Decision-making with hierarchical representations in humans

Juan del Ojo Balaguer

A thesis submitted for the degree of

Doctor of Philosophy in Experimental Psychology

Trinity 2018

Wadham College, University of Oxford

Decision-making with hierarchical representations in humans

Juan del Ojo Balaguer

Wadham College Michaelmas Term, 2018

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Experimental Psychology in the University of Oxford.

Abstract

We can make good decisions by capturing and exploiting the structure of the natural world. It is thought that the formation of hierarchical representations allows humans to encode some of this structure. This thesis aims to investigate the behavioural and neural mechanisms underlying hierarchical representations. For this, I make use of diverse behavioural tasks, each with an intrinsic hierarchical structure, to ask if and how hierarchical representations are manifested in behaviour. I further use functional resonance magnetic imaging (fMRI) to pinpoint the brain regions associated with such hierarchical representations.

In the first set of experiments, I studied hierarchical representations in the context of learning the composition of multi-feature visual stimuli that can be recursively grouped by mutual association. In a behavioural experiment, I found that participants learned a set of hierarchical relationships faster if they had previously learned stimuli with the same structure (compared to a baseline, non-hierarchical structure). A subsequent fMRI study revealed the neural concomitants of features associated with each hierarchical level.
In a second imaging study, I studied the role of hierarchical representations during navigation in a virtual subway environment where stations are grouped into lines. I found behavioural costs (reaction time) that increased with the hierarchical description length of the plan to the goal. This was accompanied by neural costs (univariate activity) in a region of interest where I also could decode the current hierarchical context using a multivariate analysis.
In a third imaging study, I modelled behaviour and imaging data in a rule discovery task, where object categorisation was determined by an unknown verbalisable rule. This allowed me to reveal the learning rates associated to different conditions, as well as neural signals related to the uncertainty of the rule.

Finally, I discuss limitations of the studies, and the generality and specificity of the neural mechanisms underlying hierarchical representations.

This work was funded by DeepMind Technologies Limited, and was conducted under the supervision of Prof. Christopher Summerfield.

Table of Contents

Acknowledgements

Sometimes it's a little better to travel than to arrive.

Zen and the Art of Motorcycle Maintenance, Robert Pirsig (1974).

First and foremost, I want to thank Chris Summerfield for being a tremendous supervisor and an even better friend. Thank you for taking me in and letting me grow with your team, and in particular for taking such good care of me all these years. You have always gone out of your way. Your relentless enthusiasm for science is both inspiring and infectious, and I'm truly lucky to have been able to absorb some of the energy, enthusiasm and enjoyment that you show at every step of the scientific journey.

Thank you also to the past and present members of the Summerfield laboratory: for afternoon coffees and hot chocolates, walks next to the canal, birthday cakes, that's-what-she-said jokes, advent calendars, school trips to Paris, and whole days in a basement in Granada. I cannot name each and every one of you but suffice say you know who you are and you're all in my mind. I have learnt a great deal from each of you — please keep up the good work.

I want to give a special thanks to DeepMind for giving me a one-in-a-lifetime experience. I never thought a first job could be so exciting. Thank you for supporting me financially and intellectually throughout my DPhil, for interesting conversations around the kitchen, and in particular thank you to the members of the Neuroscience team.

A massive thank also goes to Demis Hassabis, Shane Legg and Mustafa Suleyman for making me a part of the DeepMind family. I feel very inspired by your constant endeavour to make things right, to care and to take responsibility.

Thank you to old friendships that still keep in touch. You've been a big part of my life despite being spread across the globe. Thank you to my parents and to my wider family, for always having both my development and my well-being (which are sometimes conflicting factors) on the top of their priority list.

Finally, thank you to Christiane for your patience, and for being at the centre of everything — scientific or otherwise.

Contribution

This thesis presents the results of a few original experiments reported in chapters 2, 3, 4 and 5. My thesis supervisor, Prof. Christopher Summerfield, provided feedback in the conceptualisation, design, analysis, and writing of all these. Additionally, many others contributed in the various phases of the experiments.

In Chapter 2, Experiment 2.1 was designed together with Dr. Dharshan Kumaran. Experiments 2.2 and 2.3 were designed together with Prof. Timothy Behrens. I additionally obtained help during the data collection for Experiments 2.2 and 2.3 from Stephanie Hall and Will McCarthy.

The design of the experiment in Chapter 3 followed closely that of Experiment 2.3 and was also in collaboration with Prof. Timothy Behrens. I obtained additional help from Prof. Maria Ruz during the data collection.

In Chapter 4, I present an experiment that was designed in collaboration with Prof. Hugo Spiers and with Dr. Demis Hassabis. This work has been published in the journal *Neuron*.

In Chapter 5, I present an experiment for which I only performed the data analysis and writing of the chapter. The experiment was designed by Prof. Christopher Summerfield and Prof. Maria Ruz, who also implemented the task and collected and preprocessed the data. Monika Wasczuk helped with the data collection.

Chapter 1. General introduction

1. First definitions

I would like to start by laying down some concepts and definitions. I will illustrate these with simple examples. These examples aim to provide the reader with an intuitive understanding of what is a representation, and in particular a hierarchical representation. A more formal summary of historical research into hierarchical representations for decision-making will be provided later on.

1.1 Structured observations, and structured representations

The world is structured, meaning that it is partially consistent over space and time. This consistency allows us to rely on past experiences in order to predict the future. For example, I have seen many cats in my life and I have heard most of them meow. I have never heard a cat bark. Thus, when I see an unfamiliar cat, I expect it to meow sometimes, and I do not expect it to bark. To form this expectation, I require a *mental representation* of what is a cat, which encodes knowledge of the properties of cats. This allows me to identify new items as cats and make predictions about how they will behave.

Mental representations need not represent the relational structure of the world. For example, I could learn about the properties 'can meow', 'has whiskers' or 'is friendly when hungry' in a completely unrelated fashion. Such a class of representation would not allow me to predict that something that meows has whiskers, even if it would allow me to conceive of a cat with all these properties. By contrast, a representation that accurately captures the co-occurrence of meows and whiskers will allow me to

understand how entities in the world are related. Such representations can also said to be *compressed*, because they reduce the amount of information redundancy (MacKay, 1992; Murray & Ghahramani, 2005).

1.2 Transitive relationships

More complex mental representations can capture the relationship among a set of multiple elements. For example, let us consider the relationship between a dog, a cat, and a canary. I may learn from experience that on average "cats are smaller than dogs", and that "canaries are smaller than cats". Note that the relationship 'smaller than' is directional: it is not the same to say "cats are smaller than dogs" or "dogs are smaller than cats".

Relationships are called *transitive* when one can combine directional relationships to embed a sense of order and thus infer new relationships. Formally, transitivity means that if A goes before B and B goes before C, then A must go before C. Taking the previous example, from knowing that "canaries are smaller than cats" and that "cats are smaller than dogs" I can conclude that "canaries are smaller than dogs".

Note that the relationship "smaller than" is transitive, but not all relationships among entities are transitive. For example, knowing that "canaries are afraid of cats" and that "cats are afraid of dogs", I shouldn't conclude that "canaries are afraid of dogs". Note also that the notion of transitivity can be applied recursively as many times as required.

1.3 Hierarchy as a transitive structure

In its most general definition, a hierarchical representation is any representation that respects *transitivity*, i.e. a sense of order. For example, the natural numbers (1, 2, 3, …) are hierarchical because they follow an order defined by the operation "bigger than" (or smaller than; $1 < 2 < 3 < \ldots$). Natural numbers are a particular case of a hierarchical representation termed a "magnitude" said to be one-dimensional because their relationship (e.g. rank) can be represented along a line.

Another special case of a hierarchical representation is a *nested tree* structure. A nested tree structure is one that starts with an element (root) connected to many others (children), with each successive child being connected to other child elements of their own, thus branching out. For example, my grandfather is an ancestor of my father, and my father is my ancestor (thus my grandfather is my ancestor). However, my grandfather is also the ancestor of my uncle and aunt who are not my ancestors (and who have children of their own).

1.4 Intrinsically hierarchical relationships

So far, I have focused on hierarchical mental representations, i.e. representations that support transitivity. As I have mentioned, the world itself is structured and good representations (i.e. representations that make sense of the past but also make accurate future predictions) are typically the simplest ones that capture this structure from past observations. A hierarchical representation (i.e. assuming transitivity) may lead to the simplest accurate representation when the world is hierarchically structured, but otherwise it will not. I therefore shall distinguish between a *hierarchical mental representation* (that assumes that the world is hierarchically structured) and an *intrinsic hierarchical relationship* (objectively true, grounded in the natural structure

of the world). In my own terminology, I may at times simply use the term *hierarchy* to imply both things at the same time: a hierarchical mental representation that is efficiently applied in order to represent an intrinsically hierarchical relationship.

1.5 Concluding remarks

As I have shown, hierarchical representations can take many forms and depend strongly on the relationship that connects the elements to each other. From now on this thesis will mostly focus on hierarchies defined by a particular type of relationship: grouping and inclusion, with "all A are B" transitive relationships. For example, in the taxonomic hierarchy of animals, knowing that "a squirrel is a rodent", "a rodent is a mammal" and that "a mammal is an animal", I can conclude that a "squirrel is an animal". In the next section I will focus in this example.

I have chosen this example because the taxonomy of animals is a paradigmatic example of an intrinsic hierarchy (Linnaeus, 1758). It is thus a clear case where hierarchical representations can be efficient, in terms of building simple (compressed) representations and in terms of improving the prediction of future observed events.

2. An example: the taxonomy of animals

As mentioned above, mental representations can be recursively nested to produce a hierarchy. Consider the first time that a child sees a squirrel, e.g. in a neighbourhood where squirrels are not a common species. One would expect the child to be able to make some accurate predictions about its behaviour. For instance, she may expect that the squirrel can walk on its four legs, or that it will get scared and run away quickly if she moves towards it. She may also believe she can lure the squirrel with food, and that she should be careful because a squirrel can bite. These predictions are derived from mental representations built from previous experiences with other species

(similar to how we usually expect something that meows to also have whiskers). For instance, she can reason that a squirrel is an animal (because it moves on its own and has four legs and fur, like dogs and cats) and thus a squirrel behaves similarly to how other animals she has encountered generally do, i.e. it runs and forages. Furthermore, she may assume that this unfamiliar creature shares certain properties with specific subsets of animals she knows of, such as rodents (e.g. hamsters) and thus it may also exhibit rodent properties (e.g. eats seeds and has sharp teeth) that not all animals display. Conversely, it would be unreasonable to expect that a squirrel reflects all sorts of animal behaviours, e.g. fly like a pigeon (bird) or swim like a fish, because it is less similar to these other animals. Many of these predictions would be indeed correct.

Nonetheless an infant is unlikely to perfectly predict everything that defines a squirrel at first exposure, because squirrels also have properties particular of its species. She may for instance be surprised to discover that a squirrel can climb with ease on a tree, a feat she has never seen other rodents achieve beforehand. After this first exposure, she should be able to build from her experience a concept of a squirrel and incorporate all the squirrel's relevant properties in her knowledge, some of which are shared with other species and some which are not. Thanks to a hierarchical mental representation of animals, the child can keep a very simple description of her experience: a squirrel is a rodent (with all the rodent, mammal and animal properties that rodents entail) that can climb on trees.

2.1 Short overview

This example illustrates a few of the core properties of hierarchies. I suggested a mental hierarchical representation of the squirrel by relating it to other animals: all squirrels are rodents; all rodents are mammals; all mammals are animals. By the

principle of transitivity, one can deduce that all squirrels are animals. This allows the child, who sees a squirrel for the first time, to perform *inductive reasoning* by considering the relevant properties at each level of the hierarchy (rodents, mammals, and animals). This representation would also allow the child to ignore knowledge about other specific, irrelevant properties that do not apply to all rodents, mammals or animals (e.g. knowledge that birds can fly). I will next expand on each of these points.

2.2 Compression and categorisation

Hierarchical representations can be efficient by reducing the amount of information redundancy, i.e. by storing the knowledge represented economically (compressing). For example, it would be redundant to store the information that "squirrels need to forage" because this is already accounted by knowing that a squirrel is an animal, and that all animals forage. This redundancy would lead to an unnecessarily complex representation. For instance, squirrels, rats, mice, hamsters, guinea pigs, and chipmunks are relatively small, have a tail, short limbs, short fur and eat seeds. Representing all this knowledge independently for each species would lead to high levels of redundancy. Much more easily, one could just remember that such species can be grouped as 'rodents', and that most rodents share these properties. This grouping strategy, which is the basis of (inclusive) hierarchical representations, is known as *categorisation*.

In this section I will discuss the concept of categorisation outside of the scope of hierarchical representations. This is necessary because categorisation is the key relationship over which transitivity is performed in the hierarchical representations of interest to this thesis. I will postpone until then a description of how hierarchical

representations emerge by performing categorisation recursively, i.e. by building categories over categories.

Categorisation (or classification) is the process of recognising and differentiating things by placing them into categories or classes. Note that under this general definition, categories may be learnt in a supervised fashion (i.e. taught), and may not rely on the perceived similarities. For example, medical doctors need to classify regularly mammograms as normal or as tumorous (which may look similar to the naïve eye), and manipulation of the supervision can aid them in making better decisions (Hornsby & Love, 2014). Categorisation and category learning are thought to be a cognitive mechanism key to human intelligence and have been extensively studied in humans (Ashby & Maddox, 2005, 2011; G. Murphy, 2004) and other animals (Freedman, Riesenhuber, Poggio, & Miller, 2001; Seger & Miller, 2010; J. D. Smith, Minda, & Washburn, 2004).

Still, there is current disagreement about how humans and other animals represent categories neurally and psychologically. There are many theories in the past literature, but a few have dominated. I describe these here.

The "exemplar" theory considers categories as sets of exemplars (Brooks, 1978; Estes, 1986, 2008; Hintzman, 1986; Lamberts, 2000; Medin & Schaffer, 1978; Nosofsky, 1986). Under this theory, the category of 'rodents' would be defined by a set including squirrels and hamsters, but also many other exemplars such as rats, mice, degus, guinea pigs, marmots and others. Under this theory, a new species is to be categorised as a rodent if it is (on average) similar to all other rodent species (exemplars).

Alternatively, the "prototype" theory instead corresponds each category with a representative (prototypical) member of the category, such that the membership of an exemplar to the category is defined by the distance to its prototype (E. H. Rosch, 1973). In this case, a rodent would be defined by a non-existent small animal with short fur that resembles most rodents on average.

The "decision boundary" theory does not define the categories per se but asserts that only the boundary between categories is necessary in order to classify an exemplar (Ashby & Gott, 1988; Ashby & Townsend, 1986; Maddox & Ashby, 1993). For example, to classify squirrels from hamsters one would compare an exemplar against the multi-dimensional bound between the two. Note that the bound would depend on the categories. For instance, it would be different for classifying squirrels from guinea pigs.

Categories can also be defined by rules that are verbalisable (Ashby & Maddox, 2005). In this case, one could envisage a rule such that "an animal is a squirrel if it has a fluffy long tail" (this is not entirely accurate).

Progress has been made in this field despite all the open questions that surround the process of categorisation and category learning. For example, it is generally accepted that categorisation is subserved by a network of brain regions differentially involved as a function of the modality (e.g. visual, auditory) and of the amount of expertise (Ashby & Maddox, 2005; Seger & Miller, 2010).

2.3 Recursive categorisation and hierarchies

Categorisation can be performed recursively. For example, the rodent species category (e.g. squirrels, hamsters, etc) can be subsequently grouped together with other categories like primates (e.g. gorillas, orangutans and chimpanzees) and carnivora (e.g. cats, dogs and bears) into a higher-level category of 'mammals',

because their species are more similar with each other (e.g. they all have fur) than they are with any fishes, birds or reptiles. This way, increasingly more general categories like mammals, animals, and living beings can be defined from basic observations. Such representation is called *hierarchical* because categories build on top of each other, increasingly reducing the redundancy about what is known and allowing an affordable encoding of all this information. Note that the relationship required for recursive categorisation ("all A are B") is transitive.

2.4 Concluding remarks

Generalisation is the term given to the mental process whereby extant knowledge is applied to new stimuli. Hierarchical grouping representations make generalisation straightforward. Once a specimen from a new species is encountered (e.g. a squirrel), one can easily categorise it simultaneously at multiple levels of abstraction (e.g. animal, mammal, and rodent) from a few features (e.g. it moves, it is small, it has short fur and a long tail) and generalise many other unseen properties from what is known about the category at each level. Analogously, whether the species encountered is familiar or not, this representation has the advantage that one can easily locate all the knowledge that is relevant to the encounter (i.e. knowledge related to animals, mammals and more specifically to rodents) rather than considering irrelevant knowledge (e.g. knowledge about sharks' teeth).

3. Hierarchical representations in semantic knowledge

3.1 Evidence from reaction times

What evidence is there that humans use hierarchical representations? The pioneering work of Collins and Quillian (Collins & Quillian, 1969) showed that human participants reported that an exemplar was included into a subset ("a canary is a bird")

faster compared to a superset category ("a canary is an animal"). Similar results were reported for properties that were specific or general (e.g. "a canary can fly" faster than "a canary can move"). These results were originally explained through a "hierarchical-network" model (see **Figure 1.1**) that can be represented as a nested-tree representation (graph). This graph is assumed to be known (provided by the researcher) rather than learned autonomously by the model. The nodes in the graph correspond to the exemplars ('canary') and categories ('bird'). The edges in the graph link the nodes, representing the exemplars that belong to each category (e.g. a canary is a bird). Nodes are additionally associated with the properties of each category or exemplar. Under this model, the reaction time required for recall is then assumed to increase linearly with the "hierarchical distance", i.e. the number of edges that must be traversed within the network in order to connect two nodes.

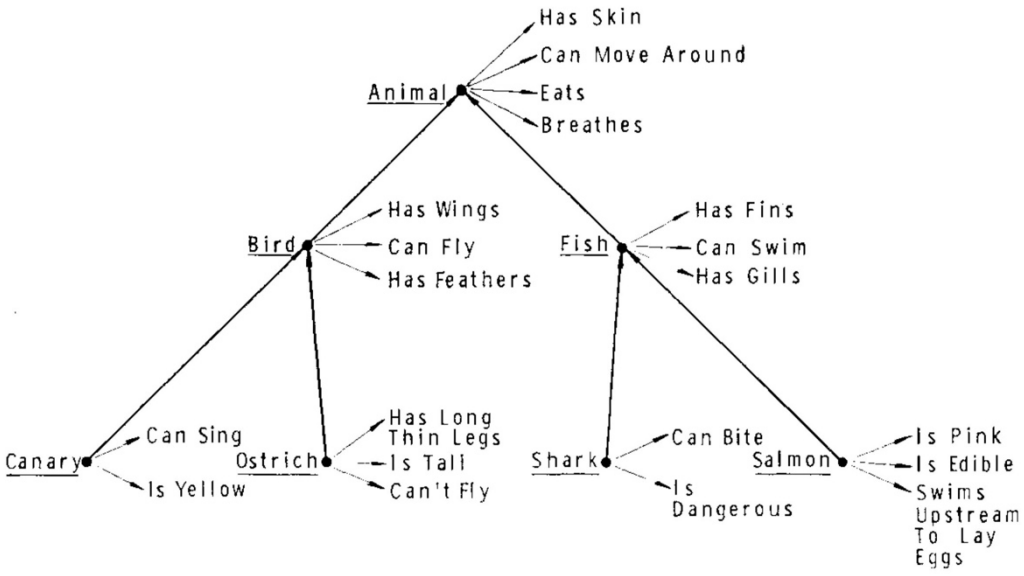

Figure 1.1. The "hierarchical-network" model of semantic knowledge (reproduced from (Collins & Quillian, 1969))

This “hierarchical network” model remained influential for the decades following its publication, despite it failed to explain certain behavioural effects. For example, reaction times are sensitive not only to the category but also to the exemplar representativeness. Indeed, responses are faster if an exemplar is more representative of a category (e.g. ‘robin’ is more representative than ‘chicken’ for the category ‘bird’) (Chang, 1986; E. E. Smith, Shoben, & Rips, 1974), in line with the “prototype theory” of categorisation mentioned earlier. Additionally, some exemplars of a category are systematically taken as an exception (e.g. penguins are birds but cannot fly, and pines are trees without leaves) and this may lead to interesting patterns in reaction times (e.g. faster reaction times for verifying that “a penguin is an animal” than for “a penguin is a bird”) (T. T. Rogers & McClelland, 2004; E. E. Smith et al., 1974). These results led to all flavours of alternative models and sparked controversy as to the nature of the representations, and mechanisms of retrieval during semantic categorisation (Close & Pothos, 2012; Collins & Loftus, 1975; Glass & Holyoak, 1974; McCloskey & Glucksberg, 1979; Meyer, 1970; E. E. Smith et al., 1974). Interestingly, a relatively recent article described a set of experiments that failed to replicate the classical findings of (Collins & Quillian, 1969) with a hierarchical structure learned in a laboratory setting (G. L. Murphy, Hampton, & Milovanovic, 2012). They found in this case that participant’s reaction times were better explained in terms of a feature-comparison heuristic when the task allowed for it.

3.2 Evidence from memory performance

In a different approach, Bower et al. (Bower, Clark, Lesgold, & Winzenz, 1969) showed that providing study items in a hierarchical format aids subsequent retrieval. Participants were shown a list of item words (e.g. minerals) for later recall. In the condition of interest, the items were visually arranged in a tree diagram similar to **Fig**

1.1 (e.g. bronze is an alloy, which is a metal, which is a mineral). In the control group, items were also arranged in tree diagram, but this did not reflect the natural hierarchical relationship of the items (i.e. the words were "shuffled" within the tree diagram). Participants in the hierarchical presentation condition recalled two to three times more words than the control group that was not provided with a coherent hierarchical relationship between the items.

More generally, semantic representations that capture the structure of language (including hierarchical but not exclusively) may also be responsible for false memory beliefs. It has been shown that participants exposed to related words (e.g. 'snow', 'ice', 'winter' and 'warm') also tend to falsely misremember seeing a lure word that is related but has not been seen (e.g. 'cold') (Chadwick et al., 2016; Roediger & McDermott, 1995).

3.3 Evidence from developmental psychology and clinical research

Evidence that humans use hierarchical representations has also been complemented by findings of semantic knowledge acquisition during development (infancy and childhood), as well as by clinical research. Children are thought to first learn the most abstract categories before their subcategories, allowing them to gradually make finer and finer distinctions. For example, during development infants may learn that animals have bones and over-generalise this belief to worms (Keil, 1979; Mandler, 1992, 2000; Mandler & McDonough, 1993; Pauen, 2002). Additionally, a clinical condition known as *semantic dementia* that affects the anterior temporal lobe has been characterised by a deterioration of semantic knowledge in the reverse direction, from specific to general (Warrington, 1975; Warrington & Shallice, 1984). For instance, patients with semantic dementia may forget to draw the stripes of a tiger (specific property that differentiates tigers) but not its four legs (general property that applies to

many animals). Overall, these results imply that specific semantic knowledge corresponding to the lower (more specific) levels of a hierarchy are learned after and forgotten before the general information corresponding to higher levels of the hierarchy.

3.4 The limits of hierarchical representations

Subsequent research has also reported other observations that are harder to reconcile with hierarchical representations. This includes the basic-level advantage, where a more commonly-encountered hierarchical level of abstraction has a special status and is preferred over others (for example, one would consider her pet to be a 'dog' more naturally than thinking of it as a 'labrador' or as an 'animal') (E. Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), and effects of intransitivity in semantic categorisation (e.g. "a scrub-oak is an oak" and "an oak is a tree" but "a scrub-oak is not a tree") (Hampton, 1982; Randall, 1976). Despite these controversial results, the general gist that hierarchical representations are central to semantic memory has remained a core and unshakeable assumption in the field.

3.5 The Parallel Distributed Processing (PDP) approach

A radically different family of models for explaining the hierarchical structure of semantic knowledge was proposed by D. E. Rumelhart and J. L. McClelland, among others, in what is known as the Parallel Distributed Processing (PDP) approach (Rumelhart, McClelland, & Group, 1987). PDP models differ from other models in that they do not consider concepts to be explicitly represented as nodes in a graph. Instead, knowledge is represented in a distributed fashion across the synaptic connection weights of an "artificial neural network" (see **Fig 1.2**). In the simplest case, a feed-forward network receives as input nodes a set of items (such as "canary")

and a set of relations (e.g. “is a” or “can”) and is expected to produce the output that correctly matches the input as well as the relation (“bird” or “fly”). At any point in time, only some of the input nodes are active, and it can activate any number of output nodes. The network is then trained progressively (by update of its weights) through supervision in order to learn the target response, by reducing the error of the output. The development of the back-propagation optimisation algorithm (Rumelhart, Hinton, & Williams, 1986) allowed the possibility of more complex networks, eventually leading to a PDP approach for semantic cognition best described in (McClelland & Rogers, 2003; T. T. Rogers & McClelland, 2004). Note that in this framework the network is trained via supervised learning, and the learnt target labels (e.g. “canary is a bird”) conform to the more general definition of categories and are not discovered by grouping exemplars with statistically overlapping properties.

The PDP approach is appealing for multiple reasons. First of all, the hierarchical relationship is not specified explicitly in the definition of the model. Instead, the network learns internal representations (a vector of latent activations relating inputs to outputs, see “hidden” layer in **Fig. 1.2**) through its optimisation procedure while learning to produce the target output. This is advantageous because it does not require a prior assumption that the relationships among items are intrinsically hierarchical. Instead, it proposes a mechanism that is systematically and autonomously sensitive to the hierarchical relationship of the dataset, presumably by finding a compressed representation of the relationship between the inputs and outputs (targets). This representation will capture the intrinsic hierarchical structure of the dataset and thus be hierarchical itself.

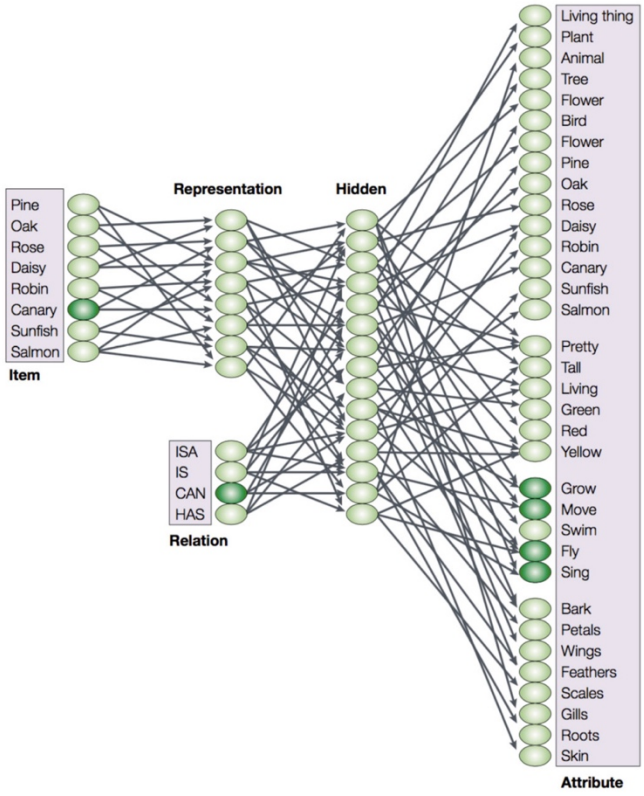

Figure 1.2. Diagram representing the connectionist PDP model of semantic knowledge (reproduced from (McClelland & Rogers, 2003))

Second, it is biologically meaningful because it resembles in spirit the connectionist nature of the networks of neurons in biological cognitive systems.

Third, simulations with this model are shown to replicate qualitatively some hierarchical biases such as the general-to-specific gradient of human learning, as well as a bias to over-generalise (e.g. “a pine has leaves”) during early training (Keil, 1979; Mandler, 1992, 2000; Mandler & McDonough, 1993; Pauen, 2002). This was observed by analysing the predictions of the network half-way through the training. For example, when the inputs were active for ‘canary’ and ‘can’, the network responded early that a canary can ‘grow’ and ‘move’ (general properties shared by most animals) and only later on that it is also ‘yellow’ (specific property of canaries in their example). Additionally, the model was able to generalise its knowledge after

learning was complete by appending a new input node (e.g. a sparrow) and retraining the network with only a limited information (e.g. "sparrow is a bird"). As a result, the network correctly generalised new knowledge, such as the fact that a sparrow can fly.

Fourth, this model also explained other effects that were in conflict with the hierarchical account of semantic knowledge. For instance, it was able to capture the observation of a basic-level advantage (mentioned before) and how it is affected in semantic dementia patients, by manipulating the distribution of the data such that information related to a given category (dog) was overweighted. In that case the network quickly learned the label for dog, and over-generalised during early training such that it categorised other animals (goats) as dogs (but not unrelated exemplars such as a pine tree), just as children would.

Fifth, the PDP theory of semantic knowledge further suggested a functional relationship between regions in the human brain and the model. Semantic knowledge retrieval is thought to be distributed throughout the cortex, and to "evoke a pattern in a brain region dedicated to that type of information" (e.g. knowledge related to colour or form in the ventral regions of the occipital cortex) (McClelland & Rogers, 2003). By this account, the temporal pole (the critical region affected in semantic dementia) works as a hub that ties together the information retrieved into a single construct. Recent imaging research in humans is consistent with this prediction (Chadwick et al., 2016). The effects of false memory beliefs mentioned earlier can be explained by the overlapping similarity between the neural pattern of activations in the temporal pole for concepts (such as 'cold' of 'snow'), and the distributed representational patterns from a computational model.

3.6 Complementary Learning Systems

The PDP theory of semantic knowledge acquisition only aims to explain how semantic knowledge is acquired over long periods of time (months or years). It assumes that knowledge is slowly learned through modest and progressive updates of the synaptic weights, and does not explain the human ability to rapidly acquire new knowledge even with a single exposure. This observation is better accounted for by extending the model to include complementary learning systems (CLS) (Kumaran, Hassabis, & McClelland, 2016; McClelland, McNaughton, & O'reilly, 1995) . In CLS theory, the neural biological systems rely on a second, episodic learning system implemented in the medial temporal cortex that is able to rapidly store information. This information is episodic in the first instance, but is slowly consolidated into structured knowledge in the neocortex — a process that is thought to depend on sleep. In line with this theory, recent work (Schapiro et al., 2017) has shown that sleep benefits memory for general object features that are shared among members of a category, i.e. a signature that consolidation prioritises the "general" properties corresponding to the higher level of a hierarchy.

3.7 Concluding remarks

It is worth noting that the most part of the empirical research described above has relied on natural hierarchical representations, i.e. hierarchical representations learned in the real-world, in contrast to those learned from artificial stimuli in the context of a laboratory experiment. This limits the experimental control over the hierarchical structure as well as the stimuli, and it is thought that many of the effects originally attributed to hierarchical representations are at least partially explained by nuisance factors such as visual saliency as well as familiarity (Chang, 1986).

4. Hierarchical representations in temporal and spatial domains

4.1 Hierarchical representations in syntax

Hierarchical representations have been shown to be critical in many domains other than semantic categorisation. For example, it is believed among linguists that hierarchical nested tree representations are an essential component of any theory of how humans represent the structure of sentences (language syntax), and that at the same brain regions engaged during language syntax processing are also engaged during processing of other temporally structured domains, such as musical syntax, and motor observation and production (Collard & Povel, 1982; Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015; Koelsch & Siebel, 2005; Restle, 1970). I will borrow an example from (Dehaene et al., 2015) to illustrate this point. The sentence "he drove to this big house" can be flexibly reduced to "he drove to it", "he drove there", or even just "he did". Grouping the words in this example ("to this big house" as "there") is similar to how objects (or species) can be grouped into categories in two ways: i) a set of elements (exemplars; related words) are grouped together into a common compressed representation (category; syntagm) and ii) in both cases grouping can be performed recursively. This induces a hierarchical representation (i.e. a transitivity principle) similar to the one explained for recursive categorisation such that if "big" is part of "the big house", and if "the big house" is part of "drove to the big house", then "big" is part of "drove to the big house".

This intrinsic hierarchical relationship could be autonomously revealed thanks again to the PDP approach. In a seminal article by J. L. Elman, a distributed neural network

model for language that included recurrent connections (effectively endowing the model with short term memory and enabling it to exploit the temporal structure of language) could learn a distributed representation that respected the hierarchical structure of syntax in language (Elman, 1990).

4.2 Temporal structure and statistical learning

It is worth noting that the temporal context (i.e. other events that occur just before or after the to-be-encoded stimulus) plays an important role in the hierarchical structure of syntax — whether it applies to words, musical notes, or action steps. In the extreme, a hierarchical decomposition may be fully determined by the temporal sequence itself, rather than by the properties of the elements. Structure in the temporal domain can be studied in isolation, by using arbitrary stimuli (e.g. images, sounds, pseudo-words) that have no 'real-world' meaning, and where the relationship between the elements can be captured by the temporal association (i.e. transition) between its elements. This has been studied in depth by a subfield of cognitive psychology known as "statistical learning". There is abundant evidence that humans can discover and learn temporal structure, even subliminally and with a variety of stimuli (Ahlheim, Schiffer, & Schubotz, 2016; Ahlheim, Stadler, & Schubotz, 2014; Fiser & Aslin, 2002; Furl et al., 2011; Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012; Perruchet & Pacton, 2006; Petersson, Folia, & Hagoort, 2012; Saffran, Aslin, & Newport, 1996; Turk-Browne, Scholl, Chun, & Johnson, 2009).

4.3 Hierarchical representations in statistical learning

Notably, Schapiro et al. investigated whether human participants could learn a hierarchical temporal representation (see **Fig 1.3)** (Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013) when a stream of images had an intrinsically hierarchical

temporal relationship. Note that unlike for language or music, in this case the hierarchical structure was fixed, so that the same images would always be associated with the same temporal context (~category). In this experiment, participants became familiar with the temporal structure of a stream of images (fractals or symbols). The sequence of images was generated from a Hamiltonian path in a hierarchical graph (i.e. a pseudo-random walk, but visiting each node once on every pass). This graph was hierarchical because its community structure could be decomposed efficiently into three "clusters" that were fully connected (coloured in orange, purple and green). Critically, this hierarchical decomposition could not be explained by the constant number of adjacent nodes (four in each case, thus controlling for the uncertainty of prediction the upcoming image in the stream).

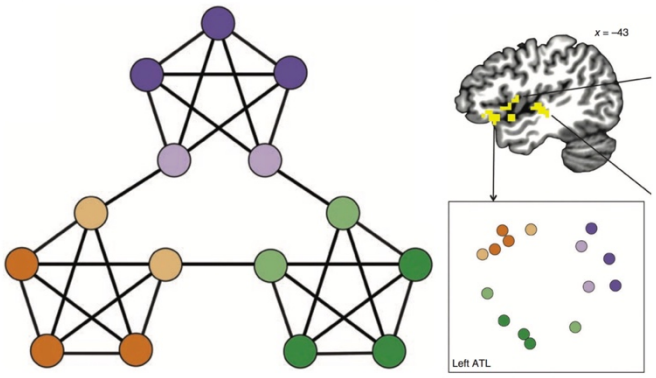

Figure 1.3. Statistical learning of a hierarchical structure (reproduced from (Schapiro et al., 2013))

They discovered that the participants' neural representation reflected this hierarchical structure and was sensitive to the transition between the hierarchical clusters, and showed similar patterns of activity in the anterior temporal lobe for nodes of the graph (i.e. images) that were strongly connected. This is best illustrated in the two-

dimensional representation in **Fig 1.3** (right panel), where the multi-dimensional space of possible relationships between the nodes is compressed into a ring, with strongly connected nodes (i.e. belonging to the same cluster) placed near each other. Later analyses found that the hippocampus (within the medial temporal lobe) was also sensitive to the hierarchical structure of the graph (Schapiro, Turk-Browne, Norman, & Botvinick, 2016). In summary, participants learned to represent hierarchically the transitions in the stream of images by encoding similarly the images within a cluster that were likely to be temporally adjacent to each other. This is consistent with other studies that ascribe a critical role to the hippocampus for (temporal) statistical learning (Bornstein & Daw, 2012; Schapiro, Gregory, Landau, McCloskey, & Turk-Browne, 2014; Turk-Browne et al., 2009; Turk-Browne, Scholl, Johnson, & Chun, 2010).

4.4 Hierarchical representations in motor plans

Hierarchical representations in the temporal domain also play a role in motor plans (i.e. sequences of actions). It has been known for a long time that behaviour that spans multiple, sequential actions can be roughly speaking hierarchical in that it is "organised simultaneously at several levels of complexity" (Lashley, 1951; Miller, Galanter, & Pribram, 1960). There is a wealth of evidence from imaging experiments supporting this view. Micro-stimulation of primary motor cortex in the macaque has revealed neurons that code for action endpoints independently of the current starting state, instead of simple actions (Graziano, Taylor, & Moore, 2002). In the premotor cortex, neurons seem to code for multi-componential movements (Tanji & Shima, 1994). In the prefrontal cortex, neurons seem to code for more abstract (i.e., categorical) motor plans (Shima, Isoda, Mushiake, & Tanji, 2007). Evidence in human imaging, and cell-recording in songbirds, supports the claim that a high-level

sparse coding can trigger specific simpler programs learned from experience (Fiete & Seung, 2007; Hahnloser, Kozhevnikov, & Fee, 2002; Koechlin & Jubault, 2006; Schneider & Logan, 2006). This clustering of simple actions into increasingly more complex motor programs has sometimes been referred to as "temporal abstraction".

4.5 Hierarchical representations of space

Signatures of hierarchical representations have also been found in spatial cognition (Hirtle & Jonides, 1985; McNamara, Hardy, & Hirtle, 1989). For instance, humans may incorrectly believe that Philadelphia is located northerly of Rome. Although this finding is often explained in terms of perceptual grouping (Tversky, 1981), it may also depend on the influence of subjectively defined regions that are independent from the geographical organisation. For example, Philadelphia is considered to be in the north of the US, and the US in the north of America, whereas Rome is in the middle of Italy and Italy is in the south of Europe (Friedman, Brown, & Mcgaffey, 2002). This effect, sometimes referred to as *regionalisation*, may also influence navigational strategy: during wayfinding, humans prefer routes that permit a spatial context boundary to be crossed earlier rather than later (Wiener & Mallot, 2003).

4.6 Hierarchical representations can alleviate planning complexity

The observations of hierarchical representations in time, space and motor behaviour are conceptually meaningful when one considers the problem of goal-directed planning. Goal-directed planning can be defined as the problem of making a correct sequence of temporally-dependent decisions in order to reach a goal. This sequential dependency makes planning more complex than categorisation, where each decision is independent and where feedback can be received for every decision. Examples of goal-directed behaviour include playing board games (such as chess where one aims

to checkmate) or in navigation where the shortest (fastest) trajectory between two different locations of a city is desirable. Planning is thought to be achieved through a model of the environment termed a 'cognitive map' (originally framed within the context of spatial navigation) (Tolman, 1948) over which multiple trajectories can be simulated until the optimal one is found. Computationally, such simulations are expensive because the number of possible trajectories to account for grows exponentially with the number of possible steps (depth) given the number of possible actions on each step (branching factor).

In machine learning and computational neuroscience, it is widely recognised that the computational demand associated with planning can be reduced by exploiting a hierarchical structure in the environment, with states clustered into larger "contexts" (Badre, Kayser, & D'Esposito, 2010; M. M. Botvinick, Niv, & Barto, 2009; Koechlin & Jubault, 2006; Sutton & Barto, 1998). It has been suggested that sub-goal decomposition (the ability to hierarchically decompose a task into easier sub-tasks) is an essential feature of problem-solving (Anderson, 2000).

As an illustration of why this is the case, imagine yourself in an environment composed of four rooms (**Fig 1.4, left panel**). This environment can be approximated as a graph with a finite number of nodes, such that on any step you can displace from your current position to one of the neighbour nodes. In this example, the starting point is the green node on the top-right corner, and the goal location where you want to navigate is the red node on the bottom-left region. Thus, in order to plan the optimal trajectory, one could consider all the possible trajectories, devoting similar computational time to each neighbour node on each step of the simulated trajectory

("flat" policy). Alternatively, one's policy could be complemented by knowledge of how to reach specific key locations, such as the doorways (light blue nodes in the graph; "doors" policy) independently of the goal. This latter approach, which relies on *subgoals*, can drastically reduce the computing time required for planning, and lead to a hierarchical decomposition of the environment into rooms (**Fig 1.4, right panel**). Note however that augmenting the policy with subgoals is not always be beneficial. For instance, learning a routine that navigates to the corners (orange nodes) may be harmful and increase the planning computing time (**Fig 1.4, middle panel**).

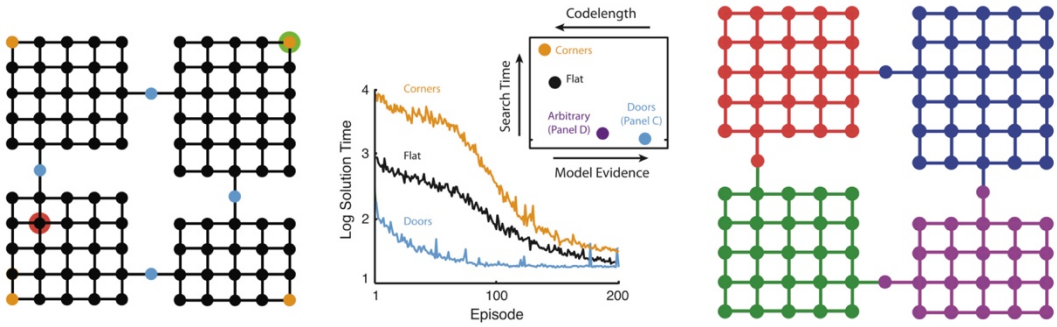

Figure 1.4. Hierarchical decomposition of an environment for efficient planning (reproduced from (Solway et al., 2014))

Planning, and hierarchical planning, have been extensively studied in neuroimaging and clinical studies. Patients with lesions to the prefrontal cortex often exhibit disordered action sequences that fail to achieve the specified goal (Owen, Downes, Sahakian, Polkey, & Robbins, 1990; Shallice, 1982; Shallice & Burgess, 1991). Hippocampal patients have difficulty imagining the future states entailed by a plan of action (Schacter et al., 2012). Moreover, functional neuroimaging has confirmed the involvement of human prefrontal and limbic structures in forming and executing plans, particularly in spatial environments (Howard et al., 2014; Schacter & Addis,

2007; Unterrainer & Owen, 2006). Nevertheless, linking these macroscopic neural findings to the underlying computational mechanisms that subserve planning remains an open challenge for psychologists and neuroscientists, and it is still debated which of many computational perspectives are pertinent, including tree-search, heuristic search, inference, and evidence integration (Bonet & Geffner, 2001; M. M. Botvinick et al., 2009; M. Botvinick & Toussaint, 2012; Huys et al., 2012; Huys, Lally, et al., 2015; Kaplan & Friston, 2017; Keramati, Smittenaar, Dolan, & Dayan, 2016; Korn & Bach, 2018; Mattar & Daw, 2018; Solway & Botvinick, 2015).

4.7 Reinforcement Learning

One promising framework for understanding human planning is Reinforcement Learning (RL) (Dayan & Niv, 2008; Niv, 2009; Sutton & Barto, 1998). RL is a framework for Machine Learning that was derived from early ideas in psychology, whereby an agent is optimised to maximise the long-term cumulative reward over a sequence of states and actions (i.e. a planning problem). The promise of RL as a model for human hierarchical planning is two-fold.

First, RL has already been validated as a useful model for understanding dopaminergic signals in the mammalian midbrain. Activity in dopaminergic neurons are thought to reflect the *reward prediction error* (RPE) between the observed and the expected reward, a quantity that is essential for learning to maximise the long-term reward in reinforcement learning algorithms (Schultz, Dayan, & Montague, 1997). These signals were originally discovered in the midbrain, but comparable results from neuroimaging studies have since been reported in striatum and the orbitofrontal cortex (Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Niv & Schoenbaum, 2008; Nobre, Coull, Frith, & Mesulam, 1999; O'Doherty, 2004; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Tobler, O'Doherty, Dolan, & Schultz, 2006), and unsigned

prediction errors are often reported in the medial prefrontal and particularly in the dorsomedial cingulate cortex (Alexander & Brown, 2017; Hyman, Holroyd, & Seamans, 2017; Silvetti, Alexander, Verguts, & Brown, 2014). Note that standard algorithms for RL, such as temporal-difference (TD) learning (Sutton & Barto, 1998) learn very slowly, first by accidentally finding an unexpected reward or punishment (i.e. a reward prediction error; e.g. at the location of the goal) and then by slowly and recurrently propagating the RPE to adjacent states that were visited prior to the rewarded state.

Second, a wide set of reinforcement learning algorithms are equipped with hierarchical capabilities (Barto & Mahadevan, 2003; Dayan & Hinton, 1993; Sutton, Precup, & Singh, 1999; Wiering, urgen Schmidhuber, & Elvezia, 1997). For example, the *options* framework of reinforcement learning (Sutton et al., 1999) allows RL agents to learn complex motor program (e.g. "go to the door" in the four rooms environment) by providing them an auxiliary reward, and these motor programs can be later executed like any other basic action available for an extended period of time. In line with this hypothesis, reward prediction errors at different hierarchical levels of plan abstraction have been observed in the anterior cingulate cortex (ACC) (Fernandes, Shahnazian, Holroyd, & Botvinick, 2018; Ribas-Fernandes et al., 2011) leading to the interpretation this region may be critical for maintaining hierarchical representations during plan formation and execution.

4.8 Hierarchical decomposition of an environment

The hierarchical reinforcement learning account of human planning is still far from complete. Among the most pressing issues, it is unclear (also within the Machine Learning community) how a computationally efficient hierarchical representation can be extracted from experience. Computational simulations have shown that exploiting

a hierarchical representation that does not match the environment can be harmful, increasing the difficulty of planning (Solway et al., 2014). One possibility is that participants rely on ecologically valid heuristics in order to exploit the hierarchical structure of the environment, for instance by detecting "bottlenecks" in a graph that represents the connectivity of the states in the environment (Botvinick, 2012; Schapiro et al., 2013).

Alternatively, recent research (Stachenfeld, Botvinick, & Gershman, 2014) has highlighted a set of analytical tools (derived from "*spectral graph theory*") (Chung, 1997; Spielman, 2009) that are able to hierarchically decompose the environment by learning a set of basis functions called 'eigenvectors'. This is a complex approach, but it is roughly implemented as follows: first, the connectivity of the environment is represented as a matrix, e.g. encoding the transition probability from one given state to another; then, a standard mathematical procedure known as 'diagonalisation' (or 'eigen-decomposition') is used in order to uncover a set of latent (compressed) components that represents the transitions between states in the environment. These latent components have been shown to be sensitive to the hierarchical connectivity of the environment (Kulkarni, Saeedi, Gautam, & Gershman, 2016; Stachenfeld et al., 2014). A related approach has also suggested that the spectral 'eigen-vector' function bases can be leveraged to form efficient representations for reinforcement learning in large spaces, because they can facilitate the structural decomposition of the environment that can be learned even without reward (Mahadevan, 2005; Mahadevan & Maggioni, 2007).

