process, and therefore not apt with regards to the true generative process; this leads the agent to continue making incorrect responses throughout the remainder of the trials. Agent 17 on the other hand, continues with the prior until trial 9, where it selects hypothesis 5 (i.e., the generative process), allowing this agent to make correct responses for 100% of the remaining trials. Please note that, for the agents in this group (i.e., SL group), there is no *correct* or *incorrect* model (i.e., alternative hypothesis), only more or less informative models (hypotheses) up until that point in time (i.e., based on the observations and actions up until that point).

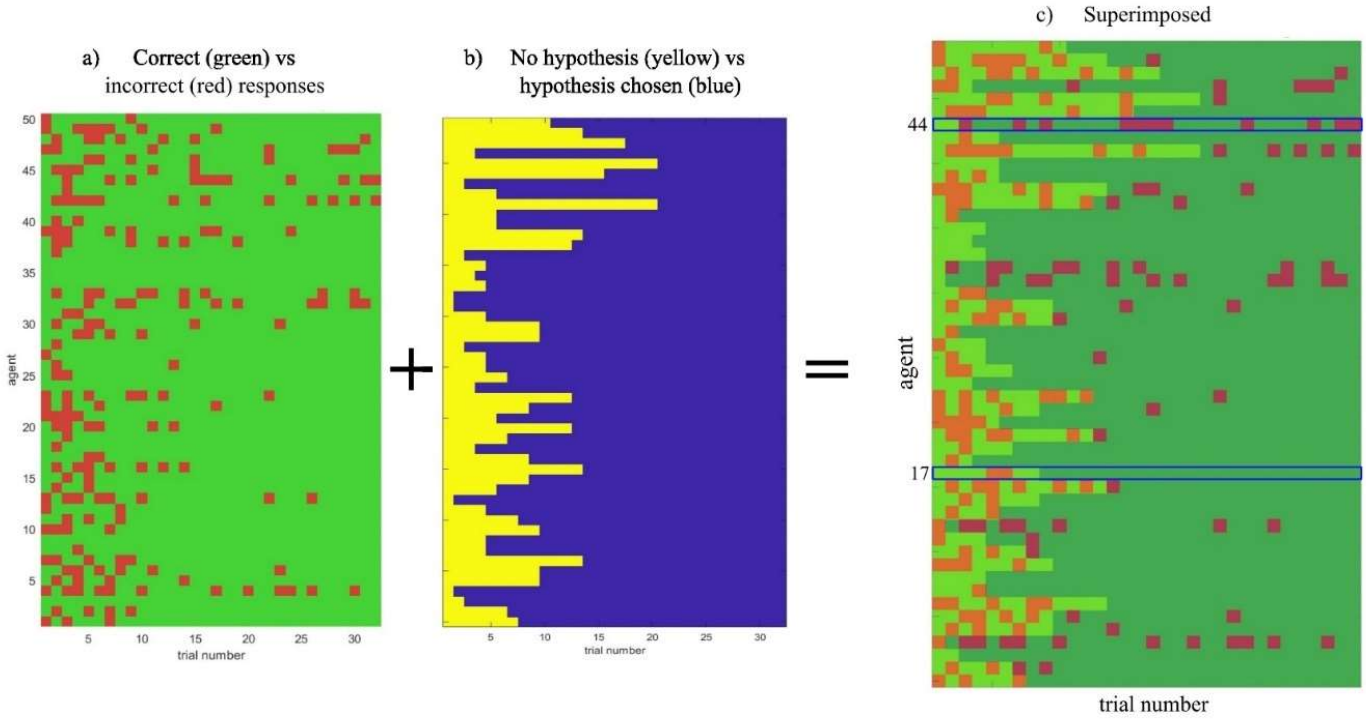


Figure 5.17 Performance and hypothesis-selection for agents in the SL condition. a) Patterns of correct vs incorrect responses for agents in the SL group, where the number of successive correct responses is increasing. b) Matrix showing the trial at which a hypothesis is selected, where yellow indicates no hypothesis chosen, and blue indicates that an alternative hypothesis was selected. c) correct vs incorrect responses superimposed with the matrix showing when hypotheses were selected. Here, light colours indicate responses before a hypothesis was selected (i.e., green-correct, red-incorrect) and dark colours indicate responses after a hypothesis was chosen. These results suggest that it is possible to apply SL (i.e., BMR) and not obtain 100% accuracy thereafter, which would depend on the agents' previous observations, when an alternative hypothesis was selected, and specifically which hypothesis was selected. The patterns of correct vs incorrect responses under this condition resemble the patterns of behaviour observed in human subjects, where rules were discovered.

Nevertheless, these results suggest that equipping (synthetic) agents with BMR (Bayesian Model Selection) results in higher performance on average, and a higher succession of correct trials after 'discovering' the rules (as compared to simulating agents without BMR (i.e., AI-only condition). The patterns observed with synthetic agents in the SL condition resemble those observed in human subjects in the 'rules' discovered' group.

This phenomenon becomes more evident (and the patterns more similar to human behaviour) when simulating agents with BMR, but containing a set of hypotheses that only includes a stark contrast between informative hypotheses (i.e., hypotheses that are in line with the generative process) and less informative hypotheses. The next set of numerical experiments involved simulating 50 agents in the SL condition, but this time with only four alternative hypotheses: the generative process, and three alternative models that combine elements of the prior (i.e., *a*) and the generative process, i.e., alternative hypotheses 5, 16, 17, and 18 from Table 5.1) . The resulting performance in terms of correct vs incorrect responses, and trial at which an alternative hypothesis was selected, are shown in Figure 5.18 below.

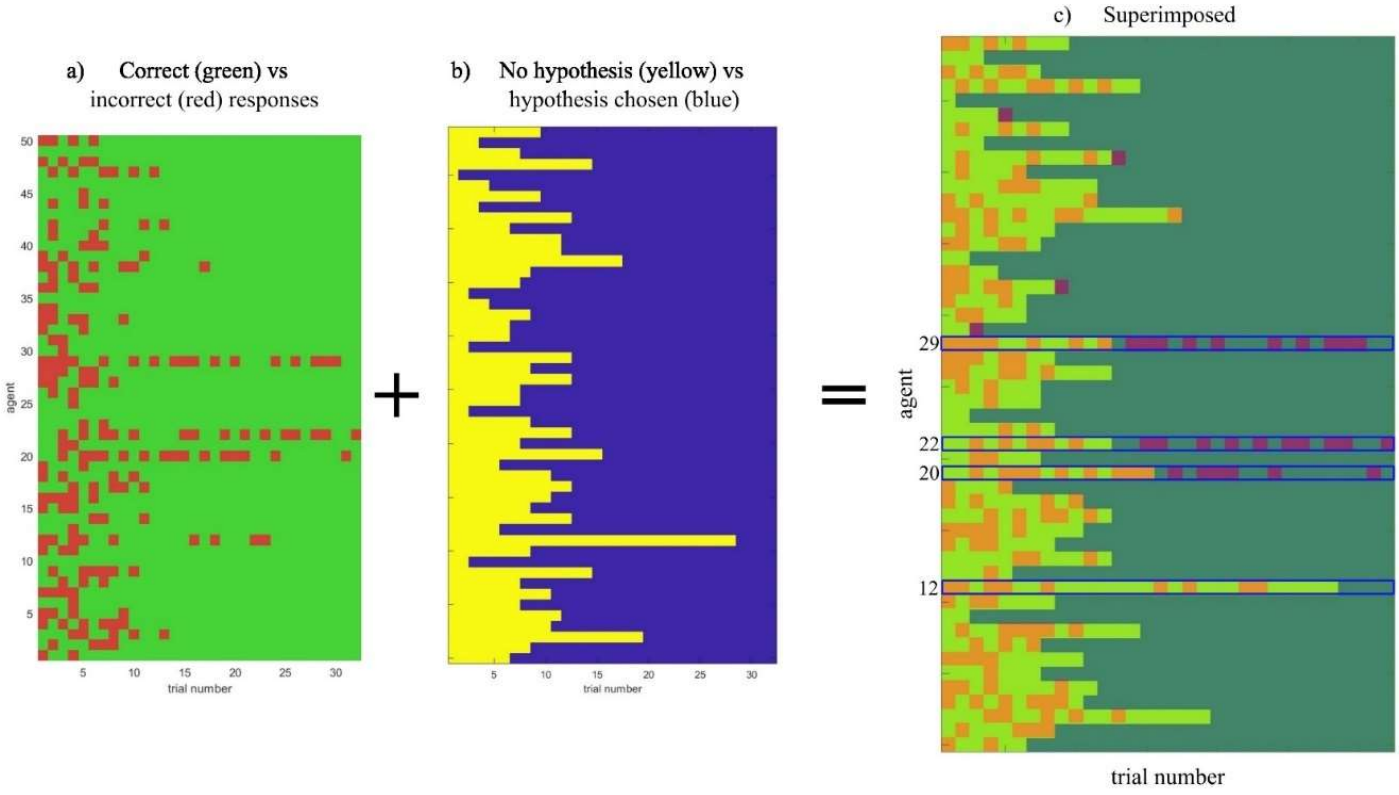


Figure 5.18 Performance and hypothesis-selection for agents in the SL condition, 4 hypotheses only. a) Patterns of correct vs incorrect responses, where trial-and-error is followed by a number of successive correct responses. b) Matrix showing the trial at which a hypothesis is selected, where yellow indicates no hypothesis chosen, and blue indicates that an alternative hypothesis was selected. c) correct vs incorrect responses superimposed with the matrix showing when hypotheses were selected. Here, light colours indicate responses before a hypothesis was selected (i.e., green-correct, red-incorrect) and dark colours indicate responses after a

hypothesis was chosen. These results suggest that it is possible to apply SL (i.e., BMR) and not obtain 100% accuracy thereafter, which would depend on the agents' previous observations, when an alternative hypothesis was selected, and specifically which hypothesis was selected. However, applying SL entails that it is more likely to 'discover the rules' along with the implicit increased performance. The patterns of correct vs incorrect responses under this condition resemble the patterns of behaviour observed in human subjects, where rules were discovered. In this simulation, synthetic agents show a clearer trend of initial trial-and-error followed by correct successive responses for the remaining trials.

In this numerical experiment, only three out of 50 agents selected an alternative hypothesis that was not the generative process (agents 20, 22, and 29). All other agents selected the hypothesis corresponding to the generative process, including agent 12, where this hypothesis was selected at trial 29 of 32. The patterns of correct vs incorrect responses in these simulations resemble the behaviour observed in human participants more closely, when compared to model selection (BMR) using 20 (informative and non-informative) alternative hypotheses (c.f. Figure 5.18a left panel, Figure 5.5b, and Figure 5.14b). That is, there is a clearer trend of initial trial-and-error, followed by successive correct responses for the remaining trials.

Overall, the results from numerical experiments shed light on notable features of this kind of Structure Learning, foregrounding the results from fitting both Experiments. Simulated data in both conditions is reminiscent of behavioural results in Experiment 1 Block 1, and Experiment 2, where the behaviour and performance of SL agents mimics that of subjects who discovered the rules; and the behaviour of agents in the AI-only condition mimic results observed in subjects who did not discover the rules. In the computational simulations, agents engaged in (Active) Inference (i.e., inverting a generative model given a sequence of outcomes), PL (i.e., updating model parameters), and SL (maximizing model evidence using BMR) whilst aiming to discover the rules that underlie the generative process. That is, obtaining a correct response entailed 'learning' the rules that generate observable outcomes.

With these aspects in mind, having evidence for SL during the process of fitting behavioural responses — where rules were discovered — can be interpreted as subjects engaging in SL successfully. Conversely, there are instances of unsuccessful engagement in SL: for example, by selecting an unhelpful hypothesis (i.e., a hypothesis that diverges from the true generative process), by selecting a hypothesis prematurely (i.e., after insufficient observations, and if these observations bias agents' beliefs towards unhelpful hypotheses), and also by selecting a hypothesis too late (where the prior remains in place, precluding agents from 'learning' about the contingencies that allow them to increase their performance), or not selecting one altogether. In essence, the notion that if subjects discover the rules, *it implies the use of SL* can be updated to '*it implies the successful application of SL'*.

5.9 Interim discussion

The focus of this chapter was to present evidence for Structure Learning in a cognitive task in humans, and to offer a more detailed explanation of how this process guides behaviour. The overarching hypothesis for this chapter was that rapid learning entails Structure Learning, over and above Parametric Learning and Active Inference. An abstract rule learning task was employed in the empirical and synthetic experiments, whose results highlighted a significant performance gap: between subjects who discovered the rules and subjects who did not, and in parallel, between synthetic agents where SL was applied, vs. where SL was denied. The results overall emphasize the need for an information-processing mechanism beyond mere associative (i.e., Hebbian) learning.

The task used involved presenting different arrangements of three different stimuli (tool, house, and face) for the duration of 32 trials (for Experiment 2 and numerical

experiments) or 96 trials (for Experiment 1). The goal for this task was to provide correct responses, and this could only be achieved if agents, synthetic or human, discovered the rules that generate the observable outcomes. Whereas for human subjects, inspecting the images entailed clicking to reveal the masked images, for synthetic agents, this involved taking actions (e.g., sample left).

In section 5.4, the results showcase rapid learning where rules are discovered, indicated by a significant difference between subjects who discovered the rules, and subjects who did not. This difference remained consistent across the three different blocks. Furthermore, there was a predilection for sampling novel cues, which abated with experience, and a reduction in reaction times as trials progressed. Section 5.5 addressed the overarching hypothesis for this chapter: that the behaviour of subjects who discovered the rules is better explained by a model that includes SL; and the behaviour of subjects who did not discover the rules would be better explained by a model without. Following model comparison, for subjects who discovered the rules, the best explanation was provided by Model 1 (i.e., the model with SL), providing *decisive* evidence that subjects who discovered the rule engage in a form of SL. For subjects who did not learn the rules, the results weakly favoured Model 3 (i.e., AI-only).

For subjects who did not discover the rules, there was a markedly lower log-likelihood of responses for Blocks 1 and 2. This implies that subjects in this condition might have been employing strategies not accommodated for in the generative model, at least until Block 3. The inconclusive (weak) results for subjects who did not discover the rules could also be explained by the mixture of hypotheses that were selected during BMR for the fitting procedure. Under Model 1 (i.e., SL), some subjects' data was fitted with hypotheses less informative than the generative process, but more informative than the prior, which entailed a higher log-likelihood, purely by virtue of that alternative hypothesis explaining the subjects' behaviour data. In other words, the behavioural data was better explained by alternative models that were not

informative, but they were more informative than the prior, resulting in higher evidence for Model 1 (SL). Other subjects' data was better explained by Model 3 (AI-only), but overall, these log-evidence values were cancelled out by those subjects whose behaviour was better explained with Model 1.

To evince a stronger underlying distinction — between discovering and not discovering the rules — Experiment 2 essentially coupled the tool and face pair probabilistically, in a way that was congruent with one of the underlying rules (i.e., if tool is at centre, the correct image is on the left). Comparable to Experiment 1, there was a significant difference in performance between subjects who discovered the rules, and subjects who did not. However, in Experiment 2, subjects who discovered the rules performed significantly better than subjects in Experiment 1 (who discovered the rules). Furthermore, subjects who discovered the rules made significantly less reveals in Experiment 2 compared to Experiment 1; subjects who did not discover the rules on the other hand, made significantly more reveals in Experiment 2 (compared to Experiment 1). Overall, this indicates that introducing this probabilistic association increases the difference in behaviour between the two groups. The behavioural responses were fitted using the Active Inference scheme, with results that provided *decisive* evidence for Model 1 (SL) for subjects who discovered the rules, and *strong* evidence for Model 3 (AI-only) for subjects who did not discover the rules.

Section 5.8 endeavoured to recreate the human behaviour observed with this task in synthetic agents, using numerical experiments. One hundred agents were simulated, half of which were in the SL group, with the other half in an AI-only group. The equivalent of discovering rules in human subjects was the inclusion of SL in the computational simulations. The behavioural trends in performance, reaction times, and correct *vs*. incorrect responses were similar to those observed in human participants, with a significant difference in accuracy between the conditions, and an overall decrease in reaction times as trials progressed. The

simulations revealed some notable insights regarding Structure Learning. Firstly, employing BMR following each trial does not necessarily result in selecting an alternative hypothesis. This is because alternative hypotheses are only selected when the evidence for an alternative model exceeds that of the existing model. Moreover, selecting an alternative hypothesis does not guarantee that a correct model will be selected. Correct responses are most likely when the selected hypothesis captures the relevant contingencies in the generative process. That is, it is possible to select an alternative hypothesis that is more useful than the prior, but much less useful than other hypotheses that are closer to the true contingencies. Considering these factors, the evidence for Structure Learning during the process of fitting behavioural responses where rules are discovered, can be interpreted as a *successful* deployment of Structure Learning. Conversely, there are instances of unsuccessful deployment of Structure Learning, such as selecting an unhelpful hypothesis (i.e., a hypothesis that diverges from the true generative process), selecting a hypothesis prematurely (i.e., after sparse observations, which biases agents' beliefs toward unhelpful hypotheses), or as a result of delaying the selection of a hypothesis (e.g., under contradictory or mixed observations, where the prior remains in place, thus hindering agents' ability to learn). Essentially, the idea that subjects discovering the rules implies the use of Structure Learning can be nuanced: that discovering the rules involves the *successful* deployment of Structure Learning, where the term *successful* depends on all the agent-environment contingencies at play.

5.9.1 Limitations and future directions

One limitation concerns the scheduling of the three distinct levels of belief updating. For the current simulations and fitting procedure, the order was Active Inference and PL, followed by SL (i.e., BMR) at the end of each trial (after an assumed number of sampled observations).

However, it is conceivable that there would be a different scheduling. For instance, SL could be preceded by 5 'episodes' of engagement with the environment via AI and PL. This raises an empirical question: 'Is there an optimal scheduling (i.e., timing) for Structure Learning?'. This could be evaluated by simulating agents and fitting behavioural data using a variety of SL schedules, such as implementing SL after every 2, 5, 10, or 20 trials.

There is, however, a notable aspect of SL implementation using the set-up described in this chapter: although SL is implemented at the end of each trial, it does not automatically imply the selection of an alternative hypothesis. This is because during SL, a comparison is made between alternative hypotheses and the running hypothesis, meaning that the running hypothesis can in principle continue having more evidence when compared to other hypotheses. That is, in some cases, it may not be necessary to test the scheduling empirically, since SL naturally selects an alternative model when there is more evidence for one, as compared to the current set of beliefs. This aspect is of course subject to the contingencies specific for that environment, and the way in which alternative hypotheses are specified. For every agent-environment interaction, (i.e., for every generative model situated in its generative process), there would be an appropriate timing to select an alternative explanation (e.g., 'change one's mind'); this optimal timing would occur when the model evidence for an alternative explanation exceeds that of the current posterior (and that of other alternative explanations). For the current implementation, this would be when any of the alternative explanations becomes ~20 times (or above) more likely than the current explanation (i.e., model).

In the current experiments, for the SL group, twenty hypotheses were generated for simplicity, to illustrate the power of Structure Learning in both synthetic and human subjects. Neither the set of alternative hypotheses, nor their structure, change during simulations or fitting, an aspect which would most likely differ in reality. That is, although the agents' beliefs

change and update as a result of the three levels of evidence accumulation, the same hypotheses are always compared with the running posterior concentration parameters. Furthermore, even if at trial 1 for example, an agent dismissed hypothesis 5 as uninformative, this hypothesis will nevertheless be included in the Model Comparison procedure at the end of trial 2. One can speculate that this assumption (i.e., that all hypotheses are used for Model Comparison at the end of each trial) means that our computational simulations entertain very open-minded agents, who do not (implicitly) dismiss any hypotheses even if they have already inferred 31 times in the past that one or more of these hypotheses were unlikely. Current implementations of BMR in the Active Inference Framework have considered a principled way to best define alternative hypotheses (Friston, Da Costa et al. 2023). Here, the alternative hypotheses are created in an ongoing fashion, using Bayesian Model Expansion, where the likelihood mappings essentially expand (bottom up) to accommodate new content. In this special kind of Structure Learning, the comparison is between a hypothesis where each outcome is generated by a previously unseen set of contingencies and (the most likely) previously encountered set of contingencies. There are several empirical questions that could be derived in terms of alternative hypotheses. For example, are the weights of different alternative hypotheses asymmetrical? That is, are some hypotheses more likely to be selected *a priori*? If that is the case, how would each individual's cognitive set contribute to the likelihood of that alternative hypothesis being more likely *a priori*? And conversely, how does the structure of the environment itself contribute to the likelihood of that alternative hypothesis being more likely a priori?

Other future experiments could explore how various parameters influence the process of SL in other tasks, for example, the precision of policy selection, the learning rate, or prior preferences. For example, in the current context, decreasing the precision of policy selection resulted in some of the simulated agents selecting more than one hypothesis during the 32 trials for which they were engaging with the task. This aspect of Active Inference and Structure

Learning could be especially useful in paradigms that investigate when agents or subjects are likely to 'change their mind', which could be the equivalent of selecting an alternative hypothesis using SL.

Another limitation of the task employed in the current experiments is that the task design contains information (that can be learnt) other than the underlying rules: although the proportion of the three stimuli was approximately 33% at each location (excluding Experiment 2, with the added 70-30 feature), the occurrence of each type of stimulus as a correct response did not have a 33% chance. This is due to the rules themselves: for Block 1, house was at the centre 33% of the time. This implies that already 33% of the time the correct response was 'house'; however, when tool was at the centre location (and the rule entailed 'look left'), if house was at the left location, it entailed an additional correct response for house. Overall, the actual correct response for Block 1 was overwhelmingly 'house'. In other words, subjects could learn to select 'house' more often during Block 1 without discovering the rules; with the current design, it is not possible to disentangle between subjects selecting a response as a result of learning this feature (i.e., that 'house' is more likely to be correct) and selecting a response as a result of discovering the underlying (hidden) rule(s). However, with the current implementation, the results were at their most interpretable. One solution could have been to allow a stimulus to occur more than once per trial. For instance, picture a trial where 'face' was at the centre, and face was also 'right' and 'left'. Selecting 'face' as a response in this case, is uninformative to both the subject, and the interpreter: it is unclear whether the subject selects 'face' because it is at the central location or 'right/left' locations, resulting in a false positive. That is, the subject would be more likely to select the correct response, despite being unaware of the rules that underlie the observations. In general, any format that involves symmetrical rules will result in an asymmetrical number of each stimulus being correct, because of the underlying rules. Altogether, this means that for this task, there was additional information

(peripheral to the rules) that subjects could take advantage of to provide correct responses. In future work, one could explore how varying features of this task result in changes in performance. One example could be to implement a 70-30 rule that is incongruent with the underlying rules.

5.9.2 Summary

In conclusion, this chapter reports three experiments, two empirical, and one involving computational simulations. We considered an abstract rule-learning task in two settings: subjects who discovered the rules (SL condition for synthetic agents), and subjects who did not (AI-only condition for synthetic agents). The discovery of rules (via Structure Learning) resulted in a marked increase in performance, a decrease in reaction times, and a decrease in sampling of novel cues. We fitted the empirical choice behaviour by optimising an Active Inference Framework model. The results of fitting indicate the presence of Structure Learning when rules are discovered, and its absence when rules were not discovered. In other words, the behaviour of subjects who discovered the rules is best explained by a model that includes Structure Learning. For subjects who did not discover the rules, the behaviour was best explained by an AI-only model (i.e., a model without SL). Across the two empirical experiments, subjects were recruited world-wide, suggesting that the Structure Learning mechanisms illustrated in this section could potentially capture a universal feature of human cognition.

Chapter 6

Overarching discussion

Structure learning and the learning of structures are fundamental aspects of information processing in humans and ethology, attracting an enduring interest for theoretical and computational modelling research. In Chapter 1, we saw how scholars from various fields apply the term *structure learning* to describe a variety of phenomena such as statistical learning (Emberson, Misyak et al. 2019, Monroy, Meyer et al. 2019), adaptive generalisation (Mulavara, Cohen et al. 2009), learning to learn (Braun, Mehring et al. 2010), concept learning (Smith, Schwartenbeck et al. 2020), and probabilistic learning (Lake, Salakhutdinov et al. 2015). Conversely, other work employs structure learning mechanisms and concepts without directly referring to the process as such: predictive motor activation (Ghilardi, Meyer et al. 2023), motor learning and imagery (Di Rienzo, Debarnot et al. 2016), causal reasoning (Gopnik, Glymour et al. 2004), abstract reasoning (Jung-Beeman, Bowden et al. 2004), replay (Stoianov, Maisto et al. 2022), and deep (meta) reinforcement learning (Hu, Ma et al. 2021).

These diverse frameworks offer different perspectives and predictions regarding structure learning, with variations in properties and mechanisms depending on the specific framework. However, it allowed me to propose a temporary, but inclusive, definition of SL – as the ability to internalise a model of contingencies (i.e., conditional (in)dependencies) among different elements in space-time, that can be generalised and leveraged with ease. While inclusive, this definition is too generic, and lacks a (unified) mechanistic account. This brought us to Chapter 2, where I introduced the Active Inference Framework, foregrounding an extensive and specific definition for Structure Learning, operationalised as Bayesian Model Selection.

In Chapter 2, I introduced the three main levels of optimisation in the AIF, with (Active) Inference as the inference of latent causes (Friston, FitzGerald et al. 2017), Parametric Learning as a gradual evidence accumulation linked with associative (Hebbian) learning (Friston, FitzGerald et al. 2016), and Structure Learning as the selection of models using BMS for

optimising model evidence – associated with synaptic homeostasis (Kiebel and Friston 2011, Hobson and Friston 2012, Friston, Lin et al. 2017). In light of the AIF, *Structure Learning* was endowed with an augmented definition: the comparison and selection of models (i.e., hypotheses) with the purpose of maximising model evidence via a process called Bayesian Model Selection. BMS can follow two approaches, where one approach entails re-simulating experienced events (and assessing the model evidence under different prior models) – as seen with fitting during Chapter 5; and the other involves a comparison of post-hoc beliefs (i.e., models, hypotheses) with alternative models, without the need to reinvert the model (i.e., to re-simulate experiences or events) – as seen in Chapters 4 and 5. That is, Structure Learning as BMS can involve reinversion or re-simulation of events – in which case it is referred to as simply BMS (Friston and Penny 2011), or no reinversion – in which case it is referred to as BMR or BME (which are subsets of BMS). Structure Learning occurs offline (i.e., it happens in the absence of novel evidence), and can be interleaved with episodes of (Active) Inference and Parametric Learning, or it could occur after a longer period of active engagement with the environment. Structure Learning goes beyond (gradual, associative) Parametric Learning of structures (as seen in Chapter 3), granting agents the capacity for rapid learning. It is important to note here that Structure Learning is contextualised by – and cannot occur without – an underlying Parametric Learning component (e.g., Dirichlet concentration parameters in discrete state-space models), which is in turn contextualised by (Active) Inference. This is because the optimisation implied in the AIF relies on a factorisation of the variational density over the three main levels of unknowns presented here: model selection contextualises learning, which in turn contextualises inference.