4.9 Concluding remarks

Spectral analyses are conceptually interesting because they have also been related to the learning dynamics of artificial neural networks such as those discussed in the PDP framework (see also the Introduction section to **Experiment 2.1** in **Chapter 2**) (Saxe, McClelland, & Ganguli, 2013). The relationship between spectral analyses and deep neural networks will be described in more detail in the *Introduction* section of **Experiment 2.1** in **Chapter 2**. To preview this discussion, standard deep neural networks learn faster the 'strongest' components that carry more information, which correspond to the top level when the dataset learned has an intrinsic hierarchical structure.

This makes the promise of a common framework for understanding hierarchical representations in semantic categorisation and goal-directed, sequential behaviour. Furthermore, recent studies have shown that deep neural networks can mimic the well-described spatial representations found in the medial temporal cortex of a cognitive map (i.e. mental model of the environment) (Banino et al., 2018). These findings are complemented by other commonalities. Machine Learning has undergone impressive advances recently by coupling function approximation methods (e.g. deep neural networks) with reinforcement learning algorithms, achieving super-human performance in complex tasks such as videogames, and board games such as 'Go' and chess, proving that such algorithms can scale to complex problems and behaviour (Mnih et al., 2015; Silver et al., 2016, 2017). Furthermore, it has been suggested that the episodic memory mechanism implemented in the medial temporal lobe for rapidly acquiring knowledge can serve an analogous role during goal-directed navigation, supporting navigation in environments whose structure has not yet been fully learnt (Lengyel & Dayan, 2008). Overall, this suggests that the neural and algorithmic

mechanisms responsible for learning hierarchical representations in semantic categories and in goal-directed sequential behaviour may overlap to some extent.

5. Decision-making mechanisms underlying generalisation

5.1 Inference and generalisation in hierarchical representations

As I have illustrated with the earlier example of the squirrel, hierarchical representations can permit generalisation on the basis of similarity (e.g. between a squirrel and other animals). I have also explained that this can be achieved through a structured and/or compressed representation, e.g. for example by using a deep neural network. However, one weakness of artificial deep neural networks is that (unless augmented with additional components) they are only expected to generalise to (i.e. perform accurate predictions about) new experiences that are similar to those that they were trained on. This is at least in part because they do not build causal models and cannot harness compositional representations (Lake, Ullman, Tenenbaum, & Gershman, 2017). In **Chapter 5** I will explore one such case where generalisation requires learning a more complex relationship than grouping by similarity. For this, I will use a categorisation task where the rule is verbalisable and taught via supervision. In the next sections I provide an overview of the problem of learning, inductive inference and generalisation.

5.2 Inductive inference in sequential tasks

The problem of data-efficiency and generalisation has been acknowledged in Reinforcement Learning for a long time, leading to a distinction between model-free and model-based learning (Kuvayev & Sutton, 1996). Model-free agents are optimised by learning the long-term value of performing an action in a given state and require very few assumptions about the environment. Model-based agents learn or are

given a 'model' that captures the temporal structure of the environment, over which they can simulate promising future sequences of actions. Model-based agents have the benefit that, when provided with an accurate model of the world, they can learn faster and generalise better their predictions in novel situations (or after changes in the environment) compared to model-free agents. Research in the past decade has shown that humans are able to exploit a model of the world in order to plan efficiently, and that this has an effect on learning and decision-making neural signals such as the aforementioned reward prediction error (Bornstein & Daw, 2013; Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Doll, Simon, & Daw, 2012).

A classic experiment that explicitly tested whether humans engaged in model-based or model-free learning in humans is (Daw et al., 2011) (see **Fig 1.5**). Participants performed a "two-step" task where they had to learn the value of each action in order to maximise the outcome. Critically, each trial was composed of two steps. In the first step (green), participants selected one of two stimuli, which had a high probability of leading to one of two subsequent decisions (blue and purple) respectively. Interestingly, in this task model-free and model-based learning could be dissociated by looking at the response in the first step (green) of the following trial, as a function of the transition and the outcome. A model-free approach would ignore any information related to the transition of states (i.e. whether the "common" 70% or the "rare" 30% transition happened) and would simply select a green action more often if it had been rewarded. The model-based approach would associate the reward with the second step, and then plan how it could most likely reach that context again. Daw et al. found that human behaviour was better explained by a combination of both

approaches, and this effect was complemented by model-based reward prediction error signals in the striatum.

Further evidence for both model-free and model-based behaviour in humans and other animals has since been reported, and the overall picture seems to be better described by a combination of both (Huys, Cruickshank, & Seriès, 2015; Keramati et al., 2016; Lee, Shimojo, & O'Doherty, 2014; O'Reilly, Jbabdi, Rushworth, & Behrens, 2013; Russek, Momennejad, Botvinick, Gershman, & Daw, 2017).

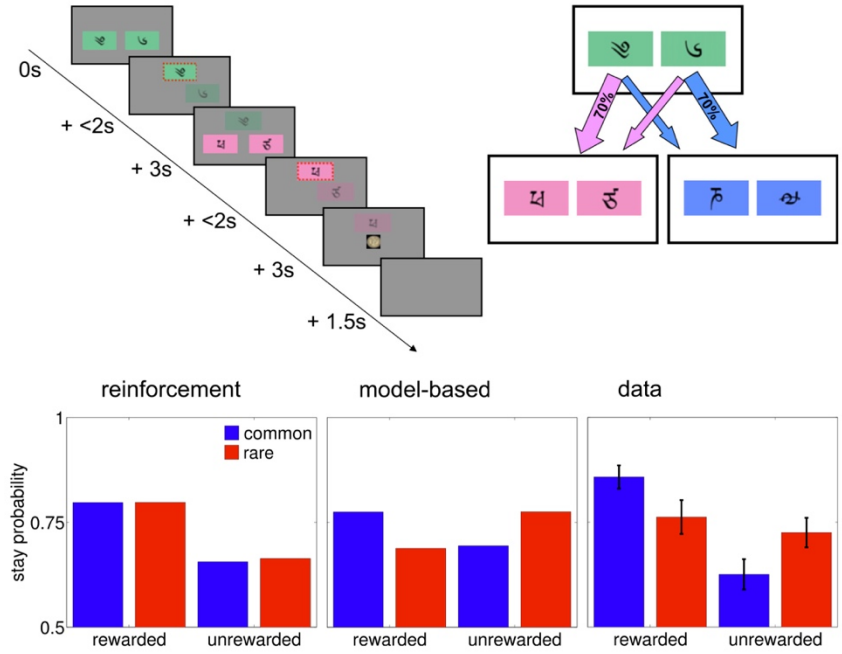

Figure 1.5. Model-free and model-based behaviour in the two-step task (reproduced from (Daw et al., 2011))

5.3 Generalisation as meta-learning

It is unclear how such temporal models (of transitions between states) are learned or exploited at the computational level, or how model-based inference can be extended to non-sequential tasks (without a notion of causality). Inductive inference is indeed

not restricted to exploiting temporal structure. Monkeys have been shown to leverage abstract relationships between objects in order to generalise to novel situations, an approached known as 'learning to learn' or 'meta-learning'. This was most clearly demonstrated in an experiment with monkeys by H.F. Harlow (Harlow, 1949). In this experiment, monkeys were shown pairs of objects and learned to select by trial-and-error the one that was associated with the highest reward. In this case, trials were not sequentially dependent because every trial would show two objects, independently of the choice selected by the monkey previously. In each of a series of blocks of trials, a new pair of objects was used, and only one of the objects was determined as rewarding. Monkeys learned and exploited this reward structure over blocks, displaying an increase of reward obtained over blocks and achieving ceiling performance from the second trial in late (but not early) blocks. In other words, monkeys were able to infer which of the two was the rewarding object independently of the reward obtained in the first trial (generalisation). Preliminary work has aimed to incorporate model-based reasoning with distributed representations (e.g. deep neural networks) (Wang et al., 2018) but this area of research is still in early stages.

6. Aims and structure of the thesis

In this thesis, I aim to explore questions related to the behavioural and neural signatures of hierarchical representations in humans during decision-making. The thesis is structured in four experimental chapters, each aiming to answer a specific question and describing the results of an experiment or set of experiments.

Chapter 2. *Do humans exploit the hierarchical structure of a newly learned artificial set of relationships?*

In this chapter, I develop two behavioural paradigms inspired by the literature in semantic categorisation and aim to reveal signatures of hierarchical representations in humans.

The first paradigm was inspired by the PDP approach, that I describe more extensively. We found that participants' learning was not sensitive to the hierarchical structure that underlay the properties of a set of items, and that participants engaged instead in holistic approaches best explained by episodic mechanisms.

In the second paradigm, we modified our paradigm in the light of the initial results, and found that participants were able to learn an abstract hierarchical structure and transfer it between two behavioural sessions. This latter approach laid down the ground work for an imaging experiment.

The main contribution of this chapter is the development of a behavioural paradigm that asserts the learning of a hierarchical structure. This work enables us (as experimenters) to perfectly control for the hierarchical structure, thus removing potential confounds from the stimuli and the structure, as it is often the case with natural hierarchical representations.

Chapter 3. *What are the neural correlates of hierarchical representations in the human brain?*

In this chapter, I describe a functional resonance imaging study that was adapted from our previous behavioural paradigm and investigate the neural representation of an artificial hierarchical structure. Participants performed two different tasks while undergoing functional magnetic resonance imaging. We found, in both tasks, that the left parietal cortex encoded the hierarchical level of the stimuli in a parametric

univariate fashion, a result that is reminiscent of the encoding of abstract magnitude representations.

The main contribution of this chapter is the investigation of neural representations of newly formed hierarchical representations analogous to those of semantic categorisation and learned within a controlled setting.

Chapter 4. *What are the behavioural and neural correlates of hierarchical representations during goal-directed planning?*

In this chapter, I describe the results of a functional resonance imaging study we conducted, where we examined the behavioural and neural costs of hierarchical representations during goal-directed navigation in a virtual environment. We found that during plan execution, reaction times and imaging signal in prefrontal regions increased linearly with the hierarchical description length of the plan to the goal. Furthermore, we were able to decode the hierarchical context from the dorsomedial prefrontal cortex.

The chapter makes the unique contribution of investigating the behavioural and neural costs of goal-directed hierarchical representations in a task where participants were required to form and execute a plan.

Chapter 5. *What are the behavioural and neural correlates of model-based generalisation and rule discovery in decision making?*

In this chapter, I take a leap and explore human behaviour and functional resonance imaging brain data in a task where a simple underlying rule can be discovered, thus permitting model-based reasoning to generalise beyond simple stimulus-response contingencies. I explain human behaviour in terms of a probabilistic (Bayesian)

model and reveal some learning biases captured by the model fitting procedure. I also describe a set of neural regions (the dorsolateral prefrontal cortex most prominently) associated with the discovery and maintenance of these rules.

This chapter's contribution is to investigate the behavioural and neural mechanisms underlying model-based reasoning that allow humans to generalise their knowledge to novel situations.

Chapter 2. Exploratory behavioural studies on the discovery of latent hierarchical relationships in multi-dimensional datasets

Chapter abstract

The main contribution of this chapter is to describe a novel laboratory-based approach to teaching human participants a hierarchical representation of multi-dimensional stimuli that are (or are not) intrinsically hierarchically structured. I present three behavioural designs together with experimental data. The first experiment aimed to explore the order in which humans learn the levels of an intrinsically hierarchical dataset, inspired by recent research showing that deep neural networks learn the levels of the hierarchical dataset in an ordered fashion, from coarse to fine. However, we found that behaviour of human participants was better described by a memory strategy that simultaneously learned all features within an exemplar, rather than learning higher-level before low-level features. Subsequently, we designed a second experiment where participants learned multiple hierarchical structures in blocks. Two different stimulus sets were used on consecutive days. We found that hierarchical learning on the second day (i.e. new stimulus set) was improved when the dataset on day 1 was also hierarchically organised, compared to a control group that had previously learnt a non-hierarchical dataset. We interpret this as evidence that participants abstracted the hierarchical structure by repeated exposure (across multiple blocks) and 'transferred' this knowledge to the second stimulus set to accelerate learning. We then repeated this experiment with further changes that adapted the design as an imaging study. Building on these findings, we performed an imaging

experiment (described in Chapter 3) that explored the neural signatures of hierarchical representations for novel stimuli.

Chapter introduction

In this chapter, I begin by describing some basic properties of nested tree structures and introduce the concept of singular value decomposition (SVD). Here, we use this mathematical tool to decompose the knowledge contained in a neural network linearly, and to quantify the amount of variance explained by each level in a hierarchical dataset. I will then illustrate through simple simulations the findings of Saxe et al. (Saxe et al., 2013) showing that deep neural networks can qualitatively reproduce the hierarchical dynamics of semantic knowledge acquisition in infants from coarse to fine, as well as their tendency to over-generalise (Keil, 1979; Mandler, 1992, 2000; Mandler & McDonough, 1993; Pauen, 2002). In **Experiment 2.1**, we drew inspiration from these findings to ask whether human participants learned the hierarchical levels of a novel nested tree structure at different rates. We found that unlike for neural networks, in this task learning was not biased towards the higher levels of the hierarchy. Instead, the data suggested that humans relied on a different strategy that involved memorising the exemplars holistically.

In **Experiment 2.2** we used a different approach. In this case, we introduced between-trial temporal correlations in the presentation of hierarchical exemplars, such that stimuli characterised by a common higher level occurred in sequence. We used this approach to teach participants multiple hierarchical structures over consecutive days. We used two datasets on different sessions (fruits and vegetables). Each hierarchical structure thus differed in terms of the visual features that composed each exemplar,

but they all respected the same abstract structure (hierarchical relationship). We found that in this case participants learned the hierarchical structure faster in the second session compared to a control group that learned a non-hierarchical structure during session one, implying that they had indeed acquired knowledge about the underlying structure and reused it to speed-up learning.

We then used a variant of this design (without the control group) that removes a potential spatial confound in our design (**Experiment 2.3).** Here, we found that participants were still able to learn the hierarchical structure, and remembered the specifics of the structures learnt after the experiment in a later probe test. This paradigm was later adapted for a functional neuroimaging experiment (Chapter 3).

In summary, this piloting chapter explored two distinct experiments in which participants could potentially exploit a hierarchical representation to drive learning, and found that this was only the case in one of them. This research laid down the ground-work necessary to design a functional magnetic resonance imaging study described in Chapter 3 where we investigated the neural code of hierarchical representations in the human brain.

Experiment 2.1. Learning dynamics across hierarchical levels in multi-dimensional stimuli

Introduction

Hierarchy discovery through singular value decomposition

Imagine a set of objects (exemplars) such as those in **Fig 2.1a**: a red cog, a red triangle, a green cross and a green square. A hierarchical representation of such a set might, for the sake of compression, group the two red exemplars together on the one hand, and the two green exemplars on the other. Grouping exemplars that share similar features is efficient because it compresses knowledge about the features that characterise each exemplar. This gives rise to high-level representations (e.g. the notion of red-ness and green-ness) that consider some features (e.g. "is red" and "is green") while abstracting over others (e.g. the shape features). Note that features corresponding to the higher levels are common (e.g. two objects are red) while low-level features are rare (e.g. only one exemplar is a triangle). This is an intrinsic property of nested tree hierarchies that are unambiguous (i.e. that only allow one efficient hierarchical decomposition).

In this section I describe a well-known approach, known as Singular Value Decomposition (SVD), to retrieve such a hierarchical representation from a dataset in the form of a matrix (**Fig 2.1a**), with rows corresponding to exemplars, and columns corresponding to features. SVD is an generalisation of what I referred in the *General*

Introduction in **Chapter 1** to as a 'spectral' method, 'diagonalisation', or 'eigen-decomposition'.

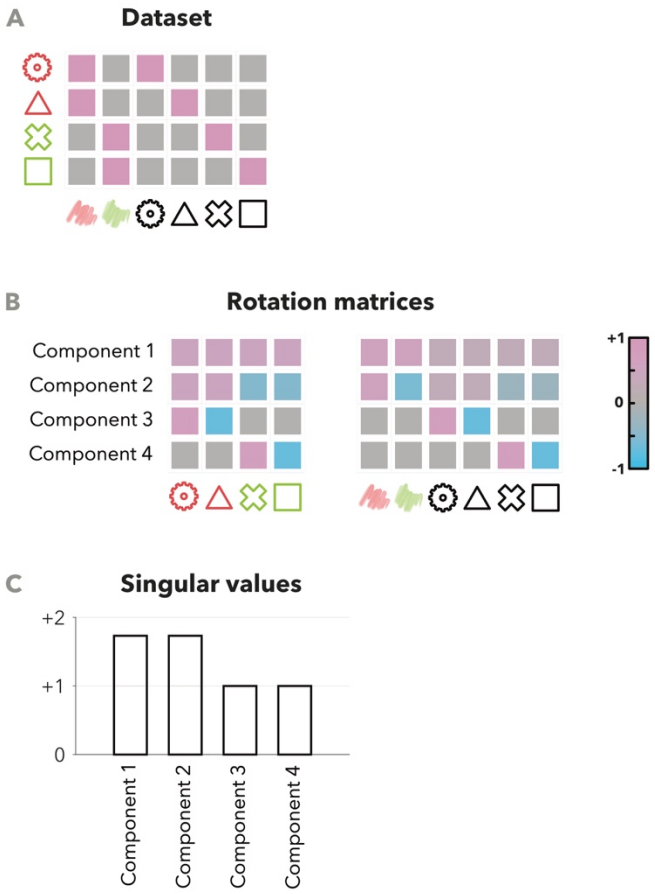

Figure 2.1. Hierarchy discovery through Singular Value Decomposition
A. Matrix encoding the exemplars (rows) and their features (columns). A value of one (pink) implies that an exemplar possesses a given feature, while a value of zero (gray) implies that the exemplar does not possess the feature. **B.** The rotation matrices produced from the SVD method discover the relationship between a set of latent components and the exemplars (left panel) and features (right panel). The first component is constant throughout the exemplars, and reflects the overall frequency of each feature. The second component is positive with red exemplars and negative with blue exemplars. Components 3 and 4 reflect the differences of the exemplars within each category respectively.

For simplicity, in this example I will assume that all features are binary (i.e. yes or no, one or zero), but these could in fact take any value. SVD is a standard mathematical tool for decomposing matrices. Broadly, it extracts a set of latent components that describe the exemplars and their features, such that the original matrix can be

reconstructed as a linear combination of these components. Furthermore, SVD has the property of compressing the matrix, in the sense that a subset of SVD components achieve the best linear approximation of the matrix (similar to low-dimensional projections). This property is common to hierarchical representations, and this is why SVD allows to discover hierarchical structure. Components that explain away more signal in the matrix will thus correspond to higher levels of the hierarchy, while the remaining components will correspond to lower levels.

SVD takes a matrix as its input (e.g. the matrix that encodes all the features of all the exemplars), and computes three outputs. The first two are orthonormal matrices that map the latent components with the exemplars and the features respectively (**Fig 2.1b**). In the example, the first component estimates for all exemplars the average value of the feature, while the second component distinguishes between red and green exemplars and the features associated with each. SVD also yields a third output (**Fig 2.1c**) that estimates the amount of signal explained by each of the components, called the 'singular value'.

This method will find a number of components equal to the number of exemplars, (under the assumption that the dataset has sufficient features). Note that the components can be associated with a hierarchical level. For example, in **Fig 2.1b** components 1 and 2 (with high singular value) correspond to the higher level of the hierarchy, while components 3 and 4 (with low singular value) correspond to the lower level of the hierarchy. However, when each hierarchical category splits in more than two sub-categories the mapping between SVD components and the tree representation may become non-trivial.

SVD can be used to formulate hypotheses concerning how hierarchical representations are built and represented in artificial and biological systems. Within psychology and neuroscience, this method has been used to make predictions about the representation of semantic knowledge in neural networks (e.g. revealing the distinction animal classes) (Saxe et al., 2013) and the representation of space during navigation, in biological brains (Stachenfeld et al., 2014). Other approaches have been brought forward for clustering (Rokach & Maimon, 2005) and for hierarchically decomposing environments in order to improve planning efficiently (McNamee, Wolpert, & Lengyel, 2016; Solway et al., 2014). However, SVD has multiple benefits over these. It is conceptually simple; computationally efficient; it is general in the sense that it can be applied to static multi-dimensional stimuli as well as to sequential decision processes (e.g. for navigation); and it allows continual hierarchical representations (i.e. some exemplars may be encoded more strongly than others within a given component). Note also that SVD is insensitive to the order in which the exemplars or the features are encoded in the matrix.

In **Experiment 2.1** we aimed to understand how an intrinsic hierarchical structure influences learning in biological systems. As we shall see, a deep neural network (DNN) is a computational model that makes very specific and testable predictions about the progression of learning for general and specific features in such a hierarchical structure. We will use SVD in two ways, namely i) to understand the intrinsic structure of the dataset, and ii) to understand the learning dynamics of deep neural networks in the dataset that they trained on.

Learning dynamics of humans and deep neural networks

Deep neural networks (DNNs) are a state-of-the-art Machine Learning model that aim to learn an input-output mapping (LeCun, Bengio, & Hinton, 2015). In the past few years, a body of research has pointed an increasing number of similarities in behaviour and representation between deep neural networks and monkeys/humans (Kriegeskorte, 2015; Ritter, Barrett, Santoro, & Botvinick, 2017; D. L. Yamins, Hong, Cadieu, & DiCarlo, 2013; D. L. K. Yamins et al., 2014). Most interestingly for our purposes, Saxe et al. (Saxe et al., 2013) recently analysed the learning dynamics of artificial deep neural networks (DNN) trained with supervision to predict the features of items arranged hierarchically (e.g. species taxa; similar to **Figure 1.2** in **Chapter 1**, where the network would take "canary" as input, and learn to output "fly"). The authors focused on a special case of DNNs known as deep linear network, for which they could derive mathematically the learning dynamics throughout training. In their work, they found that DNNs learned the SVD components sequentially, such that strongest singular values of a matrix that related the inputs (e.g. an exemplar) and the outputs of the network (e.g. a set of features) were learned earlier during training. In other words, learning of the higher hierarchical level preceded learning of the lower ones in a DNN. Deep linear networks have practical limitations and are mostly of interest for theoretical purposes, but their learning dynamics have shown to generalise to other sorts of DNNs (Saxe et al., 2013).

An example can be seen in **Fig 2.2** where I replicate the simulation in (Saxe et al., 2013). This is not an original contribution of this thesis.

I built a 2-layer neural network (a "deep linear" network) composed of an input layer (4 units, coding object identity), one hidden layer (4 units), and an output layer (6

units, coding object features) (**Fig. 2.2a**). The network's input layer receives information about the current exemplar, with a value set to 1 in the position corresponding to the object identity (e.g. the red triangle) and zeros everywhere else ('one-hot encoding'). Thus, the input does not directly code information about the features corresponding to the exemplar; these are predicted at the output layer. The network is trained in a supervised fashion to output the correct features associated with each exemplar. For example, when the input corresponds to a red triangle, the correct output should be 1 for "is red" and for "is a triangle" and zeros everywhere else. The network computes the values in the hidden layer by multiplying the inputs with a matrix of weighs w_1, and the values in the output layer by multiplying again with a second matrix of weights w_2. Note that this *deep linear* network does not include any non-linearities. At first, weights were initialised with a normal distribution around zero (standard deviation 0.001). The network's output was thus approximately zero for all features before any learning occurred (**Fig 2.2b**).

The network's weights were updated with back-propagation (learning rate 0.03) (Rumelhart et al., 1986), in order to decrease the Euclidean distance between the network's output and the true features of each exemplar. I trained the network in 50 independent simulations with random weight initialisation in order to obtain statistics. The error decreased until convergence after around 300 steps on average (**Fig 2.2b**).
This learning was also reflected by the timecourse of the singular values computed from the output of the network (**Fig 2.2c**). For this, at each timestep I performed SVD on a matrix that encoded the features predicted by the network for each exemplar.

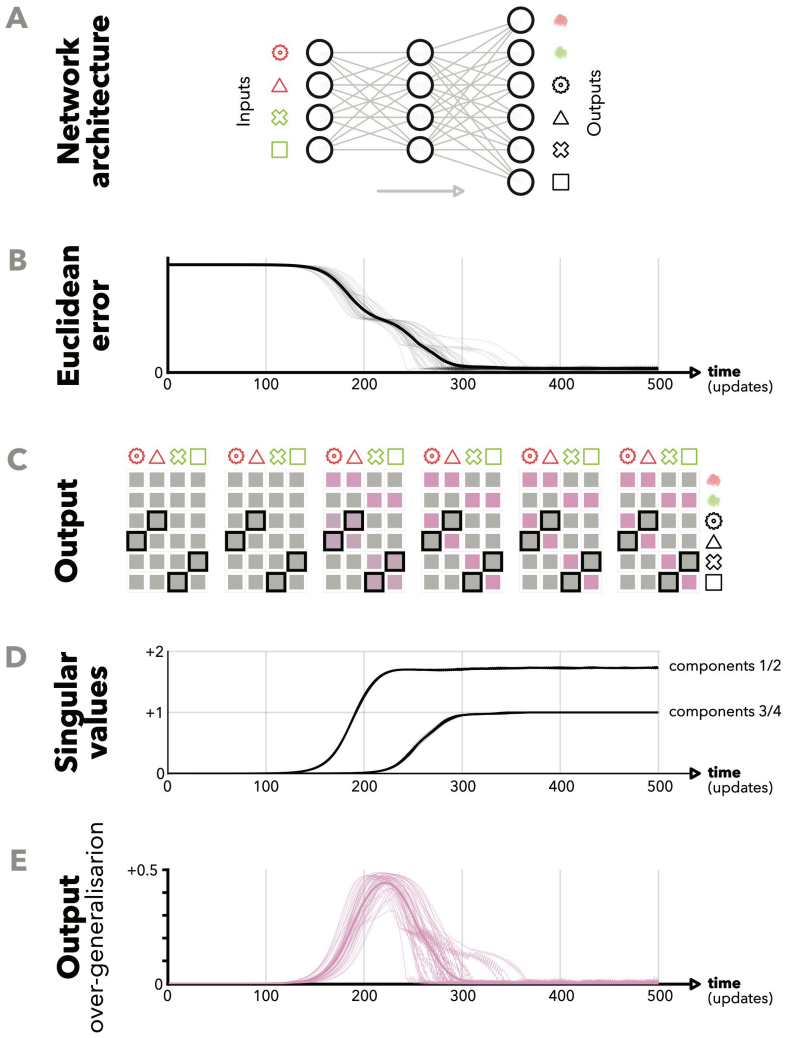

Figure 2.2. Hierarchical learning dynamics in deep neural networks
A. Architecture of the 'deep' neural network. The network received the identity of the exemplar as its input and was trained to predict the features associated with each exemplars. B. The euclidean error between the predicted and the target features decreased with every update until convergence. C. Predicted features of the exemplars by the network throughout the training. Initially, the network predicted that exemplars had no features. D/E. Computed singular values of the network's output. After 200 updates, the network learnt the difference between red and green exemplars but not the difference within each of the category. This induced a pattern of over-generalisation (E) where the network responded positively to features associated with the other exemplar of the same category (e.g. a red cog is a triangle).

Interestingly, this analysis revealed that during learning the two components with the strongest singular values converged earlier (around ~230 updates) than the two weakest components (around ~300 updates; **Fig 2.2d**). This was consistent with a pattern of over-generalisation for a subset of outputs (**Fig 2.2e**). For example, in the interval where the network had learnt the higher but not the lower hierarchical levels, it responded that a red cog had the feature "is a triangle", because it had learnt that red things but not green are often triangles.

This simulation replicates findings that were previously reported in the context of Machine Learning (Saxe et al., 2013). Neural networks have also been proposed as a computational model for well-known hierarchical patterns of semantic knowledge acquisition during infancy. It is well-known that infants tend to over-generalise, for example by stating that worms have bones (like other animals) and that penguins can fly (like other birds) (Keil, 1979; Mandler, 1992, 2000; Mandler & McDonough, 1993; Pauen, 2002). This has often been taken as evidence that knowledge is thus acquired hierarchically, from general to specific. The work of Saxe and others previously (McClelland et al., 1995; Plunkett & Sinha, 1992; Quinn & Johnson, 1997; T. T. Rogers & McClelland, 2004; Rumelhart & Todd, 1993; Saxe, 2013) has shown that 'distributed representation' systems (e.g. deep neural networks) can reproduce these learning dynamics. We reasoned that if these are general principles that apply to hierarchically structured facts, then we could extrapolate these ideas to other domains. For example, could we uncover similar hierarchical learning dynamics i) in adults, and ii) over a short experimental session rather than over the whole developmental trajectory? We designed a memory task based on a set of multi-dimensional

exemplars in order to look at the pattern of responses. We predicted that learning of higher levels of the hierarchy would precede those of lower levels.

Experiment conceptualisation

In **Experiment 2.1** our goal was to test if the dynamics of semantic knowledge acquisition in adult humans as described above also apply to shorter timescales. We were particularly interested in whether learning of the higher level of the hierarchies (that are more general, and less sparse) preceded that of the lower levels. In order to test this question, we devised a behavioural task that involved approximately one hour of training. We defined a hierarchical dataset where each exemplar was defined by a set of features as well as a label. Unlike in most category tasks (but like in the simulation above), participants were cued with the label, and where they were instructed the select the multiple features corresponding to the exemplar. This allowed us to estimate, per trial, the amount of learning for each of the hierarchical levels.

As I have explained in this chapter, (hierarchical) nested tree structures can be unambiguously decomposed but this requires more diverse and sparser features in lower levels of the hierarchy. This has the consequence that features in the higher levels of the hierarchy also occur more frequently (e.g. in **Fig 2.1** more exemplars have the feature "is red" than the feature "is a triangle"). Research in statistical learning has shown that more frequently-occurring features are likely to be learned faster and more effectively (Zhao, Al-Aidroos, & Turk-Browne, 2013). Hence, in order to make any claim about hierarchical learning beyond the effect of frequency, we were required to introduce a control group that learnt a structure with matched frequencies but that did not respect a hierarchical organisation.

Methods

Participants

We recruited a total of 60 human participants via Amazon Mechanical Turk (http://www.mturk.com) in accordance with ethical guidelines approved by the Oxford University Medical Sciences Ethics Board. Participants were randomly and evenly allocated into two groups termed "hierarchical" and "control" (see *stimuli and task design* section). They were paid $8 for participation in a one-hour session. A bonus monetary incentive of up to $5 proportional to performance was added to the previous amount. Six participants were excluded due to poor performance on the task (2 in the *hierarchical* and four in the *control* groups; accuracy not significantly higher from chance in the "evaluation" phase of the experiment ($p < 0.01$, under a cumulative binomial probability distribution assuming chance level).

Hierarchical and control sets of stimuli

In **Experiment 2.1**, participants were asked to learn the features associated with exemplars (eight unique visual stimuli, 'ponies'). Each pony was composed of a label (a pseudo-random word, its name) and of visual features organised across seven *dimensions*: mane, skin pattern, tail, sunglasses, hooves, bracelets, and smile. Each dimension could be present or absent; if present, it could take on one of two features (e.g. skin patterns in stripes or patches). At the beginning of the experiment, we instructed participants of the possible features that each dimension could take, and they were allowed to observe them. In the *hierarchical* group, we defined a hierarchical set of ponies by defining conditional relationships among the features, such that a specific feature would determine the presence of another dimension (e.g. ponies with skin patches have sunglasses) or absence (e.g. ponies with skin stripes do

not have sunglasses). The hierarchy included three levels (**Fig 2.3a**; red: top, yellow: middle, green: bottom) and all pony dimensions belonged to one of the three hierarchical levels. In the *control* group, we constructed a set of ponies that respected the frequencies of the dimensions and features but was not hierarchical. We use the terms top, middle, and bottom levels for the *control* group (in terms of equivalent frequencies) even if it does not have the same meaning in terms of a hierarchical structure. In both groups, we constructed a total of 8 ponies per participant, each including 3 present dimensions and a label. The dimensions and features were randomly permuted for each participant. Names were generated pseudo-randomly, with the constraints that they should include three consonants interspersed with two vowels. An example of the visual stimuli composing the hierarchical and control structures is illustrated in **Fig 2.3a**, and a table describing their underlying structure at an abstract level can be found in **Table T2.1**.

Note that the hierarchical and control structures we made use of here differ slightly from the example given in the deep neural network simulation (**Fig 2.1**). First and most importantly, we grouped the features into dimensions because this gave us a more explicit representation of the hierarchical levels. This means, for instance that ponies with a stripy skin could not have a patchy skin. This notion of "grouping" multiple, mutually exclusive features into a dimension is neither embedded in a matrix representation nor captured by the SVD analysis (for which a pony could in principle have both patchy and stripy skin), but it can play a role in how humans learn the task. Intuitively, we expected that this change may be beneficial to the *hierarchical* (where a feature was predictive of the presence of another dimension)

but not the *control* group. Second, each dimension was trinary (present and taking one of two possible features, or absent).

Task design

Participants were not made aware in either of the groups of the potential for a hierarchical structure in the stimulus set. Instead, they learnt by trial-and-error the features associated with each of the ponies (see **Fig 2.3b**). On each trial, they were cued with the name of a pony and were asked to select three features (from mutually exclusive dimensions) using a set of buttons on the left side of the screen. A visual stimulus in the centre of the screen showed the features selected at any one point. Once three features had been selected, participants could submit the response by clicking on the pony. Responses were allowed within a window of 60 seconds (responses beyond this deadline were deemed incorrect) before the experiment moved on. The response was followed by a feedback screen of 2 seconds that displayed the correct response before a new trial started. Ponies were selected pseudo-randomly across trials, such that all names would be cued within covert batches of eight consecutive trials.

The experiment was divided in two phases. The first "training" phase had a total length of 160 trials. At the end of the first phase, participants were asked to complete a second phase ("evaluation") of 48 trials, and were made aware that the response accuracy during this second phase alone would determine the monetary incentive received at the end of their participation. The evaluation phase was otherwise identical to the learning phase, and in particular it also provided feedback. We included this second evaluation phase in order to obtain independent data to discard participants that failed to perform above chance (as stated in the *Participants* section).

A demo of the task can be found at:

http://185.47.61.11/sandbox/tasks/jan/thesis/pony/?skip=1

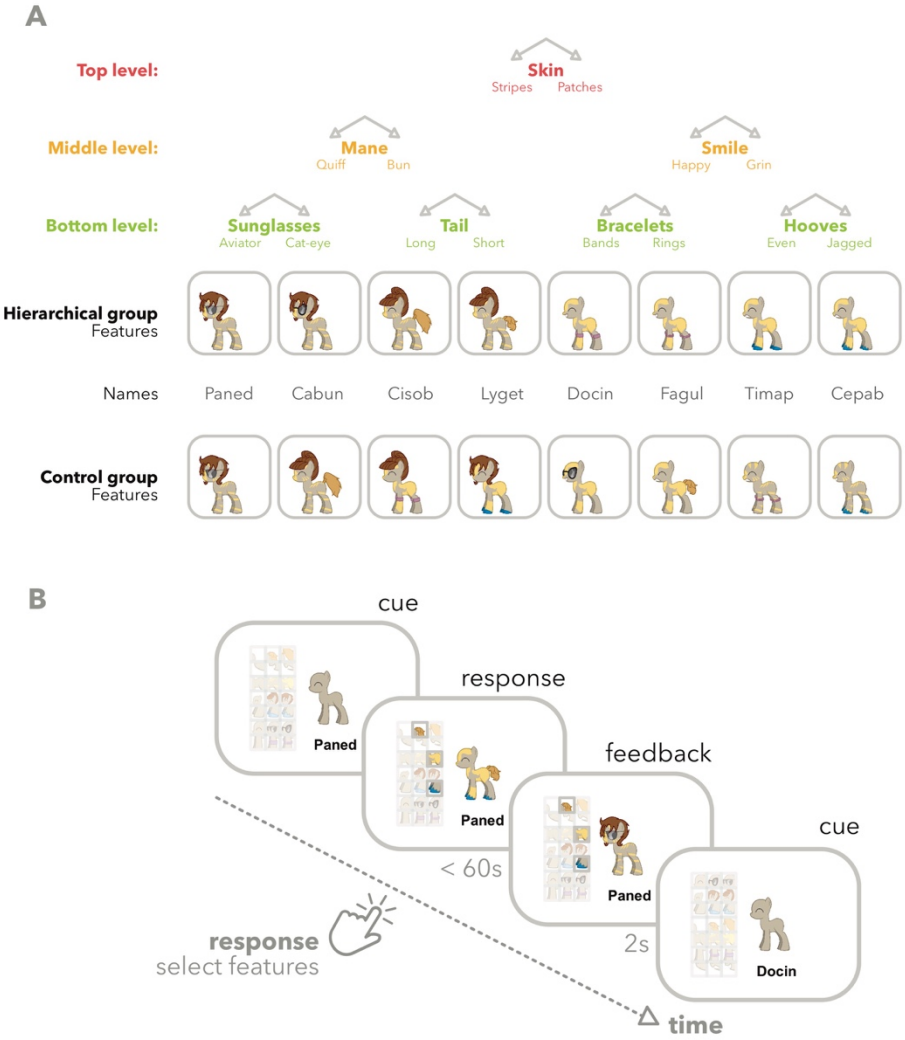

Figure 2.3. Experiment 2.1 — Task design and stimuli
A. For each participant, we defined a set of eight ponies defined by a name and a visual stimulus. Pony stimuli were composed of seven dimensions that could be either present (and take one of two feature values) or not. In the "hierarchical" group the features were organised hierarchically, such that the feature on a given dimension (e.g. skin stripes) would determine the presence or absence of the dimensions in lower levels of the hierarchy. Dimensions in the "hierarchical" group could be categories as belonging to the top (red), middle (yellow) or bottom (green) levels of the hierarchy. We defined an alternative set of ponies for a "control" group of participants, that respected the frequencies the features but did not follow a hierarchical relationship. **B.** Illustration of a trial. Participants were cued with a name and selected three present features from mutually exclusive dimensions and submitted their response by clicking on the central pony. This was followed by a feedback screen that displayed the features that were actually associated with the name for 2 seconds, before a new trial began.

Task and analysis implementation

We implemented the task with HTML and JavaScript. We used Raphaël JS (http://dmitrybaranovskiy.github.io/raphael/) to display all visual stimuli. The visual stimuli ('ponies') were obtained from a web application named Pony Creator (https://generalzoi.deviantart.com/art/Pony-Creator-Full-Version-254295904). All behavioural and computational analyses were done with custom scripts for Matlab (Mathworks, Natick, MA, United States of America).

Results

Evaluation accuracy

We discarded 6 participants whose overall accuracy was not significantly higher than chance level during the evaluation phase, leaving us with a total of 54 participants. For the remainder, accuracy in the evaluation was generally high: 88.5 ± 3.5% correct responses in the "hierarchical" group and 86.0± 3.9 % correct responses in the "control" group. After discarding participants at chance, performance did not differ significantly between the two groups ($t_{(52)} = +0.47$, $p = 0.64$).

Accuracy during learning

We then looked at the pattern of responses during the "training" phase. We first computed the proportion of responses where all the features selected (i.e. all dimensions and their features; e.g. a patchy skin pattern) were correct. This accuracy measure increased from chance level (~0.3%) to 63.0 ± 4.8% correct responses (including all features, averaged across both groups) over the course of training, indicating that participants had significantly learnt the structure, and that similar amount of learning had been achieved for both groups (**Fig 2.4** top panels, gray line;

hierarchical: 67.9 ± 6.0%; control: 57.7 ± 7.5%; unpaired t-test: $t_{(52)} = 1.05$, $p = 0.3$). We then examined the accuracy for each hierarchical level individually by estimating the proportion of 'hits' per trial and per dimension. For this, we calculated the proportion of trials where a participant correctly selected the true feature independently for each of the three present dimensions (corresponding to the top/middle/bottom hierarchical levels), ignoring the responses of all the other dimensions. We tested for statistical significance using a mixed-effects analysis of variance (ANOVA) with 3 hierarchical levels (top, middle, and bottom; fixed effect) and 2 groups (hierarchical and control; random effect). Note that the chance level was matched for all hierarchical levels (~21.4%). This yielded a significant difference across levels (**Fig 2.4** top panel, lines red/yellow/green), with faster learning for the top level ($F_{(2,52)} = 3.18$, $p = 0.045$) but no difference between both groups ($F_{(1,52)} < 1$, $p = 0.77$) and critically no interaction ($F_{(2,52)} = 1.73$, $p = 0.18$). This suggests that although there was (weak) evidence for faster learning of the dimension we had designated "higher", this did not depend on the structure of the task, and was probably thus due to the higher frequency of occurrence of features at the higher level.

In order to understand the pattern of errors in the participants' responses, we also looked at the average number of times per trial that an incorrect feature was selected but whose dimension was correct (**Fig 2.4**, middle panel) or not (**Fig 2.4**, bottom panel). We reasoned that participants in the "hierarchical" group could be focusing on the dimensions rather than the features and would display more incorrect responses for correct dimensions, compared to the "control" group. A second mixed-effects ANOVA yielded a main effect of hierarchical level ($F_{(2,52)} = 12.2$, $p < 0.001$), implying that participants learned to select the dimension corresponding to higher levels of the

hierarchy (e.g. always select a skin pattern) even when the feature selected was incorrect (e.g. sometimes they selected a patchy instead of the correct stripy skin). However, we found no difference between the two groups ($F_{(1,58)} < 1$, $p = 0.96$) and no interaction ($F_{(2,58)} < 1$, $p = 0.55$). For completeness, we provide an analysis of the reaction times in the supplementary information (**Supplementary Figure SF2.1**), although this was not of key interest to our hypothesis.

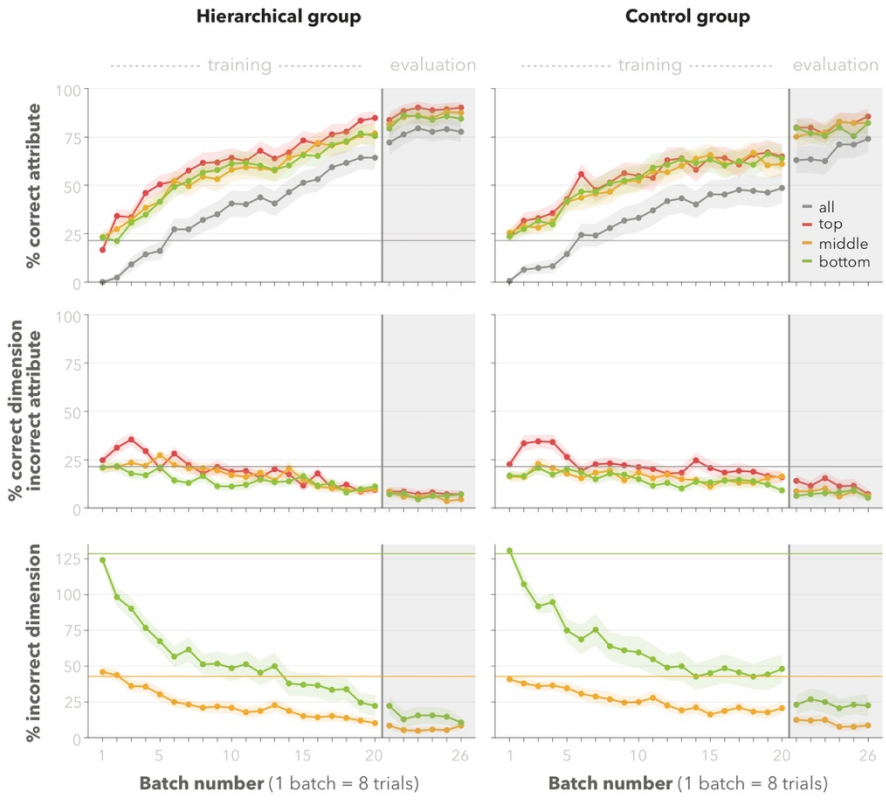

Figure 2.4. Experiment 2.1 — Results
Pattern of responses for participants in the hierarchical (left panels) and control (right) groups, independently for the top (red), middle (yellow) and bottom (green) levels of the hierarchy. Top panels show the proportion of times that a feature was correctly selected (grey lines represent the proportion of trials where all features were correct). Middle panels represent showed, similarly, the proportion of trials where the incorrect feature of the correct dimension was selected. Bottom panels show, for middle and bottom levels of the hierarchy, the proportion of trials where an alternative dimension of the same hierarchical level was selected. On every trial, participants could only provide up to three features from mutually exclusive dimensions. Horizontal lines represent the chance level, i.e. the proportion of responses under uniformly random policy.

Statistical dependencies of learning across hierarchical levels

In these analyses, we have looked at the learning curves for each of the hierarchical levels while disregarding the performance achieved for the other two. In other words, we have assumed that participants learned each of the levels independently. We found that learning in the top-level generally reached higher accuracy, but that this also the case in the non-hierarchical group and thus was likely to reflect an overall effect of feature frequency, rather than a hierarchical representation.

One possibility is that participants learnt each of the ponies independently. In this case, one would expect a positive correlation in the accuracy of feature reports for the three hierarchical levels across ponies, such that if a participant correctly responded to a feature, she should also be likely to respond correctly to the other features. In what follows we explore the dependency of accuracy during learning across the levels. We followed two distinct, but mathematically equivalent approaches to quantifying this dependency.

In our first approach, we compared the empirical proportion of trials where participants correctly selected all the features against that predicted by assuming independence. Indeed, under mathematical independence one could predict the probability of responding correctly to all levels by multiplying the probability of being correct for each of the levels. When we compared these two values, we found that they differed significantly both in the hierarchical ($t_{(27)} = +11.0$, $p < 0.001$) and in the control group ($t_{(25)} = +9.6$, $p < 0.001$) without any significant different across groups ($t_{(27)} = +1.4$, $p = 0.167$). We interpreted this result as a positive correlation in the accuracy across groups, such that participants were more likely to correctly respond to multiple hierarchical levels (or none) than to just one or the other. One possible

explanation for this would be that participants structured their learning by exemplars (i.e. learn one pony at a time) instead of by dimension (i.e. learn one dimension at a time).