Whereas some of the features of structure learning (in the literature) are accounted for in existing work using the AIF, others require the use of SL as implemented with BMS. For example, we saw how statistical learning (Monroy, Gerson et al. 2019) can be described using

a concept called Bayesian surprise (a.k.a., information gain, mutual information, etc.), without the need to invoke SL as BMS. Motor learning via online motor imagery (Toth, McNeill et al. 2020) has been illustrated with habit formation, where agents learn a probability distribution over policies, implemented with Parametric Learning (Friston, FitzGerald et al. 2016). Yet offline motor learning necessitates a mechanism beyond Parametric Learning: learning in the absence of novel evidence. Going by the formal definition of SL (in the AIF), offline motor learning (via motor imagery) can be investigated and formalised using BMS. This aspect – i.e., Structure Learning of policies, implemented using BMS – has yet to be explored with the AIF.

In another example, we saw how a spontaneous emergence of replay (of events) using the AIF can be thought of as a replay of previous actions in the attempt to make sense of them (Parr and Pezzulo 2021). In this case, replay emerged naturally during the simulation of behaviour, because of the uncoupling inherent in the AIF, where the *actual* actions taken are different from actions *considered* (i.e., actual vs. counterfactual actions). However, replay events have been shown to occur both online (Eysenbach, Salakhutdinov et al. 2019, Tambini and Davachi 2019) and offline (Gruber, Ritchey et al. 2016, Stoianov, Maisto et al. 2022). Additionally, replay was suggested to not only involve a rehashing of previous experience, but also as a form of compositional computation that synthesises information into relational structures to derive new knowledge – i.e., it involves any possible re-arrangement of sequences (Kurth-Nelson, Behrens et al. 2023). In light of this research, offline and compositional replay cannot be fully accounted for using online models that preclude learning – e.g., as seen in (Parr and Pezzulo 2021). This is where Structure Learning can step in – we saw in Chapters 2, 4, and 5 that SL (as BMS, BMR, BME) happens in the absence of novel evidence (i.e., it occurs offline); and SL (as BMR, BME) is compositional in that it can reorganise and flexibly (re)combine individual elements into novel relational structures (i.e., hypotheses), and select a most informative one (given model evidence). Furthermore, SL (in AIF) has been associated

with physiological processes such as the synaptic homeostasis observed during periods of sleep (Tononi and Cirelli 2006, Friston, Lin et al. 2017). In sleep, endogenous activity – resembling the neural message passing in wakefulness – has been interpreted as the generation of fictive data to evaluate model evidence (Hinton, Dayan et al. 1995). In other words, fictive episodes are replayed in the absence of sensory evidence, with the purpose of optimising (generative) models. The field of replay can therefore benefit from employing the AIF, and more specifically SL (implemented as BMS, BMR, BME) since it introduces aspects of epistemic gain, which is absent from current (computational) implementations of replay using reinforcement learning; and the ability to characterise replay during both online (e.g., planning) as inference, as well as offline hypothesis (i.e., model) comparison and selection. In relation to benefits of introducing notions of epistemic gain or model evidence, let us consider (Liu, Mattar et al. 2021), where replay (here modelled using reinforcement learning) prioritises events in terms of *need* (the probability that a specific event will be visited in the future given its frequency) and *gain* (a function of rewards). One could augment the *gain* metric to include information gain – where utility is balanced with predictive power – to ascertain whether replay also favours highly informative events without a reward (e.g., events that make an agent certain about what is *not* a good action to take). In summary, in light of the AIF, what is being replayed could be counterfactuals (e.g., alternative courses of actions and their consequences) and alternative explanations of experienced events.

Furthermore, the dual nature of learning found in the AIF is present in various other descriptions concerning the learning of structures. This is both in terms of mechanistic similarity and phenomenological similarity. The mechanistic similarity concerns the differentiation between Parametric Learning and Structure Learning as found in the meta-learning literature (Braun, Mehring et al. 2010), where learning involves a gradual adaptation by steady evidence accumulation (i.e., PL), and forming new structures by associating existing

elements in novel ways (i.e., SL). The phenomenological similarity – between the general literature and AIF in terms of the 'dual nature of learning' – concerns the effects of learning during problem solving (Jung-Beeman, Bowden et al. 2004), where agents can discover the solution using a trial-and-error incremental approach (i.e., as seen with PL in Chapter 3), or the insight-based approach, entailing instantaneous learning – as seen with SL as BME (Smith, Schwartenbeck et al. 2020, Friston, Da Costa et al. 2023) and SL as BMR (Chapters 4 and 5).

Chapter 3 was centred on Parametric Learning, which was one of the two main approaches to learning structure in the literature, and a requisite level of processing for establishing Structure Learning. Here, we saw how Parametric Learning of likelihood and transition (arrays) using AIF reproduced two (phenomenologically different) types of foraging behaviour: goal directed navigation, and epistemically driven exploration. In the first type of foraging, agents simply follow trajectories to their (known) goal, given preferences for a target location. In the second type of foraging however, the preference-seeking motivation is contextualised by explorative (i.e., epistemic) imperatives. In this sense, the selected policies (i.e., courses of action) can be thought of as a product between the prior preferences and the epistemic value. If a policy has a negligible probability of securing a preferred outcome, it will be discarded even if it has high epistemic value (e.g., salience). These findings are reflected in work on planning and navigation using the AIF (Kaplan and Friston 2018). The work by Kaplan and Friston (2018) shows how the explore-exploit dilemma can be dissolved using Expected Free Energy principles: epistemic foraging implicitly entails a component of securing prior preferences. Nevertheless, Parametric Learning sets the scene for the need for Structure Learning. The goal-directed navigation (i.e., one of two examples of foraging behaviour) in Chapter 3 started with precise prior beliefs about contingencies (implying minimal need for exploratory behaviour) – but how did beliefs (i.e., models) become precise in the first place? One answer is found in Chapter 3: by gradually learning a probability distribution over

(likelihood and transition) contingencies. Another answer was explored in Chapters 4 and 5: via Structure Learning implemented as model comparison and selection.

The main focus of Chapter 4 was to illustrate the importance of Structure Learning as implemented (in silico) by Bayesian Model Reduction, in the context of concept formation. Synthetic agents foraged a novel environment comprised of 16 different rooms, each with 16 locations and an associated reward location, forming (precise) beliefs about contingencies gradually via Parametric Learning, and rapidly via Structure Learning. One benefit of SL was the stark improvement in performance. Other benefits were presented in terms of information gain: there was a marked difference in information gain after BMR relative to before BMR (at the time, the first work to date to my knowledge to compare between the two information gain metrics). This aspect is reminiscent of, and may provide a good explanation for the sudden subjective feeling of certainty that people report following insight in the problem-solving literature (Weisberg 2013, Webb, Little et al. 2016). Furthermore, in Chapter 5, we saw that Structure Learning via BMR may not necessarily imply that the most useful representation of contingencies is learnt (since the representation of contingencies is always contextualised by previous observations), despite the drastic change in precision. This aspect of SL provides further explanatory value for insight research showing that although the subjective feeling of certainty was present, it did not necessarily mean that the correct solution was found (Metcalfe 1986), although there is some degree of correlation between the two measures (Salvi, Bricolo et al. 2016). Interestingly, results from this work (Salvi, Bricolo et al. 2016) also show that the analytical (associated with PL in the AIF) solutions were less accurate as compared to insight (associated with SL in AIF) solutions, which was one of the observed results in Chapters 4 and 5: agents in the SL (i.e., BMR) group performed consistently better than agents in the group where BMR was precluded (i.e., PL only – Ch. 4; or AI only – Ch. 5). Future work could therefore employ the paradigm in Chapter 5 empirically, adding a component that assesses the

participants' subjective beliefs about feelings of certainty, followed by comparing this metric with measures of information gain or model evidence in the computational (AIF) paradigm.

Chapter 5 presented the first empirical evidence for Structure Learning as implemented with BMR in a cognitive paradigm with humans. More specifically, after fitting behavioural responses under three different AIF models, the results for subjects who discovered the rules illustrated decisive evidence for the model that included Structure Learning. In contrast, for subjects who did not discover the rules, the best explanation was the (Active) Inference only model (i.e., no learning). Furthermore, in silico simulations – that implement the settings derived from the empirical results – are significantly more likely to reproduce the 'discovery of rules' when a SL component is present as compared to when it is not. One aspect of the SL implementation in this chapter involved that the same set of hypotheses are compared (at the end of each trial) for model comparison and selection. Interestingly, in one study, where learning to solve a category of problems was interpreted as a search through a hypothesis space of rules (Lee, Betts et al. 2016), subjects were shown to not have a memory of past incorrect hypotheses, making it likely to retry them. Although originally unintended, this implementational feature of the task in Chapter 5 appears to have some degree of ecological validity based on the (Lee, Betts et al. 2016) study. However, over prolonged periods of time, it is unlikely that the set of hypotheses used for comparison and selection would remain the same. Future work could explore how a change in the number of hypotheses compared would influence observed behaviour in decision-making. For example, an AIF paradigm similar to the one in Chapter 5 Experiment 1 can be implemented (with 3 or more blocks), and the behaviour can be fitted with models that include a decreasing or increasing number of alternative hypotheses, to answer the question of how posteriors from one block are carried over as priors to the next block (if at all), and depending upon whether the rules have been discovered.

Structure Learning is formalised as a generic type of information processing and can therefore be applied to any generative model. The challenge then, is not to figure out how alternative explanations (i.e., models, hypotheses) are evaluated and decided upon: in AIF, they are evaluated using model comparison, and selected based on their model evidence (in relation to prior and posterior beliefs). The challenge (which remains an open question) lies in defining the alternative explanations (i.e., models, hypotheses) in the first place. One criticism of Structure Learning as BMR in (Erdmann and Mathys 2021) suggests that specifying a priori a set of hypotheses (which is usually done with BMS and BMR) is a strong assumption in general. For example, in Chapter 4, twelve alternative hypotheses were specified a priori, and in Chapter 5, there were twenty alternative hypotheses. These sets were arbitrarily chosen to illustrate Structure Learning, rather than to solve the problems of 'infinite hypothesis space' or 'building models from scratch'. In the first case (i.e., SL as a type of information processing), the empirical question is: "Is there evidence for Structure Learning, as a type of (rapid) learning that goes beyond associative (gradual) Hebbian learning?" – which this thesis addressed. Answering this question is important, particularly because associative (Hebbian) learning does not explain the type of learning observed as a result of sleep, rest, or pauses from soliciting evidence, nor does it account for the rapid learning that involves components not initially included in the original contingencies. In the other cases, the questions on structure learning are: i) "How does the brain solve the problem of infinite alternative explanations for the world?" (Ullman, Goodman et al. 2010) and ii) "How do we grow a model?" (Tenenbaum, Kemp et al. 2011).

I speculate that question i) could be addressed in the future by finding out the ways in which neuronal (or nervous system) architectures constrain what hypotheses can be (momentarily) instantiated or constructed for model comparison based on species-specific and individual-specific features or histories. In other words, this question can be answered by

finding out what types of priors and inductive biases are implied in the processing of information in humans. For instance, one such constraint (relevant to Structure Learning) could be related to the 'synaptic homeostasis hypothesis' (Tononi and Cirelli 2003, Tononi and Cirelli 2006), where synaptic potentiation is accumulated during wakeful moments, followed by synaptic downscaling during offline periods (e.g., sleep). In work based on this hypothesis, it was shown that the induction of local plastic changes (i.e., synaptic potentiation) was associated with local induction of slow wave activity during sleep (itself associated with synaptic homeostasis) (Huber, Felice Ghilardi et al. 2004). The constraint could then be about a *Structure Learning* localised in areas of earlier (online) *Parametric Learning.* As a side note, it is interesting that sleep is a ubiquitous characteristic in the animal kingdom (Cirelli and Tononi 2008). Organisms as small as the *Cassiopea* jellyfish have been observed to experience a primitive equivalent of sleep (Arnold 2017), and even the very simplest forms of life, such as non-photosynthetic bacteria have active and passive phases that correspond to the light-dark cycle (Sartor, Eelderink-Chen et al. 2019), raising the question of whether these organisms are capable of Structure Learning.

Another example constraint (for momentarily instantiated alternative hypotheses) could be related to research on replay: the 'content' being replayed (e.g., the hypotheses or models being compared) is defined in terms of priority, where priority is judged based on experienced events (Liu, Mattar et al. 2021). By answering the question of what is being replayed, one can derive rules (e.g., of priority), that can in turn be applied automatically as transformations to the probability landscape (of the contingencies in question) for generating hypotheses during model comparison and selection.

Furthermore, based on the complete class theorem (Brown 1981, Isomura, Shimazaki et al. 2022), there will always exist a better explanation or model (in this infinite set of alternative models), so it could be that this infinite set is sampled stochastically (but in a way

that is constrained by evolutionary and developmental histories). Addressing question i), however, is a huge task (and remains an open question), but its answers will contribute to the goal of digital twins (i.e., simplified models of physiology, behaviour, or dynamics that characterise specific phenomena in question). This is important because constructing digital twins could, in the future, allow the implementation of some otherwise costly predictions in a non-invasive manner, e.g., testing psychopharmacological predictions with novel treatments.

Question ii) on the other hand, has some answers already, such as in (Tenenbaum, Kemp et al. 2011, Smith, Schwartenbeck et al. 2020, Erdmann and Mathys 2021, Ellis, Wong et al. 2023, Friston, Da Costa et al. 2023). Although question ii) is less relevant for the goal of digital twins – since humans come endowed with priors shaped by evolutionary histories, rather than learning everything 'from scratch' and starting with sparse models that imply few prior constraints (Lake, Ullman et al. 2017) – answering it can nevertheless contribute to the development of Artificial General Intelligence (Ellis, Wong et al. 2023) for example in terms of computational efficiency (i.e., starting with the lowest amount of prior assumptions). However, these formulations still run into the issues presented in question i). That is, even in cases where the set of alternative hypotheses (i.e., models) is built in an ongoing manner, where adding another contingency is being compared with only the parent model (i.e., hypothesis), one could argue that the model comparison should be between the parent model *and* a model with an added contingency *and* other alternative hypotheses that transform the probability landscape of the parent model without adding a contingency. Showing that one can add a variable to a model is, in principle, as trivial as my ability to add another alternative hypothesis to the model (manually). Comparing a model with or without a variable or element technically involves the same computation (i.e., model selection) as comparing two models with alternative probability landscapes. Again, the problem here is not which model to choose (the answer is the one with highest model evidence), but which alternative models to consider. As mentioned,

the problem of 'infinite hypothesis space' remains an open question. But if the goal is to create digital twins, then the challenge is not in terms of how Structure Learning works, but in terms of how to characterise cognitive architectures and translate them into priors (and alternative priors).

To return to the criticism in (Erdmann and Mathys 2021), a priori assuming a set of alternative hypotheses to compare and select from has the purpose of characterising learning and behaviour in humans, in a way that aims to be ecologically valid (Kiebel and Friston 2011, Hobson and Friston 2012, Hobson, Gott et al. 2021, Pezzulo, Zorzi et al. 2021). Although in Chapters 4 and 5 the process of model comparison and selection (i.e., SL) was simplified to include only a set number of alternative hypotheses, its formalisation is grounded in first principles employed with BMS and BMR in the AIF landscape more generally. Furthermore, the learning via grammar-based rule induction employed in (Erdmann and Mathys 2021) is task specific, whereas SL (as implemented with BMS, BMR, BMR) is not. Characterising a different task using the grammar-based rule induction will involve a different learning process (comprised of grammar-based logic rules defined for that specific task). In other words, specifying a priori a set of logic-based grammar rules may be as strong an assumption as specifying a priori a set of alternative models. Furthermore, deciding which combination of logic-based rules to implement given a paradigm runs again into the problems of question i). The belief-update schemes for learning in the AIF do not differ depending on the content; what differs are the generative models constructed to illustrate specific behaviours. When constructing digital twins, the objective isn't necessarily to make the learning (e.g., of rules) ever faster – as in (Erdmann and Mathys 2021), but rather to characterise computational processes that mimic human-level cognition. SL as implemented with AIF is a promising framework in this sense because of its implied aims for ecological validity. Since SL (e.g., BMR) proceeds on architectures underlined by a synaptic connectivity, future work could

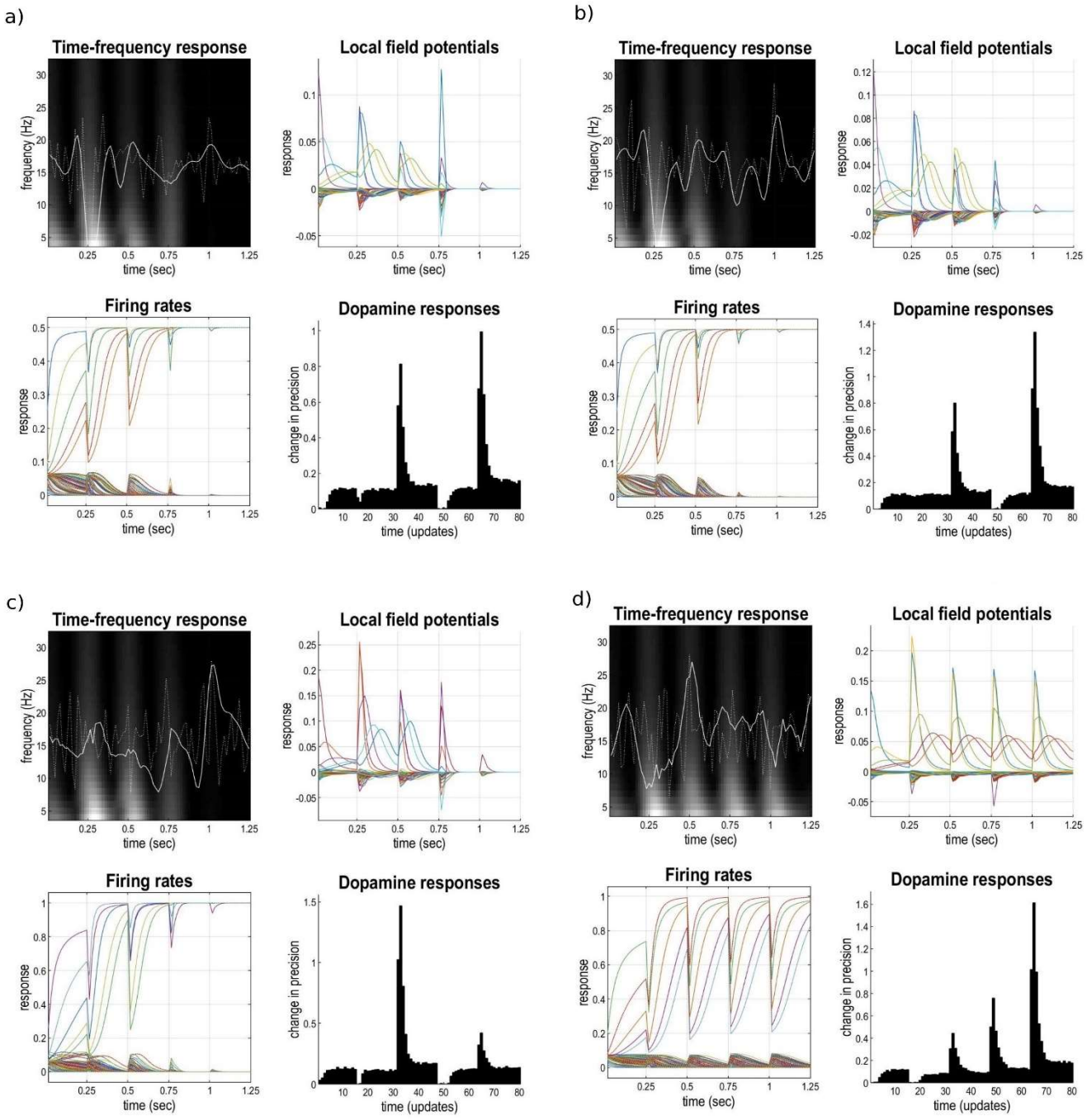
involve assessing focused (e.g., effective, or functional) connectivity before and after the learning of rules in the paradigm from Chapter 5.

Since both (Erdmann and Mathys 2021), and deciding which alternative models to compare with Structure Learning in the AIF, run into the same issues related to question i), one possible solution could be implemented by applying grammar-based rules as in (Larranaga, Kuijpers et al. 1996, Larranaga, Poza et al. 1996, Ji, Wei et al. 2013, Erdmann and Mathys 2021, Kitson, Constantinou et al. 2023) as transformations to the probability landscape, in order to generate alternative hypotheses. In other words, one could start with a given (AIF) generative model that describes contingencies of interest in a good-enough manner, whose state-space can expand, contract, or transform, based on comparing and selecting from alternative hypotheses (i.e., models) generated from the original probability landscape using transformations such as mutation, pruning, as well as conjunction, disjunction, etc. Furthermore, one can implement topology-based constraints such as the local potentiation (and its implied localised synaptic downscaling) of synaptic homeostasis mentioned in (Tononi and Cirelli 2006). The idea of implementing this potential solution to questions i) and ii) places Structure Learning not in direct competition with alternative frameworks, but rather, promotes an enhanced description of the process of Structure Learning using mechanisms derived from these frameworks.

Other open questions relate to the neural implementation of Structure Learning: how are these alternative hypotheses (i.e., explanations, models) momentarily instantiated for comparison and selection? How are they represented in terms of neural activity and activation flows in a way that connects the necessary elements? Furthermore, are some hypotheses more likely a priori? And if so which ones, and how did that come to be? How do these depend on an individual's cognitive milieu? Resolving these questions can also provide some answers to question i) mentioned earlier – in this context, I suspect that the neural implementation itself (of SL) acts as a constraint to the set of hypotheses considered. In other words, I suspect that

topology and energy requirements provide constraints for the set of alternative models considered. For instance, (Kiebel and Friston 2011), explored a potential neural implementation for Structure Learning using BMS in the context of synaptic homeostasis, offering an account of how neurons self-organise and selectively sample potential presynaptic inputs. In this account, the BMS scheme re-simulates (i.e., reinverts) the experiences under different alternative models. Future work could consider applying a similar approach to SL as implemented with BMR or BME, with the purpose of illuminating what alternative hypotheses are considered during Structure Learning, based on constraints of topology.

Appendix



Neural activity comparison for four further rooms. Panels a) and b) in the figure show Room 15 with the agent adopting the same trajectory (locations 7, 11, 10, 14, 14) at 2 different instances: a) block 28 and b) block 30. Neural activity appears to be similar, as anticipated. Panels c) and d) compare different rooms with the same trajectory (locations 7, 11, 10, 14, and 14), who's neurophysiological activity also differs in spite of having a similar trajectory. Panel c) depicts simulated electrophysiological activity for Room 4 and panel d) shows activity for Room 12.

Publications not included in this thesis:

L Da Costa, T Parr, N Sajid, S Veselic, **V Neacsu**, K Friston (2020) Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology.*

R Worden, M Bennett, **V Neacsu** (2021) The thalamus as a blackboard for perception and planning. *Frontiers in Behavioural Neuroscience.*

K Friston, L Da Costa, A Tschantz, A Kiefer, T Salvatori, **V Neacsu**, …, C Buckley (2023) Supervised structure learning. *arXiv preprint*.

W So, K Friston, **V Neacsu** (submitted 2024 - revisions) The inherent normativity of concepts. *Minds and Machines.*

Note – software used throughout the thesis for analysis and visualisation: Matlab (https://uk.mathworks.com/products/matlab.html), InkScape (https://inkscape.org/), and SPSS (https://www.ibm.com/products/spss-statistics).

References

Albus, J. (2008). "Toward a computational theory of mind." Journal of Mind Theory **1**(1): 1-38.

Andersen, R. A. (1995). "Encoding of intention and spatial location in the posterior parietal cortex." Cerebral Cortex **5**(5): 457-469.

Andrieu, C., N. De Freitas, A. Doucet and M. I. Jordan (2003). "An introduction to MCMC for machine learning." Machine learning **50**(1): 5-43.

Anselme, P. and O. Güntürkün (2019). "How foraging works: uncertainty magnifies food-seeking motivation." Behavioral and Brain Sciences **42**.

Arnold, C. (2017). "Jellyfish caught snoozing give clues to origin of sleep." Nature.

Auersperg, A. M., A. Kacelnik and A. M. von Bayern (2013). "Explorative learning and functional inferences on a five-step means-means-end problem in Goffin's cockatoos (Cacatua goffini)." PloS one **8**(7): e68979.

Baek, S., S. Jaffe-Dax and L. L. Emberson (2020). Chapter 8 - How an infant's active response to structured experience supports perceptual-cognitive development. Progress in Brain Research. S. Hunnius and M. Meyer, Elsevier. **254:** 167-186.

Baranes, A. and P.-Y. Oudeyer (2013). "Active learning of inverse models with intrinsically motivated goal exploration in robots." Robotics and Autonomous Systems **61**(1): 49-73.

Barron, H. C., R. Auksztulewicz and K. Friston (2020). "Prediction and memory: A predictive coding account." Progress in neurobiology **192**: 101821.

Barry, C. and N. Burgess (2014). "Neural mechanisms of self-location." Current Biology **24**(8): R330-R339.

Barsalou, L. W. (1983). "Ad hoc categories." Memory & Cognition **11**(3): 211-227.

Barsalou, L. W. (2009). "Simulation, situated conceptualization, and prediction." Philosophical transactions of The Royal Society B: biological sciences **364**(1521): 1281-1289.

Barto, A., M. Mirolli and G. Baldassarre (2013). "Novelty or surprise?" Frontiers in psychology **4**: 907.