We explored this possibility by estimating the accuracy for one hierarchical level, conditioned on the accuracy for different hierarchical level. If we assume independence, then these two probabilities should match. We found however, in line with our previous analysis, that accuracy increased when separately looking at trials with correct vs. incorrect responses for a different level (hierarchical group: all $t_{(27)} > +10.5$, $p < 0.001$; control group: all $t_{(25)} > +7.1$, $p < 0.001$). This is mathematically equivalent to the previous analysis, but it was nevertheless a more intuitive approach and it allowed us to confirm our interpretation of the results. This was also the case when limiting our analyses to data during the evaluation phase, or when repeating this analysis independently for each trial number. We thus conclude that participants studied each of the ponies holistically, rather than considering the hierarchical structure of the dataset.

Discussion

In this behavioural experiment, we aimed to find signatures of learning dynamics that reflect the hierarchical structure of a dataset composed of multiple dimensions. We tested the idea that hierarchical structures could be exploited in a memory task where an efficient hierarchical representation that compresses the relationship between exemplars and features can help reduce the amount of cognitive load. However, we failed to observe differences in the pattern of responses during learning between participants learning a hierarchical task and a control group.

There are many possible explanations for why we found no differences between the hierarchical and control groups. We cannot rule out the possibility that we lacked statistical power despite each group included more than twenty participants. Indeed, the plotted learning curves were qualitatively different from our deep neural network simulations, and they largely overlapped between both groups.

It is possible that the unbalanced saliency or distinctiveness across dimensions and features captured most of learning and that this carried most of the variance in what participants learnt and paid attention to. For example, the "mane" was a larger feature than some of the others (e.g. "hooves") and may have been a particularly salient cue to the pony identity. We found indeed that there was a main effect across the seven dimensions ($F_{(6,52)} = 11.0$, $p < 0.001$) that did not interact between the two groups ($F_{(1,58)} < 1$, $p = 0.91$; see also **Supplementary Figure SF2.2**). The relationship between saliency and the position within the hierarchy should however be randomised across participants and not interact with other effects of interest. For instance, we found an effect of hierarchical level in both the "hierarchical" and "control" groups that we interpret as driven by the frequency of occurrence of the features, in line with research in statistical learning.

Another possibility is that learning structured across features was impeded by the fact that all dimensions were easily bound within a unitary visual construct (object, pony). We found evidence in the response dependencies revealed across the hierarchical levels, even if the instructions explicitly outlined all the possible features (dimensions and features) and hence in principle the possibility of binding the visual dimensions

should have had little to no effect. In other words, it is possible that participants memorised the ponies holistically, rather than structuring their learning by the hierarchical relationship of the features across the ponies. One possible solution to avoiding this strategy in future participants would be to increase the cognitive load. For example, we could have expanded the hierarchical structure with an additional (fourth) hierarchical level, thus expanding the set of exemplars to 16 ponies in total. In theory, this could induce a benefit for the hierarchical group, but in practice the set of relationships would most likely have been too great for participants to learn within a single behavioural session.

We note in passing that the names (labels) of the ponies were random by design and thus did not reflect the hierarchical relationship present in the visual features. This was a choice made during the design of the task in order to provide parity with the neural network simulations, but it is possible that the different levels of the hierarchy could have been learnt at different speed if their names had also shared some properties.

Altogether, these data suggest that in our task the learning dynamics may be dominated by qualitatively different mechanisms from those of learning during development or that spans longer periods of time. The notion of two different learning systems is not novel and has been most clearly framed within the Complementary Learning Systems theory (Kumaran, Hassabis, et al., 2016; McClelland et al., 1995), where the hippocampal formation quickly encodes episodic memories before these are slowly consolidated within distributed, cortical representations. In **Experiment 2.1** participants were only exposed to a single instance with the hierarchical structure.

Indeed, previous research has shown that exposure to multiple instances with the same underlying relationship is necessary in order to exploit an abstract structure (Harlow, 1949) and that sleep may benefit structure learning (Schapiro et al., 2017). If this is the case, our findings suggest that participants did not reveal or exploit the underlying (hierarchical) structure across features of a single dataset within a behavioural session and instead learned each item (pony) hollistically.

These results seem to diverge from those of the deep linear network simulation that I presented in the introduction. The theoretical findings in (Saxe et al., 2013) dictate that the learning dynamics of deep linear networks are determined by the Singular Value Decomposition (SVD) of the dataset which is known to be different between the hierarchical and control groups. Additionally, a deep linear network should not learn the exemplars holistically and thus would not display the statistical dependencies of accuracy discussed previously.

Having said this, there were non-trivial differences between the experimental design and the introductory simulation. Among them, participants in the experiment were forced on each trial to select three features from mutually exclusive dimensions. It is not straightforward to implement such a constraint in a deep linear network, and this would limit the interpretation of a quantitative modelling of the data.

Experiment 2.2. Transfer of hierarchical representation of multi-dimensional stimuli

Introduction

The previous experiment was optimised to address the question of hierarchical learning dynamics and did not allow us to address other potentially interesting research questions related to the representation of hierarchically structured datasets. For instance, the design in **Experiment 2.1** was not optimised to investigate whether there were any behavioural biases (e.g. in reaction times) driven by the hierarchical structure after the participants' learning had converged. Additionally, a deep neural network does not explicitly represent in its hidden activations the hierarchical decomposition that drives its learning dynamics (i.e. activity in each neuron in the hidden layer is sensitive to information about the lower levels of the hierarchy as well as the higher level) and once learning has converged it would, by definition, not display any biases due to representing information hierarchically (e.g. over-generalisation).

Secondly, we wished to develop a behavioural paradigm that could be adapted as an imaging experiment later on. In the previous task, we could not address the neural representation of the different hierarchical levels in the human brain because these were confounded with the *variability* (i.e. number of different features per level) and *rate* (i.e. how many exemplars were associated with each feature) of the features associated with each hierarchical level.

We thus introduced some fundamental changes in **Experiment 2.2** that addressed most of the points raised in the discussion of **Experiment 2.1**, but we used a task that

was similar in spirit. These changes also made the behavioural design easier to adapt as an imaging experiment where we could investigate the neural signatures that might indicate the presence of hierarchical representations.

First, we disposed of the word labels (e.g. pony names) associated with each exemplar. This was motivated by the fact that labels were arbitrary and did not relate to the hierarchical structure, and so could encourage memorisation strategies. Instead, we cued participants with partial exemplars and instructed them to complete the missing feature. This change also allowed us to formulate our study as a four-alternative forced choice task.

Second, we decided to change the stimuli to artistic drawings of ingredients (fruits and vegetables) in order to ensure that they were distinctive and equally familiar. In **Experiment 2.1** we found that i) the features differed in their salience, and ii) using features that could be easily bound into a single visual construct could induce a holistic learning strategy, where the object is learned as a whole. In this case, ingredients represented the features of an exemplar, and the exemplars themselves were salads (triplets) that combined the ingredients. We expected that these stimuli would be better balanced in terms of saliency, distinctivity and familiarity, and that participants would not be able to represent them visually as a single object.

Third, we defined a distinct hierarchical dataset in which the features associated with all levels were similar in terms of variability and rate. We achieved this by making the hierarchical structure redundant, such that there were multiple possible nested tree structures over features in different blocks (see **Fig. 2.5b**). However, this change

weakened the intrinsic hierarchical structure of the dataset, because we reused the features corresponding to the lowest level of the hierarchy across blocks. To counter the ambiguity in the mapping from features to tree nodes that this entailed, we introduced a temporal correlation in the features across trials for the top but not the bottom level of the hierarchy (see **Fig. 2.5d**). This was further inspired by i) research showing that humans may learn a structure better when events are correlated in time (Garvert, Dolan, & Behrens, 2017; Schapiro et al., 2013), and ii) evidence that exposure to multiple instances with the same underlying relationship can promote learning of an abstract structure (Harlow, 1949). We achieved this by defining a temporal structure such that within each block of trials we hoped to induce a hierarchical nested tree representation of the features and exemplars. This abstract hierarchical relationship (nested trees with depth 2 and branching factor 2; see **Fig. 2.5b**) was common to all blocks, but the relevant features varied from block to block. This was unlike in **Experiment 2.1**, where participants were exposed to a single instance of a hierarchy, and thus could overfit to it rather than learn its structure (relationship) independent of the surface features. This change allowed us to teach participants multiple instances of hierarchical datasets, and critically across blocks (i.e. over the full length of the experiment) all levels were equally complex in terms of variability and rate of the features.

However, one drawback of this change was that the temporal structure of the trials limited our ability to explore the learning dynamics as we did in **Experiment 2.1**. This is because participants could rely on short-term statistics that were satisfied within a block (e.g. the top-level feature was constant within a block) but that weren't true over all blocks.

Fourth, participants performed two behavioural sessions. They were exposed to similar structures in consecutive days such that the superficial features (i.e. the ingredients) were exclusive to each session but the underlying hierarchical relationship was common to both. This was partially motivated by evidence that structure learning may occur preferentially following consolidation processes that occur during sleep (Schapiro et al., 2017). In **Experiment 2.2** we tested whether learning during the second session was faster compared to a control group that learned the exact same structure but with a control temporal correlation during the first session that prevented them from learning a hierarchical representation. To preview the findings, we found that this was the case, and interpret this as evidence that participants had learnt and transferred the underlying hierarchical structure across sessions.

Methods

Participants

Participants were recruited via Amazon Mechanical Turk (http://www.mturk.com) and paid $24 for participation in accordance with ethical guidelines approved by the Oxford University Medical Sciences Ethics Board. In **Experiment 2.2** we recruited a total of 48 human participants (22 female, 25 male, 1 unknown; age 18–60+, mean 34.8 years) and randomly allocated into two groups termed "hierarchical" ($N_{hierarchical}$ = 22) and "control" ($N_{control}$ = 26).

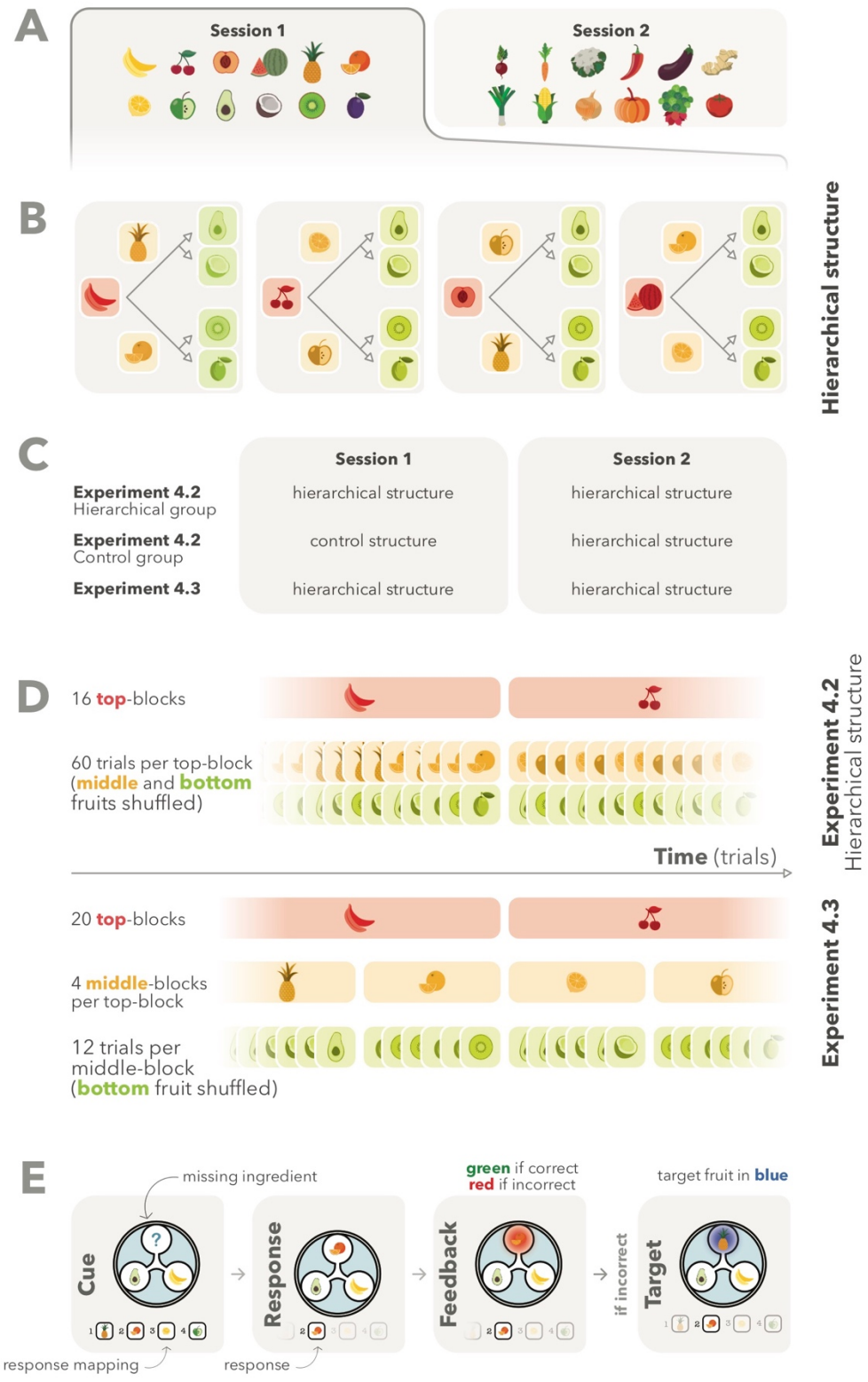

Figure 2.5. Experiments 2.2 and 2.3 — Feature-completion task design.
A. Two sets of 12 ingredients (fruits or vegetables, mutually exclusive) were used for each learning session on sessions one and two. B. We constructed a total of 16 triplets from the 12 ingredients corresponding to each session, such that ingredients would belong to a hierarchical level (top, middle, or bottom). In the hierarchical structure,

triplets were grouped in four types of blocks that formed nested tree representations. The arrows illustrate the hierarchical relationship. For example, a banana could be paired with a pineapple or with an orange, and a pair banana-pineapple could be associated with an avocado (hence the triplet banana-pineapple-avocado) or with a coconut. C. Table describing the different experiments and groups, as well as the structure that each group learnt on each session. D. Trials in the feature-completion task followed a temporal structure. In Experiment 2.2, trials were grouped into blocks with a constant ingredient corresponding to the top hierarchical level. In Experiment 4.3, trials were grouped into middle-blocks (with a constant ingredient corresponding to the middle hierarchical level) that were in turn grouped into top-blocks (with a constant ingredient corresponding to the top hierarchical level). E. Each trial was composed of a cue, a response, a feedback/target screen. At the time of the cue, a triplet was shown with missing ingredient, together with four alternative responses. After participants selected the response, the feedback informed participants if their response was correct (in green) or not (in red). If the response was incorrect, one last screen displayed the correct triplet (in blue).

Experiment overview

In each experiment, participants performed two learning sessions (**Fig 2.5a**) on consecutive days. On each session, the task only differed in the set of stimuli used (ingredients; either fruits or vegetables). They were performed over the internet for a total duration of less than 90 minutes per session. Participants were instructed to learn via trial and error to associate triplets of ingredient stimuli. On each trial, they were asked to complete a missing ingredient from one of the triplets (*feature-completion* task; **Fig 2.5e**), receiving deterministic feedback. For all experiments and groups, participants completed a total of 960 trials in the feature-completion task.

A demo of the tasks is available at:

Experiment 2.2: http://185.47.61.11/sandbox/tasks/jan/fruit_2a/

Stimuli and hierarchical structure

We pre-defined a set of hierarchical structures that participants were instructed to learn (**Fig 2.5b** and **Tables T2.2** and **T2.3**). These structures were composed of triplets of cartoon images of fruits and vegetables ("ingredients" from now on)

contained within three placeholders (top, left and right; see **Fig 2.5e**). The stimuli followed a hierarchical structure in the sense that there was a sequential dependency between the ingredients. This dependency was two-fold: it was reflected in the stimuli (**Fig 2.5b**) but also in the temporal association across trials (**Fig 2.5d**; see *Temporal structure* section below). For example, if a triplet included a banana then it also included a pineapple or an orange. Similarly, if a banana and a pineapple were members of a triplet, then the third ingredient could only be an avocado or a coconut. Hierarchies were thus composed of three levels: *top level* ingredients (e.g. banana), *middle level* ingredients (e.g. pineapple), and *bottom level* ingredients (e.g. avocado). The position of the ingredients during the feature-completion task was fixed, such that the top-level ingredients were shown in the top placeholder, middle-level ingredients were shown in the left placeholder, and bottom-level ingredients were shown in the right placeholder.

In each session we constructed 16 triples from a set of 12 distinct ingredients, such that each ingredient could only belong to one level, and such that the frequency was balanced (four ingredients per hierarchical level, occurring equally often across triplets). Each learning session made use of distinct, non-overlapping sets of ingredients (e.g. fruits on session 1 and vegetables on session 2; **Fig 2.5a**). The identity of the ingredients within the triplets was randomised for each participant, but the relationship of the triplets (abstracted over the ingredients themselves) was fixed and identical for all participants and between the two learning sessions. We created an additional control structure (**Table T2.2**) for the first session of the control group in **Experiment 2.2** that respected the overall frequencies but that could not be easily represented hierarchically. Critically during the second session, both groups were exposed to the hierarchical structure such that any difference in performance could

only be explained as a transfer of knowledge from the first session (see **Fig 2.5c**), and the failure to observe such a difference would be indicative of a memorisation strategy. The fruit and vegetable stimuli were obtained from the ConceptDraw Food Court solution (http://www.conceptdraw.com/samples/food-beverage-food-court) under Attribution-NonCommercial-NoDerivs 3.0 Unported license (CC BY-NC-ND 3.0; https://creativecommons.org/licenses/by-nc-nd/3.0/).

Temporal structure

We introduced temporal structure to the trials of the feature-completion task (**Fig 2.5d**). We expected that the temporal structure would induce a hierarchical representation of the triplets except when for learning of the control structure (i.e. first session of the *control* group in Experiment 2.2). In **Experiment 2.2**, we made use of temporal blocking for both groups, and we expected this would induce a hierarchical representation during the first session for the "hierarchical" but not for the "control" group. Each session included 16 blocks composed of 60 trials formed by the four triplets (see **Table 2.2**). When the exemplars followed a hierarchical structure, exemplars within a block shared the same top-level ingredient (on the top placeholder), and the middle-level (left placeholder) could only be one of two possible ingredients. When the exemplars followed the control structure, each placeholder (top, left and right) could contain four possible exemplars.

Notice that the temporal structure manipulated locally, by controlling the frequency of the ingredients for each hierarchical level only within a block. The temporal structure did not have an effect over the length of the task, where participants saw exactly 240 times each ingredient, independently of the hierarchical level it was associated with.

Feature-completion task

On each trial (**Fig 2.5e**) two ingredients and a question mark appeared on screen within a disk composed of three placeholders. The question mark would indicate the position of the missing ingredient. Participants could then choose among four alternative ingredients (all belonging to the same hierarchical level of the relevant session) by pressing one of four keys in the keyboard. Participants could provide a response within 20 seconds, after which a lack of response was registered as incorrect and the experiment moved on to the next trial. After a response was registered, the selected ingredient would be displayed on screen surrounded in green for 500 milliseconds (if correct) or for 2000 milliseconds in red (if incorrect). When the ingredient missing corresponded to the bottom level, two responses were accepted as correct (e.g. a banana and a pineapple could be combined with an avocado or with a coconut). Thus, the chance level in these trials was 50% (two correct responses) instead of the usual 25% chance level (one correct out of four alternative responses). The position of the response mapping was randomised independently on each trial for both experiments.

Note that accuracy (% correct) for chance on the trials where the bottom level was missing was 50% (two correct responses out of 4), compared to 25% (once correct response) when the top or middle levels were missing. This was also the case for the control structure, where we artificially accepted two responses as correct.

Behavioural analyses

All behavioural analyses were performed using mixed-effects ANOVA (meANOVA) and t-tests with the statistics toolbox from Mathworks MATLAB. All significant results were reported under a p-value threshold of 0.05.

Results

Performance on the first session

We first examined the performance in the first session in order to assess whether participants were able to learn the structure. We performed a 3-way mixed-effects ANOVA on accuracy (% correct) with factors (i) group, (ii) hierarchical level of the missing ingredient (top, middle or bottom) and (iii) early vs late blocks (early: blocks 1 to 8, late: 9 to 16). In other words, we decided to split in half the data for each session so that we could look for signatures of learning between blocks. Accuracy was higher in the second compared to the first half of the experiment ($F_{(1,46)} = 45.5$, $p < 0.001$) implying that participants learned something about the structure. Additionally, this analysis revealed that accuracy was higher for the hierarchical group than for the control group ($F_{(1,46)} = 31.9$, $p < 0.001$), and thus that the hierarchical structure was easier to learn in session 1 than the control structure. We also found an effect of the level of the missing ingredient ($F_{(1,46)} = 75.8$, $p < 0.001$) such that accuracy was highest when the missing ingredient corresponded to the bottom level (as expected by a chance level of 50% in the bottom level, compared to 25% in the top and middle levels) and lowest when corresponded to the middle. These two effects interacted, such that the gap in accuracy between groups was particularly accentuated for trials with the top level missing (interaction group × level, $F_{(2,46)} = 74.8$, $p < 0.001$) as expected by the fact that the top level was constant within a block exclusively for the hierarchical group. In the second half of the experiment performance was above chance for all groups and missing levels (six independent t-tests against chance level; all t-statistic $> +7.8$, $p < 0.001$) implying that some learning had been achieved in both groups and for all placeholders (top, left and right).

Structure consistency improves performance on the second session

We then compared the performance between the hierarchical and control groups during the second session, when they performed identical tasks. This was our analysis of interest because any between-groups effect could only be explained in terms of knowledge carried over from learning during the first session. Indeed, an identical mixed-effects ANOVA on accuracy on the second session yielded a significant effect between groups ($F_{(1,46)} = 5.0$, $p = 0.029$) implying that participants in the hierarchical group had transferred the underlying structure between sessions and re-used it in order to speed-up learning (see **Fig 2.6a**). In line with this interpretation, we also found a significant interaction between group and early/late blocks (group × block, $F_{(1,46)} = 4.9$, $p = 0.031$). Similar to the first session of the hierarchical group, participants' accuracy was highest, and learning was steepest, for the top level (hierarchical group: 96.5±0.3%; control group: 94.2±0.7%). Accuracy was lower for the middle-level (hierarchical group: 79.1±3.2%; control group: 69.8±3.0%) compared to the bottom level (hierarchical group: 90.2+1.6%; control group: 84.9±2.0%) but this result was hard to interpret because the chance level differed between these conditions. I refer to the interested reader to **Table T2.4** for a list of all effects from this ANOVA.

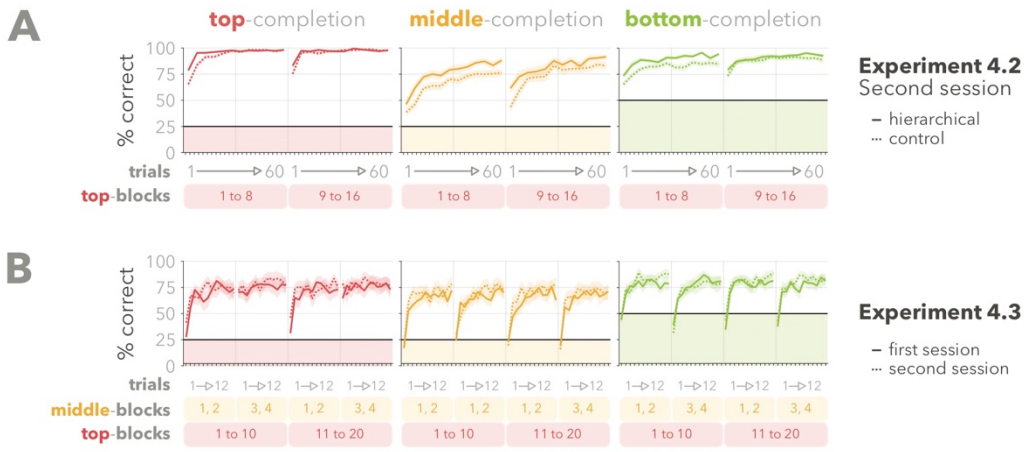

Figure 2.6. Experiments 2.2 and 2.3 — Feature-completion task results.
A. Learning curves with accuracy (% correct; on the y-axis) across time (x-axis) for the second session of Experiment 2.2, for the hierarchical group (continuous line) and the control group (dashed line; with shaded standard error of the mean). Each point on the x-axis corresponds to a bin of 5 trials uniformly distributed. Blocks were averaged for the first half (blocks 1–8) and the second half (blocks 9–16) of the experiment. Learning curves were calculated independently for trials with a missing top-level (left panel, in red), middle-level (middle panel, yellow) and bottom-level ingredient (green, right panel). **B.** Learning curves of Experiment 4.3, for the first session (continuous line) and second session (dashed line). Top-blocks were averaged for the first half (top-blocks 1–10) and the second half (top-blocks 11–20) of the experiment. Middle-blocks were averaged for the first half (middle-blocks 1 and 2) and the second half (middle-blocks 3 and 4) of each top-block.

Discussion

Here I have presented an experimental design that was more sophisticated than **Experiment 2.1**. It required training participants over multiple blocks on multiple hierarchical datasets; and over multiple sessions so that we could compared performance between groups that only differed on the first session. This was required in part by our hypotheses, where we expected that participants abstracted the structure away from the specific features; and in part so that we could balance the features associated with each hierarchical level.

The results of this experiment give preliminary evidence that humans can learn a hierarchical structure and re-use it to speed-up learning in a subsequent session. We interpret these results in terms of structure learning. Indeed, if participants learned a hierarchical structure that was common across the blocks of the first session, they could reuse it speed-up learning on the second session. However, the design was such that the spatial position was confounded with the hierarchical level. This implied, for instance, that participants could learn that the top placeholder was easier (because the ingredient was constant within a block) and this could affect learning and accuracy. However, as we have discussed in **Experiment 2.1**, this is not a property intrinsic of hierarchies and is potentially related to a more basic spatial statistical learning.

In light of this consideration, we decided to run a further pilot **Experiment 2.3**. In this case we only made use with the hierarchical group, and the position of the ingredients varied in a trial basis such that given hierarchical level could not be associated with a placeholder. The aim of this follow-up experiment was to assess whether this change would disrupt learning. Additionally, we attempted to further improve the task in a number of ways. Firstly, we modified the temporal structure of the task to emphasise the hierarchical relationship between the middle and bottom level, in the hope that this would lessen the gap of accuracy between trials with missing top- and middle-level ingredient. Secondly, we included a second task at the end of each session in order to test how much the participants could remember about the triplets when the temporal structure was disrupted.

Experiment 2.3. Replication and spatial control

Introduction

We performed an additional **Experiment 2.3** in which we introduced further changes to our experimental design. These modifications were necessary to address some of the issues discussed previously, and also to adapt the paradigm for a later imaging study (described in **Chapter 3**). We found that these had minor consequences in behaviour, and in particular that participants still learned the structure and improved performance on the second session.

Methods

Participants

Just like in **Experiment 2.2**, participants were recruited via Amazon Mechanical Turk (http://www.mturk.com) and paid $24 for participation in accordance with ethical guidelines approved by the Oxford University Medical Sciences Ethics Board. In **Experiment 2.3** we recruited a total of 17 human participants (8 female, 9 male; age 20–39, mean 32.1 years) within a single group. We blacklisted the participants from Experiment 2.2 such that only naïve participants were allowed to take part.

Experiment overview

Participants in **Experiment 2.3** performed two sessions with the hierarchical structure just as in the "hierarchical" group in **Experiment 2.2**. At the end of each session we included an additional second task where we tested their knowledge by asking them if an exemplar (i.e. a set of three ingredients) was valid or not (*triplet-validity* task; **Fig 2.7a**). This task had a duration of 288 trials (20–25 minutes on average)

A demo of the tasks is available at:

Experiment 2.3: http://185.47.61.11/sandbox/tasks/jan/fruit_2b/

Stimuli and hierarchical structure

We made use of the same structure and stimuli as in **Experiment 2.2**. However, in this case the position on screen of the ingredients was randomised independently on each trial for **Experiment 2.3**, such that ingredients belonging to the top, middle or low bottom were equally likely to appear on the top, left or right placeholders.

Temporal structure

In **Experiment 2.3**, we made use of a slightly more sophisticated temporal structure that better reflected the hierarchical structure of the task. Trials were organised in *top*-blocks and *middle*-blocks (see **Table 2.3** and **Fig 2.5d**) in order to also reflect on the middle level of the hierarchy. We constructed *middle*-blocks of 12 trials where the top- and middle-level ingredients of the triplet were constant (two unique triplets; e.g. triplets always included a banana and a pineapple). Similarly, we also constructed top-blocks of 4 *middle*-blocks, where the top ingredient (e.g. the banana) was constant but the middle ingredient (e.g. the pineapple) was allowed to change (four unique triplets).

Triplet-validity task

After participants completed the feature-completion task on each session, we evaluated how much participants had learnt in a second phase. In this task, we instructed participants to respond if a set of three ingredients was a valid triplet (that is, if they learned this triplet during the feature-completion task) or not. Each trial started with the presentation of two ingredients placed horizontally at the centre of the screen ("sample"), that remained on the screen for 2.5s. Then the sample disappeared,

and a third ingredient would appear ("probe"). Participants were instructed to respond whether the combination of the three ingredients was a *valid* triplet (by pressing the [F] key on the keyboard) or not ([J] key). Participants could respond within 20s from the moment they saw the probe before the new trial started. In this task no feedback was provided, in order to prevent any further learning. This allowed us to evaluate the level of learning that had occurred during the *feature-completion task*, using a task where the features were not temporally correlated and where participants could not rely on short-term memory strategies. Our goal was to assess that participants could still remember the triplets, which was a requirement before we adapted the design as an imaging experiment.

Trials only included ingredients corresponding to the relevant learning session (e.g. fruits on session 1 and vegetables on session 2). There was an equal number of valid and invalid triplets. Valid trials were selected pseudo-randomly, such that there was equal number of triplets from each session, and all levels were equally likely to appear in the sample and probe. The same ingredient (e.g. banana) was never shown twice within the same triplet. Invalid trials were constructed such that they would repeat and omit one of the hierarchical levels (e.g. two top-level ingredients and one middle-level ingredient).

Behavioural analyses

All behavioural analyses were performed using repeated-measures analysis of variance (rmANOVA), t-tests, and logistic regression with the statistics toolbox from Mathworks MATLAB. All significant results were reported under a p-value threshold of 0.05.

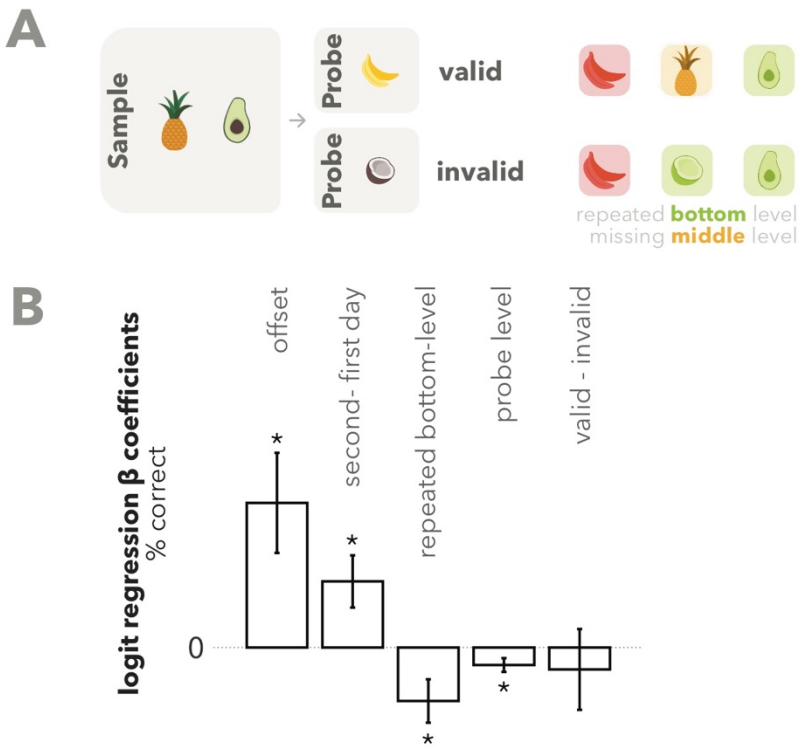

Figure 2.7. Experiment 2.3 — Triplet-validity task design and results.
A. Participants in Experiment 4.3 performed an additional *triplet-validity* task. On each trial, participants saw a two ingredients ("sample" screen) followed by a third ingredient ("probe" screen). Participants were instructed to respond if the triplet had been learnt during the previous task ('valid') or not ('invalid'). **B.** Logistic regression on accuracy (% correct) with 5 z-scored regressors. A star (*) denotes a significant effect of a regressor ($p < 0.05$). The scale of the beta coefficients is in arbitrary units.

Results

Spatial consistency is not necessary for significant learning and performance increase between sessions.

In **Experiment 2.3** we ran a single group trained with the hierarchical structure in both sessions, allowing us to analyse our data by comparing the performance between the first and the second session. We performed a 4-way mixed-effects ANOVA on accuracy (% correct) between session × hierarchical level of the missing ingredient (top, middle or bottom) × between *top*-blocks (early: 1st to 10th, late: 11th to 20th) ×

and between *middle*-blocks within a *top*-block (early: blocks 1st and 2nd, late: 3rd and 4th) (see **Fig 2.6b**). I refer to the interested reader to **Table T2.5** for a list of all effects from this ANOVA, and here I will only report effects that I deem of interest. We found, in line with our hypotheses, an effect of session whereby accuracy was higher for the second compared to the first session ($F_{(1,16)} = 6.1$, $p = 0.025$). We additionally found that performance also increased across *middle*-blocks ($F_{(1,16)} = 4.8$, $p = 0.043$) and this was so in particular when the top level was missing (interaction level × middle-block: $F_{(1,16)} = 22.8$, $p < 0.001$) reflecting the temporal structure of the task.

Interestingly, accuracy did not differ significantly between halves of the task (early vs late top-blocks: $F_{(1,16)} < 1$, $p = 0.44$). This was surprising and made us consider whether participants re-learned the features on each block based on the same underlying abstract hierarchical structure. This hypothesis indeed would be in line with the participants' performance equal in early and late top-blocks, but would additionally predict that participants failed to remember the triplets after they finished the task. Participants performed a second task immediately after the end of the feature-completion task that allowed us to test this question.

Performance in the triplet-validity task

The *triplet-validity* task did not have any temporal structure (all trials were independent from each other). On each trial, participants saw three ingredients (first two, then one) and responded whether the triplet was seen previously during the feature-completion task. Participants responses were not biased towards valid or invalid responses (50.9 ± 2.3% valid responses; $t_{(16)} = +0.38$, $p = 0.71$ against 50%). Accuracy (% correct) was significantly above chance for the first session (55.1

± 2.2%; $t_{(16)}$ = +2.20, p = 0.021 one-tailed against chance level 50%) and for the second session (61.8 ± 3.3%; $t_{(16)}$ = +3.41, p = 0.002 one-tailed against chance level 50%). Additionally, performance increased in between sessions ($t_{(16)}$ = +2.58, p = 0.020). Performance was significantly above chance independently for valid triplets (59.3 ± 2.7%; $t_{(16)}$ = +3.35, p = 0.002 one-tailed against chance level 50%) and for invalid triplets (57.5 ± 4.0%; $t_{(16)}$ = +1.82, p = 0.044 one-tailed against chance level 50%). Additionally, performance for invalid triplets changed as a function of what hierarchical level was repeated (one-way ANOVA: $F_{(2,32)}$ = 5.6, p = 0.008), such that accuracy when a triplet included two bottom-level ingredients was lower and not significantly above chance (51.8 ± 4.1%; $t_{(16)}$ = +0.42, p = 0.341 one-tailed against chance level 50%). This was evidence that participants had learnt the top- and middle-level ingredients equally well, but surprisingly better than the bottom-level ingredients. We found a similar effect for the level of the probe ingredient (one-way ANOVA: $F_{(2,32)}$ = 11.0, p < 0.001) where performance was worse when the probe ingredient was associated with bottom level of the hierarchy.

We summarised these effects within in a logistic regression (see **Fig 2.7b**), where we introduced 5 regressors: a constant offset; the difference between the second and the first session (session); whether the bottom-level was repeated within the trial (+1) or not (0); the hierarchical level of the probe ingredient (top: -1, middle: 0, bottom: +1), and whether the triplet was valid (+1) or invalid (-1). All regressors were standardised (by z-transformation) before entry into the design matrix; they competed with each other in order to explain the variance of the accuracy. We found that overall performance was significantly above chance (offset: $t_{(16)}$ = +2.8, p = 0.013) and higher for the second session (session: $t_{(16)}$ = +2.46, p = 0.026). Performance was lower both when the bottom level was repeated ($t_{(16)}$ = -2.39, p = 0.030) and when it was

associated with the probe ingredient ($t_{(16)} = +2.45$, $p = 0.026$). Finally, the true validity of the triplet had no effect on accuracy ($t_{(16)} = -0.53$, $p = 0.606$).

We additionally looked into the proportion of participants whose individual accuracy was significantly above chance ($p < 0.05$) under a binomial distribution. We found that 6 and 9 participants (out of 17; for the first and second session respectively) remembered the triplets significantly above chance, with a large overlap on the participants that performed above chance on both sessions.

Discussion

Here I have presented a follow-up of **Experiment 2.2** where we found that participants could still learn the structure, and that about half of the participants remembered the triplets in a second task without feedback where they couldn't rely on short term memory.

It is worth pointing out that in **Experiments 2.2** and **2.3** we did not filter out any participants as a function of their performance. There are many reasons why many participants performed at chance during the evaluation. For instance, it has been reported that learning experiments, but not simpler decision-making tasks, can be particularly difficult to replicate with participants drawn from the pool of Amazon Mechanical Turk users (Crump, McDonnell, & Gureckis, 2013). We replicated these results with participants that performed the task physically from within the Oxford premises, and confirmed that behaviour was similar albeit more accurate.

Ideally, we would have run **Experiments 2.2** and **2.3** as a single experiment. Indeed, the results would be more convincing if we could compare both groups while controlling for the spatial position. However, at the time we ran these experiments we could not know whether participants would be able to learn the relationships and it seemed sensible to pilot a version first where we expected learning to be easier. We plan to run an improved design that combines insights from **Experiments 2.2** and **2.3** post-thesis.

Chapter discussion

In this chapter we have explored two behavioural paradigms in which participants were trained to learn a set of multi-dimensional exemplars that follow a hierarchical structure. These pilot exploratory experiments aimed at verifying whether participants could learn and reuse a hierarchical data structure within the short time frame provided by a 2-day experiment.

In **Experiment 2.1** we found that the participants' pattern of errors was driven by frequency statistics, but this was similar between a hierarchical and a control group. Additionally, participants responses were such that the features were learnt jointly within each exemplar, rather than independently for each dimension.

We then performed a different **Experiment 2.2** where participants performed two similar learning sessions and found that learning in the second session improved when it matched the abstract structure learnt on the first session. This was interpreted as a behavioural signature of structure transfer of knowledge.

We performed a follow-up pilot **Experiment 2.3** where we refined the experimental design, namely by controlling for the spatial position, by modifying the temporal structure of the task, and by adding a second phase that allowed us to test for the amount of knowledge that participants acquired after learning. We found that performance increased in between sessions, similar to the previous task, and that at least half of the participants performed above chance in the evaluation task, meaning that they had learnt the structure to a reasonable amount.

To my knowledge, there are surprisingly few studies reported that investigate the acquisition of novel hierarchical representations within a learning session. One of them (G. L. Murphy et al., 2012) found that participants often relied on feature-

comparison strategies, rather than building a hierarchical representation of the exemplars and their features. It is possible that sleeping and time play a critical role for building such representations, as predicted by the Complementary Learning Systems (CLS) framework (Kumaran, Hassabis, et al., 2016; McClelland et al., 1995). Our experiments were not designed in order to look into the effect of sleep on the formation of hierarchical representations but suggested a novel approach to teach participants a hierarchical representation of multi-dimensional exemplars within a behavioural session.

In Chapter 3, I describe a full replication of **Experiment 2.3** with naïve participants. In addition, we performed a third imaging session after participants had built these hierarchical representations. They performed two new tasks that allowed us to investigate the neural code associated with hierarchical representations and to pinpoint the regions relevant to these representations in the human brain.

Tables

Table T2.1. Experiment 2.1. Description of hierarchical and control structures

The tables below describe the hierarchical structure of Experiment 2.1. Letters A–G refer to different dimensions (e.g. mane, sunglasses, skin) and the numbers 1–2 refer to the different features that a feature can take in a given dimension (e.g. patchy or stripy for the skin dimension). Columns correspond to exemplars (e.g. exemplar I1 is composed of features A1, B1 and D1, and all other dimensions are absent). Rows correspond to the hierarchical level assigned to each feature.

Hierarchically structured exemplars:

Exemplar	e1	e2	e3	e4	e5	e6	e7	e8
Top	A1	A1	A1	A1	A2	A2	A2	A2
Middle	B1	B1	B2	B2	C1	C1	C2	C2
Bottom	D1	D2	E1	E2	F1	F2	G1	G2

Control exemplars:

Exemplar	e1	e2	e3	e4	e5	e6	e7	e8
Top	A1	A1	A2	A2	A2	A2	A1	A1
Middle	B1	B2	B2	B1	C1	C2	C2	C1
Bottom	D1	E1	F1	G1	D2	F2	E2	G2

Table T2.2. Experiment 2.2. Description of exemplars and temporal blocking

Below is a description of the exemplars of **Experiment 2.2**. The control structure was used for the first session of the *control* group. The second session of the control group, as well as both sessions of the *hierarchical* group made use of the hierarchical structure. The 12 ingredients within each behavioural session were mapped to features that followed the abstract relationship described in the tables below. The letters correspond to the position on screen (A: top, B: left, C: right). In the hierarchical group, these also corresponded to the hierarchical levels (A: top, B: middle, C: bottom). The numbers 1–4 correspond to the four different ingredients that were mapped with each hierarchical dimension. In total there were 12 different ingredients combined in 16 different triplets.

Hierarchically structured exemplars:

Exemplars	**e1**	**e2**	**e3**	**e4**	**e5**	**e6**	**e7**	**e8**	**e9**	**e10**	**e11**	**e12**	**e13**	**e14**	**e15**	**e16**
Top	A1	A1	A1	A1	A2	A2	A2	A2	A3	A3	A3	A3	A4	A4	A4	A4
Left	B1	B1	B2	B2	B3	B3	B4	B4	B1	B1	B4	B4	B3	B3	B2	B2
Right	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4

Control exemplars:

Exemplars	**e1**	**e2**	**e3**	**e4**	**e5**	**e6**	**e7**	**e8**	**e9**	**e10**	**e11**	**e12**	**e13**	**e14**	**e15**	**e16**
Top	A1	A2	A3	A4	A2	A3	A4	A1	A3	A4	A1	A2	A4	A1	A2	A3
Left	B1	B2	B3	B4	B3	B4	B1	B2	B2	B1	B4	B3	B4	B3	B2	B1
Right	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4

Temporal structure (blocks): We defined four different types of blocks such that only four unique exemplars could be probed in each. Each block type was seen exactly four times (16 blocks of 60 trials in total) and all exemplars were seen equally often within a block (15 trials per exemplar per block).

Block	a1				a2				a3				a4			
Exemplars	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11	e12	e13	e14	e15	e16

Table T2.3. Experiment 2.3. Description of exemplars and temporal blocking

Below is a description of the exemplars of **Experiment 2.3**. and how they were blocked.

Structure: The structure of Experiment 2.3 was identical to the hierarchical structure of Experiment 2.2. The ingredients were mapped to features. The letters correspond to the hierarchical level (A: top, B: middle, C: bottom) and in this case ingredients could appear randomly in any placeholder (top, left or right) on any given trial. The numbers 1–4 correspond to the four different ingredients that were mapped with each hierarchical dimension. In total thus there were 12 different ingredients combined in 16 different triplets.

Exemplars	**i1**	**i2**	**i3**	**i4**	**i5**	**i6**	**i7**	**i8**	**i9**	**i10**	**i11**	**i12**	**i13**	**i14**	**i15**	**i16**
Top	A1	A1	A1	A1	A2	A2	A2	A2	A3	A3	A3	A3	A4	A4	A4	A4
Middle	B1	B1	B2	B2	B3	B3	B4	B4	B4	B4	B1	B1	B2	B2	B3	B3
Bottom	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4

Temporal structure (blocks): Trials were blocked in *middle*-blocks, that were in turn blocked into *top*-blocks. *Top*-blocks were similar to those of Experiment 2.2, except that we included 20 *top*-blocks of 48 trials (instead of 16 blocks of 60 trials each). There were four different type of *top*-blocks that differed in terms of the exemplars cued. The 48 trials from a *top*-block were also divided in four *middle*-blocks. Each top-blocks included two different types of *middle*-blocks, and each middle-block was shown twice within a top-block. Middle-blocks included 12 trials and were composed of two different exemplars that were cued equally often (6 trials per *middle*-block per exemplar).

***top*-block**	a1				a2				a3				a4			
***middle*-block**	*b11*		*b12*		*b21*		*b22*		*b31*		*b32*		*b41*		*b42*	
Exemplars	i1	i2	i3	i4	i5	i6	i7	i8	i9	i10	i11	i12	i13	i14	i15	i16

An example of a history of trials is given for illustration purposes. Marked in bold is the transition between blocks. Note that new *middle*-blocks start in trials 1, 13, 25, 37 and 49, but *top*-blocks change only in trials 1 and 49 after four *middle*-blocks have been completed. The first row corresponds to the total trial number in the experiment. Rows two, three and four count the total number of top-blocks (up to 20), the number of middle-blocks per top-block (up to 4), and the number of trials per middle-block (up to 12) respectively. The last three rows correspond to the type of top-block, middle-block and exemplar.

#trial	1	2	…	12	13	…	24	25	…	36	37	…	48	49	…
#top-b	1	1	…	1	1	…	1	1	…	1	1	…	1	**2**	…
#mid-b / top-b	1	1	…	1	**2**	…	2	**3**	…	3	**4**	…	4	**1**	…
#trial / mid-b	1	2	…	12	**1**	…	12	**1**	…	12	**1**	…	12	**1**	…
top-block	a1	a1	…	a1	a1	…	a1	a1	…	a1	a1	…	a1	**a4**	…
middle-block	b11	b11	…	b11	**b12**	…	b12	**b11**	…	b11	**b12**	…	b12	**b42**	…
Exemplar	e1	e1	…	e2	e4	…	e3	e2	…	e1	e3	…	e4	e15	…

Table T2.4 Experiment 2.2. ANOVA effects

This is an exhaustive list of the statistical effects resulted from the mixed-effects ANOVA analyses in **Experiment 2.2**. This includes two different analyses (one for each session).