Behrens, T. E., T. H. Muller, J. C. Whittington, S. Mark, A. B. Baram, K. L. Stachenfeld and Z. Kurth-Nelson (2018). "What is a cognitive map? Organizing knowledge for flexible behavior." Neuron **100**(2): 490-509.

Blass, J. A. and K. D. Forbus (2016). Modeling Commonsense Reasoning via Analogical Chaining: A Preliminary Report. CogSci.

Blei, D. M., T. L. Griffiths, M. I. Jordan and J. B. Tenenbaum (2003). Hierarchical topic models and the nested Chinese restaurant process. NIPS.

Blischke, K. and D. Erlacher (2007). "How sleep enhances motor learning-a review." Journal of human kinetics **17**: 3.

Bowman, C. R. and D. Zeithamova (2018). "Abstract Memory Representations in the Ventromedial Prefrontal Cortex and Hippocampus Support Concept Generalization." The Journal of Neuroscience **38**(10): 2605-2614.

Bowman, C. R. and D. Zeithamova (2020). "Training set coherence and set size effects on concept generalization and recognition." J Exp Psychol Learn Mem Cogn **46**(8): 1442-1464.

Bragin, A., G. Jando, Z. Nadasdy, J. Hetke, K. Wise and G. Buzsaki (1995). "Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat." J Neurosci **15**(1 Pt 1): 47-60.

Braun, D. A., A. Aertsen, D. M. Wolpert and C. Mehring (2009). "Learning optimal adaptation strategies in unpredictable motor tasks." Journal of Neuroscience **29**(20): 6472-6478.

Braun, D. A., A. Aertsen, D. M. Wolpert and C. Mehring (2009). "Motor task variation induces structural learning." Curr Biol **19**(4): 352-357.

Braun, D. A., C. Mehring and D. M. Wolpert (2010). "Structure learning in action." Behavioural brain research **206**(2): 157-165.

Brown, L. D. (1981). "A complete class theorem for statistical problems with finite sample spaces." The Annals of Statistics: 1289-1300.

Brown, T. H., Y. Zhao and V. Leung (2009). Hebbian Plasticity. Encyclopedia of Neuroscience. L. R. Squire. Oxford, Academic Press**:** 1049-1056.

Bruner, J. S., J. J. Goodnow and G. A. Austin (1956). A study of thinking. Oxford, England, John Wiley and Sons.

Buzsáki, G. (2015). "Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning." Hippocampus **25**(10): 1073-1188.

Buzsaki, G. and E. I. Moser (2013). "Memory, navigation and theta rhythm in the hippocampal-entorhinal system." Nature Neuroscience **16**(2): 130-138.

Calhoun, A. J. and B. Y. Hayden (2015). "The foraging brain." Current Opinion in Behavioral Sciences **5**: 24-31.

Caruana, R. (1997). "Multitask learning." Machine learning **28**(1): 41-75.

Cirelli, C. and G. Tononi (2008). "Is sleep essential?" PLoS biology **6**(8): e216.

Clark, A. (2015). Surfing uncertainty: Prediction, action, and the embodied mind, Oxford University Press.

Clayton, N. S. and N. J. Emery (2007). "The social life of corvids." Current Biology **17**(16): R652.

Colby, C. L. and J.-R. Duhamel (1996). "Spatial representations for action in parietal cortex." Cognitive Brain Research **5**(1-2): 105-115.

Collins, A. G. and M. J. Frank (2013). "Cognitive control over learning: creating, clustering, and generalizing task-set structure." Psychological review **120**(1): 190.

Combs, K., H. Lu and T. J. Bihl (2023). "Transfer Learning and Analogical Inference: A Critical Comparison of Algorithms, Methods, and Applications." Algorithms **16**(3): 146.

Conant, R. C. and W. R. Ashby (1970). "Every Good Regulator of a system must be a model of that system." Int. J. Systems Sci. **1**(2): 89-97.

Constant, A., M. J. D. Ramstead, S. P. L. Veissière and K. Friston (2019). "Regimes of Expectations: An Active Inference Model of Social Conformity and Human Decision Making." Frontiers in Psychology **10**(679).

Constantino, S. M. and N. D. Daw (2015). "Learning the opportunity cost of time in a patch-foraging task." Cogn Affect Behav Neurosci **15**(4): 837-853.

Conway, C. M. (2020). "How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning." Neuroscience & Biobehavioral Reviews **112**: 279-299.

Corcoran, A. W., G. Pezzulo and J. Hohwy (2020). "From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition." Biology & Philosophy **35**(3): 32.

Cox, J. and I. B. Witten (2019). "Striatal circuits for reward learning and decision-making." Nature Reviews Neuroscience **20**(8): 482-494.

Cristol, D. A. and P. V. Switzer (1999). "Avian prey-dropping behavior. II. American crows and walnuts." Behavioral Ecology **10**(3): 220-226.

Da Costa, L., K. Friston, C. Heins and G. A. Pavliotis (2021). "Bayesian mechanics for stationary processes." Proc Math Phys Eng Sci **477**(2256): 20210518.

Da Costa, L., T. Parr, N. Sajid, S. Veselic, V. Neacsu and K. Friston (2020). "Active inference on discrete state-spaces: A synthesis." Journal of Mathematical Psychology **99**: 102447.

Davidson, J. D. and A. El Hady (2019). "Foraging as an evidence accumulation process." PLoS computational biology **15**(7): e1007060.

Dayan, P. and C. J. Watkins (2002). "Reinforcement learning." Stevens' handbook of experimental psychology **3**: 103-129.

Dean, T. and K. Kanazawa (1989). "A model for reasoning about persistence and causation." Computational intelligence **5**(2): 142-150.

Dehaene, S., F. Meyniel, C. Wacongne, L. Wang and C. Pallier (2015). "The Neural Representation of Sequences: From Transition Probabilities to Algebraic Patterns and Linguistic Trees." Neuron **88**(1): 2-19.

Deuker, L., J. Olligs, J. Fell, T. A. Kranz, F. Mormann, C. Montag, M. Reuter, C. E. Elger and N. Axmacher (2013). "Memory consolidation by replay of stimulus-specific neural activity." Journal of Neuroscience **33**(49): 19373-19383.

Di Rienzo, F., U. Debarnot, S. Daligault, E. Saruco, C. Delpuech, J. Doyon, C. Collet and A. Guillot (2016). "Online and offline performance gains following motor imagery practice: a comprehensive review of behavioral and neuroimaging studies." Frontiers in human neuroscience **10**: 315.

Donchin, O., J. T. Francis and R. Shadmehr (2003). "Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control." Journal of Neuroscience **23**(27): 9032-9045.

Doya, K. (2002). "Metalearning and neuromodulation." Neural Netw. **15**(4-6): 495-506.

Doya, K. (2008). "Modulators of decision making." Nat Neurosci. **11**(4): 410-416.

Doyon, J. and H. Benali (2005). "Reorganization and plasticity in the adult brain during learning of motor skills." Current opinion in neurobiology **15**(2): 161-167.

Drton, M. and M. H. Maathuis (2017). "Structure learning in graphical modeling." Annual Review of Statistics and Its Application **4**: 365-393.

Efron, B. and R. Thisted (1976). "Estimating the Number of Unseen Species: How Many Words Did Shakespeare Know?" Biometrika **63**(3): 435-447.

Eichenbaum, H. (2017). "Prefrontal–hippocampal interactions in episodic memory." Nature Reviews Neuroscience **18**(9): 547-558.

Ellis, K., L. Wong, M. Nye, M. Sable-Meyer, L. Cary, L. Anaya Pozo, L. Hewitt, A. Solar-Lezama and J. B. Tenenbaum (2023). "DreamCoder: growing generalizable, interpretable knowledge with wake–sleep Bayesian program learning." Philosophical Transactions of the Royal Society A **381**(2251): 20220050.

Emberson, L. L., J. B. Misyak, J. A. Schwade, M. H. Christiansen and M. H. Goldstein (2019). "Comparing statistical learning across perceptual modalities in infancy: An investigation of underlying learning mechanism (s)." Developmental science **22**(6): e12847.

Emberson, L. L., J. E. Richards and R. N. Aslin (2015). "Top-down modulation in the infant brain: Learning-induced expectations rapidly affect the sensory cortex at 6 months." Proceedings of the National Academy of Sciences **112**(31): 9585-9590.

Erdmann, T. and C. Mathys (2021). Rule Learning Through Active Inductive Inference. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer.

Evans, T. and N. Burgess (2019). "Coordinated hippocampal-entorhinal replay as structural inference." Advances in Neural Information Processing Systems **32**.

Eysenbach, B., R. R. Salakhutdinov and S. Levine (2019). "Search on the replay buffer: Bridging planning and reinforcement learning." Advances in Neural Information Processing Systems **32**.

Finlayson, N. J., V. Neacsu and D. S. Schwarzkopf (2020). "Spatial heterogeneity in bistable figure-ground perception." i-Perception **11**(5): 2041669520961120.

Finley, T. and T. Joachims (2008). Training structural SVMs when exact inference is intractable. Proceedings of the 25th international conference on Machine learning.

Fiorillo, C. D., P. N. Tobler and W. Schultz (2003). "Discrete coding of reward probability and uncertainty by dopamine neurons." Science **299**(5614): 1898-1902.

Fishburn, P. C. (1970). Utility theory for decision making, Research analysis corp McLean VA.

Fleck, J. I. and R. W. Weisberg (2013). "Insight versus analysis: Evidence for diverse methods in problem solving." Journal of Cognitive Psychology **25**(4): 436-463.

Fourment, M., A. F. Magee, C. Whidden, A. Bilge, F. A. Matsen IV and V. N. Minin (2020). "19 dubious ways to compute the marginal likelihood of a phylogenetic tree topology." Systematic biology **69**(2): 209-220.

Friedman, N. and D. Koller (2003). "Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks." Machine Learning **50**(1-2): 95-125.

Friston, K. (2010). "The free-energy principle: a unified brain theory?" Nature reviews neuroscience **11**(2): 127-138.

Friston, K. and G. Buzsáki (2016). "The Functional Anatomy of Time: What and When in the Brain." Trends Cogn Sci **20**(7): 500-511.

Friston, K., T. FitzGerald, F. Rigoli, P. Schwartenbeck, O. D. J and G. Pezzulo (2016). "Active inference and learning." Neurosci Biobehav Rev **68**: 862-879.

Friston, K., T. FitzGerald, F. Rigoli, P. Schwartenbeck and G. Pezzulo (2017). "Active Inference: A Process Theory." Neural Computation **29**(1): 1-49.

Friston, K., J. Mattout and J. Kilner (2011). "Action understanding and active inference." Biological cybernetics **104**: 137-160.

Friston, K., J. Mattout, N. Trujillo-Barreto, J. Ashburner and W. Penny (2007). "Variational free energy and the Laplace approximation." NeuroImage **34**(1): 220-234.

Friston, K., R. J. Moran, Y. Nagai, T. Taniguchi, H. Gomi and J. Tenenbaum (2021). "World model learning and inference." Neural Networks **144**: 573-590.

Friston, K. and W. Penny (2011). "Post hoc Bayesian model selection." Neuroimage **56**(4): 2089-2099.

Friston, K., F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald and G. Pezzulo (2015). "Active inference and epistemic value." Cognitive neuroscience **6**(4): 187-214.

Friston, K., P. Schwartenbeck, T. Fitzgerald, M. Moutoussis, T. Behrens and R. Dolan (2013). "The anatomy of choice: active inference and agency." Frontiers in Human Neuroscience **7**(598).

Friston, K., P. Schwartenbeck, T. FitzGerald, M. Moutoussis, T. Behrens and R. J. Dolan (2014). "The anatomy of choice: dopamine and decision-making." Philosophical Transactions of the Royal Society B: Biological Sciences **369**(1655): 20130481.

Friston, K. J., L. Da Costa, A. Tschantz, A. Kiefer, T. Salvatori, V. Neacsu, M. Koudahl, C. Heins, N. Sajid and D. Markovic (2023). "Supervised structure learning." arXiv preprint arXiv:2311.10300.

Friston, K. J., G. Flandin and A. Razi (2022). "Dynamic causal modelling of COVID-19 and its mitigations." Sci Rep **12**(1): 12419.

Friston, K. J., M. Lin, C. D. Frith, G. Pezzulo, J. A. Hobson and S. Ondobaka (2017). "Active Inference, Curiosity and Insight." Neural Computation **29**(10): 2633-2683.

Friston, K. J., T. Parr and B. de Vries (2017). "The graphical brain: Belief propagation and active inference." Netw Neurosci **1**(4): 381-414.

Friston, K. J., T. Parr and P. Zeidman (2018). "Bayesian model reduction." arXiv: Methodology.

Friston, K. J., R. Rosch, T. Parr, C. Price and H. Bowman (2017). "Deep temporal models and active inference." Neuroscience & Biobehavioral Reviews **77**: 388-402.

Gabay, A. S. and M. A. J. Apps (2020). "Foraging optimally in social neuroscience: computations and methodological considerations." Social Cognitive and Affective Neuroscience **16**(8): 782-794.

Geeraerts, D. (2006). "Prototype theory." Cognitive linguistics: Basic readings **34**: 141-165.

George, D. and J. Hawkins (2009). "Towards a mathematical theory of cortical micro-circuits." PLoS Comput Biol **5**(10): e1000532.