Session 1:

Effects	d.f.	F-value	p-value
level	(2,46)	75.8	<0.001
group	(1,46)	31.9	<0.001
block	(1,46)	45.5	<0.001
level × group	(2,46)	74.8	<0.001
level × block	(2,46)	8.6	<0.001
group × block	(1,46)	7.3	0.01
level × group × block	(2,46)	18.0	<0.001

Session 2:

Effects	d.f.	F-value	p-value
level	(2,46)	92.1	<0.001
group	(1,46)	5.0	0.029
block	(1,46)	33.4	<0.001
level × group	(2,46)	2.6	0.083
level × block	(2,46)	17.6	<0.001
group × block	(1,46)	5.1	0.029
level × group × block	(2,46)	1.8	0.168

Table T2.5 Experiment 2.3 ANOVA effects

This is an exhaustive list of the statistical effects resulted from the repeated-measures ANOVA analysis in **Experiment 2.3**.

Feature-completion task

Effects	d.f.	F-value	p-value
level	(2,32)	25.4	<0.001
day	(1,16)	6.1	0.025
top-block	(1,16)	0.6	0.439
middle-block	(1,16)	4.8	0.043
level × day	(2,32)	0.6	0.573
level × top-block	(2,32)	0.9	0.389
day × top-block	(1,16)	0.5	0.486
level × middle-block	(2,32)	22.8	<0.001
day × middle-block	(1,16)	3.5	0.079
top-block × middle-block	(1,16)	0.0	0.848
level × day × top-block	(2,32)	0.4	0.693
level × day × middle-block	(2,32)	0.4	0.674
level × top-block × middle-block	(2,32)	1.5	0.247
day × top-block × middle-block	(1,16)	0.2	0.644
level × day × top-block × middle-block	(2,32)	0.8	0.477

Chapter 3. Neural correlates of hierarchical level

Abstract

In this chapter I describe the results of a fMRI experiment in which participants learned a set of hierarchical structures over two behavioural sessions. On the third day, they performed two different tasks in the scanner, each of which aimed to probe for differential neural signals as a function of hierarchical level. We observed a univariate effect with parietal and rostrolateral prefrontal cortices being more active when features (fruits) corresponding to the higher levels of the hierarchy were both expected and observed. We obtained independent data from a second task, where we found an effect of repetition suppression for the top level of the hierarchy in these regions. These two effects (encoding of the top level, and repetition suppression) correlated positively across participants. Overall this experiment looks into the neural signatures of hierarchical representations. I close this chapter by discussing some similarities with the literature in magnitude representation.

Introduction

In Chapter 2 we explored two behavioural paradigms where participants were given a chance to learn a hierarchical representation. I found evidence that in one of them (**Experiments 2.2** and **2.3**) participants learned the structure and further emphasised the potential to adapt the design as an imaging study. In this chapter, I will describe a functional magnetic resonance imaging (fMRI) study. Here, we trained participants in two sessions that were almost identical to **Experiment 2.3** before we brought them to the MRI scanner. In this case however we addressed a different research question. Namely, what are the brain regions that code for hierarchical relationships? How is

the level of the hierarchy represented? To the best of our knowledge, this question has never been addressed with hierarchies with more than two levels and whilst controlling for the difference of complexity across different levels intrinsic to hierarchical structures.

The two behavioural (learning) sessions included minor changes. First, we disposed of the vegetables and instead made use of a larger set of fruit stimuli for both sessions. This was of importance for one of the tasks that participants performed in the scanner (see *session-matching task*) where they were instructed to report if two stimuli came from the same session or not. Second, participants were explicitly told through instruction that the task would follow a hierarchical structure, which we expected would help participants discover and represent the hierarchical relationship of the fruits.

Participants completed two similar learning sessions in consecutive days before we brought them into the magnetic resonance imaging (MRI) scanner. After two learning sessions, participants had learnt the association between fruit stimuli and one of three hierarchical levels. Participants subsequently performed two different tasks in the scanner: one that required participants to recall the specific relationships they had learnt during the learning (i.e. the *triplet-validity* task in **Experiment 2.3**); and an incidental task where we relied on repetition suppression in order to elicit effects related to the hierarchical level, even when this was irrelevant to the actual task that participants were instructed to perform. We expected that Blood-Oxygen-Level Dependent (BOLD) signal (captured by fMRI) in medial and lateral prefrontal cortices would encode the higher levels of the hierarchy. We found that these regions,

together with left parietal cortex, encoded the hierarchical level in both tasks in a univariate fashion, irrespective of whether this information was necessary for the task at hand.

Methods

Participants

35 healthy participants (16 female; age 18-32, mean 21.7 years) were recruited into the study in accordance with ethical guidelines approved by the Oxford University Medical Sciences Ethics Board and by the ethics committee of the University of Granada. Participants were recruited through the University of Granada. No participants reported a history of psychiatric or neurological illness, and all had normal or corrected-to-normal vision. They were paid 30€ for participation in two behavioural ("learning") sessions in consecutive days. Those participants that performed significantly above chance ($N_{behaviour}$ = 26) were invited to participate in a functional magnetic resonance imaging (fMRI) session on the subsequent day, and were paid an additional 20€ plus a performance-based bonus of up to 15€. From the participants that were invited to the scanner session, 5 failed to perform significantly above chance in the *triplet-validity task* (see below) and 1 moved excessively during the scanning due to reported coughing, thus leaving us with N_{fMRI} = 20 imaging datasets.

Learning sessions: overview

Participants performed two identical learning sessions prior to the imaging session (**Fig 3.1a**). These sessions were performed over the internet for a total duration of around 90 minutes and was divided in two tasks. During the first *feature-completion* task, participants learned via trial and error to associate 12 fruit stimuli into 16 triplets

that respected a hierarchical organisation by completing the missing fruit from a triplet (see **Fig 3.1**). In the second *triplet-validity* task, we tested their knowledge by asking them if a set of three fruits was valid or not (see **Fig 3.2a**). A demo of the learning session is available: http://185.47.61.11/sandbox/tasks/jan/fruits_3/

We experienced technical issues with the recording of the data from the learning sessions of three participants. Thus, our behavioural analyses of the *fruit-completion* and *triplet-validity* task were performed with N = 23 participants in order to relate performance during the imaging session to that one of the learning sessions.

Stimuli and hierarchical structure

The hierarchical structure was identical to that of **Experiment 2.3**, and only differed in that we used a set of 24 distinct fruit stimuli random and evenly split between the two sessions. This removed the categorical distinction between stimuli across the two sessions (i.e. fruits and vegetables previously).

The fruit stimuli set was expanded from the same source where we originally obtained the fruit stimuli. The stimuli belong to the ConceptDraw Food Court solution (http://www.conceptdraw.com/samples/food-beverage-food-court) under Attribution-NonCommercial-NoDerivs 3.0 Unported license (CC BY-NC-ND 3.0; https://creativecommons.org/licenses/by-nc-nd/3.0/).

Learning sessions

The two learning sessions were performed on consecutive days over the internet and were identical to **Experiment 2.3**. The experimental design is illustrated in **Figure 3.1** and summarised below in the interest of readability.

Each learning session started with a "training" phase. As in **Experiment 2.3**, we introduced a temporal structure to the "training" phase in order to induce a hierarchical representation of the triplets (**Fig 3.1c**), such that fruits corresponding to lower levels within the hierarchy changed more frequently. We constructed *middle*-blocks of 12 trials where the top and middle fruits of the triplet were constant (e.g. triplets always included a banana and a pineapple). Similarly, we also constructed top-blocks of 4 *middle*-blocks, where the top fruit (e.g. the banana) was constant but the middle fruit (e.g. the pineapple) was allowed to change. The "training" phase had a total duration of 960 trials and around 60 minutes. On each trial (**Fig 3.1d**) two fruits and a question mark appeared on screen within a disk composed of three placeholders. The question mark would indicate the position of the missing fruit. Participants could then choose among four alternative fruits (all corresponding to the same hierarchical level) by pressing one of four keys with the right hand. Participants could provide a response within 20 seconds, after which the response was registered as incorrect and the experiment moved on to the next trial. After a response was registered, the selected fruit would be displayed on screen surrounded in green for 500 milliseconds (if correct) or for 2000 milliseconds in red (if incorrect). When the ingredient missing corresponded to the bottom level, two responses were accepted as correct (e.g. a banana and a pineapple could be combined with an avocado or with a coconut). Thus, the chance level in these trials was 50% (two correct responses) instead of the usual 25% chance level (one correct out of the four alternative responses). The position of the fruits and the response mapping were randomised independently on each trial.

After participants completed the *feature-completion* task, we evaluated how much participants had learnt in a second phase with a different task similar to the *triplet*-

validity task (**Fig 3.2a**, see below for a full description) that they subsequently performed in the third day while undergoing fMRI. This phase had a duration of 288 trials (less than 30 minutes). Unlike the imaging session, trials only included fruits corresponding to the relevant learning session, and participants pressed the [F] key for valid responses and the [J] key for invalid responses. We invited participants that had successfully learnt the triplets (accuracy significantly higher than chance, $p < 0.01$) to complete the imaging session in the third day.

Imaging session

After completing two learning sessions participants had been exposed to a total number of 24 fruits associated across 32 triplets (**Fig SF3.1**). In the third session, participants completed an imaging session of up to 6 scanner runs. Each run started with a screen that cued participants with one of the two tasks they were instructed to perform. For all participants odd runs (1, 3, and 5; less than 7 minutes, 144 trials long each) corresponded to the *triplet-validity* task and even runs (2, 4 and 6; less than 14 minutes, 432 trials long each) corresponded to the *session-matching* task. All runs were buttressed by lead-in and lead-out durations of 10 and 15 seconds respectively. The total functional scanning time was just over an hour.

Participants were allowed to see a list with all fruits (images and labels displayed for each fruit individually) and practice one run of each task before going into the scanner. Participants could respond by pressing a button (in the scanner) or a key (in the practice) with their left or their right hand. The response mapping was counterbalanced across participants but was consistent across tasks, such that valid responses (for the *triplet-validity* task) and session match responses (for the *session-matching* task) were associated with the same hand. Participants did not receive feedback at any point during the third day.

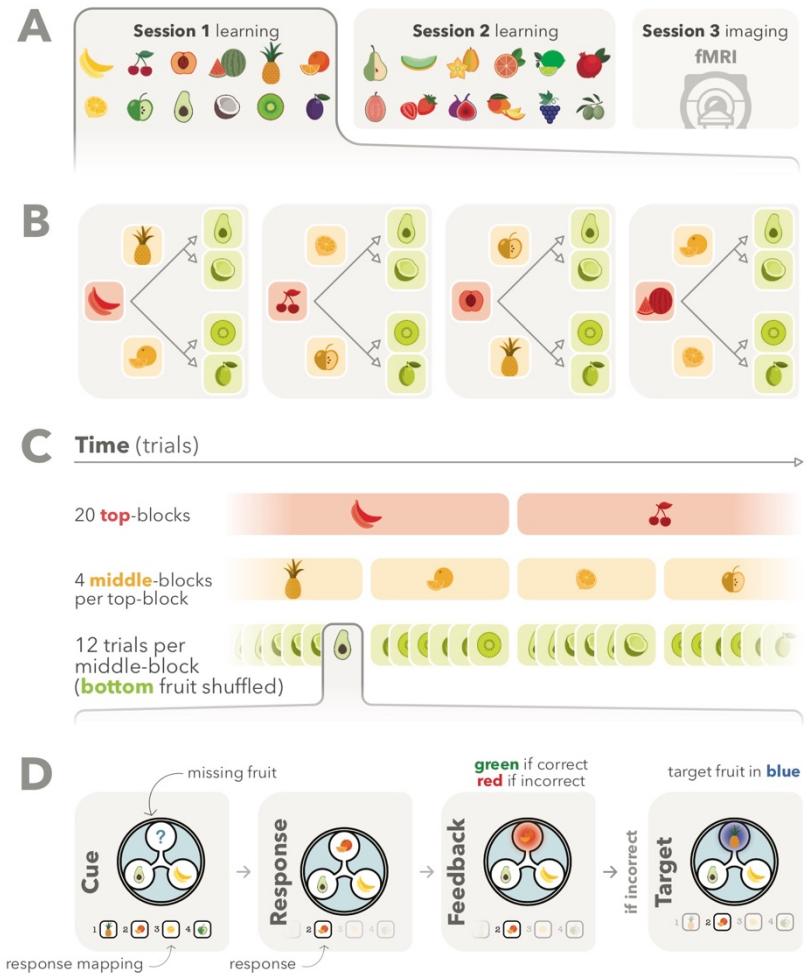

Figure 3.1. Learning sessions design
Illustration of the training phase during the two learning sessions. **A.** Two sets of 12 fruits mutually exclusive were used for each learning session on two consecutive days. **B.** We constructed a total of 16 triplets from the 12 fruits corresponding to each day, such that fruits would belong to a hierarchical level (*top*, *middle*, or *bottom*). The arrows illustrate the hierarchical relationship. For example, a banana could be paired with a pineapple or with an orange, and a pair banana-pineapple could be associated with an avocado (hence the triplet banana-pineapple-avocado) or with a coconut. **C.** The "training" phase followed a temporal structure, such that trials were grouped into *middle*-blocks (with a constant fruit corresponding to the middle hierarchical level) that were in turn grouped into *top*-blocks (with a constant fruit corresponding to the top hierarchical level). **D.** Each trial was composed of a cue, a response, a feedback/target screen. At the time of the cue, a triplet was shown with missing fruit, together with four alternative responses. After participants selected the response, the feedback informed participants if their response was correct (in green) or not (in red). If the response was incorrect, one last screen displayed the correct triplet (in blue).

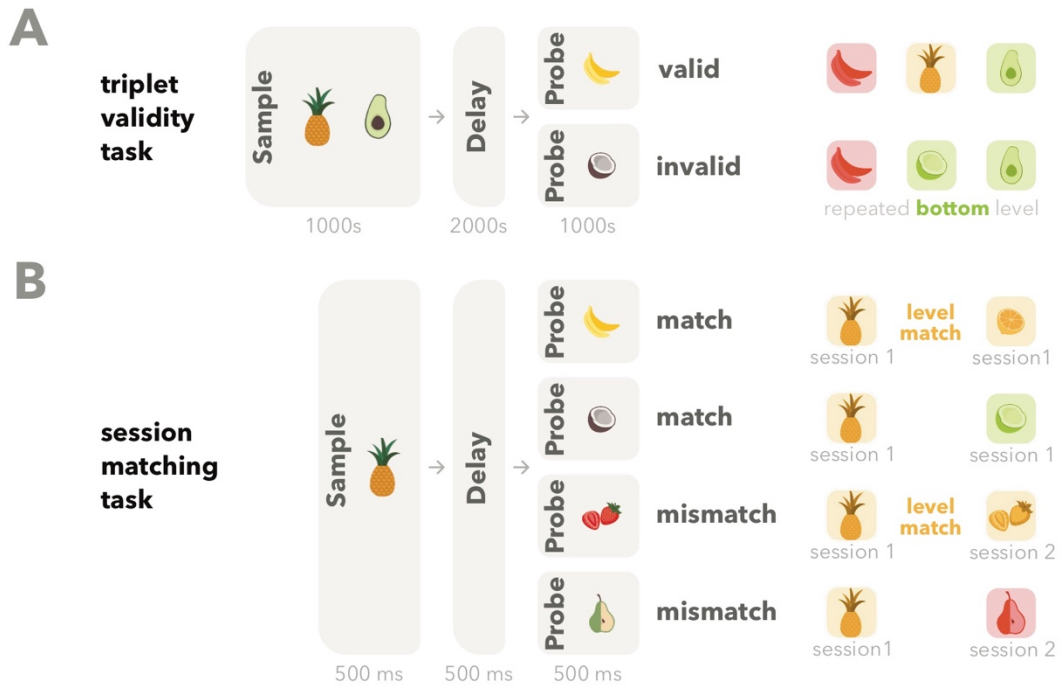

Figure 3.2. Task design for the imaging session
Participants completed three runs of two different tasks. Both tasks consisted of a 'sample', a 'delay', and a 'probe' period. Participants could respond after the probe appeared on display. **A.** In the *triplet-validity* task, the sample included two fruits. Participants were instructed to respond if the set of three fruits (sample plus probe) corresponded to a valid triplet as presented during the two learning sessions. Invalid triplets were characterised by a hierarchical level repeated in the probe. **B.** In the *session-matching* task, participants only saw two fruits in succession (one in the sample and one in the probe). They were instructed to respond if the two fruits belonged to the same day ('session match'; see **Fig 1a**), but we implicitly designed the task in order to compare trials where the sample and the probe belonged to the same hierarchical level ('level match'; see **Fig 1b**). In the task, 'session match' and 'level match' were orthogonal.

Imaging session: triplet-validity task

Participants performed the *triplet-validity* task (**Fig 3.2a**) in the imaging session as well as in the "evaluation" phase of the two learning sessions. During the imaging session, a given trial could include fruits from the first or the second learning session (but not from both). Each trial started with the presentation of two fruits placed horizontally at the centre of the screen ("sample"), that remained on the screen for 1s. Then the fruits disappeared, and a fixation cross was displayed instead on the centre

of the screen. After this delay period that lasted for 2 seconds, a third fruit would appear for 1 second ("probe") before it was replaced again by the fixation cross. Participants were instructed to respond whether the combination of the three fruits was a *valid* triplet (that is, if they learned this triplet during the training phase of the learning sessions) or not. Participants could respond from the moment they saw the probe, but also up to 2 seconds after it disappeared. After the trial finished, a period of 0-3s (jittered) was introduced before the new trial started.

The trials were selected as follows. There were an equal number of valid and invalid triplets. Valid trials were selected pseudo-randomly, such that there were equal numbers of triplets from each day, and all levels were equally likely to appear in the sample and probe. Invalid trials were designed from valid trials by shuffling the third fruit pseudo-randomly across triplets, such that within all triplets the third fruit would be of the same level than one of the first two fruits. Invalid triplets could not be inferred from the fruits in the sample screen alone.

Imaging session: session-matching task

Interleaved with the valid/invalid task, participants also performed 3 runs of a *session-matching* task (**Fig 3.2b**). In these runs, participants were instructed to respond if a pair of fruits were drawn from the same learning session ('session match') or not ('session mismatch'). Each trial began with the presentation of one fruit placed in the centre of the screen for 500ms. Then a delay with a fixation cross was introduced for 1s. Finally, a second fruit would appear for 500ms, before it was replaced again by the fixation cross. Participants could respond anytime during the presentation of the second fruit or after by pressing a response button with the index fingers of each hand, but only up to 2 seconds after the second fruit disappeared.

fMRI data acquisition and preprocessing

Magnetic resonance images were acquired with a 3T Siemens scanner using a standard echo-planar imaging sequence. Whole-head T_2*-weighted echo-planar images were continuously acquired in descending sequence with a voxel resolution of 3.5 isotropic, slice spacing of 4.2mm, with a repetition time of 2 s, and echo time of 30 ms. Each volume included 64 × 64 × 32 voxels. A high-resolution T_1-weighted structural image was also obtained (voxel size = 1 × 1 × 1 mm). For standard preprocessing and univariate statistical analyses, we used SPM12 (Wellcome Department of Cognitive Neurology, London, United Kingdom). All other analyses were done with custom scripts for Matlab (Mathworks, Natick, MA, United States of America). We also used xjview (http://www.alivelearn.net/xjview) to construct mask images. For each participant, we first realigned all functional images, then we co-registered (rigid body transformation) the participant's anatomical scan to the mean functional image, and then co-registered the participant's data to the Montreal Neurological Institute (MNI) template brain. We then normalized each participant's data to the template brain space, using segmented probabilistic maps for grey matter, white matter, and cerebro-spinal fluid. Functional images were resampled (4 × 4 × 4 mm voxels) and spatially smoothed (6-mm full-width half-maximum (FWHM) Gaussian kernel).

fMRI analysis

Our univariate analyses used a generalized linear model (GLM) approach. A 128-s temporal high-pass filter was applied to remove low-frequency scanner artifacts. Temporal autocorrelation in the time series data was estimated using restricted maximum-likelihood estimates of variance components using a first-order autoregressive model (AR-1), and the resulting non-sphericity was used to form

maximum-likelihood estimates of the activations, consistent with standard approaches in SPM8 and SPM12 (Penny et al., 2006). Our GLM included regressors coding for onsets and durations of stimuli or events, which were then convolved with the canonical hemodynamic response function (HRF; itself convolved with a duration window from trial onset until response) and regressed against the observed fMRI data. We modelled all scanner runs using shared regressors (by concatenation) but constant terms for each run were included in order to model any difference in offset. We report peak voxels from clusters that responded at thresholds that were corrected for multiple comparisons, using a false discovery rate of $p < 0.05$. Voxelwise statistics were rendered onto the MNI template brain using xjview 9.5 (http://www.alivelearn.net/xjview/) and our own Matlab scripts.

Behavioural and imaging analyses

All behavioural and region-of-interest (ROI) imaging analyses were performed using a repeated measure analysis of variance (ANOVA), pearson correlation and t-tests with the statistics toolbox from Mathworks MATLAB. All significant results were reported under a p-value threshold of 0.05. ROI regions were defined from orthogonal contrasts using a threshold of $p < 0.005$. For whole-brain analyses, we applied cluster-level Family Discovery Rate (FDR) multiple comparisons correction for clusters larger than 10 voxels, with an initial threshold at 0.001 uncorrected.

General Linear Models

We performed a total of three different general linear models (GLMs) in our imaging data. The full set of task regressors for each GLM can be found in the **Supplementary Tables ST3.1–3**. Additionally, the activation tables for the contrasts of interest can be found in the **Supplementary Tables ST3.4–5**. **GLM3.1** and

GLM3.2 corresponded to the imaging data during the *triplet-validity* task. We made use of **GLM3.1** in order to run the contrast of valid>invalid (**Fig 3.4a**), to run ROI analyses as well as to display the nature of the interaction (**Fig 3.4b**). **GLM3.2** made use of an alternative approach in order to look for interactions between triplet validity and hierarchical level of the probe. **GLM3.3** was performed on the *session-matching* task in order to retrieve effects of repetition suppression of the hierarchical level (**Fig 3.4c**).

Results

Training learning curves

We first explored the amount of learning that participants achieved during the learning session (**Fig 3.3a**) using a similar approach as for **Experiment 2.3**. We ran 5-way ANOVA against accuracy (% correct) by crossing the factors trial number (1 to 12) × *middle*-blocks (two bins; early 1–2 vs. late: 3–4) × *top*-blocks (two bins; early 1–10 vs. late 11–20) × day (first vs. second) × missing level (top, middle or bottom). Accuracy improved across the twelve trials that composed a *middle*-block ($F_{(11,242)}$ = 202.0, $p < 0.001$), independently for trials where the missing fruit belonged to the top level of the hierarchy (top-completion trials 1–6 vs. 7–12; $t_{(22)} = +13.6$ $p < 0.001$), the middle level (middle-completion trials; $t_{(22)} = +12.1$, $p < 0.001$) or the bottom level (bottom-completion trials; $t_{(22)} = +10.5$ $p < 0.001$). At the beginning of a new *middle*-block, where the fruits corresponding to the middle and low levels of the hierarchy (but not the top) changed, accuracy dropped for middle- and bottom-completion but less so for top-completion trials (*middle*-blocks × level interaction from a 4-way ANOVA including only first trials for each *middle*-block: $F_{(2,44)} = 18.7$, $p < 0.001$).

Accuracy also increased across blocks and throughout the training. We found that there was a higher mean accuracy for late middle-blocks compared to early ($F(1,22) = 99.8$, $p < 0.001$); for late top-blocks compared to early ($F(1,22) = 81.6$, $p < 0.001$); and for the second learning session (day 2) compared to the first (day 1; $F(1,22) = 20.8$, $p < 0.001$). These effects interacted such that the difference in accuracy was particularly pronounced in early trials, in line with the possibility that participants were learning the structure of the task (trial × top-blocks interaction: $F(11,242) = 7.1$, $p < 0.001$; and trial × day interaction: $F(11,242) = 3,3$, $p < 0.001$), and this effect was particularly strong for middle-completion trials (missing level × top-blocks × day interaction: $F(2,44) = 5.2$, $p = 0.01$). All in all, participants performance increased throughout training for all levels, in a manner that reflected the temporal structure of the task.

In comparing these results with those of **Chapter 2**, it is worth noting that accuracy was overall higher (87.96 ± 1.20%) than in **Experiment 2.3** (72.94 ± 4.14%), possibly because participants were drawn from a different population (they were recruited through the University of Granada instead of via Amazon Mechanical Turk).

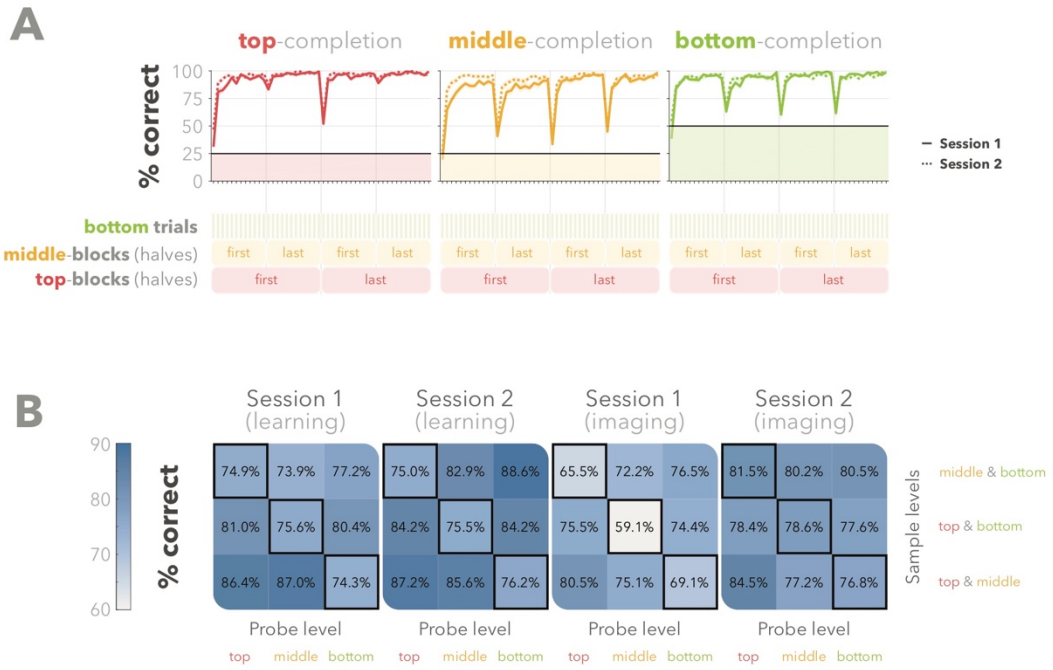

Figure 3.3. Behavioural results for the triplet-validity task
A. Learning curves for the first session (continuous line) and second session (dashed line) during the triplet-validity task. Top-blocks were averaged for the first half (top-blocks 1–10) and the second half (top-blocks 11–20) of the experiment. Middle-blocks were averaged for the first half (middle-blocks 1 and 2) and the second half (middle-blocks 3 and 4) of each top-block. **B**. Accuracy (% correct) for the triplet-validity task as a function of the sample and probe hierarchical levels for the three sessions. The imaging session is shown separately for trials with fruits corresponding to the first and the second session.

Behavioural performance in the triplet-validity task

Participants achieved high performance during the training phase, but this could be driven by the fact that the correct fruit on a given trial had often been presented recently (i.e. in the previous trials), as determined by the temporal correlation of the learning trials. We tested participants' knowledge in a second (*triplet-validity*) task where the order of the triplets was randomised. Participants performed this task also on the third day in the scanner, and so we chose to analyse jointly the behavioural data for the learning and the imaging sessions. Note that participants' accuracy in this

task was above chance by design for the learning sessions (only participants that responded significantly better than chance were invited to the third session).

The performance during the *triplet-validity* task is illustrated in **Fig 3.3b**. Accuracy was generally high, both in the learning session (78.5±2.9%) and in the imaging session (74.9±3.5%). We performed a four-way analysis of variance (ANOVA) regressing the sample (levels top-middle, top-bottom or middle-bottom) × probe (levels top, middle or bottom) × learning session (in which the fruits were learnt first or second) × task session (when they performed the triplet-validity task; directly after learning or in the imaging session). Participants' accuracy during the imaging session was significantly lower than in the learning session directly after the training ($F_{(1,22)}$ = 10.9, $p < 0.001$). Accuracy for triplets from learned in session 1 was marginally lower compared to triplets from session 2 ($F_{(1,22)}$ = 3.4, p = 0.075) but under detailed inspection this effect was driven by the imaging session (learning session × task session interaction: $F_{(2,44)} = 4.0$ $p < 0.055$). It is thus likely that participants successfully learned the triplets by the end of each learning session (as assessed by their performance on the day), but forgot them to some extent before the imaging session. This forgetting was more pronounced for triplets learned in session 1, for which the interval between learning and imaging session was the longest.

We performed a similar ANOVA with the factors validity (valid or invalid) × probe × learning session × task session in order to test for differences in learning between the hierarchical levels. We exploited the fact that invalid triplets (e.g. top-bottom-top) were characterised by a repeating level (top) and a missing level (middle) but had the exact number of occurrences of the remaining level (bottom). We expected that any unbalancing in the learning of the hierarchical levels would be reflected in a difference in performance across the different invalid trials. The interaction between

validity × probe level was not significant ($F_{(2,44)} < 1$, $p = 0.59$) implying that all hierarchical levels had been learnt to a similar extent. Finally, we also found a significant three-way interaction between validity × learning session × task session ($F_{(1,22)} = 7.2$, $p = 0.014$), that captured a low accuracy for valid (not invalid) triplets from "day 1" during the imaging session. This was in agreement with a bias towards "invalid" responses ($t_{(22)} = +4.87$, $p < 0.001$) that accrued with triplets learned in session 1 while performing the task during the imaging session (learning session × task session interaction: $F_{(1,22)} = 6.2$, $p = 0.02$ from an independent 2-way ANOVA on response choice). We interpret this last result as evidence that participants followed a strategy to respond 'valid' when they recalled a triplet and 'invalid' otherwise, and that triplets from the second day were more correctly recalled as these had been learnt more recently.

Functional magnetic resonance imaging data

So far, the behavioural analyses have assessed the amount of learning, as well as provided insight into the strategy that participants are following during the *triplet-validity* task. In particular, participants were sensitive to the presence (not absence) of a valid triplet composed of one fruit from each level (top, middle and bottom) of the hierarchy. In contrast, accuracy did not depend on the level of the probe for invalid triplets where the hierarchical level of the third fruit (probe) repeated one of the levels in the first two (sample) fruits, (interaction not significant). In what follows, we describe some analyses that explore the effects of valid and invalid triplets in the imaging data as a function of the hierarchical level of the probe.

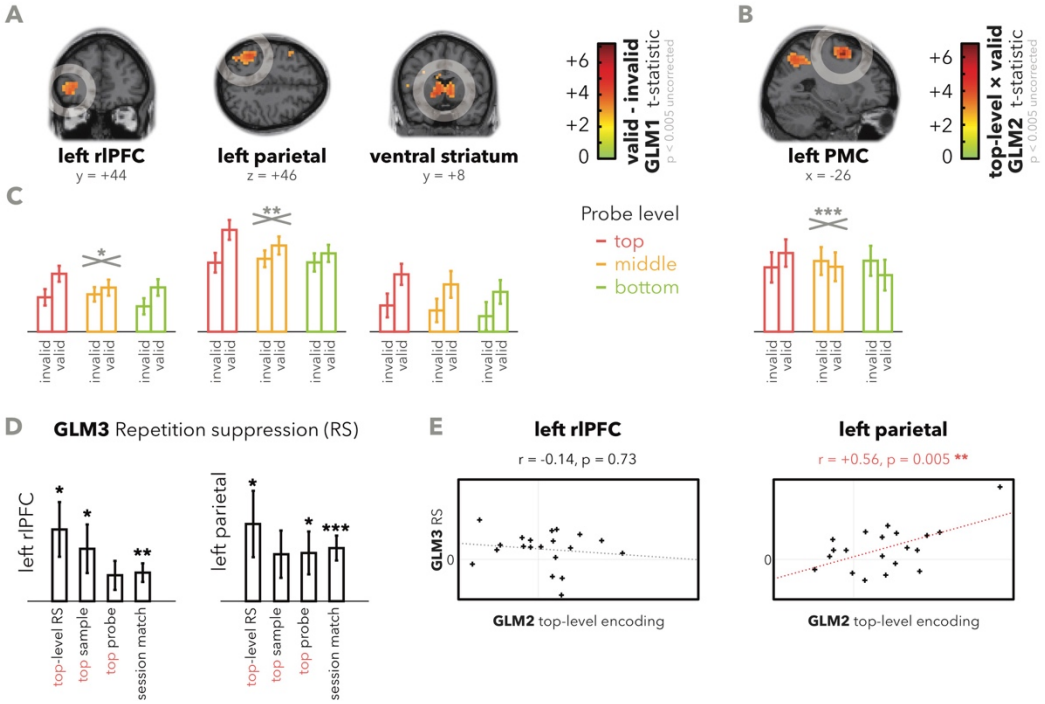

Figure 3.4. Imaging results
A. Results of the contrasts valid>invalid contrast in **GLM3.1** during the *triplet-validity* task, and highlighting three regions of interest. **B.** Results of the level × validity contrast in **GLM3.2** during the *triplet-validity* task, and highlighting the PMC (premotor cortex) as a region of interest. **C.** Region of interest (ROI) analyses showing the mean univariate response of each region (y-axis in arbitrary units) to the probe level and the validity of the triplet. Significant interactions are marked by * ($p < 0.05$), ** ($p < 0.01$) and *** ($p < 0.001$). **D.** ROI analyses on **GLM3.3** during the *session-matching* task, showing the univariate effect of the regressors. Note in particular that the effect of the repetition suppression regressor (top-level RS) is significantly above zero. **E.** Scatter plot and best-fitting linear trends for the ROI effects in **GLM3.2** and **GLM3.3**. Each cross corresponds to one participant. The correlation across participants was significant in the parietal region but not in rlPFC.

Encoding of validity in left parietal and left rostrolateral prefrontal cortices

We analysed imaging data from the three runs in which participants performed the *triplet-validity* task a general linear model. We assessed the relative BOLD signal on correct trials that were valid or invalid, and with a probe that could belong to the top, middle or bottom hierarchical level (thus 6 conditions in total; see **GLM3.1** in the Methods section). We first looked for the effect of valid over invalid triplets, that we would later use in order to define a regions of interest (ROI) (**Fig 3.4a**). Our ROI-

based approach was motivated by the relatively low levels of data that were acquired for each task (further data collection is in progress). Each participant only performed 20 minutes (144 trials) of the *triplet-validity* task in the scanner task, which limited our statistical power.

Nevertheless, we found that a network of regions that were more active for valid than for invalid trials, including a cluster in the left rostrolateral prefrontal cortex (rlPFC; peak -38, +44, -6; $t_{(19)} = +4.7$, $p < 8 \times 10^{-5}$ uncorrected), left parietal cortex (peak -42, -48, +46; $t_{(19)} = +5.6$, $p < 2 \times 10^{-5}$ uncorrected) and bilaterally in ventral striatum (left peak -10, +8, -2, $t_{(19)} = +5.8$, $p < 7 \times 10^{-6}$ uncorrected; right peak +10, +12, -2, $t_{(19)} = +5.7$, $p < 9 \times 10^{-6}$ uncorrected). Note that we failed to identify any significant clusters within in the right hemisphere.

BOLD signal is modulated by the hierarchical level of the fruits in parietal, lateral prefrontal, and premotor cortices

Next, we ran an analysis of variance (ANOVA) on the average beta coefficients from **GLM3.1** within each of the regions of interest, crossing the validity of the triplet (valid, invalid) with the hierarchical level of the probe (top, middle or bottom; see **Fig 3.4b**). We found a significant interaction in left parietal cortex ($F_{(2,38)} = 5.8$, $p = 0.006$) and in left rlPFC ($F_{(2,38)} = 4.2$, $p = 0.022$) but not in ventral striatum ($F_{(2,38)} < 1$, $p = 0.85$). We post-hoc assessed that the direction of the interaction both in left parietal and rlPFC was driven by valid triplets (one way ANOVAs on probe level: $F_{(2,38)} > 4.1$, $p < 0.023$ for both regions) and not by invalid ones ($F_{(2,38)} < 2.0$, $p > 0.150$).

With this first set of results, we performed a second analysis in **GLM3.2** where we explicitly included a regressor that coded for interactions between valid triplets (valid: +1, invalid: 0) and probe fruits from the top hierarchical level (top: +1, middle and bottom: 0) whilst controlling for the main effects of level and validity. In line with our

previous results, we found a significant cluster in left parietal cortex (peak -30, -48, +42, $t_{(+19)} = +4.87$, $p < 5.3 \times 10^{-5}$) and a cluster in right parietal cortex that failed to reach significance (peak 30, -68, +38, $t_{(+19)} = +4.76$, $p < 6.9 \times 10^{-5}$). We found additional clusters significantly in left premotor cortex (PMC; peak -26, +4, +54, $t_{(+19)} = +6.80$, $p < 8.4 \times 10^{-7}$ uncorrected) and without statistical significance in right premotor cortex (peak +30, 0, +58, $t_{(+19)} = +4.71$, $p < 7.7 \times 10^{-5}$ uncorrected). A second interaction regressor between valid triplets and middle level in the probe didn't yield significant clusters. Finally, we performed ANOVAs in the BOLD signal within left PMC and found that this interaction was, similar to left parietal and rlPFC, driven by valid triplets ($F_{(2,38)} = 12.1$, $p < 0.001$) and not by invalid ones ($F_{(2,38)} = 1.9$, $p = 0.17$).

We thus conclude that the left parietal, left rostrolateral prefrontal and left premotor cortices are sensitive to the order of presentation of the hierarchical levels in our task, and were in particular more active when they the top level was last seen. One potential explanation for this is that upon seeing two fruits, participants could predict the level of the third missing fruit (assuming the triplet was valid), and that activity in this region is enhanced when reasoning about the higher level of the hierarchy. We will test this hypothesis in the second, *session-matching* task that participants performed in the scanner.

Behaviour: performance in the *session-matching* task

In the *session-matching* task, participants saw two fruits consecutively ('sample' and 'probe') and reported if they belonged to the same learning session ('session match') or during different learning sessions ('session mismatch'). Responses were highly accurate in the *session-matching* task during the imaging session (92.1±2.1%). We ran a 4-way ANOVA using the hierarchical level and the day of the sample and the probe against accuracy (% correct) but found no significant main effects or

interactions. There was also no bias to respond 'match' compared to 'mismatch' in this task ($t_{(20)}$ = +1.32, p = 0.203). We ran a second, identical ANOVA against (log-) reaction times and found an interaction between sample day × probe day ($F_{(1,20)}$ = 54.3, p < 0.001) that was driven by 'session match' responses being faster than 'mismatch'. We found an additional, unexpected effect of the hierarchical level of the sample ($F_{(1,20)}$ = 3.5, p = 0.040) and a marginal interaction between the level × the day of the sample ($F_{(1,20)}$ = 3.0, p = 0.055), such that responses were faster when the sample belonged to the top hierarchical level and to the second day.

Repetition suppression

In the *session-matching* task participants performed 432 trials (around 40 minutes) in the scanner. On each trial, they saw two fruits ('sample' and 'probe') quickly in succession. We designed this task in order to look at effects of repetition suppression (RS). Repetition suppression refers to the fact that the neural response to information tends to decrease with its amount of exposure. For instance, after exposure to multiple faces in a row, the fusiform face area that encodes for face information tends to react less strongly (Pajani, Kouider, Roux, & De Gardelle, 2017). RS has been successfully used as a technique to reveal neural regions that encode for similar information across different stimuli, with perceptual tasks but also with cognitive tasks (Barron, Garvert, & Behrens, 2016).

Given our original findings that the left parietal and lateral PFC encode for the higher levels of the hierarchy in the *triplet-validity* task, we expected that this would also be the case in the *session-matching* task even if this information was not required to perform the task. We further expected that we would see an effect of repetition suppression, such that fruits in the higher levels of the hierarchy would react less

strongly as a function of the hierarchical level of the probe when they matched the level of the sample.

Effects of repetition suppression of the hierarchical level in left parietal and lateral prefrontal cortices

In the *triplet-validity* task we found that the parietal cortex was bilaterally encoding the higher level of the hierarchy. We thus expected that in the *session-matching* task we would find an effect of repetition suppression in this region when, at quick presentation of two fruits, both belong to the highest level of the hierarchy, but not if either of them belongs to the middle or bottom levels of the hierarchy.

We performed a general linear model (**GLM3.3**) in order to test this hypothesis. Of interest, we included two parametric regressors that affected all correct trials where a 'top' fruit was present (at the sample or at probe respectively). These regressor made a linear assumption such that if both fruits in a trial corresponded to the top level the expected BOLD signal modulation would have double the amplitude. We included an additional regressor that modelled the effect of repetition suppression, namely when both the sample and the probe belonged to the top level of the hierarchy. The full set of regressors for **GLM3.3** can be found in the Methods section, as well as in the **Supplementary Table ST3.3**.

The results are illustrated in **Fig 3.4d**. We found an effect of repetition suppression in left parietal cortex ($t_{(19)} = -2.27$, $p = 0.035$), and in left rostrolateral PFC ($t_{(19)} = -2.56$, $p = 0.019$). We note in passing that BOLD signal in these regions was also sensitive to the regressor encoding the presence of a *top*-level fruit (both $t_{(19)} > +2.25$, $p < 0.036$) and were more active when the two fruits were from the same day ('session match'; both $t_{(19)} > +2.9$, $p < 0.007$). These results support the finding that these regions are active when reasoning about the highest level of the hierarchy.

Encoding of the hierarchical level encoding correlates across the cohort

We assumed that if the effects in the *triplet-validity* and in the *session-matching* tasks were having a common source, namely the representation of the hierarchical level, then participants with the strongest encoding in one task should also display the strongest encoding in the second task. Thus, we decided to look for between-participant correlates in the left parietal and left rostrolateral PFC by using the contrasts of interest from **GLM3.2** (valid × probe top level) and **GLM3.3** (repetition suppression). This analysis yielded a significant correlation in the left parietal region ($r = +0.56$, $p = 0.005$ one-sided) but not in the left rostrolateral prefrontal cortex ($r = -0.14$, $p = 0.727$ one-sided). This final result validated the left parietal region has encoding the hierarchical level of the relationships acquired during the learning sessions.

Discussion

In this experiment, we investigated the neural signatures of hierarchical representations in the neural brain. We carefully designed the experiment in order to control for the frequency and the variability within each hierarchical level. Behaviourally, we found that participants had successfully learnt the relationships by the end of two learning sessions, and despite some forgetting during the interval before the imaging session, participants performed the task above chance and similarly for all levels of the hierarchy.

At the neural level, we found that two regions (left parietal and left rostrolateral prefrontal cortices) encoded the level of the hierarchy. In the *triplet-validity* task, these regions were more active when participants expected a probe fruit belonging to

the top level of the hierarchy compared to the bottom level. This difference could not be explained by difficulty, because performance was matched in these trials. Similarly, it could also not be explained by any differences in learning between levels, because valid trials only differed in order and always included one fruit from each level. In the *session-matching* task, these two regions were more active when a fruit from the top level of the hierarchy was presented, but further displayed an effect of repetition suppression whereby this effect was not additive when both the sample and the probe included (distinct) fruits from the top level. Further, the magnitude of the neural effects in these two tasks correlated across participants only within the left parietal cortex.

The parietal cortex has previously been related with representations of magnitude coding, such that larger magnitudes displayed more activity parametrically in the presence of symbolic digits (Spitzer, Waschke, & Summerfield, 2017) and non-symbolic number stimuli (Teichmann, Grootswagers, Carlson, & Rich, 2018). Incidentally, magnitude can be understood as the simplest hierarchical structure. During the imaging session, participants were only required to remember the hierarchical level of the fruits (top, middle, or bottom) and this could have been encoded linearly similar to any other magnitude. In this analogy, we found that the top level of the hierarchy would be "higher" than the middle and bottom levels.

Additionally, the frontopolar cortex (FPC; which partially overlaps with the rostrolateral prefrontal cortex) has been suggested to be critical for 'cognitive branching', a function that is critical for representing nested tree structures and hierarchies (Koechlin & Hyafil, 2007).

Interestingly, in the *triplet-validity* task this was only the case for valid trials. This might be explained by an interaction between the expectation (set by the sample screen) and its occurrence (in the probe), such that rostrolateral PFC and parietal cortex only encode the hierarchical level when it is both expected and observed. An alternative (not mutually exclusive) interpretation is that there was an effect of repetition suppression in invalid trials with two top-level fruits (i.e. when participants saw two top-level fruits out of three within a trial), which is why we didn't see any effect of the probe (suppression of the top-level probe fruit). We could test this possibility cleanly in the *session-matching* task where participants saw two fruits, and found an effect of repetition suppression that applied only to trials where both fruits belonged to the top level of the hierarchy. The effect across the two tasks correlated positively across participants in left parietal cortex, making these results even more convincing.

I note in passing that in this experiment I could not distinguish between alternative neural codes for the hierarchical level. For example, is the parietal region encoding only the top-level, or is the middle-level also encoded such that it is significantly higher than the bottom-level and significantly lower than the top-level? Additionally, one can assume that repetition suppression works at the level of individual neurons (Barron et al., 2016), in which case one could use this method to test whether the top and middle levels are represented by overlapping or distinct populations of neurons. However, the data was noisy enough that we could reliably answer these questions.

Chapter 4. Neural mechanisms of hierarchical planning in a virtual subway network

Abstract

Planning allows actions to be structured in pursuit of a future goal. However, in natural environments, planning over multiple possible future states incurs prohibitive computational costs. To represent plans efficiently, states can be clustered hierarchically into "contexts". For example, representing a journey through a subway network as a succession of individual states (stations) is costlier than encoding a sequence of contexts (lines) and context switches (line changes). Here, using functional brain imaging, we asked humans to perform a planning task in a virtual subway network. Behavioural analyses revealed a hierarchical cost, with slower responses when further away from the goal in number of contexts (lines). Brain activity in the dorsomedial prefrontal cortex and premotor cortex scaled with the cost of hierarchical plan representation, and unique neural signals in these regions signalled contexts and context switches. These results suggest that humans represent hierarchical plans using a network of caudal prefrontal structures.

Introduction

As I mentioned in the *General introduction* in **Chapter 1**, hierarchical representations are not only helpful for representing static observations, but also to represent relationships that span across space, time, and in particular in the context of planning. In this chapter, I revise the background on planning and describe a study where I aimed to understand the neural underpinnings of hierarchical planning.

By forming and executing plans, humans can engage in complex behaviours such as preparing a cup of coffee, or organising a trip to London. When asked to perform multistep tasks such as these, patients with lesions to the prefrontal cortex (PFC) often exhibit disordered action sequences that fail to achieve the specified goal (Owen et al., 1990; Shallice, 1982; Shallice & Burgess, 1991), and hippocampal patients have difficulty imagining the future states entailed (Schacter et al., 2012). Moreover, functional neuroimaging has confirmed the involvement of human prefrontal and limbic structures in forming and executing plans, particularly in spatial environments (Howard et al., 2014; Schacter & Addis, 2007; Unterrainer & Owen, 2006). Nevertheless, linking these macroscopic neural findings to the underlying computational mechanisms that subserve planning remains an open challenge for psychologists and neuroscientists.

Planning is often described as mental exploration of a network of interlinked, internally represented episodes (or 'states'). According to one conception, future states belong to a decision "tree" in which each node is a decision point and each branch a possible response. Plans are representations of trajectories through the tree, selected on the basis of their long-term cumulative outcome (Daw et al., 2011; Daw, Niv, & Dayan, 2005; Huys et al., 2012; Russell & Norvig, 2016). Computer-based algorithms have successfully exploited this strategy to achieve expert levels of performance in board games such as chess and weiqi (Go) (Silver et al., 2016, 2017). However, because the number of possible action sequences grows exponentially with each additional step in the planning horizon, this approach is computationally intractable in many natural environments (Gershman, Horvitz, & Tenenbaum, 2015).

For example, a visitor would probably not plan a trip to London by envisaging every unique interim step en route to the destination, but might rather imagine attaining only a subset of key states, such as reaching an airport or other transport hub.

In machine learning and computational neuroscience, it is widely recognised that the computational demand associated with planning can be reduced by exploiting hierarchical structure in the environment, with states clustered into larger "contexts" (Badre et al., 2010; M. M. Botvinick et al., 2009; Koechlin & Jubault, 2006; Sutton & Barto, 1998). To understand how a hierarchical representation may alleviate the computational burden of planning, consider a metropolitan rail (subway) network, in which stations (i.e. states, e.g. King's Cross, Oxford Circus) are organised into lines (i.e. contexts, e.g. the Victoria Line; see **Fig. 4.1a**). Unlike planning in a "flat" (non-hierarchical) environment, plans formed in a hierarchical environment need not specify each and every state linking the current position and goal. Rather, it is sufficient to identify the current context, and the (termination) conditions that allow the next context to be reached; for example, when planning a journey from Marble Arch to King's Cross on the London Underground, one should "take the Central Line to Oxford Circus, and from there, switch to the Victoria line". Humans seem to represent locations hierarchically in spatial memory: for example, we have a bias to judge cities belonging to a common region (e.g. Nevada) as geographically closer than those crossing a region boundary (Newcombe & Liben, 1982; Stevens & Coupe, 1978). Regionalisation may also influence navigational strategy: during wayfinding, humans prefer routes that permit a context boundary to be crossed earlier rather than later (Wiener & Mallot, 2003). In machine learning, states that offer privileged access to a new context (such as Oxford Circus allowing access to the Victoria Line) are

considered “bottlenecks”, and hierarchical learning models successfully predict that visiting these should elicit unique patterns of behaviour and neural activity (Holroyd & Yeung, 2012; Ribas-Fernandes et al., 2011; Solway et al., 2014).

Here, thus, we taught participants to navigate a novel subway network in which stations (states; e.g. Mandela, Budapest) were organised hierarchically into lines (contexts) defined by their colour (**Fig. 4.1b**). Following training, participants were asked to complete journeys within the network without viewing the map, pressing keys to move from one station to another. We analysed behaviour and functional magnetic resonance imaging (fMRI) data in order to determine whether humans represented plans in a hierarchical fashion (over lines, or contexts) or a “flat” fashion (over stations, or states). On the neural level, an extensive literature has implicated both the medial and lateral prefrontal cortex in planning on multi-step decision tasks such as the Tower of London (Unterrainer & Owen, 2006), but the relative contribution of these different regions remains unclear. Some studies have found that the BOLD signal in dorsolateral prefrontal cortex scales with the number of moves required to attain goal state (van den Heuvel et al., 2003; Wagner, Koch, Reichenbach, Sauer, & Schlösser, 2006), but neutral structures encoding hierarchical plan complexity have yet to be identified. One theoretical perspective has suggested that the dorsomedial prefrontal cortex (dmPFC) may play a particular role in representing contextual information for future behaviour (Holroyd & Yeung, 2012). During passive observation of trajectories through a structured environment, the dmPFC is less active at bottleneck states (Schapiro et al., 2013), but by contrast, a more caudal medial prefrontal region shows a positive “pseudo-reward” signal when a

subgoal is attained (Ribas-Fernandes et al., 2011). It thus remains unclear how the medial and lateral prefrontal cortex might contribute to hierarchical planning.

To preview our findings, we identified two frontal cortical regions that encoded the cost of representing a hierarchical plan: a bilateral anterior premotor region, and the dmPFC. These regions also became differentially active at bottleneck states ("exchange" stations, where participants could switch from one context to another). Using multivariate analyses, we found that the dmPFC additionally encoded or monitored the current context (i.e. the subway line that was currently being taken), a key quantity that is required for executing a hierarchical plan. By contrast, the rostromedial prefrontal cortex and hippocampus encoded the proximity to a goal state. Together, these findings suggest that during planning, humans encode the subway network and formulate plans in a hierarchical fashion.

Methods

Subjects

A total of 22 healthy participants (10 female, 12 male; age 19-34, mean 25.6 years; one was the first author of the study) were recruited into the study in accordance with local ethical guidelines. No participants reported a history of psychiatric or neurological illness, and all had normal or corrected-to-normal vision. Participants were paid £35 for participation in both a practice and a scanner session on two separate days. A monetary incentive of up to £10, proportional to performance was added to the previous amount. Two participants were excluded due to poor performance on the task (more than 20% of the journeys included a move in the wrong direction during the main experiment).

Stimuli and task design

The same subway map was used for all participants, but the names of the stations and the colours of the lines were randomly shuffled, and the map was randomly rotated by 0º, 90º, 180º or 270º (example shown in **Fig. 4.1b**). Following training (see below), participants performed the main task, which involved navigating in a virtual subway environment, in the MRI scanner. Each journey involved a start station and a destination station that were randomly selected with the constraint that the journey would require at least one change of line (17.8% of journeys) or one change of direction without changing lines (10.7%) or both (71.5%). Participants navigated through the subway map by pressing buttons (see below). On each trial, there was a constant probability that the journey was cancelled, engineered such that cancellation probability was independent of the length of the optimal journey and led to approximately 50% of journeys being cancelled; cancellation probability was independent of the hierarchical aspects of the task. Overall, 52.9% of journeys were successfully completed. Each journey was rewarded with a monetary value (either one or five virtual coins, signalled during navigation) which were converted to real incentives (normalised to a maximum of £10) that were paid out as a bonus at the end of the experiment. Behavioural performance did not differ as a function of the incentives offered, so we collapsed over this factor for all analyses.

Procedure

The main task is depicted in **Fig. 4.1c**. Each journey began with the presentation of a cue screen for 3 seconds that indicated the starting point and the destination (stations and lines). After a period of 2-5s (jittered) of blank screen, on each of the successive trials a navigation screen was displayed for 3 seconds. This screen provided multiple pieces of information: the names of the current and destination stations; the line

colour of the destination station; the reward at stake for the current journey; the cumulative reward so far; and the cardinal directions (North, South, East, West) available from the current station. Critically, no information about the current line nor about the line associated with each action were shown. At each step, participants had to choose the direction they wanted to take by pressing one out of four buttons. If no key was pressed, the same station was shown again in the next step. Each navigation screen was followed by a blank screen of 1-3s (jittered); no feedback was provided during navigation.

The journey ended either when the participant reached the destination or when the journey was cancelled. After the journey was finished or cancelled, a feedback screen informed whether the destination had been reached or not and the reward that had been obtained. This screen was displayed for 2s and followed by a blank screen of 2-5s (jittered) before the next cue screen occurred. Participants completed as many journeys as possible in 4 successive runs buttressed by lead-in and lead-out durations of 10s and 5s respectively. The total scanning time, including anatomical and localisers scans was around 75 minutes per participant.

Training task

All participants were trained in a separate behavioural session that took place outside the scanner exactly two days before the main experimental task. This training session was similar to the main task, with the following exceptions. Firstly, the map (e.g. **Fig. 4.1b**) was shown for 10s prior to the start of each journey. Secondly, participants were allowed unlimited time to respond, moving on to the next screen only after a key press had been initiated. Thirdly, the available actions were shown in the colour of the corresponding line, and a picture matching the name of each station was displayed

consistently in the background to facilitate the learning of the map. An additional key press (Space bar) was required to switch between lines, and during a line switch, an animated clock was shown on screen and a delay of 1s was imposed. On each journey, the starting and destination stations were selected uniformly, permitting a larger number of possible journeys, and at the end of each journey, a feedback screen informed the participant of (i) the total length of the journey, and (ii) the minimum length that could have been be achieved (i.e whether their journey had been optimal or not). During training journeys were never cancelled, and no monetary outcomes were associated with successful journeys. Lastly, we introduced 10 "quizzes" at homogeneous times during the training session (always between journeys), each including 10 "questions" where the current station and the goal were cued, but participants were only required to respond the first step towards the goal. Participants were informed of the scores obtained at the end of each quiz, and were instructed to learn during the whole session as to maximise their scores during the quiz. They completed as many journeys as possible over a period of 45 minutes.

On the day of scanning, before entering the MRI, participants performed a practice block identical to one of the main task scanner runs. They were allowed to see the map one last time before the beginning of this second training session. Data from this session were not included in the analyses.

fMRI Acquisition

Magnetic resonance images were acquired with a 3T Siemens VERIO scanner with a 32-channel head coil using a standard echo-planar imaging sequence. Whole-head T2*-weighted echo-planar images were continuously acquired with a repetition time of 2 s, echo time of 30 ms. We acquired fMRI data in 4 runs (~17 minutes each) of

between 456 and 510 volumes, plus 3 dummy scans discarded before the analyses. For technical reasons, three participants completed only three runs. Each volume included 64 × 64 × 36 voxels of 3 × 3 × 3 mm. A high-resolution T1-weighted structural image was also obtained (voxel size = 1 × 1 × 1 mm). For standard preprocessing and univariate statistical analyses, we used SPM12 (Wellcome Department of Cognitive Neurology, London, United Kingdom). All other analyses were carried out with custom scripts for Matlab (Mathworks, Natick, MA). We also used XjView (http://www.alivelearn.net/xjview) to visualize the data and to construct mask images and impose an FDR correction for multiple comparisons (Genovese et al., 2002). For each participant, we first realigned all functional images, then we co-registered (rigid body transformation) the anatomical scan to the mean functional image. We then segmented each subject's co-registered anatomical scan, using segmented probabilistic maps for grey matter, white matter, cerebro-spinal fluid, bone, soft tissue and air/background in the Montreal Neurological Institute (MNI) space. The parameters obtained were applied to normalise the subject's functional scans to the template brain MNI space. Functional images were resampled (3 × 3 × 3 mm voxels) and spatially smoothed (6-mm full-width half-maximum (FWHM) Gaussian kernel). For all analyses, a 128-s temporal high-pass filter was applied to remove low-frequency scanner artifacts. Temporal autocorrelation in the time series data was estimated using restricted maximum-likelihood estimates of variance components using a first-order autoregressive model (AR-1), and the resulting non-sphericity was used to form maximum-likelihood estimates of the activations, consistent with standard approaches in SPM (Penny et al., 2006).

Behavioural analyses

We analysed log reaction times with linear regression as described in the main text, and significant contribution of each regressor was validated through a t-test using an alpha of $p < 0.05$. All regressors and interactions were z-scored before being introduced in the regression. The optimal path was obtained through a generalised version of the Dijkstra algorithm that minimised multiple distances, by priority: in number of stations, number of response switches, and number of exchange stations. The U-turn cost was defined as the signed difference between the distance in number of stations and the Manhattan (City-block) distance : $DU(a,b) = DS(a,b) - |xa - xb| - |ya - yb|$, where (xi,yi) are the geometrical coordinates of a station i, $|\cdot|$ is the absolute value operator, and DS is the distance in number of stations. An illustration of how the various indices of distance to goal were computed is shown in **Fig. 4.1d**. The original Dijkstra algorithm was based on in-house code.

Univariate analyses of functional data

All univariate analyses were based on a generalized linear model (GLM) approach. Our GLM included regressors coding for onsets and durations of stimuli or events, which were then convolved with the canonical haemodynamic response function (HRF) and regressed against the observed fMRI data. Scanner runs were concatenated for univariate analyses, and constant terms for each run were included manually. Additionally, motion parameters and the average signal outside of the brain were included as nuisance variables for all GLMs. Group-level statistics were estimated from the individual β patterns, not the within-subject statistics.

The main analyses described in the paper were based on 2 GLMs. Unless otherwise specified, we only considered journeys where the participant always moved towards

the goal ('optimal' journeys), but other journeys were modelled separately. **GLM4.1** included the following conditions convolved with the canonical HRF basis function: main effect of cue screen; main effect of feedback screen; main effect of navigation screen for sub-optimal journeys. We modelled navigation screens during optimal journeys independently for (i) line changes, (ii) exchange stations without a line change, (iii) elbow stations, and (iv) regular stations without response switch. Additionally, we included the following parametric modulators for regular stations without response switch: distance to goal in number of stations (DS); distance to goal in number of line changes (DC); distance to goal in number of exchange stations (DX) and the U-turn cost (DU). **GLM4.2** included the following conditions: main effect of cue screen; main effect of feedback screen; main effect of navigation screen. Additionally, the navigation screen included the following parametric modulators: type of station (exchange > regular); type of response (switch > stay); interaction between station and response; distance to goal in number of stations (DS); and performance on the current step (1 if optimal, -1 otherwise).

All effects reported survived FDR correction for multiple comparisons, unless noted in the main text. Images and tables are thresholded at $p < 0.001$, unless otherwise noted. All the analyses described here focused on effects during the time of navigation. Peak activations are reported with coordinate system of the Montreal Neurological Institute (MNI) template brain. Regions of interest (ROI) were defined by manually selecting clusters under a threshold of $p < 0.001$ uncorrected.

The mask in rlPFC was extracted from a main effect of distance to goal in **GLM4.2**.

BOLD-RT correlation analysis

We extracted the average beta obtained from **GLM4.1** and we obtained average values for dmPFC and PMC. We also obtained similar beta values of effect of DS and DL in explaining log-reaction times (see Behavioural analyses). We then performed a non-parametric Spearman correlation across participants for each region and type of distance.

Single-trial GLM approach

We performed a single-trial analysis in order to extract the average signal in PMC before and after a line change, an elbow station, an exchange station without response switch (i.e. a line stay), or a regular station without response switch. Firstly, we constructed a design matrix in which each trial was modelled with a unique regressor. From this we obtained a single scalar BOLD estimate for each voxel on each trial. We then averaged these values within the PMC region for each station type. To avoid double-dipping, our ROI was defined based on orthogonal contrast of type of station (exchange > regular) from **GLM4.2** ($p < 0.001$ uncorrected). Secondly, we averaged the PMC signal for the neighbouring trials around each condition (i.e. line change, elbow station, line stay, and regular station) within the journey. Critically, we restricted these neighbouring trials only to regular stations without response switch.

Our prediction was that the BOLD signal in PMC would be higher before a line change than after, but that this difference would not be reflected around elbow stations or exchange stations without a line change. We calculated the difference on the trials immediately before/after each condition, and performed a statistical analysis on the main effects of type of station and type of response on this difference. For better visualisation we controlled for between-subject variability in **Fig. 4.3a**, where

we displayed the activity in PMC of all other conditions relative to the average signal in regular stations without response switch.

Representation Similarity Analysis (RSA)

For RSA, we constructed a new GLM with four regressors (per scanner run) that each encoded regular stations (without a response switch) corresponding to one subway line (context regressors), and 4 further parametric regressors that modulated each event by distance to goal (in number of stations; context distance regressors). We used unsmoothed images for this analysis. Additional regressors encoded other quantities (cue screen; feedback screen; in navigation: line changes, "elbow" stations, and "exchange" stations without a line change; and nuisance regressors). We used a searchlight approach, in which a sphere of 15mm radius was moved progressively over the brain volume, with the resulting RSA estimates allocated to the centroid voxel for localisation and display. Results obtained with a smaller radius (10mm) were qualitatively very similar. For context decoding (**Fig. 4.4b**) we estimated for each scanner run (n = 4) the pattern of resulting betas for each of the 4 context regressors, and computed their correlation distance (1-Pearson correlation) yielding a 16 x 16 neural dissimilarity matrix. This matrix was regressed against the predicted Representation Dissimilarity Matrix (RDM) shown in **Fig. 4.4a** within each searchlight, and statistic performed on the resulting betas at the second (between-subject) level. In the predicted RDM, distances were greater between lines than within lines. We excluded comparisons within a single run, to control temporal autocorrelation in the within-session BOLD signal. An identical approach was used for the context distance regressors (**Fig. 4.4c**). In the control condition (**Fig. 4.4d**), the assignment of regular and elbow stations to each line was shuffled, so that the

hierarchical structure was lost. We then repeated the estimation of beta patterns and the searchlight RSA approach as above.

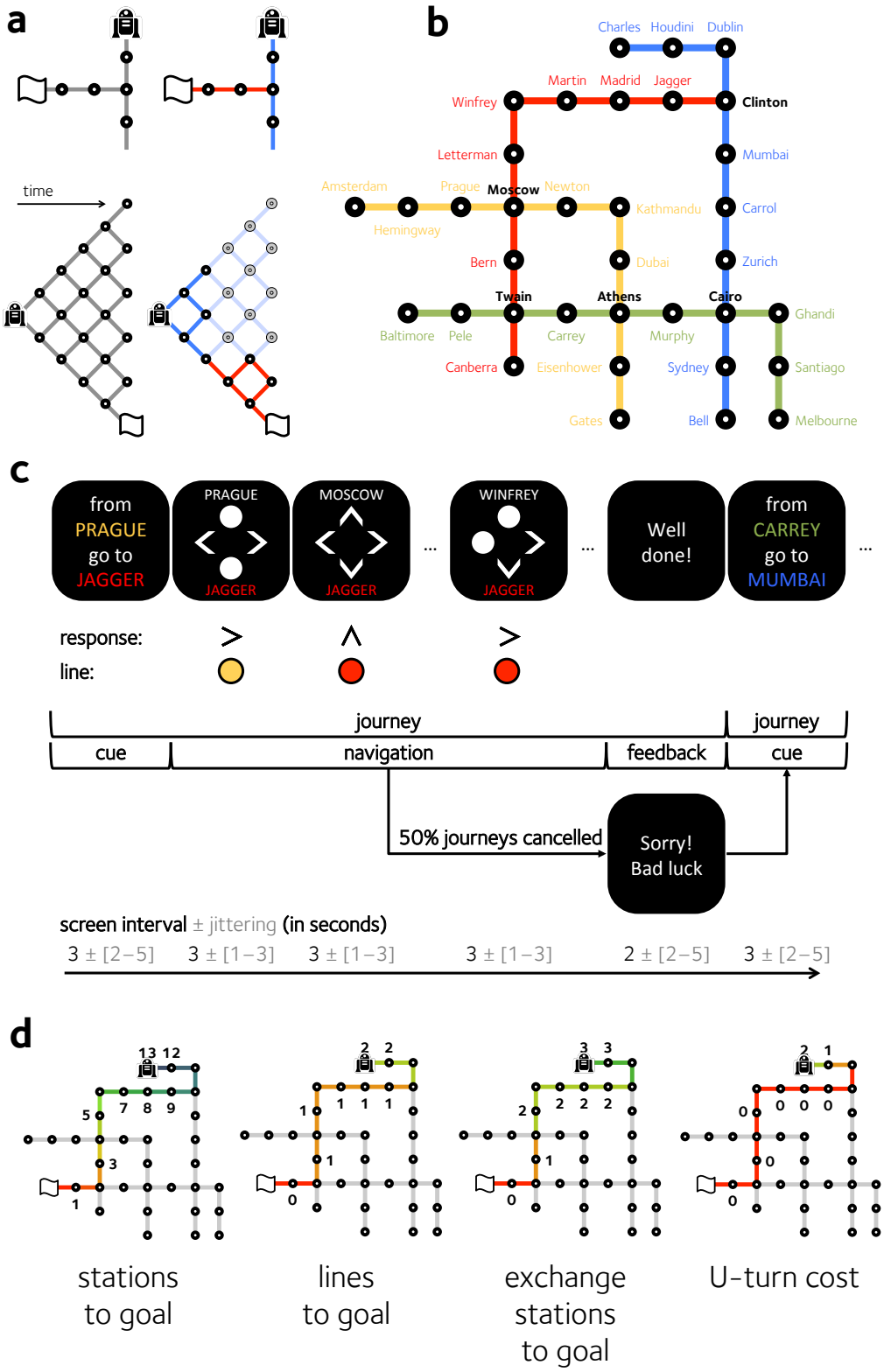

Figure 4.1. Task design
A. Schematic representation of planning under a flat (left) and hierarchical (right) policy. Each node from left (start state, shown by the robot) to right shows a possible state (i.e. station) that could be visited. The flag indicates the destination station. A hierarchical policy allows the agent to "chunk" the maze into contexts (here, a red line and a blue line). This in turn reduces the cost of planning and plan representation. **B.** The subway map that participants navigated. The map was rotated and the line colours and station names were shuffled between participants. Participants only saw the map during training. **C.** A schematic depiction of the sequence of events (trials) that occurred on an example journey. The names at the top and bottom of the screen refer to the current and destination stations respectively. The responses (arrows) and lines (coloured dots) were not shown to participants. Timings (in seconds) for the various events are shown below. **D.** Examples of how the various distances were calculated for an example map: DS (stations to goal) DL (lines to goal), DX (exchange stations to goal) and the U-turn cost DU. Numbers and blue-red colormap show the distance in each metric that was used to estimate the cost of planning. The robot shows the start point, and the flag shows the destination station.

Results

Task summary

The task is depicted in **Fig. 4.1c**. Each journey began at a pseudo-randomly chosen station (see methods). On each trial, the names of the destination and current stations were shown, and participants pressed one of four buttons (north, south, east, west) to move to an adjacent station, which was then shown on the next trial. Their goal was to navigate through the subway map from the start station to the destination station (these successive trials comprising a "journey"). During an initial training session, lines were associated with colours (red, green, yellow, blue), but at scanning, all colour information was removed. Successful journeys were rewarded with financial incentives, but there was a small but constant probability that journeys were "cancelled" on each trial and the reward was unavailable, motivating participants to make journeys in the shortest possible number of trials. Participants carried out 88.8±2 journeys in total, each consisting of an average of 5.5±0.06 trials. Of these, 78.3% were performed "optimally" (i.e. when all responses decreased the distance to

goal in number of stations). Of the remainder, 15.2% contained at least one action that led participants further away from the goal; these responses were made more slowly (t_{19} = 7.56, p < 0.000001). Additionally, 9.0% of journeys included at least one missing response (when subjects failed to respond on time, and remained in the same station as in the previous trial).

Behaviour: the cost of plan representation

The complexity (or description length) of representing a "flat" (nonhierarchical) plan is proportional to the number of remaining states (here, stations) that must be traversed to reach the goal (here, destination station). By contrast, in a hierarchical plan, this cost scales with the remaining number of contexts that must be traversed for the goal to be attained. We thus began by defining measures of plan complexity that might be computed by participants under flat and hierarchical policies. First, we calculated, on each trial, the number of steps (stations) that remained to be traversed before the goal was reached, assuming a shortest path trajectory (DS). This represents plan complexity under a flat policy (see **Fig. 4.1d**, leftmost panel). Next, we calculated the number of contexts that remained to be traversed before the goal was reached. Thus, if on the current trial there was only one change of context that would be required to reach the goal, this value would be 1; beyond that context switch, the value would be 0. This quantity DL indexes the cost of a hierarchical policy (**Fig. 4.1d**, centre left panel). Then, as a control, we computed the distance to goal in number of exchange stations to be traversed. By design, on many journeys the shortest path involved passing through an exchange station without switching context (**Fig. 4.1d**, centre right panel). This measure, which we call DX, was thus decorrelated from DL (for details of the correlation among distance measures, see **Supplementary Table ST4.3**). Finally, we computed another cost, which represented

the number of steps that had to be taken away from the goal (in cityblock space) in order to reach it by the shortest path. Thus, this measure, which we call the U-turn cost (or DU) was high for paths that required "doubling back" (**Fig. 4.1d**, rightmost panel).

We then used linear regression to ask whether (log) response times (RTs) during navigation were sensitive to the complexity of the plan as indexed by DS, DL, DX and DU. Critically, this analysis yielded significant positive coefficients for number of lines to goal (DL: $t_{19} = 3.46$, $p = 0.003$) and for the U-turn cost (DU: $t_{19} = 4.26$, $p < 0.001$; see **Fig. 4.2a**). When these predictors competed for variance within a single regression, however, the number of stations to goal failed to predict RTs (DS: $t_{19} = 1.26$, $p = 0.223$), as did the number of interchange stations (DX: $t_{19} = -0.49$, $p = 0.628$). This finding suggests that the main costs of representing the plan were contextual or structural aspects of the subway map, rather than the number of unique steps required to reach the destination station. This supports the view that plans are formed and executed in a hierarchical fashion.

We defined stations as "regular" (i.e. within a single line; e.g. Madrid in **Fig. 4.1b**) and "exchange" (i.e. "bottlenecks", occurring at the intersection between lines, e.g. Clinton). Moreover, responses were classified as either stay (i.e. travel in the same direction as the previous step) or switch (i.e. change the direction of travel). These factors were orthogonal in our paradigm, because regular stations sometimes required a direction switch, as when a single line turned a corner (e.g. Kathmandu in **Fig. 4.1b**) but participants could also pass through exchange stations without switching response (e.g. when passing through Moscow en route from Winfrey to Bern). This feature of

our design thus allowed us to further include, in the above regression, separate binary predictors encoding station type (exchange vs. regular) and response type (switch vs. stay). We observed a main effect of station type (exchange > regular, $t_{19} = 3.40$, $p = 0.003$) and of direction (switch > stay, $t_{19} = 7.92$, $p < 0.001$). The interaction between station type and response type was not significant ($t_{19} = 1.05$, $p = 0.309$). Mean RTs in each condition are plotted in **Fig. SF4.1**.

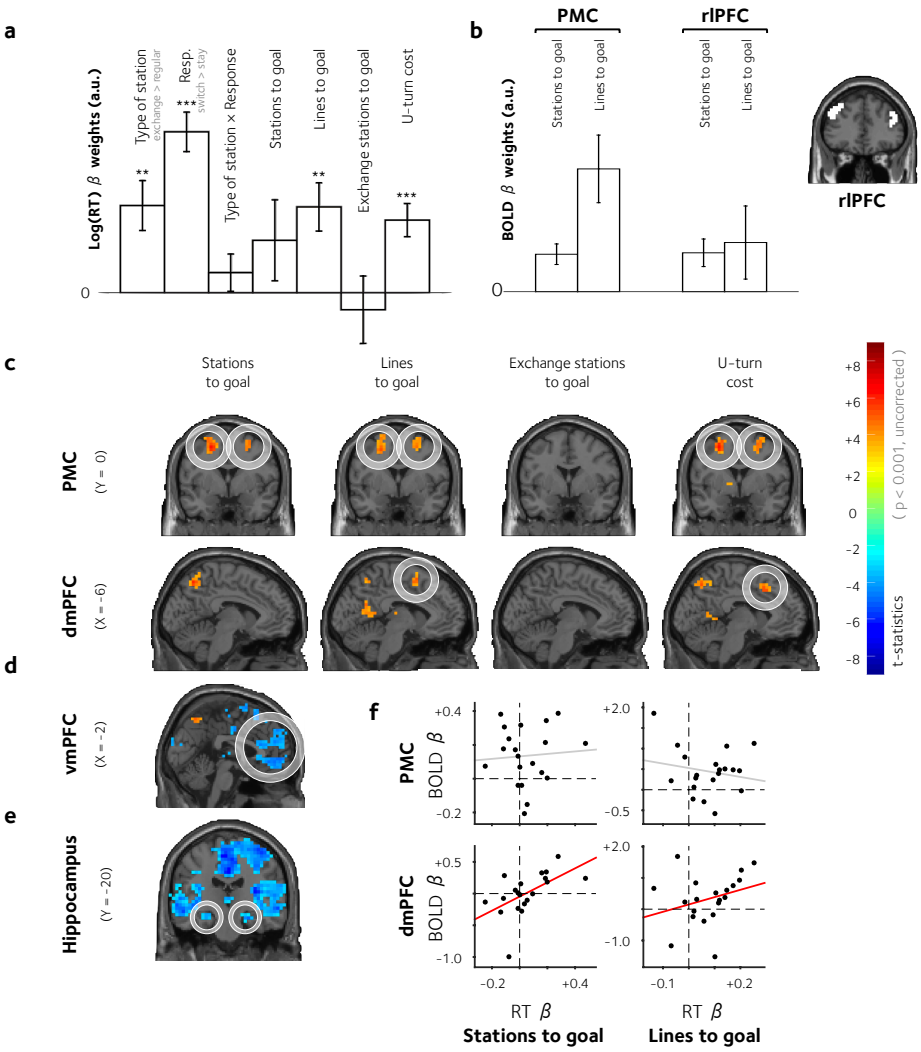

Figure 4.2. Behavioural and neural costs of plan description length
A. Regression coefficients (mean ± SEM across participants) showing the slope of the predictive relationship between experimental variables (including distance estimates) and log response times (RTs). **B.** Parametric responses (mean ± SEM) to DS and DL in the PMC and rlPFC. There is a significant condition × region interaction. The rlPFC ROI is shown on the right. **C.** Encoding of the 4 plan complexity measures

(GLM4.1) in the lateral (coronal view; upper panels) and medial (saggital view; lower panels) frontal cortices, rendered onto a template brain, thresholded at $p < 0.001$ uncorrected. **D.** Correlation with proximity to goal (GLM4.1) in the vmPFC. **E.** Correlation with proximity to goal (GLM4.2) in the hippocampus. Activations are shown that exceed $p < 0.001$, uncorrected. **F.** Correlation between parameter estimates linking log(RT) to plan complexity in units of station (left panels) and lines (right panels), with beta values encoding the corresponding distance measure in the PMC (upper panels) and dmPFC (lower panels). Dots correspond to individual subjects. Lines are to best linear fits for significant (red) and non-significant (grey) correlations respectively. Significant regions within a circle survived multiple comparisons correction.

Neural cost of plan representation

Next, we sought to identify in the brain imaging data the neural costs of representing flat or hierarchical plans. In this analysis and all that follow, all reported results survive correction for multiple comparisons using a false discovery rate (FDR) with an alpha of $p < 0.05$, unless otherwise noted. We built a design matrix (**GLM4.1**) with regressors encoding the various indices of distance to goal introduced above (DS, DL, DX and DU; **Fig. 4.2c**). Examples of how these distances were computed are shown in **Fig. 4.1d**. Regressing this design matrix against BOLD data, we found that a dorsal portion of the medial prefrontal cortex (dmPFC; BA8/32) responded positively to the cost of plan representation in units of both lines (peak: -6, 8, 58; $t_{19} = 5.21$, $p < 0.0001$) and the U-turn cost (peak: -2, 12, 46; $t_{19} = 5.63$, $p < 0.00001$). Critically, in **GLM4.1** (when all four regressors competed to explain variance in BOLD activity) no dmPFC voxels were sensitive to the distance to goal in terms of number of stations.

In the lateral prefrontal cortex, we observed a similar pattern of BOLD signals in an anterior premotor region (PMC) that straddled BA6 and BA8, where BOLD activity scaled with DL (left peak: -26, -8, 54; $t_{19} = 6.58$, $p < 0.000001$; right peak: 30, 4, 66; t_{19}

= 4.99, $p < 0.0001$) and DU (left peak: -26, 4, 54; $t_{19} = 6.51$, $p < 0.000001$; right peak: 26, 8, 46; $t_{19} = 6.30$, $p < 0.000001$). Here, we also observed an effect of distance in number of stations, DS (left peak: -22, -8, 50; $t_{19} = 6.62$, $p < 0.000001$; right peak: 30, 4, 58; $t_{19} = 6.39$, $p < 0.000001$). Notably, the number of exchange stations between the current position and the goal (DX) failed to show any consistent effect at the group level. In other words, these regions encoded the cost of representing a plan in units that reflected the structure of the subway map, over and above any encoding of the distance to goal.

Previous neuroimaging studies have noted that BOLD signals in the rostrolateral prefrontal cortex (rlPFC) scale with the number of moves that are required to solve the Tower of London task (van den Heuvel et al., 2003; Wagner et al., 2006), equivalent to our DS measure. To permit direct comparison with past studies, we created a new GLM (**GLM4.2**) that included only DS (alongside other nuisance quantities; see Methods), omitting the distance regressors in units of lines, exchange stations or the U-turn cost. Consistent with previous work, this analysis identified not only the premotor cortex (PMC) but also a portion of bilateral rlPFC (left: -42, 32, 34; $t_{19} = 7.87$, $p < 0.000001$; right: 42, 40, 34; $t_{19} = 4.81$, $p < 0.0001$; see **Fig. 4.2c**). Plotting the average beta parameters across the cohort for DS and DL confirmed that the PMC, but not the rlPFC, encoded the cost of a hierarchical plan, as demonstrated by a region (PMC, rlPFC) x distance (DS, DL) interaction ($F1,19 = 4.71$, $p < 0.05$; see **Fig. 4.2b**).

Proximity to goal

Consistent with previous findings (Howard et al., 2014), using **GLM4.1** we also observed a signal that reflected a negative correlation with distance in stations to goal

(DS) in the ventromedial prefrontal cortex (vmPFC, peak: 10, 48, -6; $t_{19} = 5.80$, $p < 0.00001$; in other words, this region became more active the closer to the goal. In this region, distance was encoded in units of stations only, with no evidence for encoding of hierarchical distance (**Fig. 4.2d**). Including only DS (**GLM4.2**) identified a number of other regions, including the hippocampus, where BOLD signals have previously been found to scale with distance to goal during navigation (Howard et al., 2014). In our task, the hippocampus reflected distance to goal bilaterally in the same direction as the ventromedial prefrontal cortex (**Fig. 4.2e**). A full range of regions that correlated with each of these distance estimates is reported in **Supplementary Tables ST4.1 and ST4.2**.

Correlation of neural and behavioural costs across the cohort

Next, we aimed to understand the relationship between the neural and behavioural effects so far observed (see **Fig. 4.2f**). For each measure of planning cost (DS, DL, DX and DU) we calculated the correlation across the cohort of participants between its influence on RT (regression coefficient from **Fig. 4.2a**) and its influence on BOLD signals in (i) the PMC and (ii) the dmPFC. We found the correlation was significant in dmPFC for both distance in number of stations (DS: $R = 0.6$, $p < 0.005$) and in number of line changes (DL: $R = 0.39$, $p < 0.05$). However, neither of these correlations was significant in the PMC (DS: $R = -0.05$, $p = 0.57$; DL: $R = 0.07$, $p = 0.379$). No brain-behaviour correlations were observed in either region for DX or DU. However, we did observe a correlation between the behavioural cost of DU and the encoding of DU in a dlPFC region shown in Fig. **SF4.4** (DU: $R = 0.33$, $p < 0.05$ one-tailed).

Neural signals associated with bottleneck states

The analyses described above suggest that both dmPFC and PMC encoded the hierarchical cost of representing a plan, over and above any cost of plan representation computed in units of discrete states. Next, we investigated neural signals in these regions more closely, by plotting the activity that accompanied the moment in which a "bottleneck" state occurred - when participants were offered the opportunity to switch from one context to another. We once again capitalised on the factorial design of our task, asking if there were unique neural signals that varied with station type (exchange > regular, now including all trials; **Fig. 4.3b**). This analysis also included a regressor encoding DS, as well as a further nuisance predictor that signalled whether the action chosen was optimal or not (**GLM4.2**).

We observed increases in BOLD signals associated with exchange stations in both the dmPFC (peak: 6, 16, 46; $t_{19} = 4.09$, $p < 0.001$) and PMC, overlapping with the region described above (left peak: -26, 8, 54, $t_{19} = 7.24$, $p < 0.000001$; right peak: 26, 12, 54, $t_{19} = 6.56$, $p < 0.00001$). Across the subject cohort, the strength of this latter neural effect predicted the RT difference between exchange and regular stations ($r = 0.40$, $p < 0.04$) but not between switch and stay trials ($p = 0.70$). A further effect of exchange > regular stations was observed in a more anterior prefrontal region, in bilateral BA 46 (left peak: -42, 24, 30, $t_{19} = 4.48$, $p < 0.0001$; right peak: 46, 32, 22, $t_{19} = 5.38$, $p < 0.0001$).

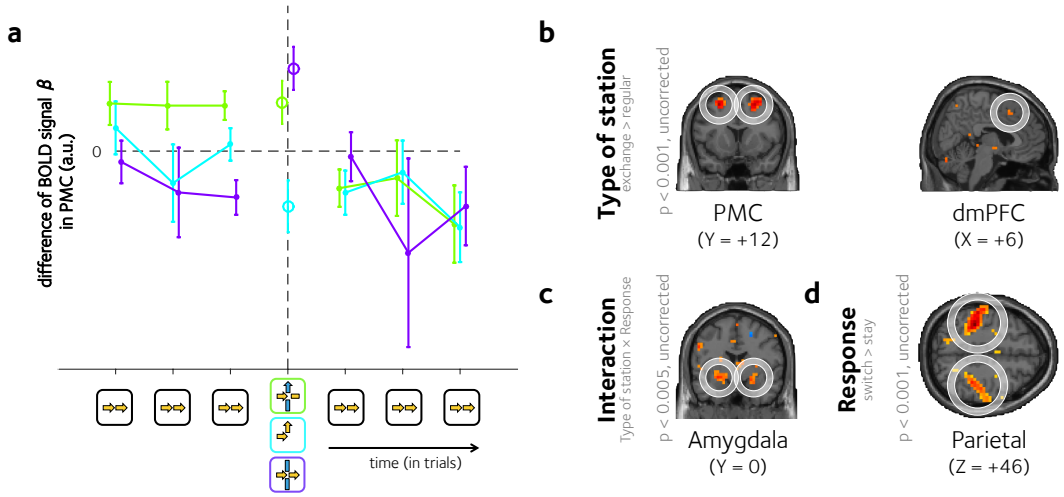

Figure 4.3. Trial by trial analyses per condition
A. BOLD signal β values (mean ± SEM) from single-trial GLM approach in the PMC on 3 regular stations preceding (leftmost points) and following (rightmost points) a context switch (green lines), an exchange station without line change (purple lines) or an elbow station (cyan lines). Activation at the context switch, exchange station or elbow are shown with a single point in the corresponding colour. The averaged BOLD signal β in regular stations is represented by the horizontal dashed line. **B.** Voxels responding to the main effect of station type (exchange > regular) in the PMC (left panel) and dmPFC (right panel). **C.** Voxels in the amygdala responding to the interaction between station type and response. **D.** Voxels in the parietal cortex responding to the main effect of response switch. Coordinates in MNI space are provided under each slice. Significant regions within a circle survived multiple comparisons correction.

Next, we plotted how the BOLD signal varied on those regular stations that both preceded and followed an exchange or an elbow station. A brain region encoding the hierarchical representation of a plan might be expected to show tonically higher BOLD signals in the trials preceding an exchange station (where the cost of plan representation in units of lines remains high) followed by a reduction immediately after context switch (where the computational burden is reduced). In **Fig. 4.3a**, we plot the BOLD signal in the PMC region (extracted from the main effect of type of station) on regular stations that precede and succeed a context switch (green lines). An elevated BOLD signal is visible on those trials preceding a context switch, after which it drops off sharply (comparison between preceding and succeeding: $t_{19} = 3.24$, $p < 0.003$). Of note, a similar drop is not observed when the same analysis is

conducted on stations that precede or succeed an exchange station without a context switch (purple lines; $p > 0.9$) and only a modest drop follows an elbow station ($t_{19} = 1.87$, $p < 0.05$, one tailed). These effects were qualified by the interaction of type of station and type of response on the difference of signal (preceding, following) around each condition: $F1,19= 5.44$, $p < 0.04$. In other words, the average BOLD signal in PMC observed was higher on trials before than after a context switch, consistent with a hierarchical representation of the plan. We additionally found a main effect of type of response: $F1,19 = 4.61$, $p < 0.05$, indicating that participants also anticipated making a response switch. Signals from the dmPFC followed a similar pattern, although the interaction failed to reach significance. An equivalent analysis for RTs is shown in the supplementary materials (**Fig. SF4.3**).

Neural signals accompanying response switch and context switch

Behavioural data indicated that there was a unique cost incurred when participants switched context, i.e. at exchange stations requiring a response switch. In the fMRI data, we observed a comparable interaction between type of station and response switch in a cluster of voxels straddling the amygdala and putamen (left peak: -26,0,-10, $t_{19} = 4.46$, $p < 0.001$; right peak: 22,4,-14, $t_{19} = 5.20$, $p < 0.0001$), as well as an extrastriate region on the lingual gyrus (peak: 26,-68,-6, $t_{19} = 5.16$, $p < 0.0001$), corresponding to area V4 where responses to colour are often observed (Zeki and Marini, 1998). Plotting parameter estimates for these regions showed that this interaction was driven by higher BOLD signals for those trials where participants switched from one context to another (**Fig. 4.3c**). However, we interpret these results with caution, because they failed to reach the threshold required for correction using an FDR threshold. Finally, we also observed strong activations in the parietal cortex

that predicted whether participants switched direction or not (left peak: -38,-32, 46, t_{19} = 10.8, $p < 0.000000001$; right peak: 54, -24, 34, $t_{19} = 8.39$, $p < 0.0000001$; **Fig. 4.3d**).

Encoding of current context

To execute a hierarchical plan, an agent must be able to identify and represent the current context, in addition to the current state (i.e. on the London Underground, to know that one is on the Victoria line, not just that one is at Green Park station). We thus used a multivariate analysis technique known as representational similarity analysis (RSA) to identify brain regions in which the patterns of BOLD signal over voxels was more similar across runs within a single subway line than between two different lines (using unsmoothed data; see Materials and Methods for details; **Fig. 4.4a**). In the scanner, no indication was given as to the subway line currently being visited, and so any significant voxels must reflect an abstract encoding of the context from memory. In conjunction with a whole-brain "searchlight" approach, this analysis once again identified the dmPFC as a region where the current context was represented (peak -10, 8, 54; $t_{19} = 7.49$, $p < 0.000001$). No evidence for context encoding in the PMC was found, although evidence was found in other regions, including more anterior portions of the prefrontal cortex in BA9 (left peak: -30, 44, 34; $t_{19} = 5.32$, $p < 0.0001$; right peak: 34, 44, 30; $t_{19} = 4.9$, $p < 0.001$).

The analyses above indicated that the dmPFC encodes distance to goal in units of lines and U-turns. It could be, thus, that the pattern encoding of this quantity may depend on the current line, providing evidence for a distinct computational cost within each context. We thus repeated our RSA, but using not the raw BOLD signal observed at each station, but the parametric encoding of distance to goal (in stations). The pattern of encoding of distance to goal was also more similar within lines than it

was between lines in the dmPFC (2, 20, 54; $t_{19} = 5.38$, $p < 0.0001$); it is shown in **Fig. 4.4b**.

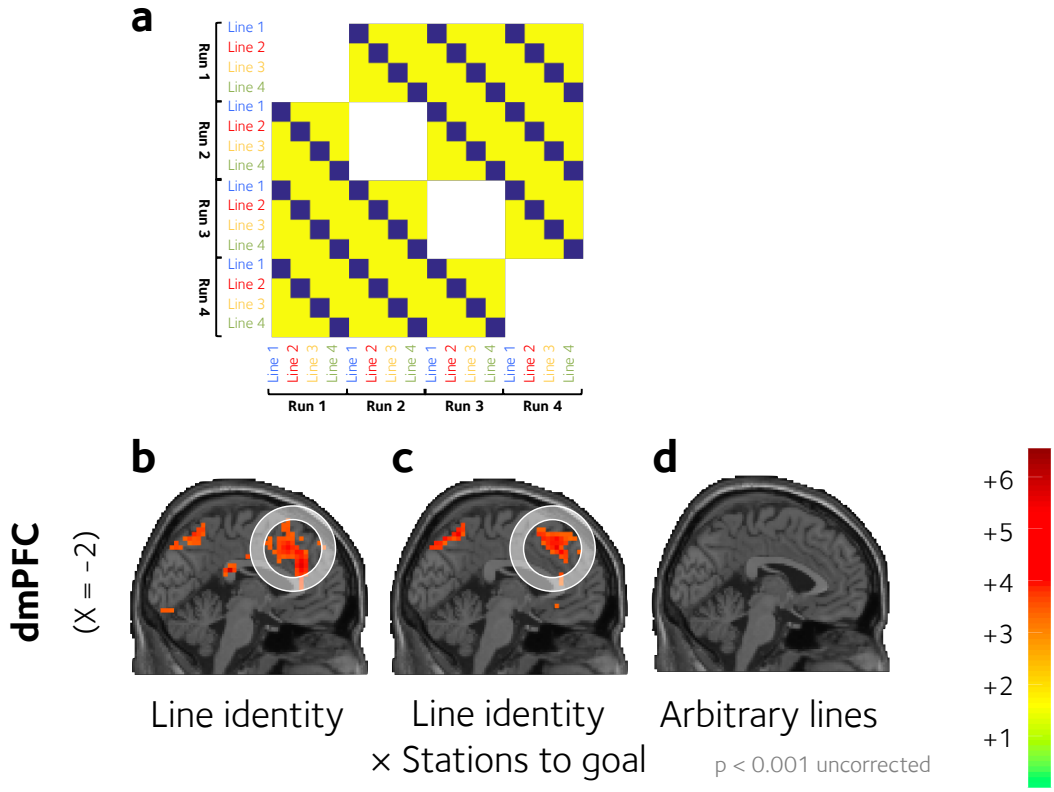

Figure 4.4. Multivariate analyses of line identity
A. A depiction of the predicted representational dissimilarity matrix that was used to identify brain regions where the similarity structure was greater within than between contexts. Blue (and yellow) squares represent low (high) dissimilarity, respectively for independent pairs of (scanner runs, lines; x and y-axis). **B.** The results of the RSA identifying voxels encoding context, i.e. where multivoxel pattern dissimilarity was greater within than between contexts (lines), identified using a searchlight approach. **C.** Voxels where the parametric of encoding distance to goal (in units of station) was more different between than within contexts (lines). **D.** The results of the control analysis for (B) involving shuffled stations-line assignments. An additional control analysis was performed to assert that the effect was not driven by line orientation (see Fig. S2). Significant regions within a circle survived multiple comparisons correction.