Gershman, S. J. (2015). "A unifying probabilistic view of associative learning." PLoS computational biology **11**(11): e1004567.

Gershman, S. J. (2017). "Dopamine, Inference, and Uncertainty." Neural Computation **29**(12): 3311-3326.

Gershman, S. J. and D. M. Blei (2012). "A tutorial on Bayesian nonparametric models." Journal of Mathematical Psychology **56**(1): 1-12.

Ghahramani, Z. (1997). Learning dynamic Bayesian networks. International School on Neural Networks, Initiated by IIASS and EMFCSC, Springer.

Gheorghe, M., M. Holcombe and P. Kefalas (2001). "Computational models of collective foraging." Biosystems **61**(2): 133-141.

Ghilardi, T., M. Meyer and S. Hunnius (2023). "Predictive motor activation: Modulated by expectancy or predictability?" Cognition **231**: 105324.

Goldvarg, E. and P. N. Johnson-Laird (2001). "Naive causality: A mental model theory of causal meaning and reasoning." Cognitive science **25**(4): 565-610.

Goodman, N. D., J. B. Tenenbaum, J. Feldman and T. L. Griffiths (2008). "A rational analysis of rule-based concept learning." Cogn Sci **32**(1): 108-154.

Goodman, N. D., J. B. Tenenbaum and T. Gerstenberg (2014). Concepts in a probabilistic language of thought, Center for Brains, Minds and Machines (CBMM).

Gopnik, A. (2011). "The Theory Theory 2.0: Probabilistic Models and Cognitive Development." Child Development Perspectives **5**(3): 161-163.

Gopnik, A., C. Glymour, D. M. Sobel, L. E. Schulz, T. Kushnir and D. Danks (2004). "A theory of causal learning in children: causal maps and Bayes nets." Psychological review **111**(1): 3.

Gopnik, A., L. Schulz and L. E. Schulz (2007). Causal learning: Psychology, philosophy, and computation, Oxford University Press.

Griffiths, T. L., A. N. Sanborn, K. R. Canini, D. J. Navarro and J. B. Tenenbaum (2011). "Nonparametric Bayesian models of categorization." Formal approaches in categorization: 173-198.

Griffiths, T. L. and J. B. Tenenbaum (2006). "Optimal Predictions in Everyday Cognition." Psychological Science **17**(9): 767-773.

Gruber, M. J., L.-T. Hsieh, B. P. Staresina, C. E. Elger, J. Fell, N. Axmacher and C. Ranganath (2018). "Theta phase synchronization between the human hippocampus and prefrontal cortex increases during encoding of unexpected information: a case study." Journal of Cognitive Neuroscience **30**(11): 1646-1656.

Gruber, M. J., M. Ritchey, S.-F. Wang, M. K. Doss and C. Ranganath (2016). "Post-learning hippocampal dynamics promote preferential retention of rewarding events." Neuron **89**(5): 1110-1120.

Grupe, D. W. and J. B. Nitschke (2013). "Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective." Nature Reviews Neuroscience **14**(7): 488-501.

Gupta, A., R. Mendonca, Y. Liu, P. Abbeel and S. Levine (2018). "Meta-reinforcement learning of structured exploration strategies." Advances in neural information processing systems **31**.

Gutiérrez, E. D. and J. L. Cabrera (2015). "A neural coding scheme reproducing foraging trajectories." Scientific Reports **5**(1): 18009.

Hafting, T., M. Fyhn, S. Molden, M.-B. Moser and E. I. Moser (2005). "Microstructure of a spatial map in the entorhinal cortex." Nature **436**(7052): 801-806.

Hall-McMaster, S. and F. Luyckx (2019). "Revisiting foraging approaches in neuroscience." Cognitive, Affective, & Behavioral Neuroscience **19**(2): 225-230.

Halpern, J. Y. (2016). Actual causality, MiT Press.

Harlow, H. F. (1949). "The formation of learning sets." Psychological review **56**(1): 51.

Harrison, W. J., P. M. Bays and R. Rideaux (2023). "Neural tuning instantiates prior expectations in the human visual system." Nature Communications **14**(1): 5320.

Hasson, U. (2017). "The neurobiology of uncertainty: implications for statistical learning." Philosophical Transactions of the Royal Society B: Biological Sciences **372**(1711): 20160048.

Haxby, J. V., C. L. Grady, B. Horwitz, L. G. Ungerleider, M. Mishkin, R. E. Carson, P. Herscovitch, M. B. Schapiro and S. I. Rapoport (1991). "Dissociation of object and spatial visual processing pathways in human extrastriate cortex." Proceedings of the National Academy of Sciences **88**(5): 1621-1625.

Haxby, J. V., B. Horwitz, L. G. Ungerleider, J. M. Maisog, P. Pietrini and C. L. Grady (1994). "The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations." Journal of neuroscience **14**(11): 6336-6353.

Hayden, B. Y. (2018). "Economic choice: the foraging perspective." Current Opinion in Behavioral Sciences **24**: 1-6.

Hayden, B. Y. and M. E. Walton (2014). "Neuroscience of foraging." Frontiers in Neuroscience **8**(81).

Hesp, C., R. Smith, T. Parr, M. Allen, K. J. Friston and M. J. D. Ramstead (2021). "Deeply Felt Affect: The Emergence of Valence in Deep Active Inference." Neural Comput **33**(2): 398-446.

Hill, S., G. Tononi and M. F. Ghilardi (2008). "Sleep improves the variability of motor performance." Brain research bulletin **76**(6): 605-611.

Hills, T. T., M. N. Jones and P. M. Todd (2012). "Optimal foraging in semantic memory." Psychological review **119**(2): 431.

Hinton, G. E., P. Dayan, B. J. Frey and R. M. Neal (1995). "The" wake-sleep" algorithm for unsupervised neural networks." Science **268**(5214): 1158-1161.

Hobson, J. A. and K. J. Friston (2012). "Waking and dreaming consciousness: neurobiological and functional considerations." Progress in neurobiology **98**(1): 82-98.

Hobson, J. A., J. A. Gott and K. J. Friston (2021). "Minds and brains, sleep and psychiatry." Psychiatric Research and Clinical Practice **3**(1): 12-28.

Hohwy, J. (2016). "The Self-Evidencing Brain." Noûs **50**(2): 259-285.

Hu, S., Y. Ma, X. Liu, Y. Wei and S. Bai (2021). Stratified rule-aware network for abstract visual reasoning. Proceedings of the AAAI Conference on Artificial Intelligence.

Huber, R., M. Felice Ghilardi, M. Massimini and G. Tononi (2004). "Local sleep and learning." Nature **430**(6995): 78-81.

Huisman, M., J. N. Van Rijn and A. Plaat (2021). "A survey of deep meta-learning." Artificial Intelligence Review **54**(6): 4483-4541.

Humphries, M. D. and T. J. Prescott (2010). "The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward." Progress in Neurobiology **90**(4): 385-417.

Hutter, S. A. and A. I. Wilson (2018). "A Novel Role for the Hippocampus in Category Learning." The Journal of neuroscience : the official journal of the Society for Neuroscience **38**(31): 6803-6805.

Isomura, T. (2022). "Active inference leads to Bayesian neurophysiology." Neuroscience Research **175**: 38-45.

Isomura, T., H. Shimazaki and K. J. Friston (2022). "Canonical neural networks perform active inference." Communications Biology **5**(1): 55.

Itti, L. and C. Koch (2000). "A saliency-based search mechanism for overt and covert shifts of visual attention." Vision research **40**(10-12): 1489-1506.

Ji, J., H. Wei and C. Liu (2013). "An artificial bee colony algorithm for learning Bayesian networks." Soft Computing **17**: 983-994.

Jo, Y. S., G. Heymann and L. S. Zweifel (2018). "Dopamine neurons reflect the uncertainty in fear generalization." Neuron **100**(4): 916-925. e913.

Jordan, M. I. (1998). Learning in graphical models, Springer Science & Business Media.

Jung-Beeman, M., E. M. Bowden, J. Haberman, J. L. Frymiare, S. Arambel-Liu, R. Greenblatt, P. J. Reber and J. Kounios (2004). "Neural activity when people solve verbal problems with insight." PLoS biology **2**(4): e97.

Jung, M. W., Y. Qin, B. L. McNaughton and C. A. Barnes (1998). "Firing characteristics of deep layer neurons in prefrontal cortex in rats performing spatial working memory tasks." Cerebral cortex (New York, NY: 1991) **8**(5): 437-450.

Kaelbling, L. P., M. L. Littman and A. W. Moore (1996). "Reinforcement learning: A survey." Journal of artificial intelligence research **4**: 237-285.

Kagan, B. J., A. C. Kitchen, N. T. Tran, F. Habibollahi, M. Khajehnejad, B. J. Parker, A. Bhat, B. Rollo, A. Razi and K. J. Friston (2022). "In vitro neurons learn and exhibit sentience when embodied in a simulated game-world." Neuron **110**(23): 3952-3969. e3958.

Kaplan, R. and K. J. Friston (2018). "Planning and navigation as active inference." Biological Cybernetics **112**(4): 323-343.

Karuza, E. A., L. L. Emberson, M. E. Roser, D. Cole, R. N. Aslin and J. Fiser (2017). "Neural Signatures of Spatial Statistical Learning: Characterizing the Extraction of Structure from Complex Visual Scenes." J Cogn Neurosci **29**(12): 1963-1976.

Kass, R. E. and A. E. Raftery (1995). "Bayes Factors." Journal of the American Statistical Association **90**(430): 773-795.

Kayser, A. S. and M. D'Esposito (2012). "Abstract Rule Learning: The Differential Effects of Lesions in Frontal Cortex." Cerebral Cortex **23**(1): 230-240.

Kemp, C. and J. B. Tenenbaum (2009). "Structured statistical models of inductive reasoning." Psychological review **116**(1): 20.

Kemp, C., J. B. Tenenbaum, S. Niyogi and T. L. Griffiths (2010). "A probabilistic model of theory formation." Cognition **114**(2): 165-196.

Kerster, B. E., T. Rhodes and C. T. Kello (2016). "Spatial memory in foraging games." Cognition **148**: 85-96.

Kiebel, S. J. and K. J. Friston (2011). "Free energy and dendritic self-organization." Frontiers in systems neuroscience **5**: 80.

Kitson, N. K., A. C. Constantinou, Z. Guo, Y. Liu and K. Chobtham (2023). "A survey of Bayesian Network structure learning." Artificial Intelligence Review: 1-94.

Klein, G. and A. Jarosz (2011). "A naturalistic study of insight." Journal of Cognitive Engineering and Decision Making **5**(4): 335-351.

Kolling, N., T. E. Behrens, R. B. Mars and M. F. Rushworth (2012). "Neural mechanisms of foraging." Science **336**(6077): 95-98.

Kounios, J., J. I. Fleck, D. L. Green, L. Payne, J. L. Stevenson, E. M. Bowden and M. Jung-Beeman (2008). "The origins of insight in resting-state brain activity." Neuropsychologia **46**(1): 281-291.

Kounios, J., J. L. Frymiare, E. M. Bowden, J. I. Fleck, K. Subramaniam, T. B. Parrish and M. Jung-Beeman (2006). "The Prepared Mind:Neural Activity Prior to Problem Presentation Predicts Subsequent Solution by Sudden Insight." Psychological Science **17**(10): 882-890.

Kurth-Nelson, Z., T. Behrens, G. Wayne, K. Miller, L. Luettgau, R. Dolan, Y. Liu and P. Schwartenbeck (2023). "Replay and compositional computation." Neuron **111**(4): 454-469.

Lagnado, D. A., T. Gerstenberg and R. i. Zultan (2013). "Causal responsibility and counterfactuals." Cognitive science **37**(6): 1036-1073.

Lagnado, D. A. and S. Sloman (2004). "The advantage of timely intervention." J Exp Psychol Learn Mem Cogn **30**(4): 856-876.

Lagnado, D. A. and S. A. Sloman (2006). "Time as a guide to cause." J Exp Psychol Learn Mem Cogn **32**(3): 451-460.

Lake, B. M., R. Salakhutdinov and J. B. Tenenbaum (2015). "Human-level concept learning through probabilistic program induction." Science **350**(6266): 1332-1338.

Lake, B. M., T. D. Ullman, J. B. Tenenbaum and S. J. Gershman (2017). "Building machines that learn and think like people." Behavioral and brain sciences **40**: e253.

Larranaga, P., H. Karshenas, C. Bielza and R. Santana (2013). "A review on evolutionary algorithms in Bayesian network learning and inference tasks." Information Sciences **233**: 109-125.

Larranaga, P., C. M. Kuijpers, R. H. Murga and Y. Yurramendi (1996). "Learning Bayesian network structures by searching for the best ordering with genetic algorithms." IEEE

transactions on systems, man, and cybernetics-part A: systems and humans **26**(4): 487-493.

Larranaga, P., M. Poza, Y. Yurramendi, R. H. Murga and C. M. H. Kuijpers (1996). "Structure learning of Bayesian networks by genetic algorithms: A performance analysis of control parameters." IEEE transactions on pattern analysis and machine intelligence **18**(9): 912-926.

Laurent, V. and Bernard W. Balleine (2015). "Factual and Counterfactual Action-Outcome Mappings Control Choice between Goal-Directed Actions in Rats." Current Biology **25**(8): 1074-1079.