RSA can yield spurious results when trials assigned to each category are not fully temporally decorrelated, and so we conducted this analysis between runs (e.g. measured the similarity between line a on run1, and line b on run 2). We additionally conducted a control analysis in which the assignments between stations and lines were shuffled; this yielded no significant results (**Fig. 4.4c**).

Finally, subway lines contained long straight sections, and so we were concerned that RSA of context might have captured similarity associated with travel in a common direction, unrelated to context per se. To test this, we conducted another RSA using the same approach, but searched for voxels where multivoxel patterns were more similar within than between directions (north, south, east, west). No activations were observed in the medial prefrontal cortex, but a large cluster of significant voxels was found in the left motor cortex (**Fig. SF4.2**).

Discussion

The behaviour of humans and other animals is controlled at least in part by a "model-based" control system that learns the structure of the world, and organises sequential behaviour in pursuit of future goals (Daw et al., 2005; Dickinson & Balleine, 2002; Dolan & Dayan, 2013; Schoenbaum, Roesch, Stalnaker, & Takahashi, 2009; Tolman, 1948). Recent work has begun to address the neural and computational substrates underlying the model-based decision-making, by constructing "two-step" decision tasks in which cached state-action values and explicit forward search strategies make opposing predictions about behaviour and brain activity (Daw et al., 2011; Gläscher, Daw, Dayan, & O'Doherty, 2010; Wunderlich, Smittenaar, & Dolan, 2012). However, these studies sidestep one key theoretical challenge associated with model-

based approaches, namely how to organise behaviour over multiple future states without incurring a prohibitive computational cost. Human cognition has evolved to meet this challenge, as exemplified by our ability to form and follow plans over multiple timescales, for example when finding an efficient route to run a series of errands, or envisaging a future career path and taking steps towards its fulfilment. Although we have known for decades that planning involves the prefrontal cortex, to date, very little has been revealed about the computational mechanisms that unfold in these regions during plan formation and execution.

Here, we drew upon a framework that has its roots in cognitive psychology (Miller et al., 1960; Norman & Shallice, 1986) but has most recently inspired advances in machine intelligence (M. M. Botvinick et al., 2009; Ponsen, Taylor, & Tuyls, 2009). This framework proposes that the space of possible states can be organised and represented hierarchically as a series of clusters or "contexts", reducing plan complexity (description length), and affording substantive increases in computational efficiency both at the time of plan formation and plan execution. In the current study, we tested a prediction arising from this hypothesis: that when planning in a complex environment, the cost of representing a plan will be expressed in units of context (or context switch) over and above any cost that is incurred in units of states themselves. Our key finding is that both response times and neural activity in the caudal frontal cortex encode the cost of representing a hierarchical plan, indicating that they participate in the hierarchical organisation of future behaviour.

The neural costs observed were identified in two frontal regions: a dorsomedial PFC region, falling in the pre-supplementary motor cortex, that is often found to be

sensitive to the difficulty (or conflict) incurred when making a choice (M. Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999), and a lateral frontal that straddles the border between the premotor and prefrontal cortices, in BA6/BA8. Both regions were also active when participants were faced with the opportunity to switch context, at an "exchange" station or bottleneck - consistent with the finding that the dmPFC responds to subgoal attainment (Ribas-Fernandes et al., 2011). However, across the participant cohort, we observed reliable brain-behaviour correlations in only the dmPFC, but not the PMC. In the dmPFC, the strength with which BOLD signals encoded distance to goal in units both of stations and contexts for a given subject predicted his or her corresponding RT cost for those plan complexity measures. We also found that the multivariate pattern of information in the dmPFC (but not PMC) was sufficient to distinguish among contexts, even though the line that was currently visited was never explicitly displayed to participants during the scanning phase. Moreover, we were also able to distinguish context-specific representations of distance to goal in the dmPFC, as if the region encoded separate costs of planning for each individual context. One interpretation of this finding is that the dmPFC is responsible for translating of a plan into behaviour, whereas the PMC participates in maintaining the plan active over the journey. However, we note that those participants showing the strongest flat cost in behaviour also showed stronger encoding of this cost in dmPFC neural signals. It may be, thus, that there are some individual differences in the way that dmPFC contributes to computing the cost of planning.

More generally, our findings are consistent with the view that the dmPFC encodes a motivational signal that is extended over time (Summerfield & Koechlin, 2009), and the complementary perspective that the dmPFC encodes "option" values under the

framework of hierarchical reinforcement learning (Holroyd & Yeung, 2012). As part of a general role in monitoring the expected value of controlled behaviour (Shenhav, Botvinick, & Cohen, 2013), the dmPFC may thus encode both the identity and value of a contextual variable over which a particular policy applies, for example when foraging from different patches (Hayden, Pearson, & Platt, 2011; Kolling, Behrens, Mars, & Rushworth, 2012).

The lateral region overlaps with the superior aspect of the caudal dorsolateral prefrontal cortex identified by Koechlin and colleagues as active when actions are selected on the basis of contextual information (Koechlin, Ody, & Kouneiher, 2003). The same region is labelled ‘pre-PMd’ by Badre and colleagues, who found that this region is active when action selection is contingent on a hierarchy of contingencies, rather than a 'flat' series of sensorimotor associations (Badre et al., 2010). In this region (as in behaviour and the dmPFC signal), the BOLD signal scaled with distance to the destination station in units of context (i.e. lines). Notably, no such effect was observed in more rostral regions that have previously been implicated in representing plan complexity in multi-step problems such as the Tower of London task (van den Heuvel et al., 2003; Wagner et al., 2006). At first glance this finding is surprising - one might have expected more anterior regions to be responsible for representing the higher hierarchical aspects of a complex plan. However, one explanation for this finding is that during hierarchical planning, potentially complex action sequences are "compressed" to a small number of steps (e.g. contexts and context switches) that can then be represented in subsidiary prefrontal regions located more caudally (Koechlin & Summerfield, 2007).

Interestingly, the cost of representing a plan was incurred in units of context. This explains the previous finding that humans seek to reach a new context earlier rather than later during navigation, as doing so reduces the computational burden of plan representations (Wiener & Mallot, 2003). This result additionally suggests that the hierarchical representation of the plan is encoded in terms of its abstract structure, rather than as a succession of macro-actions (e.g. “go straight, then go left”). Nor was the plan encoded in terms of the number of choice points, suggesting that the state space is not chunked purely on the basis of its physical properties (e.g. in terms of segments between choice points) but in a fashion that reflected the more abstract structure that they were encouraged to learn during training. What remains unclear, however, is whether context is represented as a cluster of interlinked perceptual states (i.e. stations on the yellow line), or as a series of macro-policies that dictate pursuit of a goal (e.g. keep going straight on until you reach a given switch point). A hint that participants relied on perceptual representation of context was provided by the finding that voxels in area V4 became active at context switches, as if participants were recalling the colour of the new subway line (which was not shown to them during scanning). However, the precise nature of the information that characterises a context remains an open question. For example, participants might have used information about the spatial organisation of the map (the blue line runs from north to south, or the red line is north of the green line).

Moreover, both behaviour and the premotor cortex (PMC) also encoded an additional "U-turn" cost, that indexed the extent to which plans involved "doubling back" towards the current location along a different line. In the planning literature, it has been noted that goal-subgoal conflict - for example, the need to temporarily remove

one disc from a peg and subsequently replace it in the Tower of London task - incurs a unique response time cost (Ward & Allport, 1997) and poses a particular problem for patients with lateral prefrontal lesions (Morris, Miotto, Feigenbaum, Bullock, & Polkey, 1997). Consistent with this finding, U-turn costs were visible not only in the PMC, but also in lateral prefrontal regions. The existence of a unique U-turn cost in our navigation task demonstrates that participants not only encoded plans in the subway network as a hierarchical series of contexts, but also in terms of the geometry of the map that they saw in the training session.

Although the costs of representing a 'flat' plan were minimal once variance associated with a hierarchical plan had been partialled out, there was one brain region where strong (positive) covariation with number of stations to goal was observed - the ventromedial prefrontal cortex. Previous theories have speculated that the vmPFC may be among a set of regions that tracks distance to a goal state (Holroyd & Yeung, 2012), and indeed, the vmPFC is implicated in episodic future thinking (Schacter & Addis, 2007), and has been found to track growing expected reward in decision tasks involving sequential, interdependent choices (Tsetsos, Wyart, Shorkey, & Summerfield, 2014). The hippocampus has also previously been found to covary with proximity to goal, but only in virtual reality environments that mimic much more closely the naturalistic experience of navigation (Howard et al., 2014; Viard, Doeller, Hartley, Bird, & Burgess, 2011). Here, we show that the distance to goal representation is present even when current and goal state information is devoid of the rich episodic cues that we normally use to navigate. Critically, however, the hippocampus and vmPFC showed no evidence of a hierarchical signal.

Our analyses focussed on the cost of "representing" a hierarchical (or flat plan) as participants navigated through the network. This is a general index of the cost involved in maintaining and monitoring the plan, rather than of recursively searching through all possible nodes of the decision tree (for example, via a breadth- or depth-first algorithm), or "pruning" of unpromising routes to a goal (Huys et al., 2012; Huys, Lally, et al., 2015). Whilst plan formation may have occurred mainly on presentation of the cue screen stating the start and goal stations, plans may also have been constantly updated and reformed during execution ('replanning'). Indeed, as distance to goal grows, the processing cost of these search operations will grow correspondingly. However, it is not clear that this cost would grow linearly with the number of states or contexts that must be traversed to reach a goal. One limitation of the approach taken here is that we do not have an obvious means to assess how plans are formed prior to or during navigation, or to distinguish the neural mechanisms that accompany plan maintenance and monitoring from any replanning that may be occurring. We did examine BOLD signals evoked in response to the cue screen, but they did not show convincing correlations with the various distance metrics, or predict the journeys that participants would follow. However, it is unclear whether this null finding is due a lack of statistical power, owing to the limited number of such trials. Examining the costs incurred at the time of plan formation would be an interesting avenue of research for future studies.

Chapter 5. Neural mechanisms underlying rule discovery in human decision-making

Abstract

Humans can build a structured representation of the world by learning high-level rules that explain a set of low-level observations. This allows them to generalise beyond flat stimulus-action mappings. Here, we investigated the neural correlates of learning such rules. Human participants performed a task that could be solved by learning from stimuli that verified a rule (target) or not. We relied on a Bayesian model to estimate biases in the amount of learning for different feedback (target/nontarget, correct/incorrect), and to track the subjective uncertainty about the underlying rule. Behavioural data indicated that human inference was suboptimal: they learned more rapidly in trials when they accurately verified the rule (responded target correctly) despite nontarget stimuli was more informative for discovering the rule. Functional neuroimaging revealed two distinct networks reflecting the uncertainty about the response and uncertainty about the rule. The former was reflected in motor regions as well as bilateral ventrolateral prefrontal cortex. The later included dorsolateral and dorsomedial prefrontal cortex, as well as parietal cortex. BOLD signal in these regions carried reward prediction error signals that reflected the feedback differentially as a function of the model-based estimate uncertainty of the rule. Overall, this experiment highlights a network involved in inference and rule discovery for generalisation.

Introduction

As we have seen in the previous chapters, humans can leverage the structure of the world in order to guide behaviour. In **Chapter 2** (**Experiment 2.2**) I found that participants learned a hierarchical structure that allowed them to learn faster in a subsequent session, and we found univariate correlates of the hierarchical level. In **Chapter 4**, I found behavioural (reaction times) and neural costs of plan complexity, following a compressed hierarchical representation. I also spent some time during the *General introduction* in **Chapter 1** describing the key role of categorisation in building structured representations. As Ashby and Maddox pointed out (Ashby & Maddox, 2005), categories often can be represented explicitly and verbally via 'rules' which I have not explored so far, as opposed to categories defined by superficial similarity (e.g. shared visual features, or temporal adjacency).

Consider the Wisconsin Card Sorting Test (WCST) (Heaton, 1981). In the WCST, participants (or patients) are instructed to learn to match via trial-and-error a sample card to another of the same category. All cards include geometrical patterns varying in shape, colour and number. For each rule, the category depends on a single dimension (e.g. two cards are matched if they include red patterns). Clinical research has dissociated two groups of patients affected by this type of cognitive ability. First, patients with damage to the frontal lobe display persevering errors due to an impaired flexibility to switch focus of attention (Kimberg, D'Esposito, & Farah, 1997). These results are complemented by imaging evidence implicating the frontal regions as well as the basal ganglia (caudate nucleus) during the WCST (Konishi et al., 1999; Liu, Braunlich, Wehe, & Seger, 2015; Lombardi et al., 1999; Rao et al., 1997; R. D.

Rogers, Andrews, Grasby, Brooks, & Robbins, 2000; Volz et al., 1997). Second, Parkinson's disease patients also have deficits in rule-based category learning (Ashby, Noble, Filoteo, Waldron, & Ell, 2003; Brown & Marsden, 1988; Cools, van den Bercken, Horstink, Van Spaendonck, & Berger, 1984; Downes et al., 1989). Their condition affects the dopaminergic system thought to compute the learning prediction error signals that subserve reinforcement learning (Bayer & Glimcher, 2005; Schultz et al., 1997).

It is thought that rule-based category learning requires a mechanism that explicitly generates, tests, and reasons over theories or hypotheses (Ashby & Maddox, 2011). Human reasoning is considered to be systematically flawed or limited: in the generation of hypotheses (Dasgupta, Schulz, & Gershman, 2017); in seeking information that tests a hypothetical rule (Klayman & Ha, 1987; Wason, 1968); and in integrating new evidence (Castañón et al., 2018). Particularly, humans are biased towards 'positive' inferences and confirmation biases such that they overvalue information that verifies a given rule (Nickerson, 1998).

This can be partially explained by simple heuristic approaches that approximate optimal behaviour in ecologically valid circumstances. For example, humans have a predisposition for a uni-dimensional "feature-based" strategy that promotes generalisation, such that all features are learned independently of each other before they are combined (Farashahi, Rowe, Aslami, Lee, & Soltani, 2017). However, when necessary, participants also shift to a more flexible multi-dimensional representation that considers the combinatorial relationship between the features ("object-based" learning) (Farashahi et al., 2017; G. L. Murphy et al., 2012) or can associate visually

different observations to the same category (through recruitment of the medial temporal lobe) (Davis, Love, & Preston, 2011; Love, Medin, & Gureckis, 2004; Mack, Love, & Preston, 2016, 2017). In line with this, Wutz et al. (Wutz, Loonis, Roy, Donoghue, & Miller, 2018) found that the ventro- and rostrolateral prefrontal cortices in the monkey brain were engaged distinctively during visual category learning as a function of how "abstract" (i.e. variable) was the visual category, possibly underlying a behavioural change of strategy.

Whilst cognitive mechanisms that mediate rule discovery, and in particular its overreliance on verification have been extensively investigated, much less is known about its neural implementation at the systems level. Here, we designed an experiment where we asked participants to discover the rule underlying a target category. On each trial, a stimulus composed of two coloured shapes was categorised as target (i.e. verified the rule) or nontarget (i.e. falsified the rule). We made use of a non-exclusive disjunctive rule that was spatially segregated (i.e. target if 'X' on the left *or* 'Y' on the right, or both) because it provided multiple advantages: i) it discouraged feature-based heuristic strategies, ii) it constrained the space of hypotheses, and iii) it presented a stronger challenge for generalisation. Incidentally, recent work has shown that human infants of 12 and 19-months old can reason logically over such disjunctive rules, suggesting that rule-based logical reasoning may be a core aspect of cognition acquired early during infancy (Cesana-Arlotti et al., 2018).

In our task, we asked humans to perform a rule-discovery task whilst acquiring functional magnetic resonance imaging (fMRI), allowing us to measure behavioural

and neural concomitants of rule-discovery. We instructed participants to discover by trial and error the rule that determined the correct responses within a limited set of trials. They were made aware of the general form that rules could take in order to constrain the space of possible hypotheses.

We additionally made use of visual stimuli composed of two dimensions (shape and colour) such that at any time only one would be relevant on each side for the category rule (e.g. colour on the left and shape on the right stimuli). Rules were thus additionally organised in four different classes depending on the relevant dimensions. Participants needed to keep tracking at any given time of the relevant dimensions (shape or colour of the visual stimuli); the specific rule; and thus to infer whether the current stimulus belonged to the target category or not. Furthermore, in half of the blocks we provided prior information about the relevant dimensions in order to explore how the cognitive load modulated the behavioural and neural biases underlying rule discovery. This increased cognitive load was driven by all visual dimensions being relevant a priori, and by the space of hypothetical rules being larger.

In advance of the results, we fitted human behaviour with an extended Bayesian model that could adapt the learning rates selectively as a function of the choice feedback (e.g. correct/incorrect) and the true category of the stimuli (target/nontarget). This model allowed us to interrogate the learning biases during rule-discovery. We found that participants processed the feedback better when they accurately verified a rule (i.e. responded target correctly) despite nontarget stimuli was more informative from a normative point of view. The estimated learning rates were selectively impaired for high cognitive load (no prior information about the relevant dimensions)

when participants responded target incorrectly. We exploited information-theoretic measures in order to estimate the uncertainty of the response and of the underlying rule. We found that the bilateral ventrolateral prefrontal cortex (vlPFC), which has been associated with category learning of prototypical categories, reflected the uncertainty of the response. A wider network including dorsomedial (dmPFC), bilateral dorsolateral prefrontal (dlPFC), as well as bilateral parietal cortices encoded prediction error signals of the rule.

Methods

Participants

18 healthy participants (6 female, 12 male; age 20-34, mean 25.0 years) were recruited into the study in accordance with ethical guidelines approved by the Oxford University Medical Sciences Ethics Board. No participants reported a history of psychiatric or neurological illness, and all had normal or corrected-to-normal vision. They were paid £35 for participation in both a practice and a scanner session on two separate days.

Stimuli and task

Participants performed a rule-learning task that required pairs of shapes to be classified as 'target' or 'nontarget' according to an unknown rule. Each block began with the presentation of one of 8 abstract symbolic cues (Greek letters) for 3s. After a period of 2-5s (jittered), a train of 16 pairs of stimuli appeared on the left or right of the screen at approximately 3° eccentricity. Each stimulus pair remained on the screen for 3s. Each member of the pair could be a square, circle or triangle coloured red, green or blue. Participants pressed a key (training task) or response button (scanner

task) at any point during the 3s presentation period to indicate whether the stimulus was a target or nontarget. Stimulus-response contingencies were fully counterbalanced across participants. Responses were followed by fully informative auditory feedback consisting of a pair of tones with ascending (correct) or descending (incorrect) pitch (400Hz; 800Hz) that lasted 200ms in total. An interval of 2-6s was interposed between stimuli, during which the screen was blank. Participants completed a total of 48 blocks during training, and then a further 48 blocks for the scanner session, which occurred on a subsequent day. In the scanner, the experiment was divided into 4 runs of 12 blocks each, buttressed by lead-in and lead-out durations of 10s. Each block lasted ~17 minutes, bringing total scanning time to just over an hour.

Issues during data collection

Due to an unfortunate bug in the function shuffle.m from the EEGLAB package, a significant trend was induced in the randomisation of the trials such that the probability of the stimulus being target was significantly above 50% in early trials (maximum 55.4 ± 1.4% in trial 5) and significantly lower than 50% for late trials (minimum 40.2 ± 1.2 % in trial 16). This artifact didn't change the overall proportion of trials (8 target and 8 nontarget stimuli) per block, and affected equally familiar and novel trials without significant differences between these. All results reported here carefully controlled the effect of trial number, and thus are not affected by this trend in the data.

Rules

Whether each stimulus (pair of shapes) was a target or not was determined by a disjunctive ('or') rule over its features. For example, in one block the rule might be "if

the shape on the left is a triangle, or the shape on the right is blue, the stimulus pair is a target". We denote this {left: triangle, right: blue}. Note that rules are not commutative, because the rules were sensitive to position where the shapes appeared on screen. We divided rules hierarchically into 4 classes according to the feature that was relevant on each side: {left: colour, right: colour}, {left: colour, right: shape}, {left: shape, right: colour} and {left: shape, right: shape}. The same feature was never selected in both cases (for example {left: red, right: red}), leaving 6 possible rules in each class (24 rules total). Each rule was thus repeated twice during practice and twice in the main experiment, leading to 48 blocks. Pairs of shapes were selected pseudorandomly in each block so that 8 trials were targets and 8 were nontargets, but no combinations of coloured shapes were repeated.

Cues

Humans may prefer to learn by verifying hypotheses because learning from falsification is cognitively demanding. In order to understand how human reasoning changes as the information burden is alleviated by symbolic information, we thus provided either *familiar cues* that offered partial information about the rule that was valid in that block (fewer rules possible), or *novel cues* where no such information was available (more rules possible). Four symbolic cues were randomly assigned to the four hierarchical rule classes for each participant. For example, the symbol ξ could be associated with rules where shape and colour where the relevant dimension on the left and right sides respectively. Half of the total 48 blocks were designated 'novel cues' blocks, and the remainder were 'familiar cues' blocks. In familiar blocks, the symbolic cue faithfully indicated the rule class that was relevant (but not the precise rule). Participants were fully briefed as to the meaning of the cues at the beginning of the practice session, and the cues remained unchanged during the later

scanner session. For the novel cues blocks, one of the remaining cues was chosen pseudorandomly at the start of the block (irrespective of the rule class) with the only constraint that each cue was selected an equal number of times for each rule class (thus being uninformative).

Statistical analyses of behaviour

We analysed accuracy, choice (target vs. nontarget), and learning rates with ANOVAs testing the influence of cue condition and trial number, using an alpha of $p < 0.05$.

Computational modelling

We first constructed a Bayesian model of our task (*HB model* described in detail in the **Supplementary Text SX5.1**) that inferred flawlessly the underlying rule for our task. We then made use of this model in order to describe how human inference deviates from optimality (see **Fig 5.2**). The purpose of this exercise was not to prove or disprove that humans "are Bayesian", or to arbitrate among competing models, but to make use of a normative account in order to systematically describe the biases from optimality that human display during learning as a function of the cognitive load (block cues). We estimated in a principled manner the amount of learning carried by target or nontarget trials, as well as correct and incorrect trials. Such a bias would quantify the extent of 'positive' strategies adopted by our participants. We thus generalized the HB model by drawing inspiration from the framework of reinforcement learning (RL; Sutton and Barto, 1998). RL offers a well-precedented, biologically plausible mechanism for understanding how decision values are updated, and a wealth of past imaging work allows us to make predictions about where those updates should be encoded in the BOLD signal (Rushworth et al., 2009). In standard RL models, data are usually fit with a single parameter α that controls the rate at

which beliefs change on the basis of new information. In contrast, in the HB model the posterior probability associated with each rule is set from 1 to 0 when it is violated by the current trial. We relaxed this assumption by borrowing the delta rule from RL, such that the probability of invalid rules would decrease progressively as follows:

$$\begin{aligned} P(r|\mathbf{x_n}, \mathbf{t_n}) &= P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) + \alpha\left(0 - P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}})\right) \\ &= (1 - \alpha) \times P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) \end{aligned} \quad \text{(eq. 5.1)}$$

where $P(r|\mathbf{x_n}, \mathbf{t_n})$ is the posterior probability of the rule r being the 'true' one conditioned on the history of trials seen so far; r is an invalid rule; α is the *learning rate* parameter (set to 1 in the optimal case); and $\mathbf{x_n}$ and $\mathbf{t_n}$ are the history of stimuli and feedback respectively in the current block up to trial n.

A key question in our study is the rate of learning for the different feedback conditions. For instance, higher learning rates from correct or from target feedback would reflect a confirmation bias (i.e. a 'positive' inference strategy). We extended the HB model by employing four learning rate parameters (α_{VC}, α_{VI}, α_{FC}, and α_{FI}) that controlled the learning rate at which beliefs changed following each trial with target feedback (verification; α_{V*}), nontarget feedback (falsification; α_{F*}), correct feedback (α_{*C}), and incorrect feedback (α_{*I}) such that:

$$P(r|\mathbf{x_n}, \mathbf{t_n}) = (1 - \alpha_{TC}) \times P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) \quad \text{if target and correct} \quad \text{(eq. 5.2.1)}$$

$$P(r|\mathbf{x_n}, \mathbf{t_n}) = (1 - \alpha_{TI}) \times P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) \quad \text{if target and incorrect} \quad \text{(eq. 5.2.2)}$$

$$P(r|\mathbf{x_n}, \mathbf{t_n}) = (1 - \alpha_{NC}) \times P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) \quad \text{if nontarget and correct} \quad \text{(eq. 5.2.3)}$$

$$P(r|\mathbf{x_n}, \mathbf{t_n}) = (1 - \alpha_{N}) \times P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) \quad \text{if nontarget and incorrect} \quad \text{(eq. 5.2.4)}$$

We also relaxed the assumptions of the policy of the model. On every trial, the model made a probabilistic response with a greediness that varied from trial to trial:

$$P(y_n|\mathbf{x_n}, \mathbf{t_{n-1}}) = \text{sigmoid}(\beta_n \times \text{logit}(P(t_n|\mathbf{x_n}, \mathbf{t_{n-1}}))) \quad \text{(eq. 5.3.1)}$$

where $P(y_n|\mathbf{x_n}, \mathbf{t_{n-1}})$ is the probability a target response by the model, $P(t_n|\mathbf{x_n}, \mathbf{t_{n-1}})$ is the probability of the current trial being target as estimated by the model, and β_n is the greediness in trial n. The greediness β_n was determined as follows:

$$\beta_n = 1 + \beta_A \times e^{-\beta_\tau(n-1)} \quad \text{(eq 5.3.2)}$$

$$\text{logit}(p) = \log(p) - \log(1-p) \quad \text{(eq 5.3.3)}$$

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \quad \text{(eq 5.3.4)}$$

where β_A and β_τ are the *policy* parameters controlling respectively for the greediness in the first trial and how it decayed with the trial number. This policy fitted the data very well empirically (see **Supplementary Table ST5.2** and **Fig SF5.4**) despite it has theoretical limitations. Indeed, it assumes that participants keep track of the trial number within the block, and provides little explanation for why participants would behave more greedily in early trials. It is possible that this policy is reflecting more profound factors, as is the uncertainty about the underlying rule or the response. However, our key research interest was to make inferences over the learning rates for different feedback, and this policy empirically provided the best fit to the data.

The model has a total of 6 parameters: 4 learning rates, and two policy parameters. The HB model is nested within this more general model, and thus normative behaviour could be captured, corresponding to all learning rate parameters set to 1, and the policy parameters set to 0.

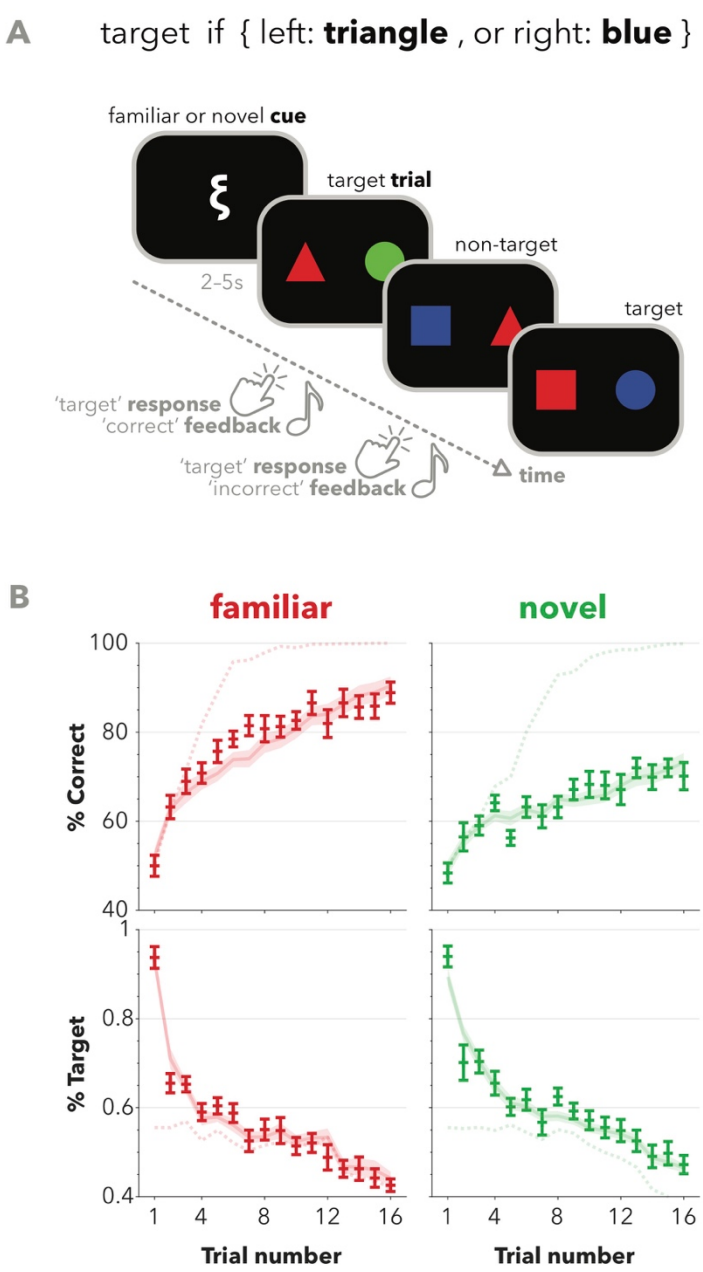

Figure 5.1. Behavioural results
A. Behavioural task. On each trial participants view a pair of coloured shapes on the left and right of the screen. They responded 'target' or 'nontarget' and were provided with positive or negative auditory feedback. Each block of 16 trials was preceded by a Greek symbol which was either fully predictive of the relevant dimensions (e.g. {colour-left, colour-right}; familiar cues condition) or not at all predictive (novel cues condition). Associations between cues and dimensions were learned in a prior training session. **B.** Behavioural data from 18 human participants (crosses with standard error of the mean SEM). Percent accuracy (top panels) and % target choice (bottom panels) over trials (1-16) are shown for familiar cues (left panels; red) and novel cues (right panels; green). Continuous lines (shaded SEM) show the same data for predictions of

the best-fitting parameterisation of the model. Dashed lines show the mean accuracy and % target of the normative HB model.

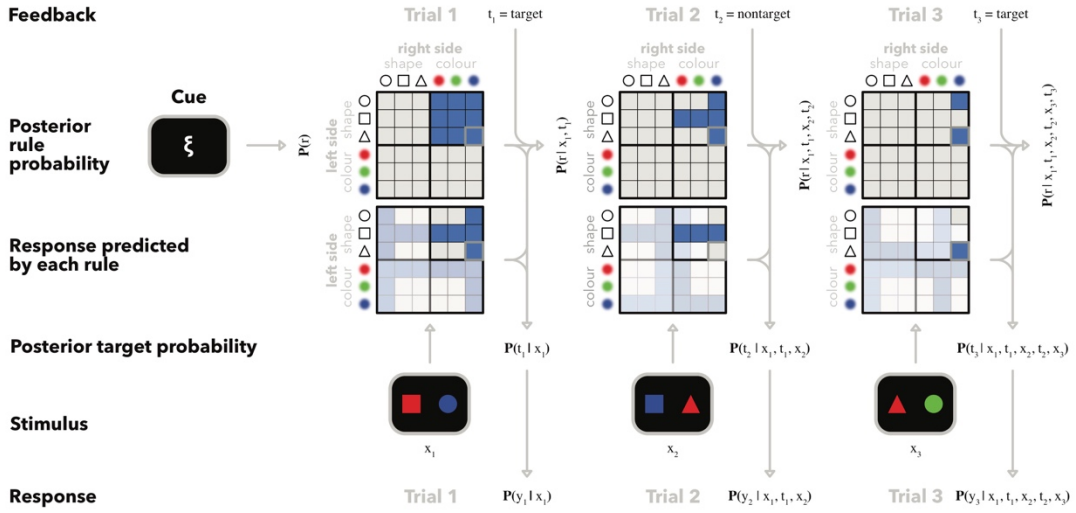

Figure 5.2. Illustration of the computational model
Diagram explaining the common workflow of the optimal (Hierarchical Bayesian; HB) model and the model fitted to human choices. In this figure, the learning rates are set to their optimal value (all 1). Cues can constrain the probability of the rules in the hypothesis space (top grids; blue: probable, white: impossible, gray square depicts the correct rule). After a stimulus is presented, a response is predicted for each possible rule (bottom grids; blue: target, white: nontarget) and the probability of the current trial being target is computed as a weighted average of the responses with the rule probabilities. This probability is passed through the *policy* of the model in order to determine a response. The target/nontarget feedback is obtained as an interaction between the response (target or nontarget) and the performance feedback (correct or incorrect). Once the feedback is observed, the probability of the rules that didn't predict the correct response are decreased through Bayes rule (for the HB model) or as a function of a learning rate (in the model of human choices), and a new trial starts.

Alternative models

In order to assess the goodness of fit of our proposed model, we compared it against two other models: a) an equivalent '*simple*' model with a single learning rate; and b) an overtly complex '*descriptive*' model with 32 parameters, each estimating the probability of responding target for each combination of familiar/novel cue and trial

number. These models yielded worse cross-validated negative log-likelihood and provided little insight unto the psychological key questions we address in this study, which is why we only made use of them to validate our model of interest. See **Supplementary Table ST5.2** for a comparison between the goodness of fit of the models.

Model fitting

Model-fitting was performed separately for familiar and novel blocks. We used gradient descent with all parameters initialised to 0, and with the learning rates constrained between 0 and 1. For more details on the fitting of the policy parameters, please refer to **Supplementary Table ST5.1**. The best-fitting model was deemed to be that which minimized the negative log-likelihood (NLL) score on choice:

$$\mathrm{NLL} = \sum_n -\log p(\varepsilon_n) \quad \text{(eq 5.4.1)}$$

$$\varepsilon_n = 1 - \|\mathrm{target}_n - P(y_n|\mathbf{x_n}, \mathbf{t_{n-1}})\| \quad \text{(eq 5.4.2)}$$

where ε_n is the probability of the model predicting incorrectly the response target_n by the participant (coded 1 for target, and 0 for nontarget responses) on a given trial n. For the reported behavioural parameter estimates and for the model-based fMRI analyses, we fitted the parameters to the full data independently for familiar and novel blocks. For the model comparison, we made use of leave-one-block-out cross-validation independently for familiar and novel blocks, which automatically down-weights model scores by their complexity (e.g. number of parameters).

Parameter recovery

Using our model fitting approach, we verified that we could successfully recover the parameters from artificial data generated under multiple underlying parameters. Specifically, the standard error across participants between the ground truth and recovered parameter estimates was smaller than 10^{-5} for all learning rates (LRs) in familiar and novel blocks, in 21 simulations with different ground truth estimates. We note in passing two consistent biases of our fitting procedure. First, the LR estimates retrieved were systematically lower than the ground truth because the fitting was initialised with all LRs set to zero and the optimizer stopped once the error criterion was reached. Second, the two *policy* parameters retrieved were afar from their real value (for familiar and novel cues) with average standard errors of 8.46 and 29.4 for β_A and β_T respectively (ground truth values between 0 and 100). Our interest lied nonetheless in the learning rate parameters, which we focused in most of our analyses, and thus the failure to recover the correct policy parameters didn't interfere with our analyses and results. In each parameter recovery simulation, we generated random parameters (shared across all participants' data) under a uniform distribution between 0 and 1 for the learning rates and between 0 and 100 for the policy parameters. We fitted independently for each participant and for familiar and novel blocks, and made use of the participants' history of trials to obtain the models' probabilistic response. The learning rates reflected the correct/incorrect feedback preserved from the participants' history of responses.

Model performance

We performed exhaustive simulations through two-dimensional grids of the learning parameters with uniformly distributed values between 0 and 1 with a step of 0.05. In each of these grids, the remaining learning rate and policy parameters were set their

optimal values. The input to the model were *not* the sequence of trials that the participants experienced, in part so that the correct/incorrect feedback specific to the participants' strategy didn't bias these results, but mostly in order to bypass the bias induced in the participants' history of trials with more target stimuli towards the beginning of the block (see *issues with the data collection* subsection in the *materials and methods* section). Instead, simulations were generated with new simulated data and correct/incorrect feedback was directly obtained from the responses of the model on each trial.

Information theoretic measures

In order to quantify the uncertainty of the model, we computed the entropy on the response, as well as that of the 'target' and 'rule' layers of our model. These are defined as follows:

$$H(Y_n|\mathbf{x_n}, \mathbf{t_{n-1}}) = H_{Bernouilli}(P(y_n|\mathbf{x_n}, \mathbf{t_{n-1}})) \quad \text{(eq. 5.5.1)}$$

$$H(T_n|\mathbf{x_n}, \mathbf{t_{n-1}}) = H_{Bernouilli}(P(t_n|\mathbf{x_n}, \mathbf{t_{n-1}})) \quad \text{(eq. 5.5.2)}$$

$$H(R|\mathbf{x_n}, \mathbf{t_{n-1}}) = -\sum_r P(r|\mathbf{x_n}, \mathbf{t_{n-1}}) \times \log(P(r|\mathbf{x_n}, \mathbf{t_{n-1}})) \quad \text{(eq. 5.5.3)}$$

with the entropy for Bernouilli distributions defined as:

$$H_{Bernouilli}(p) = -p \times \log(p) - (1-p) \times \log(1-p) \quad \text{(eq. 5.5.4)}$$

fMRI data acquisition and preprocessing

Magnetic resonance images were acquired with a 3T Siemens VERIO scanner with a 32-channel head coil using a standard echo-planar imaging sequence. Whole-head T_2*-weighted echo-planar images were continuously acquired with a repetition time of 2 s, echo time of 30 ms. We acquired 510 volumes per block, plus 3 dummy scans discarded before the analyses. Each volume included 64 × 64 × 36 voxels of 3 × 3 × 3

mm. A high-resolution T_1-weighted structural image was also obtained (voxel size = 1 × 1 × 1 mm). For standard preprocessing and univariate statistical analyses, we used SPM8 (Wellcome Department of Cognitive Neurology, London, United Kingdom). All other analyses were done with custom scripts for Matlab (Mathworks, Natick, MA, United States of America). We also used xjview (http://www.alivelearn.net/xjview) to visualize the data and to construct mask and conjunction images. For each participant, we first realigned all functional images, then we co-registered (rigid body transformation) the participant's anatomical scan to the mean functional image, and then co-registered the participant's data to the Montreal Neurological Institute (MNI) template brain. We then normalized each participant's data to the template brain space, using segmented probabilistic maps for grey matter, white matter, and cerebro-spinal fluid. Functional images were resampled (4 × 4 × 4 mm voxels) and spatially smoothed (8-mm full-width half-maximum (FWHM) Gaussian kernel).

fMRI analysis

We analysed our data using SPM12. Our univariate analyses used a generalized linear model (GLM) approach. A 128-s temporal high-pass filter was applied to remove low-frequency scanner artifacts. Temporal autocorrelation in the time series data was estimated using restricted maximum-likelihood estimates of variance components using a first-order autoregressive model (AR-1), and the resulting non-sphericity was used to form maximum-likelihood estimates of the activations, consistent with standard approaches in SPM8 and SPM12 (Penny et al., 2006). Our GLM included regressors coding for onsets and durations of stimuli or events, which were then convolved with the canonical hemodynamic response function (HRF; itself convolved with a duration window of 3 seconds) and regressed against the observed fMRI data.

We modelled all scanner runs using shared regressors (by concatenation) but constant terms for each run were included in order to model any difference in offset. We constructed a design matrix for each of the GLM analyses. All design matrices included two regressors aligned to cue and stimulus onset, and a parametric regressor of cue type (familiar vs. novel) at the time of the cue. Additionally, motion parameters were included as nuisance variables. Our GLM analyses only differed in terms of the parametric regressors at the onset of the trials. A detailed table with the exact regressors included in each GLM can be found in **Supplementary Tables ST5.3–6**, and an overview is described in the *Results and statistical analyses* section. We report peak voxels from clusters that responded at thresholds that were corrected for multiple comparisons, using a false discovery rate of $p < 0.05$. More information on the uncorrected threshold used as well as the number of voxels per cluster is reported in the **Supplementary Tables ST5.7–9**. Voxelwise statistics were rendered onto the MNI template brain using xjview 9.5 (http://www.alivelearn.net/xjview/) and our own Matlab scripts.

Results

Task and design

Eighteen healthy human participants learned to classify stimuli (each composed of a pair of coloured shapes occurring left and right of fixation) as 'target' or 'nontarget' on the basis of their shape (square, circle, triangle) and/or colour (red, green, blue), responding with a button press, and receiving fully informative trial-and-error feedback after each response (**Fig 5.1a**). Each decision rule, which remained constant over a block of 16 trials, defined targets (50% of trials) as an (inclusive) disjunctive combination of one feature on the left and another on the right. For example, in one

block stimuli were targets if the shape on the left was triangle or the colour on the right was a blue (i.e. we notate this rule as {left: triangle, right: blue}). The use of a disjunctive ('or') rule ensured that an effect (target) could have multiple possible causes (i.e. potential features, e.g. shape or colour on either side). In the previous example, correct feedback could be received for the response 'target' to the stimulus {left: red triangle, right: green circle} even though {right: green} and {right: circle} are not part of the rule.

Critically, the disjunctive task also ensured that targets and nontargets provided asymmetric information about the decision rule. Feedback that a pair of shapes was nontarget allowed participants to completely eliminate a region of the space of hypotheses (e.g. the rule doesn't include {left: blue}), whereas information that it was a target provided only incremental evidence but generally precluded definitive inclusion or exclusion of a subset of rules. This design thus allowed us to pursue our main question of interest: how participants learn differently from verification (i.e. evidence indicating that a stimulus was a target) and falsification (i.e. evidence indicating that it was a nontarget), and how this interacted with correct or incorrect feedback.

In the *familiar cues* condition, symbolic cues, whose meaning had been learned in a previous training session, disclosed the *class* of rule that applied in that block. The rule 'class' referred to the relevant dimension (e.g. shape on the left, colour on the right) but not the feature (e.g. triangle on the left, blue on the right) that was relevant for decisions. There were four rule classes: {left: colour, right: colour}, {left: colour, right: shape}, {left: shape, right: colour}, and {left: shape, right: shape}. In the *novel*

cues condition, a distinct set of cues were paired randomly with blocks, so that they offered no information about the relevant rule. An illustration of the task is provided in **Fig 5.1a**.

Model performance

We modelled the data using the framework of Bayesian induction, in which the probability of each underlying rule is estimated through Bayes' rule. Our model extended an optimal (Hierarchical Bayesian; HB) agent with parameters that reflected learning and policy biases. Concretely, it had four parameters controlling the learning rates for trials that combined target/nontarget and correct/incorrect feedback. We began by simulating how model performance varied under different parameterisations of its learning rates. This confirmed, as expected, that simulated model performance peaked for high values of all learning rates. Additionally, we found that in our task performance was most hindered by low nontarget learning, such that performance increased when learning more from nontarget than target trials ($\alpha_{T^*} < \alpha_{N^*}$) compared to the opposite ($\alpha_{T^*} > \alpha_{N^*}$), for blocks with familiar/novel cues and correct/incorrect trials (all $t_{(9999)} > +35$, $p < 10^{-200}$; blue heatmap in **Fig 5.3a**). Intuitively, this follows from the disjunctive rule, which ensures that nontarget evidence is more informative. Consider a stimulus composed of a red element on the left and a blue element on the right. Feedback that the stimulus is a 'nontarget' allows any candidate rules with component {left: red} or {right: blue} to be eliminated. Feedback that the stimulus is a 'target' implies that one component of the rule is either {left: red} or {right: blue}, but does not indicate which is. Under suboptimal learning, it is thus advantageous weighting learning from nontarget trials considerably more than from target.

Model fitting to human behaviour

Average human accuracy (% correct responses) and choices (% target responses) across the 16 trials that constituted each block are shown in **Fig 5.1b** (error bars) (see also **Fig SF5.2** for reaction times). Accuracy increased across the block overall ($F_{(15,255)}$ = 29.9, $p < 0.001$), but did so faster in blocks with familiar cues (trial × cue interaction, $F_{(1,17)} = 2.90$, $p < 0.001$). Participants began with a bias to respond 'target' that abated across the block ($F_{(15,255)} = 46.0$, $p < 0.001$) in roughly equal measure for the two conditions (trial × cue interaction, $F < 1$). Next, we adjusted the model parameters to fit the human responses on a per trial basis, separately for familiar and novel conditions. The choice and accuracy of the best-fitting model parameterisation (lines) is rendered onto equivalent human data in **Fig 5.1b**, revealing good fits across the block in each condition. The model also predicted choices independently for target and nontarget stimuli (**Fig SF5.1**) and as a function of the history of target and nontarget feedback (**Fig 5.3c**; see below), and provided a better fitting score than control alternative models (see *Alternative models* section in the methods, and **Supplementary Table ST5.2**). The normative HB model is also shown in dashed lines for a comparison of human and optimal performance.

We additionally compared the reaction time of participants with the uncertainty (entropy) of the model response (see equation 5.5), which were highly correlated ($t_{(17)}$ = +6.7 $p < 0.001$) even when we demeaned each quantity independently for each subject, block cue, trial number, target/nontarget response, and correct/incorrect feedback ($t_{(17)} = +5.06$ $p < 0.001$). Thus, our model additionally captured variance in reaction times that could only be attributed to the variability in learning between blocks irrespectively of the cue.

Model estimates of learning rates

Our model allowed us to estimate independently a learning rate for the block cues and different types of feedback (see **Fig 5.3b**). We were particularly interested in quantifying the effect of these conditions on the amount of learning per trial. We thus subjected the learning rate estimates to an analysis of variance crossing the factors cues (novel, familiar) × feedback (target, nontarget) × feedback (correct, incorrect). We found a significant main effect with faster learning from correct than incorrect feedback ($F_{(1,17)} = 133.3$, $p < 0.001$), between target and nontarget feedback ($F_{(1,17)} = +7.6$ $p = 0.013$), and interaction between the two types of feedback reflecting that participants also learnt better when their response was 'target' compared to non-target ($F_{(1,17)} = 10.4$, $p = 0.005$). Noticeably, the learning rates didn't differ significantly between familiar and novel cues as a function of the cognitive load ($F_{(1,17)} = 2.5$, $p = 0.13$). An additional interaction between cue and target/nontarget feedback reached significance ($F_{(1,17)} = 7.9$, $p = 0.012$), showing a shift in strategy between the two types of cues. We then performed ANOVAs independently for both types of blocks in order to understand this result. In familiar cues, there was no main effect of target/nontarget feedback ($F_{(1,17)} < 1$, $p = 0.41$) but an interaction between both types of feedback ($F_{(1,17)} = 11.0$, $p = 0.004$). The opposite effect was found for novel blocks, with a main effect of target/nontarget feedback ($F_{(1,17)} = 11.9$, $p = 0.004$) but no interaction ($F_{(1,17)} = 2.5$, $p = 0.129$). A more detailed exploration into the differences in learning rates between familiar and novel cues revealed that this effect was largely driven by a shift of nontarget incorrect trials, where the learning rate α_{NI} differed significantly between familiar and novel cues ($t_{(17)} = +3.9$, $p < 0.001$; $p_{FDR} = 0.0014$ corrected for multiple comparisons across all learning rates). No other learning rate varied significantly between familiar and novel cues (all $p > 0.49$).

The influence of choice history

Next, we used the best-fitting parameterisation of the model to predict how learning from the past history of target and nontarget feedback across the block might influence behaviour. We plotted how the human probability of responding "target" varied as a function of the history of feedback indicating that each of its features was part of a target or nontarget (see **Fig 5.3c.**). Specifically, for each trial, we computed the proportion of target responses as a function of the number of previous times that the features in the relevant dimensions (e.g. {left: circle} or {right: green} in a {left: shape, right: colour} block) and in the irrelevant dimensions (e.g. {left: blue} or {right: triangle}) had received 'target' feedback (i.e. feedback indicating that the stimulus contained a target) and 'nontarget' feedback (i.e. indicating that it did not). We collapsed over instances where there were 4 or more instances of target or nontarget feedback, and subjected the data for both relevant and irrelevant dimensions to an analysis of variance crossing the factors cues (novel, familiar) × feedback (target, nontarget) × bin (1–4 times) × dimension (relevant, irrelevant). A three-way interaction between the bin, the feedback and the dimension factors indicated that the slope of the line relating feedback history to choice was steeper for targets than nontargets, but that this was driven by the relevant dimensions ($F_{(3,51)} = 33.8$, $p < 0.001$). However, the four-way interaction did not reach significance, meaning that this effect wasn't modulated by the cue (familiar/novel). Note that we found a qualitatively similar pattern under normative behaviour (dashed lines in **Fig 5.3c.**), and thus this behaviour is optimal and we cannot interpret it in terms of biases of human inference. We next validated our model by performing a similar analysis on its responses (using the previously-estimated best-fitting parameterisation). We found a very similar pattern of effects to the human data, both qualitatively and quantitatively,

and with the same significant main effects and interaction — except for three-way interaction feedback × bin × dimension exclusive to the ($F_{(3,51)} = 3.028$, $p = 0.038$) and the four-way interaction ($F_{(3,51)} = 7.03$, $p < 0.001$). Human (error bars), model (continuous lines) and normative (HB model; dashed lines) data are shown for comparison in **Fig 5.3c**.

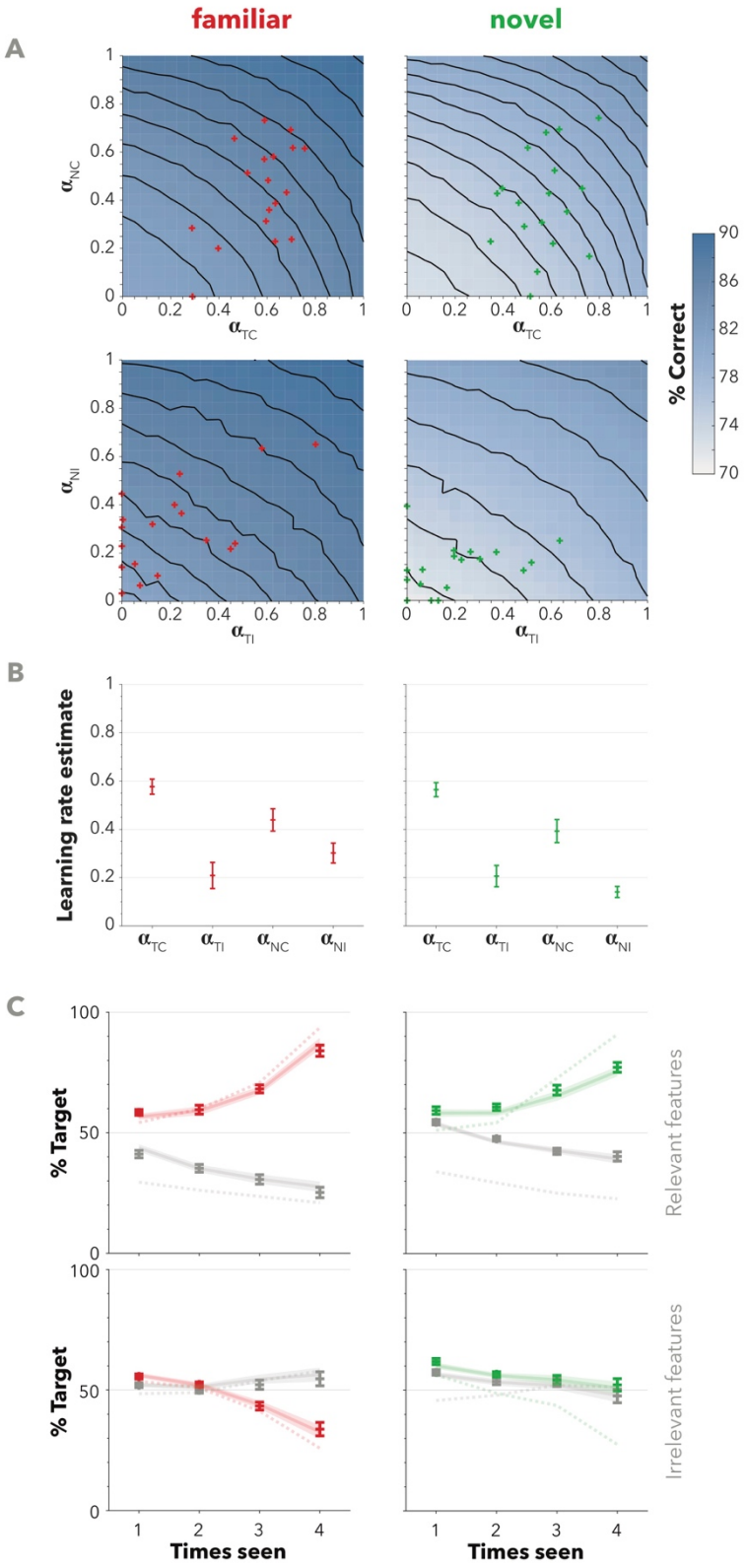

Figure 5.3. Fits of the computational model

A. Model performance (% correct) as a function of the model learning rates (blue heatmap with black contour lines) on blocks with familiar cues (left panel) and novel cues (right panel) under an optimal policy. Darker blue indicates higher performance (% correct). Red and green crosses show individual humans participants in the novel

and familiar cues conditions respectively. **B.** Best-fitting learning rates in the familiar (red) and novel (green) cues condition. Error bars represent the SEM. **C.** Proportion of "target" responses as a function of the number of previous times the rule-relevant feature (top panels) and rule-irrelevant feature (bottom panels) was associated with target (red or green crosses with SEM error bars) or nontarget (gray crosses with error bars) feedback. Data are shown separately for familiar (left panels) and novel (right panels) cues conditions. The best-fitting model is represented as a continuous line with shaded SEM. Dashed lines show normative behaviour from the HB model.

Functional neuroimaging

Together, these behavioural and modelling analyses suggest that during inductive reasoning, humans pursue a suboptimal positive strategy in which learning is over-reliant on trials where they correct and in 'target' trials, and that this tendency is particularly pronounced in the novel cues condition, when the number of possible hypotheses about the rule exceeds the likely capacity of online maintenance processes. Next, thus, we sought to characterise the neural mechanisms which might give rise to this bias, by examining blood-oxygen (BOLD) data from functional magnetic resonance imaging (fMRI) whilst participants performed the task. In what follows, all reported effects survive cluster-level false discovery rate (FDR) correction for multiple comparisons (see methods) unless otherwise stated.

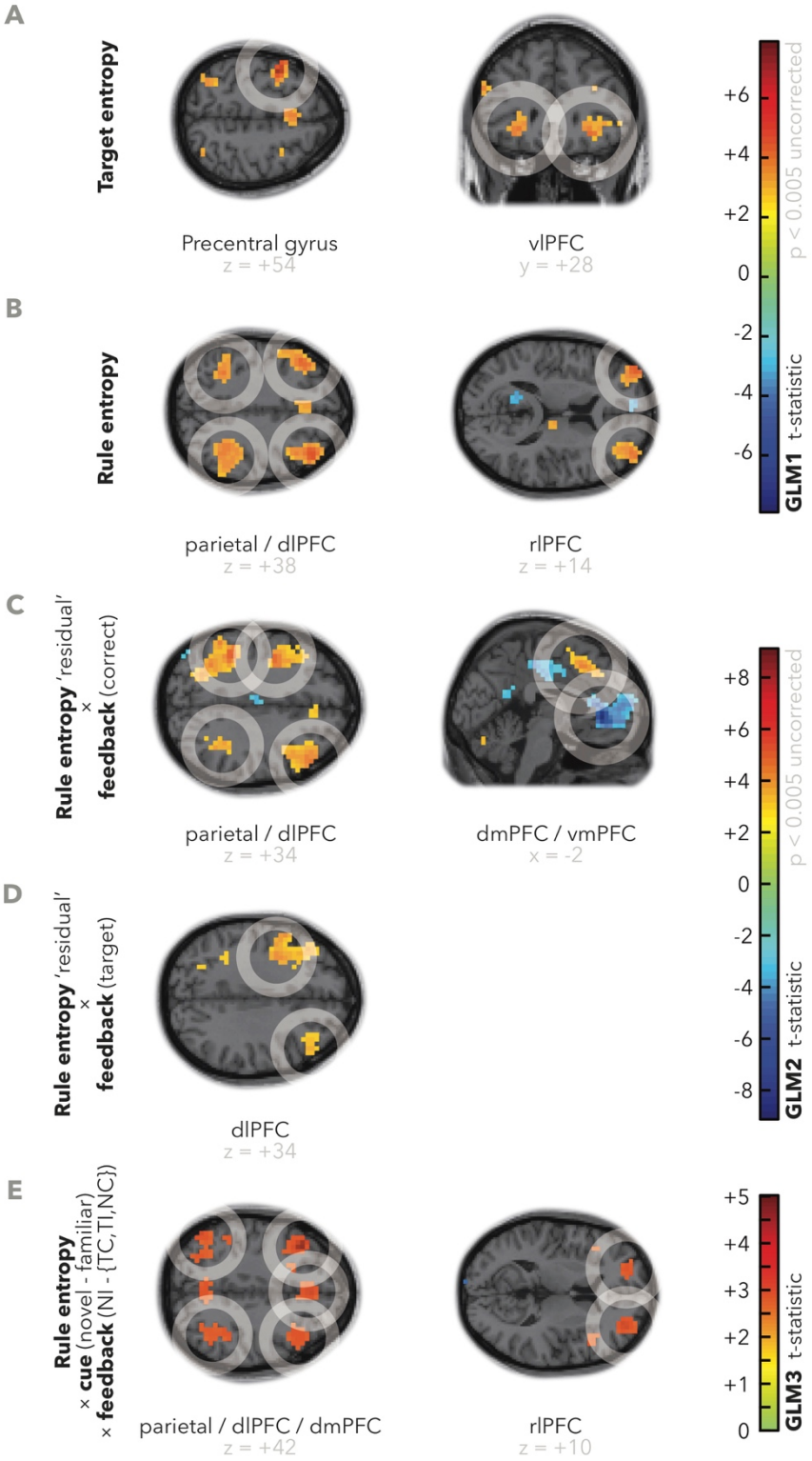

Figure 5.4. Model-based fMRI contrasts
A/B. Results of GLM5.1, looking at the dissociative effects of target (response) and rule entropies. **C/D.** Results of GLM5.2 with the interaction of rule entropy and feedback. **E.** Results of GLM5.3 retrieves a similar set of regions for the 3-way interaction between rule entropy, cue, and feedback type. All results were displayed at

thresholded at an uncorrected p-value < 0.005, and highlighted clusters survived cluster-wise FDR multiple comparisons correction.

Dissociable effects of response and rule uncertainty in the BOLD signal

Of key interest to this study was the dissociation of representation of the low-level stimulus-response association and the rule. We thus started our neural analysis by searching for brain regions involved in representing the uncertainty related to different hierarchical levels, i.e. uncertainty about the response (target or not) and uncertainty at level of the rule. Our model explicitly represents each as a probability distribution. We first looked at the effects of rule and response uncertainty in the brain data. Under the Bayesian framework (Information Theory), uncertainty can be formally quantified as the entropy of a distribution (see equations 5.5.2 and 5.5.3 in Methods, and **Fig SF5.6**). We thus regressed these two quantities together against the BOLD data (see **GLM5.1** in Methods) together with other nuisance predictors. Among the latter, we included a set of regressors that controlled for any effects related to a) the effect of response and the effect of correct/incorrect feedback, b) the difference between familiar and novel cues, c) the effect of trial number, and d) any possible interaction between cue and trial number (see **Fig SF5.7**). We relied on the fact that our modelling approach allowed us to capture the variability between blocks (even for same type of cue and trial number), and thus our results were not driven by the average tendency of the entropy regressors. The results of this analysis can be seen in **Figs 5.4a–b**. There was a positive effect of response uncertainty in the precentral gyrus that survived multiple comparisons correction in the left side (peak -42, 0, +54; $t_{(17)} = +7.4$, $p < 10^{-5}$ uncorrected) but not the right side (peak +30, +4, +42; $t_{(17)} = +3.9$, $p = 0.056$ uncorrected), and a bilateral cluster in the ventrolateral prefrontal cortex (vlPFC; left peak -26, +24, -10; $t_{(17)} = +4.9$, $p = 6.1 \times 10^{-5}$ uncorrected; right peak

+30, +28, -10; $t_{(17)}$ = +4.2, p = 2.9 × 10^{-4} uncorrected). Perhaps surprisingly, we found a distinct set of regions that reflected the uncertainty of the rule, including bilateral rostrolateral PFC (rlPFC; left peak: -38, +60, +18, $t_{(17)}$ = +5.6, p = 1.4 × 10^{-5} uncorrected; right peak: +42, +60, +14, $t_{(17)}$ = +4.14, p = 3.4 × 10^{-4} uncorrected), dorsolateral PFC (dlPFC; left peak: -46, +24, +38, $t_{(17)}$ = +4.5, p = 1.7 × 10^{-4} uncorrected; right peak: +42, +36, +38, $t_{(17)}$ = +5.6, p = 1.5 × 10^{-5} uncorrected) and parietal cortices (left peak: -34, -52, +42, $t_{(17)}$ = +4.1, p = 3.4 × 10^{-4} uncorrected; right peak: +54, -48, +50, $t_{(17)}$ = +4.7, p = 1.0 × 10^{-4} uncorrected). The full set of activations can be found in **Supplementary Table ST5.7**.

Encoding of feedback is modulated by rule uncertainty in prefrontal and parietal cortices

Once we established the regions that were modulated by the uncertainty of the rule, we looked for prediction error (learning) signals driven by each type of evidence, i.e. the difference of target/nontarget and correct/incorrect feedback. We hypothesised that these should be modulated by the rule entropy, and tested our prediction with a second regression (**GLM5.2** in Methods). Similar to **GLM5.1**, we carefully designed our analysis in order to assert that the results were not driven by the cue or the amount of time spent in the block. However, in this case we followed a different approach to **GLM5.1**. Rather than including a multitude of nuisance regressors, we instead decomposed the original rule entropy quantity $H(R|\mathbf{x_n}, \mathbf{t_{n-1}})$ into two orthogonal regressors '*mean RE*' and '*residual RE*' such that:

$$H(R|\mathbf{x_n}, \mathbf{t_{n-1}}) = mean\,RE + residual\,RE$$

We computed '*mean RE*' by estimating per participant the average rule entropy for each combination of cue and trial number, such that this value would be equal within each condition for all blocks (e.g. on all the third trials for familiar blocks). The quantity '*residual RE*' was obtained by subtracting per trial the average rule entropy within its trial/cue condition, such that the '*residual RE*' was zero on average for any given cue or trial number (or combination of these). This procedure corresponds to 'de-meaning' or 'partialling-out' the effect of cue and trial number from the rule entropy quantity. Note that the two components are orthogonal by design.

Our design matrix in **GLM5.2** included a) the two components '*mean RE*' and '*residual RE*' relating to the uncertainty of the rule (2 regressors), b) the main effects of feedback (correct-incorrect and target-nontarget feedback) and their interaction (target-nontarget response; 3 regressors), and c) the interaction of each rule entropy regressor with each feedback regressor (6 regressors). We found (**Figs 5.4c–d**) that the bilateral dlPFC region reacted positively to the interaction of '*residual RE*' with correct feedback (left peak: -34, 0, +58, $t_{(17)}$ = +5.6, p = 1.7 × 10^{-5} uncorrected; right peak: +54, +28, +30, $t_{(17)}$ = +5.3, p = 3.0 × 10^{-5} uncorrected) and also to the interaction with target feedback (left peak: -54, +8, +38, $t_{(17)}$ = +4.8, p = 8.4 × 10^{-5} uncorrected; right peak: +46, +44, +6, $t_{(17)}$ = +6.0, p = 7.3 × 10^{-6} uncorrected). Other regions yielded similar effects of the former but not the latter interaction: positively in bilateral parietal cortices (left peak: -50, -40, +38, $t_{(17)}$ = +7.5, p = 4.2 × 10^{-7} uncorrected; right peak: +46, -40, +46, $t_{(17)}$ = +5.9, p = 7.9 × 10^{-6} uncorrected), in dorsomedial PFC (dmPFC; peak +6, +20, +46, $t_{(17)}$ = +4.5, p = 1.4 × 10^{-4} uncorrected) and negatively in the ventromedial PFC (vmPFC, also termed perigenual anterior cingulate cortex or pgACC; peak -2, +36, -2, $t_{(17)}$ = -9.1, p = 3.0 × 10^{-8} uncorrected). The interactions between rule entropy and correct feedback were also present with the '*mean RE*'

regressor, significant and in the same direction. No significant clusters were found for the interaction between '*residual RE*' and response. A list with all the activations can be found in **Supplementary Table ST5.8**.

In order to assess our results, we performed an additional control analysis similar to **GLM5.2** in spirit, with the only exception that we randomly permuted the values of '*residual RE*' across blocks and within each condition cue × trial number. In this control analysis, the main effect of '*residual RE*' and its interactions vanished and no clusters were significant.

So far, the results speak towards a network of prefrontal and parietal regions that encode the feedback as a function of the necessity to infer the rule underlying the current block, together with a more posterior region encoding the uncertainty of the response to the current trial. In previous analyses, we have controlled for the block cue rather than exploring how it affected the BOLD signal. In what follows we will focus on exploring the interaction of the feedback with the type of cue.

Neural correlates of enhanced 'positive' inference in novel blocks

The parameter estimates retrieved by the model revealed an interesting effect where only processing of nontarget and incorrect feedback was hindered in novel cues, compared to familiar (see **Fig 5.3b**). In other words, estimate of the learning rate α_{NI} differed between the two cues, but not those of α_{TC}, α_{TI} or α_{NC}. We thus searched for an equivalent effect in the BOLD signal. We performed a regression (**GLM5.3**, see Methods section) with the following regressors: a) the effect of response and correct/incorrect feedback and their interaction, b) the entropy of the rule, and c) a binary regressor capturing the main effect between familiar and novel cues. We further captured the interaction between the cue and the entropy of the rule using four different regressors, i.e. one per type of feedback. We defined a contrast in order to

reveal voxels where the interaction between cue and rule entropy was stronger for nontarget and incorrect trials compared to other feedback. Consistent with our previous result, this analyses yielded the same set of regions (**Fig 5.4e**), namely rlPFC (left cluster marginally significant, peak: -18, +56, +6, $t_{(17)}$ = +3.8, p = 6.9 × 10^{-4} uncorrected; right peak: +34, +52, +6, $t_{(17)}$ = +4.3, p = 2.5 × 10^{-4} uncorrected), dlPFC (left peak: -50, +24, +22, $t_{(17)}$ = +5.0, p = 5.2 × 10^{-5} uncorrected; right peak: +54, +20, +26, $t_{(17)}$ = +4.6, p = 1.2 × 10^{-4} uncorrected), dmPFC (marginally significant, peak: +2, +32, +38, $t_{(17)}$ = +4.2, p = 2.8 × 10^{-4} uncorrected) and parietal cortex (left cluster marginally significant, peak: -50, -64, +42, $t_{(17)}$ = +4.0, p = 4.4 × 10^{-4} uncorrected; right cluster marginally significant, peak: +30, -64, +54, $t_{(17)}$ = +3.6, p = 1.1 × 10^{-3} uncorrected). A list with all the activations can be found in **Supplementary Table ST5.9**.

We performed one last GLM (see **GLM5.4** in the Methods section) in order to interpret the nature and direction of all the interactions with rule entropy. We illustrate in **Fig 5.5** the dynamics of BOLD signal within the dlPFC region as defined from the contrast of '*residual RE*' × feedback(correct - incorrect) in **GLM5.2**.

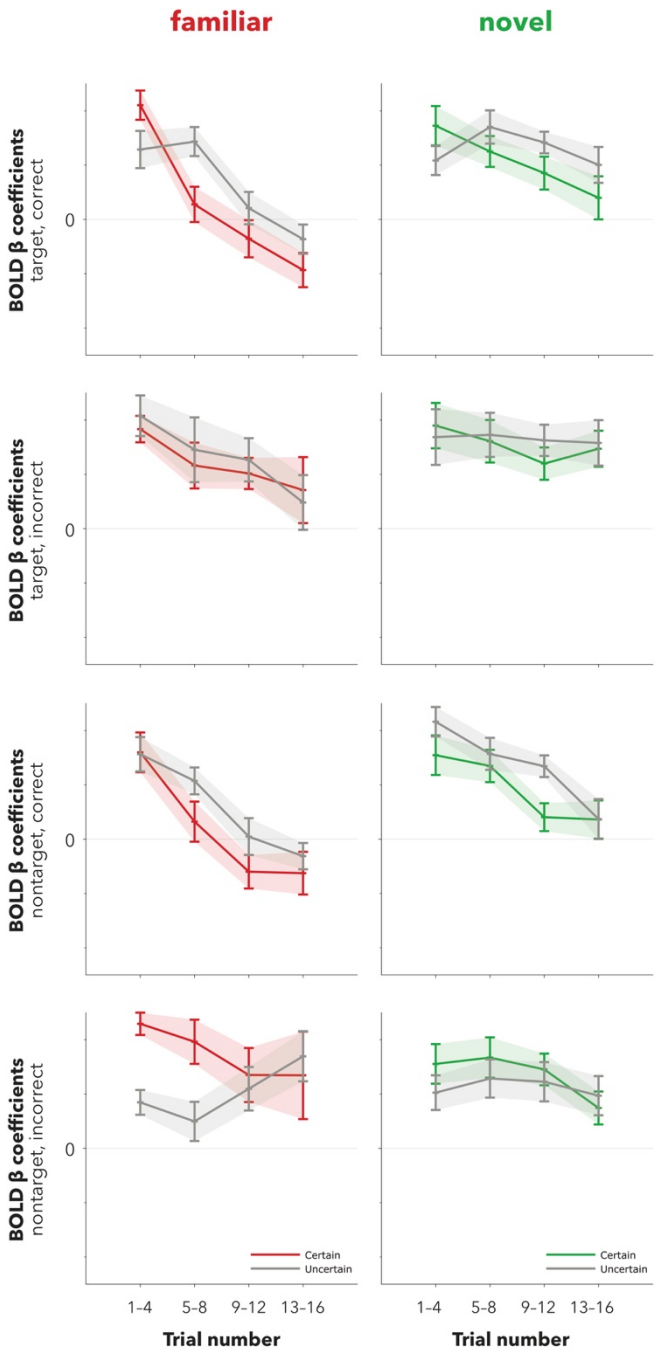

Figure 5.5. Dynamics of the BOLD signal in dorsolateral PFC
Neural effects obtained from GLM5.4 within a region of interest (ROI) in dorsolateral prefrontal cortex (dlPFC, see **Fig 4c**). We estimated a beta coefficients from the BOLD signal for each participant, independently for each combination of trial (bins: 1–4, 5–8, 9–12, 13–16) × cue (familiar, novel) × feedback(target, nontarget) × feedback (correct, incorrect) × rule entropy (bins: low, high). Rule entropy was divided evenly in two bins by the median, independently for each combination of the other conditions (trial × cue × feedbacks).

Discussion

The current data replicate the finding that humans prefer verification over falsification during rule discovery. Unlike some previous approaches, our model estimates quantitatively the objective (normative) and subjective (participants) value of verificatory and falsificatory evidence, all within a Bayesian framework where the optimal policy can be fully specified. First, we show through model simulation that nontarget trials were more informative than target trials. Second, we found from fitting the model parameters to responses that participants processed better feedback from target stimuli when they responded correctly; in other words, our participants used a positive test strategy despite this was detrimental to performance. Third, this effect was particularly pronounced when the cognitive load (the set of possible rules) is higher, and it exclusively affected trials with nontarget and incorrect feedback.

We note that the purpose of modelling human behaviour with a Bayesian model was not to argue that human neural signals explicitly represent a probability distribution. Instead, we relied on an extension of a Bayesian model in order to characterise inference biases from optimality. For example, it is possible that participants were keeping track only a few rules simultaneously, gaining evidence with time and discarding them in favour of new ones whenever they were violated. However, we believe our approach should capture the general trend of such strategies, e.g. by assigning a probability distribution to each of the rules that participants could be tracking on any given trial. One possible interpretation for the difference between familiar and novel cues is that participants were able to formulate better alternative hypotheses when their current theory was invalidated (i.e. they mistakenly classified a stimulus as target) and the space of hypothesis was smaller (familiar cues).

In the fMRI BOLD data, we first revealed two different sets of regions within the frontal and parietal cortices that responded to the uncertainty of the response and of the underlying rule. This finding has a value of its own irrespective of the inductive biases discussed above. The functional organisation of the prefrontal cortex is an active topic of research, with theories arguing a rostro-caudal gradient organised along principal dimensions such as abstraction (O'reilly, Noelle, Braver, & Cohen, 2002), time (Koechlin et al., 2003) and prediction of errors (Alexander & Brown, 2015). In this task, we found that ventrolateral prefrontal cortex (vlPFC) reflected the uncertainty about the response to the trial, but not the underlying rule. Conversely, the dorsomedial PFC (dmPFC), dorsolateral PFC (dlPFC), rostrolateral PFC (rlPFC) and parietal cortex reflected prediction error signals, namely the correct-incorrect feedback modulated by the amount of uncertainty about the rule governing the current block. This result is in agreement with theories about the dmPFC, which may solely represent the overarching goal of the hierarchy (in our task that is the rule) (Fernandes et al., 2018). Our results are harder to reconcile with current theories of dlPFC that argue for a role in action selection as opposed to monitoring (Koechlin et al., 2003). Consistent with the behavioural findings, we further observed learning signals in all regions modulated by the entropy of the rule, and these were in the same direction than the behavioural (learning rates) biases.

Overall, this experiment sheds light into how high-level rules and low-level stimulus-response associations are represented in the human brain, in accordance with the behavioural biases captured by our computational model. Notably, many of the regions found to correspond to the high-level rule overlapped with the regions

associated with hierarchical representations from the previous chapters. For instance, the dmPFC region represented the higher level also in the study described in Chapter 4. Similarly, the parietal region was also engaged in the experiment in Chapter 3, where it reflected the hierarchical level of the stimuli. The results of all experiments, and their interpretation, will be treated in the **General Discussion** in the next and last chapter.

Chapter 6. General discussion

In the previous chapters I have presented the outcome of four original experiments. In what follows, I summarise the main results of each experiment. I then discuss how they relate to each other, point out some limitations and criticisms, and suggest future lines of research that could build on these results.

1. Overview

In **Experiment 2.1** in **Chapter 2**, I found that within a single behavioural session participants do not exploit the hierarchical structure of a dataset. Instead, they learn exemplars holistically using a qualitatively distinct mechanism thought to be subserved at least in part by the medial temporal cortex.

In **Experiments 2.2** and **2.3**, as well as in **Chapter 3**, I found that when trained over longer periods of time and over two sessions, humans can transfer a learned hierarchical structure to novel stimuli, thus accelerating learning when the underlying structural relationship among stimuli is consistent. Moreover, brain regions such as the left parietal as well as the left rostrolateral prefrontal (rlPFC) are sensitive to the hierarchical level associated with each feature. I noted that the hierarchical level can be mapped onto a one-dimensional representation (low < middle < top). Under this interpretation, the univariate neural effects revealed are reminiscent of a one-dimensional magnitude coding.

In **Chapter 4**, I found that human participants display a behavioural (reaction time) cost that is proportional to the hierarchical description length of the plan over remaining steps to a goal. An equivalent effect was found at the neural level bilaterally in premotor (PMC) as well as in the dorsomedial prefrontal (dmPFC) cortices, and neural representations of plan structure could be decoded from the dACC. I interpret these results as a signature of hierarchical representations in human planning or sequential, goal-directed decision-making.

In **Chapter 5** I found that participants learned abstract rules in a categorisation task that allowed them to generalise beyond basic stimulus-outcome associations. Parameters fitted to a Bayesian model revealed that participants processed the feedback better when they accurately verified a rule (i.e. responded target correctly) despite nontarget stimuli was more informative from a normative point of view. Ventrolateral prefrontal cortex (vlPFC) reflected the uncertainty of the rule. A wider network including dorsomedial (dmPFC), bilateral dorsolateral prefrontal (dlPFC), as well as bilateral parietal cortices encoded prediction error signals of the rule.

2. Discussion

Hierarchical representations are thought to be influential for learning and decision-making in humans and other animals. The results of the experiments I have describe provide evidence for different aspects of hierarchical representations, and fit together within a bigger picture. They constrain the conditions under which hierarchical representations can have an influence in behaviour, as well as the regions engaged in representing and exploiting hierarchical representations.

2.1 The effect of time on hierarchical representations

In all experiments except **Experiment 2.1**, participants performed multiple sessions over multiple days. Coincidentally, **2.1** was the only experiment where I failed to find evidence of structure learning. One straightforward (but not conclusive) interpretation is that humans can only exploit hierarchical relationships some time after being exposed to such relationships. This can be explained by the effect of sleep-dependent consolidation (Schapiro et al., 2017).

However, this conclusion cannot be firmly drawn from the set of experiments that I have presented alone, and would need to be tested in a follow-up study. Indeed, many other aspects of the experiment differ between **Experiment 2.1** and **Experiments 2.2-2.3**, such as the stimuli, the task, the hierarchical structure, and the variables of interest. It is thus not possible to interpret meaningfully why we found signatures of hierarchical representations in some experiments but not in others.

2.2. The role of the parietal cortex in hierarchical representations

The parietal cortex was among the regions of interest in two out of three imaging experiments that I have presented in **Chapters 3**, **4**, and **5**. In **Chapter 3**, it represented in a univariate fashion the hierarchical level associated with a given feature (fruit). In **Chapter 5**, it reflected learning signals associated with the high-level rule that determined the stimulus' category. In **Chapter 4**, however, where participants planned their path throughout a subway network, the parietal cortex reflected a more basic "direction switch" effect, where it became more active following a change of direction (or response equivalently). Note however that the regions of interest labelled as 'parietal' across the different experiments are not

overlapping. Historically, the parietal cortex has been associated with many computational functions, e.g. attention (Colby & Goldberg, 1999; Posner, Walker, Friedrich, & Rafal, 1984), working memory (Jonides et al., 1998; Pesaran, Pezaris, Sahani, Mitra, & Andersen, 2002), number and magnitude encoding (Tudusciuc & Nieder, 2007), perceptual evidence integration (Shadlen & Newsome, 2001), and spatial reasoning (Duhamel, Colby, & Goldberg, 1992).

In my view, the most interesting aspect of the parietal cortex is reflected in the results of **Chapter 3**. This chapter includes the only experiment where participants are required to flexibly reason about the multiple hierarchical levels, which can be encoded with a one-dimensional representation (low < middle < top level). Note that in the design of the experiment, the level was not presented as a magnitude. This indicates that the parietal lobe may underpin mechanisms to represent transitive relationships other than one-dimensional magnitudes, and in particular takes the role of representing the position tracked within the transitive structure.

2.3. The role of the dorsomedial prefrontal cortex in hierarchical representations

In the experiments exposed in this thesis, the dorsomedial prefrontal cortex (dmPFC) is associated with a very different function compared to the parietal cortex. The BOLD signal in the dmPFC correlated with task-related variables in two out of three imaging experiments that I have presented (**Chapters 4** and **5**). In **Chapter 4**, dmPFC encoded the hierarchical cost of the plan, and also the identity of the current context. In **Chapter 5**, dmPFC was among the regions that encoded the uncertainty about the rule as well as prediction error signals modulated by the rule. Note that the regions labelled 'dmPFC' do not overlap between **Chapter 4** (more posterior) and **Chapter 5**

(more anterior). This need not stop us from considering similarities between the functional role that these neighbouring regions have across the tasks. The precise functional neuroanatomy of the dmPFC is a controversial topic. One possibility is that in both tasks dmPFC is encoding the identity represented at the highest-level of abstraction: the subway line during the navigation task in **Chapter 4**; and the current rule in the rule-discovery task in **Chapter 5**. The idea that the dmPFC may be most (or exclusively) sensitive to the highest hierarchical level is consistent with views expressed elsewhere (Fernandes et al., 2018; Holroyd & Yeung, 2012).

Interestingly, I failed to find significant effects in dmPFC related to the hierarchical level of the features (fruits) in **Chapter 3**. There are at least two differences in the design of the task that could explain this.

First, the task demands participants to focus on multiple hierarchical levels at once within each trial (e.g. by seeing three fruits in the *validity* task, or two fruits in the *session matching* task). In contrast, the feature at top hierarchical levels in the other experiments (e.g. the line of the subway; or the category rule) are constant over longer periods of time (multiple steps in the same subway line; or the full block with the same rule).

Second, the relationship between the fruit triplets (salads) is hierarchical in that it is transitive (low < middle < top level), but this sense of order is partially driven by the temporal structure during the learning session, and does not confer a strong benefit of compression (of the description of the set of triplets), unlike the experiments in **Chapter 4** (compression of the connectivity between stations) and **Chapter 5** (compression of the set of target stimuli under a given rule).

Thus, it seems like hierarchical representations may flexibly engage a different set of regions depending on more subtle aspects relating to the nature of the hierarchical structure, as well as the demands of the task to be performed.

2.4. Comparison to classic research in semantic categorisation

Notice that I failed to find any meaningful activity in the anterior temporal lobe (ATL), and more specifically in the temporal pole, in any of the tasks. This may seem at first surprising, as this is region has been historically associated to hierarchical representations (e.g. for semantic categorisation) as explained in the *General Introduction* in **Chapter 1**. However, evidence relating the ATL to hierarchical representations most often a) involve everyday knowledge that has been consolidated over a lifetime, and b) uses language stimuli, e.g. written words. It is possible that its involved is reduced in the novel and artificial stimuli employed in these experiments, but this remains a topic for future research.

3. Limitations and future lines of research

3.1 General limitations of the research

Throughout this thesis, I have mostly focused in hierarchical representations that were defined in terms of inclusion (i.e. "A is a member of B"). This is the meaning of hierarchical representations most commonly used for semantic categorisation (e.g. animal taxonomy) and for spatial representations where a hierarchical representation has the benefit of compression. However, other definitions are possible.

For instance, psychologists studying the representation of social hierarchies tend to focus on the notion of a social rank (i.e. an order or transitivity; e.g. person A is more

influential than person B) (Kumaran, Banino, Blundell, Hassabis, & Dayan, 2016; Kumaran, Melo, & Duzel, 2012). In this case, the hierarchical (transitive) structure is wholly provided by the supervisory signal (similar to **Chapter 5**) and does not relate to the observable features associated with each person.

Similarly, hierarchical relationships in the context of natural language are more flexible and dynamic than has been considered in this thesis. In natural language, the same observation (word) can be related to multiple hierarchical relationships and to semantic associations depending on the temporal context (sentence). For example, the phrase "a black taxi driver" could be parsed in two ways: "[black[taxi driver]]" or "[[black taxi] driver]" (Dehaene et al., 2015). Natural language also allows for an open-ended number of hierarchical relationships to be constructed and inferred from a single observation (e.g. hearing or reading a sentence). In the research I have described in this thesis, in contrast, a limited number of unambiguous hierarchical structures were defined across trials (as a function of observed features or of temporal structure).

3.2 Future lines of research

The research described in this thesis leaves new questions open, that could be addressed in future experiments.

One interesting question involves linking better the evidence for hierarchical relationships (traditionally derived from measures of reaction time), with more modern approaches. This was attempted by Murphy et al. (G. L. Murphy et al., 2012) where they aimed to replicate the original results from (Collins & Quillian, 1969) but they failed to find such evidence when training participants in short timescales using

balanced, artificial stimuli. Whether signatures of hierarchical representations in reaction times would arise after week- or month-long training with a hierarchical structure is an open question. Additionally, it would be interesting to measure the temporal dynamics of learning during these tasks (similar to the research question in **Experiment 2.1**) and track the associated neural representational changes. For instance, it would be informative if reaction times would first decrease with training for features or categories associated with higher levels of the hierarchical structure, because this would provide a link between the findings relating to reaction times in adults and to the learning dynamics (accuracy) during development.

The main result presented in **Chapter 3**, namely that a region within the parietal cortex tracks the hierarchical level, implies an interesting question: what would it reflect in more complex hierarchical structures (i.e. transitive structures) where the hierarchical level cannot be naturally defined, unlike nested tree structures? Can it represent a sense of transitivity that goes beyond one-dimensional representations (e.g. magnitude)? The fact that the parietal lobe underlies spatial reasoning in two- and three-dimensional spaces (Duhamel et al., 1992; Rosenberg, Cowan, & Angelaki, 2013) is a pointer that this may be the case, but new research is required in order to shed light into these questions.

Other open questions pertain to research in hierarchical representations for goal-directed behaviour (or more generally for multi-step sequential decisions). For instance, the experiment in **Chapter 4** was not optimised to investigate the costs relating to plan formation, which presumably happened at least in part before participants made their first step on each journey. Understanding how plans are

formed is an open question actively researched by psychologists and neuroscientists. Finding efficient computational mechanisms for planning is also an ongoing topic in Machine Learning. It has been shown that humans may rely on approximation heuristics that alleviate the cost of planning (Huys et al., 2012; Huys, Lally, et al., 2015). It thus would be informative to understand how hierarchical representations can enhance other planning heuristics, such as tree pruning.

Another interesting question is how the recursion (transitivity) of hierarchical representations can affect the costs of planning. The experiment described in **Chapter 4** only explored one level of grouping (from stations to lines). It would be interesting to know how the hierarchical distances affect when the environment is represented at multiple levels of abstraction simultaneously.

Additionally, it would be interesting to investigate in what space are hierarchical representations for planning grounded. As I have explained in the *General introduction* in **Chapter 1**, evidence for hierarchical representations is present in space, time, and for motor programs. In the experiment in **Chapter 4**, participants' button responses were a priori fixed with the direction of movement. It is unknown to which extent participants' hierarchical planning occurs in the state-space (e.g. subway stations), at the level of motor actions (e.g. at the level of button presses), or whether they can flexibly and efficiently switch between these.

Finally, I think that there is an interesting opportunity for linking hierarchical representations in naturally hierarchical structures, e.g. between in semantic knowledge such as animal taxonomies, and hierarchical planning. One such computational approach would be to show that a deep neural network that is trained in order to navigate can also show hierarchical effects of over-generalisation during

early training. For example, a multi-room environment could be designed where most journeys between rooms involve going through a corridor— except for one position, where a shortcut may be available. One would expect that during early training, the network (agent) would preferentially navigate to the corridor instead of exploiting the shortcut. This should be analogous to a neural network that during early training believes that a pine-tree have leaves.

In line with this, more work is also required in terms of understand how humans and other animals actually build spatial representations of novel (abstract or physical, possibly hierarchical) environments in order to efficiently navigate in them, and what interesting behaviours emerge during early exposure to such environments.

References

Ahlheim, C., Schiffer, A.-M., & Schubotz, R. I. (2016). Prefrontal cortex activation reflects efficient exploitation of higher-order statistical structure. *Journal of Cognitive Neuroscience*, *28*(12), 1909–1922.

Ahlheim, C., Stadler, W., & Schubotz, R. I. (2014). Dissociating dynamic probability and predictability in observed actions—an fMRI study. *Frontiers in Human Neuroscience*, *8*, 273.

Alexander, W. H., & Brown, J. W. (2015). Hierarchical error representation: a computational model of anterior cingulate and dorsolateral prefrontal cortex. *Neural Computation*, *27*(11), 2354–2410.

Alexander, W. H., & Brown, J. W. (2017). The Role of the Anterior Cingulate Cortex in Prediction Error and Signaling Surprise. *Topics in Cognitive Science*.

Anderson, J. R. (2000). *Learning and memory: An integrated approach*. John Wiley & Sons Inc.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 33.

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Ashby, F. G., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2003). Category

learning deficits in Parkinson's disease. *Neuropsychology*, *17*(1), 115.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154.

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, *66*(2), 315–326.

Banino, A., Barry, C., Uria, B., Blundell, C., Lillicrap, T., Mirowski, P., … Kumaran, D. (2018). Vector-based navigation using grid-like representations in artificial agents. *Nature*, *557*(7705), 429–433. https://doi.org/10.1038/s41586-018-0102-6

Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: a means to index neural representations using BOLD? *Phil. Trans. R. Soc. B*, *371*(1705), 20150355.

Barto, A., & Mahadevan, S. (2003). Recent Advances in Hierarchical Reinforcement Learning. *Discrete Event Dynamic Systems*, 1–28. Retrieved from http://www.springerlink.com/index/TL1N705W7Q452066.pdf

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141.

Bonet, B., & Geffner, H. (2001). Planning as heuristic search. *Artificial Intelligence*, *129*(1–2), 5–33.

Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience*, *35*(7), 1011–1023.

Bornstein, A. M., & Daw, N. D. (2013). Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLoS Computational Biology*, *9*(12), e1003387.

Botvinick. (2012). Hierarchical reinforcement learning and decision making. *Current*

Opinion in Neurobiology, *22*(6), 956–962.

Botvinick, M. M., Niv, Y., & Barto, A. C. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition*, *113*(3), 262–280. https://doi.org/10.1016/j.cognition.2008.08.011

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*(6758), 179.

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences*, *16*(10), 485–488.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*(3), 323–343.

Brooks, D. N. (1978). Nonanalytic concept formation and memory for instances. *Cognition and Categorization*.

Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain*, *111*(2), 323–345.

Castañón, S. H., Bang, D., Moran, R., Ding, J., Egner, T., & Summerfield, C. (2018). Human noise blindness drives suboptimal cognitive inference. *BioRxiv*, 268045.

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. L. (2018). Precursors of logical reasoning in preverbal human infants. *Science*, *359*(6381), 1263–1266.

Chadwick, M. J., Anjum, R. S., Kumaran, D., Schacter, D. L., Spiers, H. J., & Hassabis, D. (2016). Semantic representations in the temporal pole predict false memories. *Proceedings of the National Academy of Sciences*, *113*(36), 10180–10185.

Chang, T. M. (1986). Semantic memory: Facts and models. *Psychological Bulletin*, *99*(2), 199.

Chung, F. R. K. (1997). *Spectral graph theory*. American Mathematical Soc.

Close, J., & Pothos, E. M. (2012). "Object categorization: Reversals and explanations of the basic-level advantage"(Rogers & Patterson, 2007): A simplicity account. *Quarterly Journal of Experimental Psychology*, *65*(8), 1615–1632.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in the parietal cortex. *Annual Review of Neuroscience*, *22*(1), 319–349. https://doi.org/10.1146/annurev.neuro.22.1.319

Collard, R., & Povel, D.-J. (1982). Theory of serial pattern production: Tree traversals. *Psychological Review*, *89*(6), 693.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.

Cools, A. R., van den Bercken, J. H., Horstink, M. W., Van Spaendonck, K. P., & Berger, H. J. (1984). Cognitive and motor shifting aptitude disorder in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *47*(5), 443–453.

Crump, M. J. C., McDonnell, J. V, & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS One*, *8*(3), e57410.

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25.

Davis, T., Love, B. C., & Preston, A. R. (2011). Learning the exception to the rule:

Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, *22*(2), 260–273.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704.

Dayan, P., & Hinton, G. (1993). Feudal reinforcement learning. *Advances in Neural Information Processing* …. Retrieved from http://www.cs.utoronto.ca/~hinton/absps/dh93.pdf

Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current Opinion in Neurobiology*, *18*(2), 185–196.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2–19.

Dickinson, A., & Balleine, B. (2002). The role of learning in the operation of motivational systems. *Stevens' Handbook of Experimental Psychology*.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325.

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, *22*(6), 1075–1081.

Downes, J. J., Roberts, A. C., Sahakian, B. J., Evenden, J. L., Morris, R. G., & Robbins, T. W. (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific

attentional dysfunction. *Neuropsychologia*, *27*(11–12), 1329–1343.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science (New York, NY)*, *255*(5040), 90–92.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Estes, W. K. (1986). Array models for category learning. *Cognitive Psychology*, *18*(4), 500–549. https://doi.org/10.1016/0010-0285(86)90008-3

Estes, W. K. (2008). *Classification and Cognition*. *Classification and Cognition*. https://doi.org/10.1093/acprof:oso/9780195073355.001.0001

Farashahi, S., Rowe, K., Aslami, Z., Lee, D., & Soltani, A. (2017). Feature-based learning improves adaptability without compromising precision. *Nature Communications*, *8*(1), 1768.

Fernandes, J. J. F. R., Shahnazian, D., Holroyd, C. B., & Botvinick, M. M. (2018). Subgoal-and Goal-Related Prediction Errors in Medial Prefrontal Cortex. *BioRxiv*, 245829.

Fiete, I. R., & Seung, H. S. (2007). Neural network models of birdsong production, learning, and coding. *New Encyclopedia of Neuroscience*. *Eds. L. Squire, T. Albright, F. Bloom, F. Gage, and N. Spitzer*. *Elsevier*.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science (New York, N.Y.)*, *291*(5502), 312–316. https://doi.org/10.1126/science.291.5502.312

Friedman, A., Brown, N. R., & Mcgaffey, A. P. (2002). A basis for bias in

geographical judgments. *Psychonomic Bulletin & Review*, *9*(1), 151–159.

Furl, N., Kumar, S., Alter, K., Durrant, S., Shawe-Taylor, J., & Griffiths, T. D. (2011). Neural prediction of higher-order auditory sequence statistics. *Neuroimage*, *54*(3), 2267–2277.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. J. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *Elife*, *6*.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595.