Le Heron, C., N. Kolling, O. Plant, A. Kienast, R. Janska, Y.-S. Ang, S. Fallon, M. Husain and M. A. J. Apps (2020). "Dopamine Modulates Dynamic Decision-Making during Foraging." The Journal of Neuroscience **40**(27): 5273-5282.

Lee, C. and P. van Beek (2017). Metaheuristics for score-and-search Bayesian network structure learning. Advances in Artificial Intelligence: 30th Canadian Conference on Artificial Intelligence, Canadian AI 2017, Edmonton, AB, Canada, May 16-19, 2017, Proceedings 30, Springer.

Lee, H. S., S. Betts and J. R. Anderson (2016). "Learning problem-solving rules as search through a hypothesis space." Cognitive science **40**(5): 1036-1079.

Li, F., W.-Y. Cao, F.-L. Huang, W.-J. Kang, X.-L. Zhong, Z.-L. Hu, H.-T. Wang, J. Zhang, J.-Y. Zhang, R.-P. Dai, X.-F. Zhou and C.-Q. Li (2016). "Roles of NMDA and dopamine in food-foraging decision-making strategies of rats in the social setting." BMC Neuroscience **17**(1): 3.

Lindley, D. V. (1956). "On a Measure of the Information Provided by an Experiment." Ann. Math. Statist. **27**(4): 986-1005.

Lisman, J. and A. D. Redish (2009). "Prediction, sequences and the hippocampus." Philos Trans R Soc Lond B Biol Sci **364**(1521): 1193-1201.

Liu, Y., M. G. Mattar, T. E. Behrens, N. D. Daw and R. J. Dolan (2021). "Experience replay is associated with efficient nonlocal learning." Science **372**(6544): eabf1357.

Lockery, S. R. (2011). "The computational worm: spatial orientation and its neuronal basis in C. elegans." Current Opinion in Neurobiology **21**(5): 782-790.

Love, B. C. and T. M. Gureckis (2007). "Models in search of a brain." Cogn Affect Behav Neurosci **7**(2): 90-108.

Love, B. C., D. L. Medin and T. M. Gureckis (2004). "SUSTAIN: a network model of category learning." Psychol Rev **111**(2): 309-332.

Mack, M. L., B. C. Love and A. R. Preston (2016). "Dynamic updating of hippocampal object representations reflects new conceptual knowledge." Proceedings of the National Academy of Sciences **113**(46): 13203-13208.

Mack, M. L., B. C. Love and A. R. Preston (2018). "Building concepts one episode at a time: The hippocampus and concept formation." Neuroscience letters **680**: 31-38.

Mack, M. L., A. R. Preston and B. C. Love (2020). "Ventromedial prefrontal cortex compression during concept learning." Nature Communications **11**(1): 46.

MacKay, D. J. C. (1992). "Information-Based Objective Functions for Active Data Selection." Neural Computation **4**(4): 590-604.

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

Maisto, D., K. Friston and G. Pezzulo (2019). "Caching mechanisms for habit formation in active inference." Neurocomputing **359**: 298-314.

Matassi, G. and P. Martinez (2023). "The brain-computer analogy—"A special issue"." Frontiers in Ecology and Evolution **10**: 1099253.

McClelland, J. L., B. L. McNaughton and R. C. O'Reilly (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." Psychol Rev **102**(3): 419-457.

McKenzie, S., Andrea J. Frank, Nathaniel R. Kinsky, B. Porter, Pamela D. Rivière and H. Eichenbaum (2014). "Hippocampal Representation of Related and Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas." Neuron **83**(1): 202-215.

McNicholas, P. D. (2016). "Model-based clustering." Journal of Classification **33**(3): 331-373.

Meder, B., Y. Hagmayer and M. Waldmann (2009). "The role of learning data in causal reasoning about observations and interventions." Memory & Cognition **37**: 249-264.

Metcalfe, J. (1986). "Premonitions of insight predict impending error." Journal of experimental psychology: Learning, memory, and cognition **12**(4): 623.

Mihajlovic, V. and M. Petkovic (2001). "Dynamic bayesian networks: A state of the art." University of Twente Document Repository.

Miller, J. F., M. Neufang, A. Solway, A. Brandt, M. Trippel, I. Mader, S. Hefft, M. Merkow, S. M. Polyn and J. Jacobs (2013). "Neural activity in human hippocampal formation reveals the spatial context of retrieved memories." Science **342**(6162): 1111-1114.

Mirza, M. B., R. A. Adams, K. Friston and T. Parr (2019). "Introducing a Bayesian model of selective attention based on active inference." Scientific Reports **9**(1): 13915.

Mirza, M. B., R. A. Adams, C. D. Mathys and K. J. Friston (2016). "Scene Construction, Visual Foraging, and Active Inference." Frontiers in computational neuroscience **10**: 56-56.

Mitchell, M. (2021). "Abstraction and analogy-making in artificial intelligence." Annals of the New York Academy of Sciences **1505**(1): 79-101.

Mizuguchi, N. and K. Kanosue (2017). Chapter 10 - Changes in brain activity during action observation and motor imagery: Their relationship with motor learning. Progress in Brain Research. M. R. Wilson, V. Walsh and B. Parkin, Elsevier. **234:** 189-204.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland and G. Ostrovski (2015). "Human-level control through deep reinforcement learning." nature **518**(7540): 529-533.

Mobbs, D., P. C. Trimmer, D. T. Blumstein and P. Dayan (2018). "Foraging for foundations in decision neuroscience: insights from ethology." Nature Reviews Neuroscience **19**(7): 419-427.

Mok, R. M. and B. C. Love (2019). "A non-spatial account of place and grid cells based on clustering models of concept learning." Nature Communications **10**(1): 5685.

Monroy, C., M. Meyer, S. Gerson and S. Hunnius (2017). "Statistical learning in social action contexts." PloS one **12**(5): e0177261.

Monroy, C. D., S. A. Gerson, E. Domínguez-Martínez, K. Kaduk, S. Hunnius and V. Reid (2019). "Sensitivity to structure in action sequences: An infant event-related potential study." Neuropsychologia **126**: 92-101.

Monroy, C. D., M. Meyer, L. Schröer, S. A. Gerson and S. Hunnius (2019). "The infant motor system predicts actions based on visual statistical learning." NeuroImage **185**: 947-954.

Montague, P. R., P. Dayan, C. Person and T. J. Sejnowski (1995). "Bee foraging in uncertain environments using predictive Hebbian learning." Nature **377**(6551): 725-728.

Mukherjee, A., N. H. Lam, R. D. Wimmer and M. M. Halassa (2021). "Thalamic circuits for independent control of prefrontal signal and noise." Nature.

Mulavara, A. P., H. S. Cohen and J. J. Bloomberg (2009). "Critical features of training that facilitate adaptive generalization of over ground locomotion." Gait & Posture **29**(2): 242-248.

Nádasdy, Z., H. Hirase, A. Czurkó, J. Csicsvari and G. Buzsáki (1999). "Replay and time compression of recurring spike sequences in the hippocampus." Journal of Neuroscience **19**(21): 9497-9507.

Nauta, J., Y. Khaluf and P. Simoens (2020). "Hybrid foraging in patchy environments using spatial memory." Journal of the Royal Society Interface **17**(166): 20200026.

Neacsu, V., L. Convertino and K. J. Friston (2022). "Synthetic Spatial Foraging With Active Inference in a Geocaching Task." Frontiers in Neuroscience **16**.

Neacsu, V., M. B. Mirza, R. A. Adams and K. J. Friston (2022). "Structure learning enhances concept formation in synthetic Active Inference agents." PLoS One **17**(11): e0277199.

Needham, A., T. Barrett and K. Peterman (2002). "A pick-me-up for infants' exploratory skills: Early simulated experiences reaching for objects using 'sticky mittens' enhances young infants' object exploration skills." Infant Behavior and Development **25**(3): 279-295.

Needham, C. J., J. R. Bradford, A. J. Bulpitt and D. R. Westhead (2007). "A primer on learning in Bayesian networks for computational biology." PLoS computational biology **3**(8): e129.

Ngo, H., M. Luciw, A. Forster and J. Schmidhuber (2012). Learning skills from play: artificial curiosity on a katana robot arm. The 2012 international joint conference on neural networks (IJCNN), IEEE.

Nieder, A. (2013). "Coding of abstract quantity by 'number neurons' of the primate brain." Journal of Comparative Physiology A **199**(1): 1-16.

Nihei, Y. and H. Higuchi (2001). "When and where did crows learn to use automobiles as nutcrackers." Tohoku psychologica folia **60**: 93-97.

Niv, Y. (2019). "Learning task-state representations." Nature Neuroscience **22**(10): 1544-1553.

Niv, Y., M. O. Duff and P. Dayan (2005). "Dopamine, uncertainty and TD learning." Behavioral and Brain Functions **1**(1): 6.

Novick, L. R. and P. W. Cheng (2004). "Assessing interactive causal influence." Psychological Review **111**(2): 455.

O'Bryan, S. R., D. A. Worthy, E. J. Livesey and T. Davis (2018). "Model-based fMRI reveals dissimilarity processes underlying base rate neglect." eLife **7**: e36395.

O'Keefe, J. and J. Dostrovsky (1971). "The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat." Brain Research **34**(1): 171-175.

Oudeyer, P.-Y. and A. Baranes (2008). Developmental active learning with intrinsic motivation. iros 2008 workshop: from motor to interaction learning in robots.

Parkhurst, D., K. Law and E. Niebur (2002). "Modeling the role of salience in the allocation of overt visual attention." Vision research **42**(1): 107-123.

Parr, T. and K. J. Friston (2017). "Uncertainty, epistemics and active inference." Journal of the Royal Society Interface **14**(136): 20170376.

Parr, T. and K. J. Friston (2018). "Active inference and the anatomy of oculomotion." Neuropsychologia **111**: 334-343.

Parr, T. and G. Pezzulo (2021). "Understanding, explanation, and active inference." Frontiers in Systems Neuroscience **15**: 772641.

Pearl, J. (2000). "Models, reasoning and inference." Cambridge, UK: CambridgeUniversityPress **19**.

Pearl, J. (2014). Probabilistic reasoning in intelligent systems: networks of plausible inference, Elsevier.

Pearson, J. M., K. K. Watson and M. L. Platt (2014). "Decision making: the neuroethological turn." Neuron **82**(5): 950-965.

Penny, W. D. (2012). "Comparing dynamic causal models using AIC, BIC and free energy." Neuroimage **59**(1): 319-330.

Penny, W. D., K. E. Stephan, A. Mechelli and K. J. Friston (2004). "Comparing dynamic causal models." Neuroimage **22**(3): 1157-1172.

Peters, A., B. S. McEwen and K. Friston (2017). "Uncertainty and stress: Why it causes diseases and how it is mastered by the brain." Progress in Neurobiology **156**: 164-188.

Pezzulo, G., M. Zorzi and M. Corbetta (2021). "The secret life of predictive brains: what's spontaneous activity for?" Trends in Cognitive Sciences.

Pinker, S. and R. Jackendoff (2005). "The faculty of language: what's special about it?" Cognition **95**(2): 201-236.

Poincaré, H. (2022). The foundations of science: Science and hypothesis, the value of science, science and method, DigiCat.

Pouget, A., J. M. Beck, W. J. Ma and P. E. Latham (2013). "Probabilistic brains: knowns and unknowns." Nature Neuroscience **16**(9): 1170-1178.

Reznikova, Z. (2007). Animal intelligence: From individual to social cognition, Cambridge University Press.

Rigoux, L., K. E. Stephan, K. J. Friston and J. Daunizeau (2014). "Bayesian model selection for group studies — Revisited." NeuroImage **84**: 971-985.

Rouder, J. N. and R. Ratcliff (2004). "Comparing categorization models." Journal of experimental psychology. General **133**(1): 63-82.

Rubin, R. D., H. Schwarb, H. D. Lucas, M. R. Dulas and N. J. Cohen (2017). "Dynamic hippocampal and prefrontal contributions to memory processes and representations blur the boundaries of traditional cognitive domains." Brain sciences **7**(7): 82.

Rudebeck, P. H. and A. Izquierdo (2021). "Foraging with the frontal cortex: A cross-species evaluation of reward-guided behavior." Neuropsychopharmacology.

Rudy, J. W., R. M. Barrientos and R. C. O'Reilly (2002). "Hippocampal formation supports conditioning to memory of a context." Behavioral Neuroscience **116**(4): 530-538.

Rutledge, R. B., S. C. Lazzario, B. Lau, C. E. Myers, M. A. Gluck and P. W. Glimcher (2009). "Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task." J Neurosci. **29**: 15104-15114.

Salakhutdinov, R., J. B. Tenenbaum and A. Torralba (2012). "Learning with hierarchical-deep models." IEEE transactions on pattern analysis and machine intelligence **35**(8): 1958-1971.

Salvi, C., E. Bricolo, J. Kounios, E. Bowden and M. Beeman (2016). "Insight solutions are correct more often than analytic solutions." Thinking & reasoning **22**(4): 443-460.

Sartor, F., Z. Eelderink-Chen, B. Aronson, J. Bosman, L. E. Hibbert, A. N. Dodd, Á. T. Kovács and M. Merrow (2019). "Are There Circadian Clocks in Non-Photosynthetic Bacteria?" Biology **8**(2): 41.

Schillaci, G., A. Pico Villalpando, V. V. Hafner, P. Hanappe, D. Colliaux and T. Wintz (2020). "Intrinsic motivation and episodic memories for robot exploration of high-dimensional sensory spaces." Adaptive Behavior: 1059712320922916.

Schmid, D., D. Erlacher, A. Klostermann, R. Kredel and E.-J. Hossner (2020). "Sleep-dependent motor memory consolidation in healthy adults: A meta-analysis." Neuroscience & biobehavioral reviews **118**: 270-281.

Schmidhuber, J. (2006). "Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts." Connection Science **18**(2): 173-187.

Schrier, A. M. (1984). "Learning how to learn: The significance and current status of learning set formation." Primates **25**(1): 95-102.

Schwartenbeck, P., T. H. FitzGerald, C. Mathys, R. Dolan and K. Friston (2015). "The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes." Cereb Cortex **25**(10): 3434-3445.

Schwartenbeck, P. and K. Friston (2016). "Computational Phenotyping in Psychiatry: A Worked Example." eNeuro **3**(4): 0049-0016.2016.

Schwartenbeck, P., J. Passecker, T. U. Hauser, T. H. B. FitzGerald, M. Kronbichler and K. J. Friston (2019). "Computational mechanisms of curiosity and goal-directed exploration." eLife **8**: e41703.

Seamans, J. K., S. B. Floresco and A. G. Phillips (1998). "D1 receptor modulation of hippocampal–prefrontal cortical circuits integrating spatial memory with executive functions in the rat." Journal of neuroscience **18**(4): 1613-1621.

Seed, A. M., J. Call, N. J. Emery and N. S. Clayton (2009). "Chimpanzees solve the trap problem when the confound of tool-use is removed." Journal of Experimental Psychology: Animal Behavior Processes **35**(1): 23.

Seghier, M. L. and K. J. Friston (2013). "Network discovery with large DCMs." Neuroimage **68**: 181-191.

Seth, A. (2014). The cybernetic brain: from interoceptive inference to sensorimotor contingencies. MINDS project. Metzinger, T; Windt, JM, MINDS.

Seth, A. K. and K. J. Friston (2016). "Active interoceptive inference and the emotional brain." Philosophical Transactions of the Royal Society B: Biological Sciences **371**(1708): 20160007.

Seth, A. K. and M. Tsakiris (2018). "Being a beast machine: The somatic basis of selfhood." Trends in cognitive sciences **22**(11): 969-981.

Shadmehr, R. and F. A. Mussa-Ivaldi (1994). "Adaptive representation of dynamics during learning of a motor task." Journal of neuroscience **14**(5): 3208-3224.

Shapiro, A. (2003). "Monte Carlo sampling methods." Handbooks in operations research and management science **10**: 353-425.

Shen, W., Y. Tong, F. Li, Y. Yuan, B. Hommel, C. Liu and J. Luo (2018). "Tracking the neurodynamics of insight: A meta-analysis of neuroimaging studies." Biological Psychology **138**: 189-198.

Shettleworth, S. J. (2010). "Clever animals and killjoy explanations in comparative psychology." Trends in cognitive sciences **14**(11): 477-481.

Shin, H., J. K. Lee, J. Kim and J. Kim (2017). "Continual learning with deep generative replay." Advances in neural information processing systems **30**.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam and M. Lanctot (2016). "Mastering the game of Go with deep neural networks and tree search." nature **529**(7587): 484-489.

Silver, M. A. and S. Kastner (2009). "Topographic maps in human frontal and parietal cortex." Trends in cognitive sciences **13**(11): 488-495.

Skaggs, W. E. and B. L. McNaughton (1996). "Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience." Science **271**(5257): 1870-1873.

Smith, R., K. J. Friston and C. J. Whyte (2022). "A step-by-step tutorial on active inference and its application to empirical data." J Math Psychol **107**.

Smith, R., T. Parr and K. J. Friston (2019). "Simulating emotions: An active inference model of emotional state inference and emotion concept learning." Frontiers in psychology **10**: 2844.

Smith, R., M. J. Ramstead and A. Kiefer (2022). "Active inference models do not contradict folk psychology." Synthese **200**(2): 81.

Smith, R., P. Schwartenbeck, T. Parr and K. J. Friston (2020). "An Active Inference Approach to Modeling Structure Learning: Concept Learning as an Example Case." Frontiers in Computational Neuroscience **14**: 41.

Spiegler, B. J. and M. Mishkin (1981). "Evidence for the sequential participation of inferior temporal cortex and amygdala in the acquisition of stimulus-reward associations." Behavioural Brain Research **3**(3): 303-317.

Stahl, A. E. and L. Feigenson (2015). "Cognitive development. Observing the unexpected enhances infants' learning and exploration." Science (New York, N.Y.) **348**(6230): 91-94.

Stephan, K. E., W. D. Penny, J. Daunizeau, R. J. Moran and K. J. Friston (2009). "Bayesian model selection for group studies." NeuroImage **46**(4): 1004-1017.

Stephens, D. W. (2008). "Decision ecology: Foraging and the ecology of animal decision making." Cognitive, Affective, & Behavioral Neuroscience **8**(4): 475-484.

Steyvers, M., J. B. Tenenbaum, E. J. Wagenmakers and B. Blum (2003). "Inferring causal networks from observations and interventions." Cognitive science **27**(3): 453-489.

Stoianov, I., A. Genovesio and G. Pezzulo (2016). "Prefrontal goal codes emerge as latent states in probabilistic value learning." Journal of Cognitive Neuroscience **28**(1): 140-157.

Stoianov, I., D. Maisto and G. Pezzulo (2022). "The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning." Progress in Neurobiology **217**: 102329.

Sunnåker, M., A. G. Busetto, E. Numminen, J. Corander, M. Foll and C. Dessimoz (2013). "Approximate bayesian computation." PLoS computational biology **9**(1): e1002803.

Tambini, A. and L. Davachi (2019). "Awake reactivation of prior experiences consolidates memories and biases cognition." Trends in cognitive sciences **23**(10): 876-890.

Tenenbaum, J. B. and T. L. Griffiths (2001). "Generalization, similarity, and Bayesian inference." Behavioral and brain sciences **24**(4): 629-640.

Tenenbaum, J. B., T. L. Griffiths and C. Kemp (2006). "Theory-based Bayesian models of inductive learning and reasoning." Trends in Cognitive Sciences **10**(7): 309-318.

Tenenbaum, J. B., C. Kemp, T. L. Griffiths and N. D. Goodman (2011). "How to Grow a Mind: Statistics, Structure, and Abstraction." Science **331**(6022): 1279-1285.

Tik, M., R. Sladky, C. D. B. Luft, D. Willinger, A. Hoffmann, M. J. Banissy, J. Bhattacharya and C. Windischberger (2018). "Ultra-high-field fMRI insights on insight: Neural correlates of the Aha!-moment." Human Brain Mapping **39**(8): 3241-3252.

Todd, P. M. and T. T. Hills (2020). "Foraging in mind." Current Directions in Psychological Science **29**(3): 309-315.

Tolman, E. C. (1948). "Cognitive maps in rats and men." Psychological review **55**(4): 189.

Tolman, E. C. and C. H. Honzik (1930). "Introduction and removal of reward, and maze performance in rats." University of California publications in psychology.

Tolman, E. C., B. F. Ritchie and D. Kalish (1946). "Studies in spatial learning. I. Orientation and the short-cut." Journal of Experimental Psychology **36**(1): 13-24.

Tononi, G. and C. Cirelli (2003). "Sleep and synaptic homeostasis: a hypothesis." Brain research bulletin **62**(2): 143-150.

Tononi, G. and C. Cirelli (2006). "Sleep function and synaptic homeostasis." Sleep Med Rev **10**(1): 49-62.

Toth, A. J., E. McNeill, K. Hayes, A. P. Moran and M. Campbell (2020). "Does mental practice still enhance performance? A 24 Year follow-up and meta-analytic replication and extension." Psychology of Sport and Exercise **48**: 101672.

Tschantz, A., A. K. Seth and C. L. Buckley (2020). "Learning action-oriented models through active inference." PLoS computational biology **16**(4): e1007805.

Tse, D., R. F. Langston, M. Kakeyama, I. Bethus, P. A. Spooner, E. R. Wood, M. P. Witter and R. G. M. Morris (2007). "Schemas and Memory Consolidation." Science **316**(5821): 76-82.

Tu, N. A., T. Huynh-The, K. U. Khan and Y. Lee (2019). "ML-HDP: A Hierarchical Bayesian Nonparametric Model for Recognizing Human Actions in Video." IEEE Transactions on Circuits and Systems for Video Technology **29**(3): 800-814.

Ullman, T., N. Goodman and J. Tenenbaum (2010). Theory acquisition as stochastic search. Proceedings of the Annual Meeting of the Cognitive Science Society.

Van de Ven, G. M., H. T. Siegelmann and A. S. Tolias (2020). "Brain-inspired replay for continual learning with artificial neural networks." Nature communications **11**(1): 4069.

Van Merriënboer, J. J. and A. B. De Bruin (2014). Research paradigms and perspectives on learning. Handbook of research on educational communications and technology, Springer**:** 21-29.

Vapnik, V. N. (1999). "An overview of statistical learning theory." IEEE transactions on neural networks **10**(5): 988-999.

Vul, E., N. Goodman, T. L. Griffiths and J. B. Tenenbaum (2014). "One and done? Optimal decisions from very few samples." Cognitive science **38**(4): 599-637.

Waldmann, M. R. and Y. Hagmayer (2005). "Seeing versus doing: two modes of accessing causal knowledge." J Exp Psychol Learn Mem Cogn **31**(2): 216-227.

Walker, A. R., D. J. Navarro, B. R. Newell and T. Beesley (2021). "Protection from uncertainty in the exploration/exploitation trade-off." Journal of Experimental Psychology: Learning, Memory, and Cognition: No Pagination Specified-No Pagination Specified.

Wang, J. X., Z. Kurth-Nelson, D. Tirumala, H. Soyer, J. Z. Leibo, R. Munos, C. Blundell, D. Kumaran and M. Botvinick (2016). "Learning to reinforcement learn." arXiv preprint arXiv:1611.05763.

Wang, P., R. Kong, X. Kong, R. Liégeois, C. Orban, G. Deco, M. P. Van Den Heuvel and B. T. Yeo (2019). "Inversion of a large-scale circuit model reveals a cortical hierarchy in the dynamic resting human brain." Science advances **5**(1): eaat7854.

Ward, J. F., R. M. Austin and D. W. Macdonald (2000). "A simulation model of foraging behaviour and the effect of predation risk." Journal of Animal Ecology **69**(1): 16-30.

Waskom, M. L. and R. Kiani (2018). "Decision Making through Integration of Sensory Evidence at Prolonged Timescales." Current Biology **28**(23): 3850-3856.e3859.

Webb, M. E., D. R. Little and S. J. Cropper (2016). "Insight is not in the problem: Investigating insight in problem solving across task types." Frontiers in psychology **7**: 1424.

Weidman, N. (1994). "Mental testing and machine intelligence: The Lashley-Hull debate." Journal of the History of the Behavioral Sciences **30**(2): 162-180.

Weisberg, R. W. (2013). "On the “Demystification” of Insight: A Critique of Neuroimaging Studies of Insight." Creativity Research Journal **25**(1): 1-14.

Whiten, A., V. Horner and F. B. M. de Waal (2005). "Conformity to cultural norms of tool use in chimpanzees." Nature **437**(7059): 737-740.

Widloski, J. and D. J. Foster (2022). "Flexible rerouting of hippocampal replay sequences around changing barriers in the absence of global place field remapping." Neuron **110**(9): 1547-1558.e1548.

Wiering, M. A. and M. Van Otterlo (2012). "Reinforcement learning." Adaptation, learning, and optimization **12**(3): 729.

Wilson, A., A. Fern and P. Tadepalli (2012). Transfer learning in sequential decision problems: A hierarchical Bayesian approach. Proceedings of ICML Workshop on Unsupervised and Transfer Learning, JMLR Workshop and Conference Proceedings.

Wilson, M. A. and B. L. McNaughton (1994). "Reactivation of hippocampal ensemble memories during sleep." Science **265**(5172): 676-679.

Wimmer, G. E., Y. Liu, D. C. McNamee and R. J. Dolan (2023). "Distinct replay signatures for prospective decision-making and memory preservation." Proceedings of the National Academy of Sciences **120**(6): e2205211120.

Winn, J. and C. M. Bishop (2005). "Variational message passing." Journal of Machine Learning Research **6**: 661-694.

Winn, J., C. M. Bishop and T. Jaakkola (2005). "Variational message passing." Journal of Machine Learning Research **6**(4).

Wise, T., Y. Liu, F. Chowdhury and R. J. Dolan (2021). "Model-based aversive learning in humans is supported by preferential task state reactivation." Science Advances **7**(31): eabf9616.

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

Zeithamova, D., M. L. Mack, K. Braunlich, T. Davis, C. A. Seger, M. T. R. van Kesteren and A. Wutz (2019). "Brain Mechanisms of Concept Learning." J Neurosci **39**(42): 8259-8266.

Zha, D., K.-H. Lai, K. Zhou and X. Hu (2019). "Experience replay optimization." arXiv preprint arXiv:1906.08387.

Zweig, G. and S. Russell (1998). "Speech recognition with dynamic Bayesian networks."