Glass, A. L., & Holyoak, K. J. (1974). Alternative conceptions of semantic theory. *Cognition*, *3*(4), 313–339.

Graziano, M. S. A., Taylor, C. S. R., & Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron*, *34*(5), 841–851.

Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underliesthe generation of neural sequences in a songbird. *Nature*, *419*(6902), 65.

Hampton, J. A. (1982). A demonstration of intransitivity in natural categories. *Cognition*, *12*(2), 151–164.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of Neuroscience*, *28*(22), 5623–5630.

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

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*(7), 933.

Heaton, R. K. (1981). *A manual for the Wisconsin card sorting test*. Western Psycological Services.

Hintzman, D. L. (1986). " Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411.

Hirtle, S. C., & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory & Cognition*, *13*(3), 208–217.

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128.

Hornsby, A. N., & Love, B. C. (2014). Improved classification of mammograms following idealized training. *Journal of Applied Research in Memory and Cognition*, *3*(2), 72–76. https://doi.org/10.1016/j.jarmac.2014.04.009

Howard, L. R., Javadi, A. H., Yu, Y., Mill, R. D., Morrison, L. C., Knight, R., … Spiers, H. J. (2014). The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. *Current Biology*, *24*(12), 1331–1340.

Huys, Q. J. M., Cruickshank, A., & Seriès, P. (2015). Reward-based learning, model-based and model-free. In *Encyclopedia of Computational Neuroscience* (pp. 2634–2641). Springer.

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLoS Computational Biology*, *8*(3),

e1002410. Retrieved from http://dx.plos.org/10.1371/journal.pcbi.1002410

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., … Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences*, 201414219. https://doi.org/10.1073/pnas.1414219112

Hyman, J. M., Holroyd, C. B., & Seamans, J. K. (2017). A novel neural prediction error found in anterior cingulate cortex ensembles. *Neuron*, *95*(2), 447–456.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., … Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *18*(13), 5026–5034. https://doi.org/10.1523/JNEUROSCI.18-13-05026.1998

Kaplan, R., & Friston, K. (2017). Planning and navigation as active inference. *BioRxiv*, 230599.

Keil, F. (1979). *Conceptual Development*. Harvard U Press.

Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. *Proceedings of the National Academy of Sciences*, *113*(45), 12868–12873.

Kimberg, D. Y., D'Esposito, M., & Farah, M. J. (1997). Frontal lobes: cognitive neuropsychological aspects. *Behavioral Neurology and Neuropsychology. New York, NY: McGraw–Hill*, 409–418.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211.

Koechlin, E., & Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science*, *318*(5850), 594–598.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*(6), 963–974.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181–1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*(6), 229–235.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *9*(12), 578–584.

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science*, *336*(6077), 95–98.

Konishi, S., Nakajima, K., Uchida, I., Kikyo, H., Kameyama, M., & Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain*, *122*(5), 981–991.

Korn, C. W., & Bach, D. R. (2018). Heuristic and optimal policy computations in the human brain during sequential decision-making. *Nature Communications*, *9*(1), 325.

Kriegeskorte, N. (2015). Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*, 417–446.

Kulkarni, T. D., Saeedi, A., Gautam, S., & Gershman, S. J. (2016). Deep successor reinforcement learning. *ArXiv Preprint ArXiv:1606.02396*.

Kumaran, D., Banino, A., Blundell, C., Hassabis, D., & Dayan, P. (2016). Computations underlying social hierarchy learning: distinct neural mechanisms for updating and representing self-relevant information. *Neuron*, *92*(5), 1135–1147.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences*, *20*(7), 512–534.

Kumaran, D., Melo, H. L., & Duzel, E. (2012). The emergence and representation of knowledge about social and nonsocial hierarchies. *Neuron*, *76*(3), 653–666. https://doi.org/10.1016/j.neuron.2012.09.035

Kuvayev, L., & Sutton, R. S. (1996). Model-based reinforcement learning with an approximate, learned model. In *in Proceedings of the Ninth Yale Workshop on Adaptive and Learning Systems*. Citeseer.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.

Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, *107*(2), 227.

Lashley, K. S. (1951). The problem of serial order in behavior.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, *81*(3), 687–699.

Lengyel, M., & Dayan, P. (2008). Hippocampal contributions to control: the third way. In *Advances in neural information processing systems* (pp. 889–896).

Linnaeus, C. (1758). *Systema naturae*.

Liu, Z., Braunlich, K., Wehe, H. S., & Seger, C. A. (2015). Neural networks supporting switching, hypothesis testing, and rule application. *Neuropsychologia*, *77*, 19–34.

Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J.

C., & Hommer, D. W. (1999). Wisconsin Card Sorting Test performance following head injury: dorsolateral fronto-striatal circuit activity predicts perseveration. *Journal of Clinical and Experimental Neuropsychology*, *21*(1), 2–16.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: a network model of category learning. *Psychological Review*, *111*(2), 309.

Mack, M. L., Love, B. C., & Preston, A. R. (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., Love, B. C., & Preston, A. R. (2017). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*.

MacKay, D. J. C. (1992). Bayesian methods for adaptive models. California Institute of Technology.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*(1), 49–70.

Mahadevan, S. (2005). Proto-value functions: Developmental reinforcement learning. In *Proceedings of the 22nd international conference on Machine learning* (pp. 553–560). ACM.

Mahadevan, S., & Maggioni, M. (2007). Proto-value functions: A Laplacian framework for learning representation and control in Markov decision processes. *Journal of Machine Learning Research*, *8*(Oct), 2169–2231.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*(4), 587.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*(1), 3–36.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, *8*(3), 291–318.

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *BioRxiv*, 225664.

McClelland, J. L., McNaughton, B. L., & O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310.

McCloskey, M., & Glucksberg, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, *11*(1), 1–37.

McNamara, T. P., Hardy, J. K., & Hirtle, S. C. (1989). Subjective hierarchies in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 211.

McNamee, D., Wolpert, D. M., & Lengyel, M. (2016). Efficient state-space modularization for planning: theory, behavioral and neural signatures. In *Advances in Neural Information Processing Systems* (pp. 4511–4519).

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207.

Meyer, D. E. (1970). On the representation and retrieval of stored semantic information. *Cognitive Psychology*, *1*(3), 242–299.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). Plans and the organization of behavior. *NY: Holt, Rinehart and Winston*.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529.

Morris, R. G., Miotto, E. C., Feigenbaum, J. D., Bullock, P., & Polkey, C. E. (1997). The effect of goal-subgoal conflict on planning ability after frontal-and temporal-lobe lesions in humans. *Neuropsychologia*, *35*(8), 1147–1157.

Murphy, G. (2004). *The big book of concepts*. MIT press.

Murphy, G. L., Hampton, J. A., & Milovanovic, G. S. (2012). Semantic memory redux: An experimental test of hierarchical category representation. *Journal of Memory and Language*, *67*(4), 521–539.

Murray, I., & Ghahramani, Z. (2005). A note on the evidence and Bayesian Occam's razor.

Newcombe, N., & Liben, L. S. (1982). Barrier effects in the cognitive maps of children and adults. *Journal of Experimental Child Psychology*, *34*(1), 46–58.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037//1089-2680.2.2.175

Niv, Y. (2009). Reinforcement learning in the brain. *Journal of Mathematical Psychology*, *53*(3), 139–154.

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, *12*(7), 265–272.

Nobre, A. C., Coull, J. T., Frith, C. D., & Mesulam, M. M. (1999). Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. *Nature Neuroscience*, *2*(1), 11.

Norman, D. A., & Shallice, T. (1986). Attention to action. In *Consciousness and self-*

regulation (pp. 1–18). Springer.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, *14*(6), 769–776.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–337. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12718865

O'Reilly, J. X., Jbabdi, S., Rushworth, M. F. S., & Behrens, T. E. J. (2013). Brain systems for probabilistic and dynamic prediction: computational specificity and integration. *PLoS Biology*, *11*(9), e1001662.

O'reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cerebral Cortex*, *12*(3), 246–257.

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*(10), 1021–1034.

Pajani, A., Kouider, S., Roux, P., & De Gardelle, V. (2017). Unsuppressible repetition suppression and exemplar-specific expectation suppression in the fusiform face area. *Scientific Reports*, *7*(1), 160.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Statistical learning effects in musicians and non-musicians: an MEG study. *Neuropsychologia*, *50*(2), 341–349.

Pauen, S. (2002). Evidence for knowledge–based category discrimination in infancy. *Child Development*, *73*(4), 1016–1033.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

Pesaran, B., Pezaris, J. S., Sahani, M., Mitra, P. P., & Andersen, R. A. (2002). Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nature Neuroscience*, *5*(8), 805–811. https://doi.org/10.1038/nn890

Petersson, K.-M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*(2), 83–95.

Plunkett, K., & Sinha, C. (1992). Connectionism and developmental theory. *British Journal of Developmental Psychology*, *10*(3), 209–254.

Ponsen, M., Taylor, M. E., & Tuyls, K. (2009). Abstraction and generalization in reinforcement learning: A summary and framework. In *International Workshop on Adaptive and Learning Agents* (pp. 1–32). Springer.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *4*(7), 1863–1874. https://doi.org/10.1523/JNEUROSCI.04-07-01863.1984

Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: A connectionist analysis. *Journal of Experimental Child Psychology*, *66*(2), 236–263.

Randall, R. A. (1976). How Tall Is a Taxonomic Tree? Some Evidence for Dwarfism. *American Ethnologist*, 543–553.

Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J.,

Cunningham, J. M., … Binder, J. R. (1997). Functional MRI evidence for subcortical participation in conceptual reasoning skills. *Neuroreport*, *8*(8), 1987–1993.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*(6), 481.

Ribas-Fernandes, J. J. F., Solway, A., Diuk, C., McGuire, J. T., Barto, A. G., Niv, Y., & Botvinick, M. M. (2011). A neural signature of hierarchical reinforcement learning. *Neuron*, *71*(2), 370–379.

Ritter, S., Barrett, D. G. T., Santoro, A., & Botvinick, M. M. (2017). Cognitive psychology for deep neural networks: A shape bias case study. *ArXiv Preprint ArXiv:1706.08606*.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience*, *12*(1), 142–162.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Rokach, L., & Maimon, O. (2005). Clustering methods. In *Data mining and knowledge discovery handbook* (pp. 321–352). Springer.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In *Cognitive development and acquisition of language* (pp. 111–144). Elsevier.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976).

Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Rosenberg, A., Cowan, N. J., & Angelaki, D. E. (2013). The visual representation of 3D object orientation in parietal cortex. *Journal of Neuroscience*, *33*(49), 19352–19361.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533.

Rumelhart, D. E., McClelland, J. L., & Group, P. D. P. R. (1987). *Parallel distributed processing* (Vol. 1). MIT press Cambridge, MA.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*, 3–30.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLOS Computational Biology*, *13*(9), e1005768.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saxe, A. M. (2013). Precis of deep linear neural networks: a theory of learning in the brain and mind.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. *ArXiv Preprint ArXiv:1312.6120*.

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *362*(1481), 773–786.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, *76*(4), 677–694.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747.

Schapiro, A. C., McDevitt, E. A., Chen, L., Norman, K. A., Mednick, S. C., & Rogers, T. T. (2017). Sleep Benefits Memory for Semantic Category Structure While Preserving Exemplar-Specific Information. *Scientific Reports*, *7*(1), 14869.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8.

Schneider, D. W., & Logan, G. D. (2006). Hierarchical control of cognitive processes: switching tasks in sequences. *Journal of Experimental Psychology: General*, *135*(4), 623.

Schoenbaum, G., Roesch, M. R., Stalnaker, T. A., & Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews Neuroscience*, *10*(12), 885.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Seger, C. A., & Miller, E. K. (2010). Category Learning in the Brain. *Annual Review of Neuroscience*, *33*(1), 203–219. https://doi.org/10.1146/annurev.neuro.051508.135546

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936.

Shallice, T. (1982). Specific impairments of planning. *Phil. Trans. R. Soc. Lond. B*, *298*(1089), 199–209.

Shallice, T., & Burgess, P. (1991). Deficit in strategy application following frontal lobe damage in man. *Brain*, *114*(2), 727–741.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Shima, K., Isoda, M., Mushiake, H., & Tanji, J. (2007). Categorization of behavioural sequences in the prefrontal cortex. *Nature*, *445*(7125), 315–318. https://doi.org/10.1038/nature05470

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., … Lanctot, M. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., … Bolton, A. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354.

Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management to reward-based decision making: actors and critics in primate medial frontal cortex. *Neuroscience & Biobehavioral Reviews*, *46*, 44–57.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214.

Smith, J. D., Minda, J. P., & Washburn, D. A. (2004). Category learning in rhesus monkeys: a study of the Shepard, Hovland, and Jenkins (1961) tasks. *Journal of Experimental Psychology: General*, *133*(3), 398.

Solway, A., & Botvinick, M. M. (2015). Evidence integration in model-based tree search. *Proceedings of the National Academy of Sciences*, *112*(37), 11708–11713.

Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., & Botvinick, M. M. (2014). Optimal behavioral hierarchy. *PLoS Computational Biology*, *10*(8), e1003779.

Spielman, D. A. (2009). "Lecture notes on spectral graph theory.

Spitzer, B., Waschke, L., & Summerfield, C. (2017). Selective overweighting of larger magnitudes during noisy numerical comparison. *Nature Human Behaviour*, *1*(8), 145.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2014). Design Principles of the Hippocampal Cognitive Map. *Advances in Neural Information Processing Systems* *27*, 1–9. Retrieved from http://web.mit.edu/sjgershm/www/Stachenfeld14.pdf%5Cnhttp://papers.nips.cc/paper/5340-design-principles-of-the-hippocampal-cognitive-map%5Cnhttp://web.mit.edu/sjgershm/www/Stachenfeld14.pdf%5Cnhttp://papers.nips.cc/paper/5340-design-principles-of-the-hipp

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology*, *10*(4), 422–437.

Summerfield, C., & Koechlin, E. (2009). Decision-making and prefrontal executive function. *The Cognitive Neurosciences*, *4*, 1019–1030.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1). MIT press Cambridge.

Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*(1–2), 181–211.

Tanji, J., & Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. *Nature*, *371*(6496), 413.

Teichmann, A. L., Grootswagers, T., Carlson, T., & Rich, A. N. (2018). Decoding digits and dice with Magnetoencephalography: Evidence for a shared representation of magnitude. *Journal of Cognitive Neuroscience*, (Early Access), 1–12.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Human neural learning depends on reward prediction errors in the blocking paradigm. *Journal of Neurophysiology*, *95*(1), 301–310.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189.

Tsetsos, K., Wyart, V., Shorkey, S. P., & Summerfield, C. (2014). Neural mechanisms of economic commitment in the human medial prefrontal cortex. *ELife*, *3*.

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(36), 14513–14518.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural

evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience*, *30*(33), 11177–11187.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, *13*(3), 407–433.

Unterrainer, J. M., & Owen, A. M. (2006). Planning and problem solving: from neuropsychology to functional neuroimaging. *Journal of Physiology, Paris*, *99*(4–6), 308–317. https://doi.org/10.1016/j.jphysparis.2006.03.014

van den Heuvel, O. A., Groenewegen, H. J., Barkhof, F., Lazeron, R. H. C., van Dyck, R., & Veltman, D. J. (2003). Frontostriatal system in planning complexity: a parametric functional magnetic resonance version of Tower of London task. *Neuroimage*, *18*(2), 367–374.

Viard, A., Doeller, C. F., Hartley, T., Bird, C. M., & Burgess, N. (2011). Anterior hippocampus and goal-directed spatial decision making. *Journal of Neuroscience*, *31*(12), 4613–4621.

Volz, H.-P., Gaser, C., Häger, F., Rzanny, R., Mentzel, H.-J., Kreitschmann-Andermahr, I., … Sauer, H. (1997). Brain activation during cognitive stimulation with the Wisconsin Card Sorting Test—a functional MRI study on healthy volunteers and schizophrenics. *Psychiatry Research: Neuroimaging*, *75*(3), 145–157.

Wagner, G., Koch, K., Reichenbach, J. R., Sauer, H., & Schlösser, R. G. M. (2006). The special involvement of the rostrolateral prefrontal cortex in planning abilities: an event-related fMRI study with the Tower of London paradigm.

Neuropsychologia, *44*(12), 2337–2347.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., … Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *Nature Neuroscience*, *21*(6), 860–868. https://doi.org/10.1038/s41593-018-0147-8

Ward, G., & Allport, A. (1997). Planning and problem solving using the five disc Tower of London task. *The Quarterly Journal of Experimental Psychology Section A*, *50*(1), 49–78.

Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology*, *27*(4), 635–657.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–853.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology*, *20*(3), 273–281.

Wiener, J. M., & Mallot, H. A. (2003). 'Fine-to-coarse' route planning and navigation in regionalized environments. *Spatial Cognition and Computation*, *3*(4), 331–358.

Wiering, M., urgen Schmidhuber, J., & Elvezia, C. (1997). HQ-Learning. *Adaptive Behavior*, *6*(2).

Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron*, *75*(3), 418–424.

Wutz, A., Loonis, R., Roy, J. E., Donoghue, J. A., & Miller, E. K. (2018). Different Levels of Category Abstraction by Different Dynamics in Different Prefrontal Areas. *Neuron*, *97*(3), 716–726.

Yamins, D. L., Hong, H., Cadieu, C., & DiCarlo, J. J. (2013). Hierarchical modular

optimization of convolutional networks achieves representations similar to macaque IT and human ventral stream. In *Advances in neural information processing systems* (pp. 3093–3101).

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, *24*(5), 667–677.

Appendix 1. Supplementary Information for Chapter 2

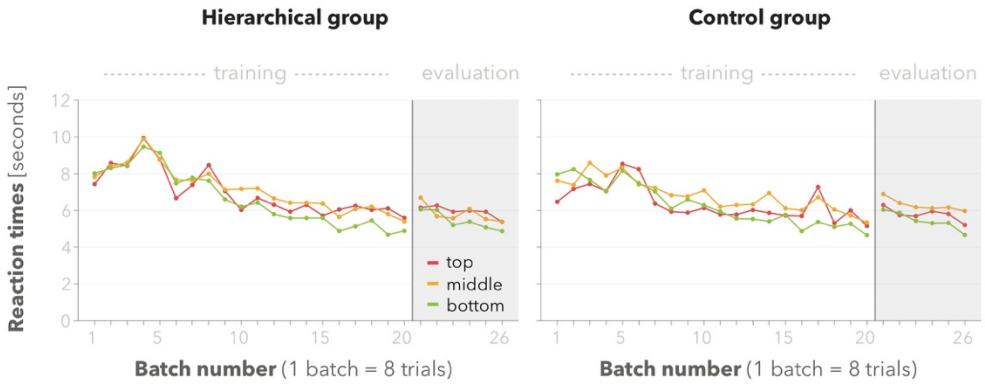

Figure SF2.1. Experiment 2.1: Reaction times
Reaction times were estimated independently for top, middle, and bottom levels of the hierarchy and for the "hierarchical" and "control" groups. On every trial, we computed the last time that a feature belonging to the relevant level was selected and calculated the temporal distance since the onset of the trial.

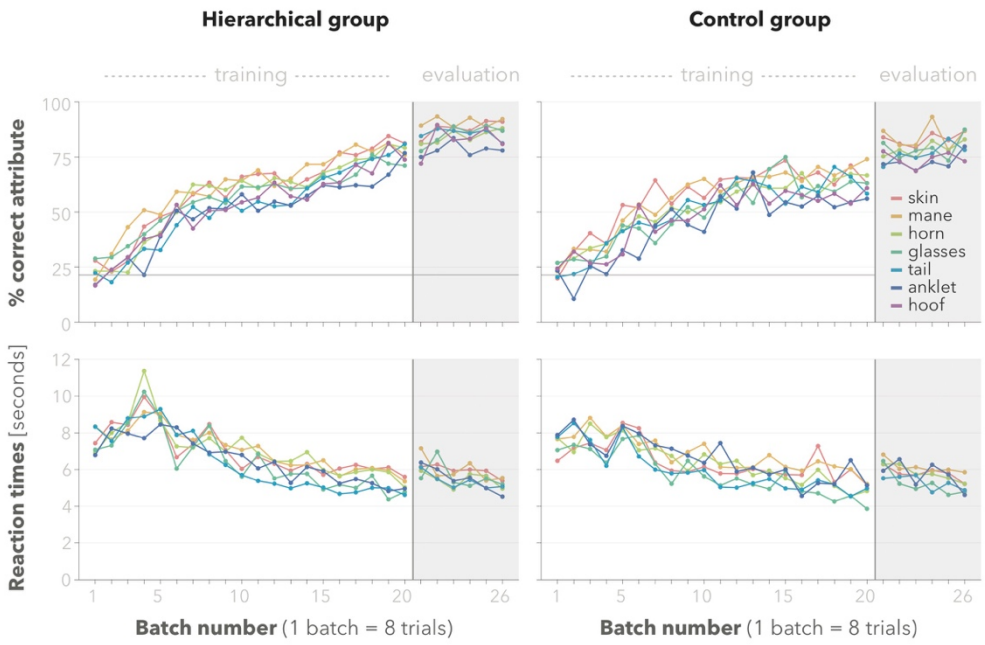

Figure SF2.2. Experiment 2.1: Learning dynamics across visual dimensions
Accuracy (% correct feature) and reaction times across trials as a function of the visual dimension (e.g. skin pattern) and independently for the "hierarchical" and "control" groups. Accuracy was computed as the probability of selecting a feature when this was relevant to the current trial. Reaction times were calculated similarly to **Fig S2.1**.

Appendix 2. Supplementary Information for Chapter 3

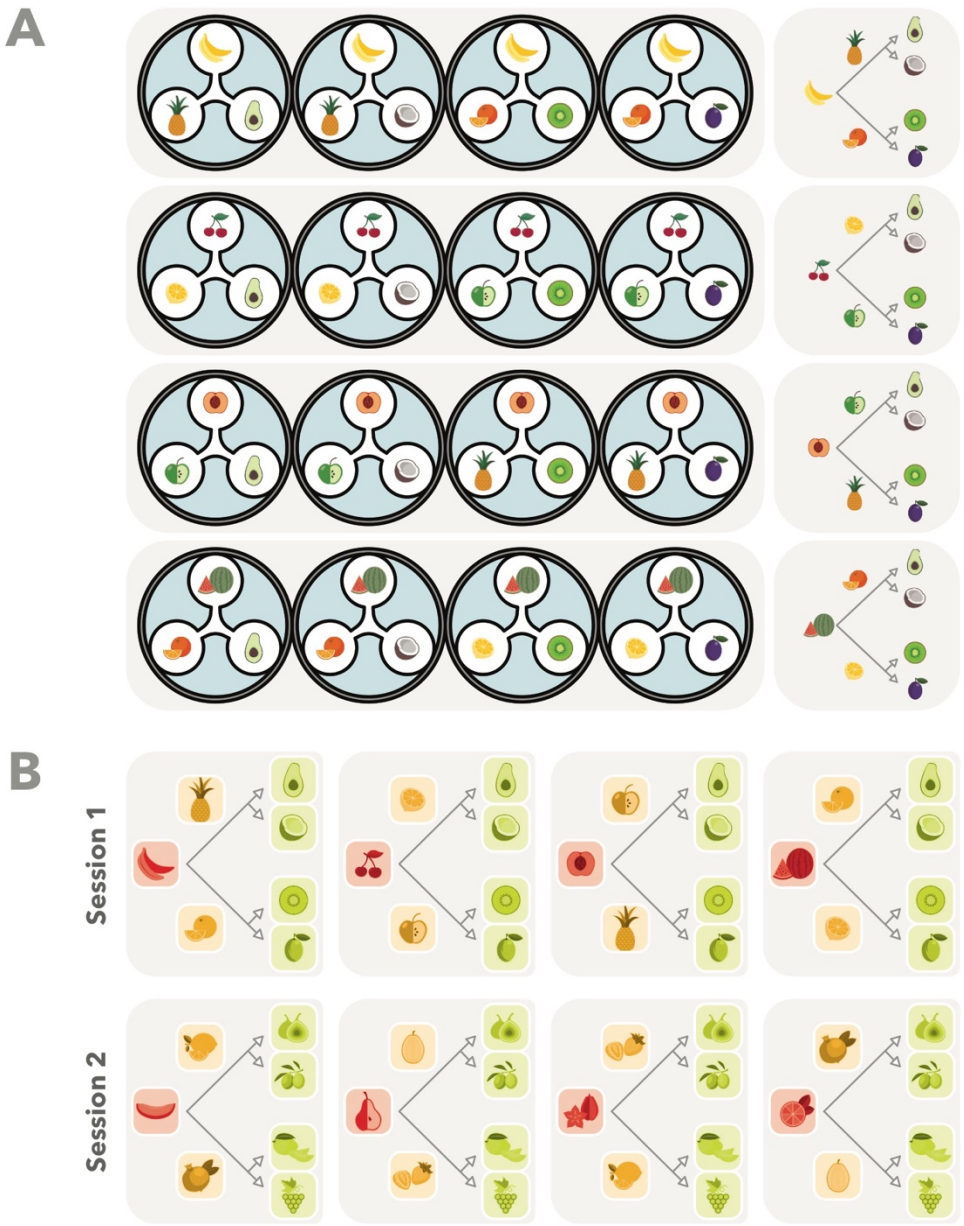

Figure SF3.1. Hierarchical structure
A. Illustration of sixteen triplets (discs on the left panels) that conform with a hierarchical relationship within a session (e.g. session one in **Fig 3.1**). The sixteen triplets are organised in four sets (rows) that could be represented as a nested tree structure (right panels). The triplets are illustrated similar to how they were displayed in the feature-completion task, except in this case are presented such that the top, left, and right placeholders correspond to the top, middle and bottom levels of the hierarchy. Each row includes the set of items from a top-block. **B.** Illustration of the structure similarity between the two learning sessions. We made use of distinct fruit stimuli, but the underlying abstract hierarchical relationship was the same.

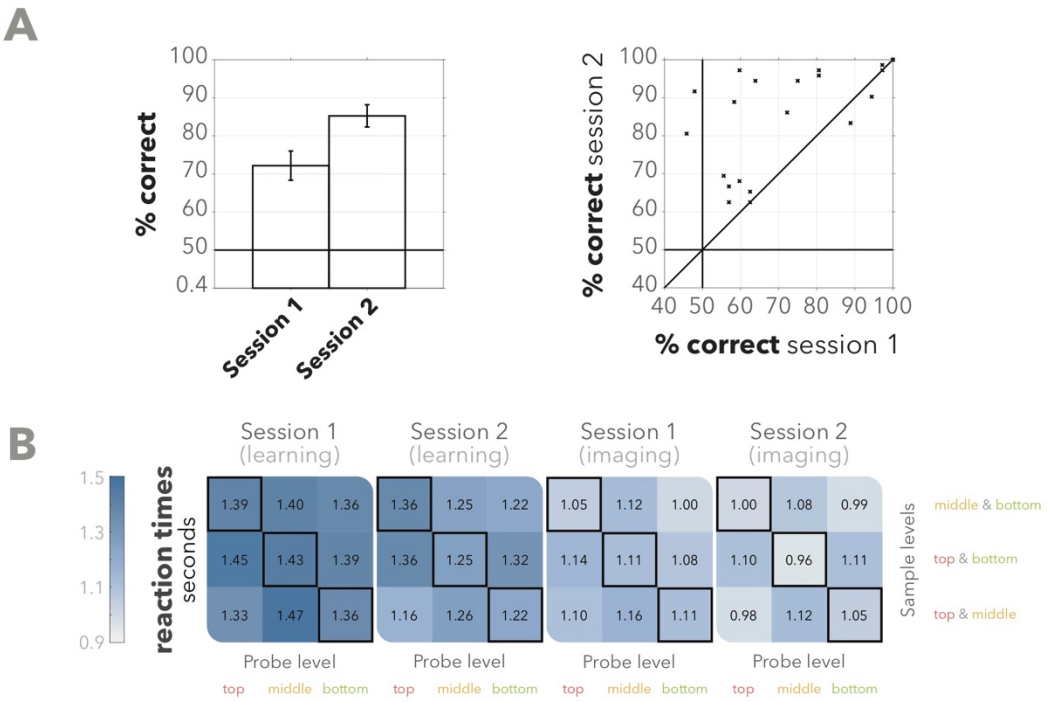

Figure SF3.2. Complementary behavioural analyses
A. Accuracy (% correct) for the triplet-validity task during the imaging session as a function of the learning session that the fruits correspond to (left panel: bars with mean ± SEM; right panel: scatter plot with participants represented by dots). **B**. Reaction times for the triplet-validity task as a function of the sample and probe hierarchical levels for the three sessions. The imaging session is shown separately for trials with fruits corresponding to the first and the second session.

Supplementary Table ST3.1. Regressors GLM3.1

This is an exhaustive list of the task regressors included in GLM3.1.

Onset	Parametric regressors
Correct trial, valid, *top* level probe	(main effect)
Correct trial, valid, *middle* level probe	(main effect)
Correct trial, valid, *bottom* level probe	(main effect)
Correct trial, invalid, *top* level probe	(main effect)
Correct trial, invalid, *middle* level probe	(main effect)
Correct trial, invalid, *bottom* level probe	(main effect)
Incorrect trial	(main effect)

Supplementary Table ST3.2. Regressors GLM3.2

This is an exhaustive list of the task regressors included in GLM3.2.

Onset	Parametric Regressors	Values
Correct trial	(main effect)	
	day2 - day1	-1 or +1
	valid - invalid	-1 or +1
	top-level probe	0 or +1
	middle-level probe	0 or +1
	(*top*-level probe) × (valid)	0 or +1
	(*middle*-level probe) × (valid)	0 or +1
Incorrect trial	(main effect)	

Supplementary Table ST3.3. Regressors GLM3.3

This is an exhaustive list of the task regressors included in GLM3.3.

Onset	Parametric Regressors	Values
Correct trial	(main effect)	
	top-level sample	0 or +1
	top-level probe	0 or +1
	top-level sample and *top*-level probe (RS)	0 or -1
	sample and probe session match	-1 or +1
Incorrect trial	(main effect)	

Supplementary Table ST3.4. Activation clusters for GLM3.1
Tables show voxel clusters (larger than 10 voxels; threshold at 0.001 uncorrected) activated by the multiple contrasts of GLM3.1 (see Methods and Results). Columns indicate (L-R): cluster corrected p-value (FDR), cluster size (Voxels), uncorrected peak t-value (Peak t), peak z-value (Peak z) and the final column gives the xyz coordinates in MNI space for each peak (Coords).

Valid > Invalid:

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0.12	13	7.19	4.94	0	-26;20;-2
0.014	40	5.85	4.37	0	-10;8;-2
0.031	27	5.68	4.29	0	10;12;-2
		4.18	3.47	0	14;16;14
0.031	24	5.62	4.26	0	-42;-48;46
0.031	23	4.72	3.79	0	-38;44;-6

Supplementary Table ST3.5. Activation clusters for GLM3.2
Tables show voxel clusters (larger than 10 voxels; threshold at 0.001 uncorrected) activated by the multiple contrasts of GLM3.2 (see Methods and Results). Columns indicate (L-R): cluster corrected p-value (FDR), cluster size (Voxels), uncorrected peak t-value (Peak t), peak z-value (Peak z) and the final column gives the xyz coordinates in MNI space for each peak (Coords).

Top level × Valid > 0:

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0.005	34	6.8	4.78	0	-26;4;54
		3.92	3.31	0	-18;-4;58
0.001	50	4.87	3.88	0	-30;-48;42
		4.8	3.84	0	-30;-64;46
		4.75	3.81	0	-18;-60;50

Appendix 3. Supplementary Information for Chapter 4

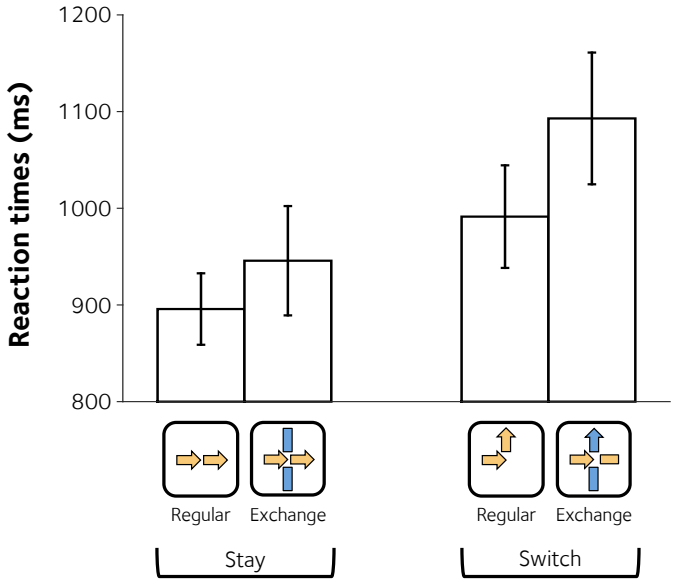

Figure SF4.1.
Reaction times (mean ± SEM) for each type of station and response. Participants showed a significant double main effect and were slower when they were required to switch their response; and in exchange stations with more than two possible responses. These two effects added linearly (interaction not significant).

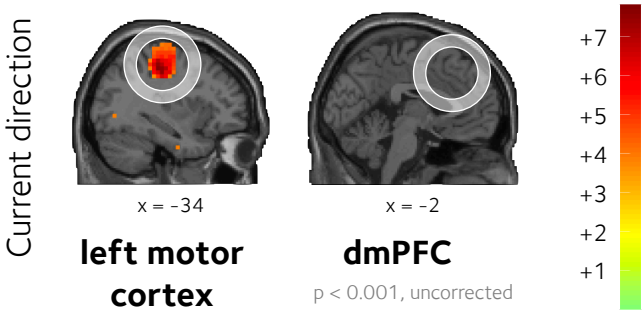

Figure SF4.2.
The results of an additional control RSA identifying voxels encoding the current direction of travel (see **Fig. 4.4**). In this task, the direction corresponds to the last response given in the current journey (but not including the one given in the current trial). The peak activation was found in left primary motor area. No significant decoding of direction was observed in the dmPFC. This RSA was performed in a very similar manner to the other ones, with the same RDM (see **Fig. 4.4a**) but comparing the multivoxel patterns associated with each direction (North, East, West, South) instead of those associated with the current line.

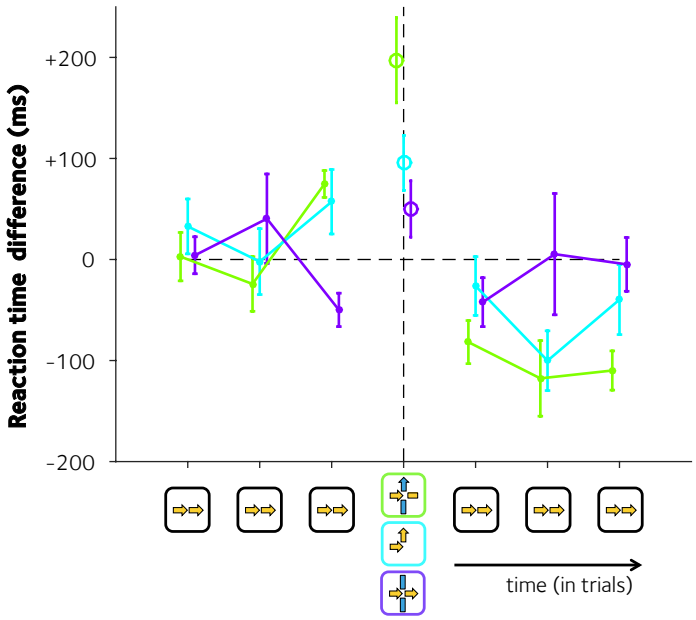

Figure SF4.3.
Reaction times (mean ± SEM; relative to average for regular stations) on regular stations (without response switch) over 3 trials that preceded (left points) or followed (right points) a context switch (green), an exchange station without line change (purple) or an elbow station.

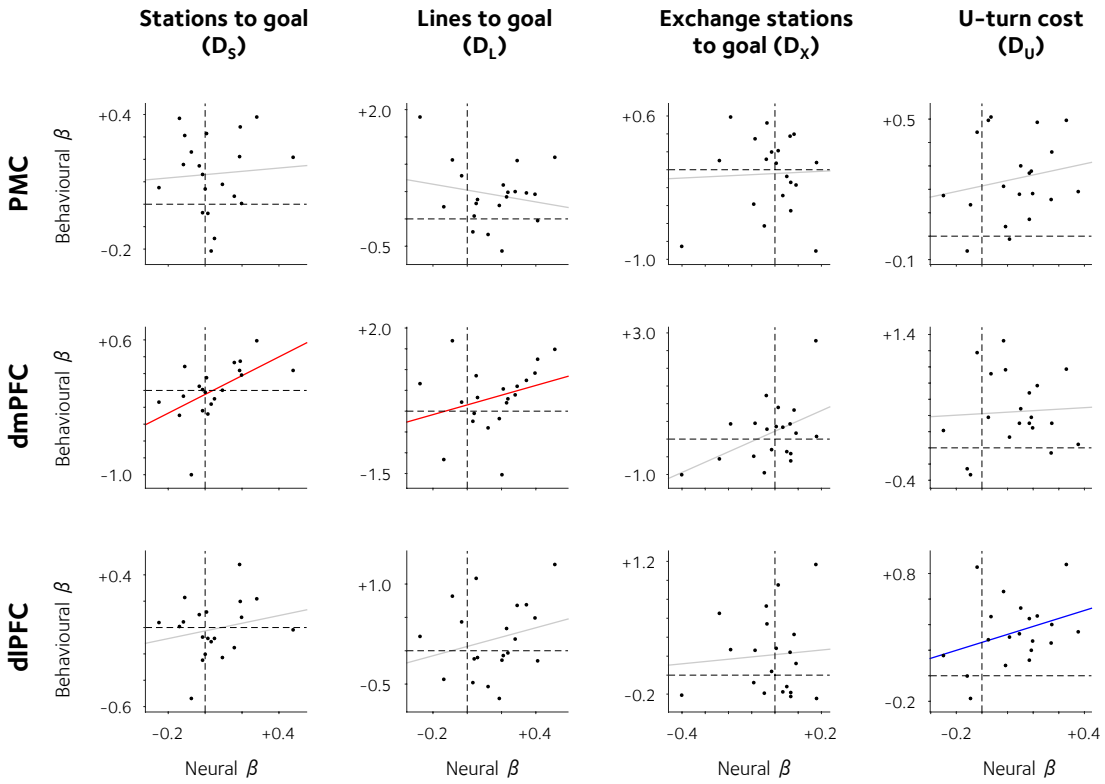

Figure SF4.4.
Between-subjects correlation between (i) the regression coefficients that best link distance measures to RT, and (ii) the regression coefficients that described the encoding of distance measures in BOLD signals recorded from the PMC, dmPFC, and dlPFC (obtained using GLM1). Best-fitting linear trends are shown for significant (red; $p < 0.05$), marginally significant (blue; $p < 0.05$ one-tailed) and not significant (grey) Spearman correlations.

Supplemental table ST4.1. Activations of GLM4.1

Tables show voxel clusters (larger than 5 voxels; threshold at 0.001 uncorrected) activated by the multiple contrasts of GLM4.1 (see Methods and Results). Columns indicate (L-R): cluster corrected p-value (FDR), cluster size (Voxels), corrected peak p-value for each cluster (Peak p), peak t-value (Peak t), peak z-value (Peak z) and the final column gives the xyz coordinates in MNI space for each peak (Coords). Positive contributions reflect clusters being more active when further away from the goal.

Number of stations to goal (D_s) - Positive contribution

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	352	0.007	9.02	5.56	-38;-36;42
		0.012	8.13	5.28	-46;-40;42
		0.051	6.67	4.73	-18;-72;46
0	85	0.051	6.62	4.71	-22;-8;50
		0.053	6.45	4.64	-18;4;54
		0.282	4.87	3.87	-26;0;66
0.001	37	0.053	6.39	4.61	30;4;58
0	69	0.065	6.1	4.49	18;-72;54
		0.118	5.55	4.23	18;-60;50
0.089	10	0.4	4.55	3.7	38;-40;50
		0.532	4.28	3.54	38;-36;42
0.022	18	0.416	4.5	3.66	42;-56;-30
		0.472	4.39	3.6	30;-60;-30
0.121	8	0.634	4.1	3.42	34;-76;34

Number of stations to goal (D_s) - Negative contribution

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	78	0.118	7.37	5.01	38;4;-18
		0.222	5.54	4.22	50;12;-26
		0.222	5.5	4.2	34;20;-22
0.005	29	0.118	7.29	4.98	-10;-48;-2
0.001	41	0.118	7.14	4.92	62;-16;6
		0.333	4.9	3.89	42;-20;10
		0.421	4.48	3.66	46;-12;10
0	112	0.118	6.75	4.76	38;-84;-14
		0.118	6.49	4.66	30;-96;-6
		0.222	5.62	4.26	42;-92;-2

0	92	0.118	6.64	4.72	-34;-92;-10
		0.166	6.01	4.44	-22;-96;-18
		0.222	5.47	4.19	-22;-100;-6
0.005	31	0.118	6.64	4.72	10;-88;42
		0.222	5.53	4.22	22;-92;30
0	124	0.118	6.55	4.68	-58;-4;-22
		0.118	6.3	4.58	-34;4;-18
		0.222	5.66	4.28	-46;20;-14
0	356	0.118	6.48	4.65	6;20;62
		0.126	6.22	4.54	10;-16;70
		0.222	5.61	4.26	46;-8;62
0.002	37	0.118	6.41	4.62	2;-12;46
		0.375	4.73	3.8	-2;-8;38
		0.421	4.46	3.64	6;-24;54
0.005	30	0.118	6.36	4.6	58;-20;-10
		0.375	4.65	3.75	58;-4;-18
		0.421	4.47	3.65	58;-4;-2
0	146	0.218	5.8	4.35	10;48;-6
		0.276	5.24	4.07	-6;44;-6
		0.31	5.12	4.01	6;60;-6
0.303	5	0.222	5.65	4.28	-10;-88;42
0.067	14	0.269	5.32	4.11	50;-36;6
		0.508	4.29	3.54	62;-36;6
0.285	6	0.324	5.01	3.95	-54;-48;6
0.303	5	0.324	5.01	3.95	14;-24;-2
0.155	9	0.375	4.69	3.78	-50;-76;-14
		0.813	3.83	3.25	-46;-76;-6
0.303	5	0.375	4.63	3.74	62;4;6
0.075	13	0.421	4.5	3.67	26;-24;58
		0.588	4.18	3.48	22;-36;58
0.106	11	0.508	4.3	3.55	-10;-96;26
0.285	6	0.508	4.29	3.55	-26;-24;-14
0.303	5	0.753	3.94	3.33	2;-84;18
		0.917	3.71	3.18	2;-80;26

Number of line changes to goal (D_L) - Positive contribution

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	98	0.262	6.58	4.69	-26;-8;54
		0.364	4.95	3.92	-26;4;46
		0.364	4.86	3.87	-18;8;66
0	90	0.316	5.75	4.32	-14;-56;18
		0.316	5.74	4.32	10;-56;18
		0.316	5.6	4.25	2;-56;22
0.001	47	0.316	5.48	4.19	-30;-44;50
0.222	5	0.343	5.32	4.11	-38;-32;-22
0.108	9	0.358	5.21	4.05	-6;8;58
0.001	41	0.364	4.99	3.94	30;4;66
		0.364	4.68	3.77	26;0;54
0.108	9	0.364	4.89	3.88	-10;-20;10
0.031	17	0.364	4.75	3.81	-14;-60;54
0.108	9	0.364	4.7	3.78	14;-68;58
0.008	26	0.394	4.55	3.7	34;-84;42
		0.398	4.45	3.64	38;-72;30
0.222	5	0.427	4.38	3.59	-34;-40;-10
0.222	5	0.577	4.11	3.43	30;-40;46

No clusters reported for negative contribution of the number of line changes to goal (D_L).

No clusters reported for positive contribution of the number of exchange stations to goal (D_X).

Number of exchange stations to goal (D_X) - Negative contribution

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0.459	5	0.439	4.79	3.83	-2;-36;-18
0.459	5	0.666	4.22	3.5	50;-44;34

U-turn cost (D_U) - Positive contribution

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	70	0.198	6.51	4.66	-26;4;54

		0.198	6.18	4.52	-18;8;58
		0.198	6.16	4.51	-30;-4;58
0.055	12	0.198	6.35	4.59	-30;24;2
0	73	0.198	6.3	4.57	26;8;46
		0.214	5.96	4.42	26;4;54
		0.219	5.54	4.22	34;-8;58
0	306	0.214	5.88	4.39	-38;-44;46
		0.219	5.67	4.28	-10;-64;54
		0.219	5.59	4.24	6;-68;54
0.002	31	0.214	5.82	4.36	14;-60;22
		0.519	4.36	3.58	10;-52;10
0	58	0.219	5.63	4.26	-2;12;46
		0.219	5.33	4.12	-6;20;42
		0.356	4.85	3.86	10;20;30
0.15	6	0.219	5.43	4.17	-58;8;30
0.168	5	0.343	4.95	3.92	14;-12;-6
0.013	20	0.348	4.91	3.9	34;-76;26
		0.899	3.68	3.16	34;-68;34
0.12	8	0.373	4.8	3.84	-42;24;26
0.08	10	0.388	4.75	3.81	30;32;30
		0.698	4	3.37	38;32;26
0.15	6	0.441	4.57	3.71	14;-4;14
		0.737	3.93	3.32	10;-12;14
0.168	5	0.441	4.56	3.7	30;48;2
0.144	7	0.449	4.53	3.69	34;28;2
		0.883	3.71	3.17	38;20;-2
0.168	5	0.519	4.34	3.57	-42;48;2
0.04	14	0.519	4.3	3.55	-6;-48;2
		0.557	4.25	3.52	-10;-60;10
0.15	6	0.567	4.2	3.49	34;-56;-30

U-turn cost (D_u) - Negative contribution

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0.215	10	0.606	5.53	4.22	54;-60;38
0.163	16	0.606	5.21	4.06	18;52;46

		0.606	4.95	3.92	6;52;46
		0.606	4.91	3.9	18;40;54
0.215	10	0.877	4.59	3.72	-6;-44;30
0.367	5	0.922	4.37	3.59	-62;-60;14
0.367	5	0.922	4.22	3.5	-54;-52;2
0.216	8	0.922	4.12	3.44	-46;-56;26
0.216	8	0.922	3.88	3.29	-10;48;6
		0.922	3.86	3.28	-6;60;10

Supplemental table ST4.2. Activations of GLM4.2

Tables show voxel clusters (larger than 5 voxels; threshold at 0.001 uncorrected) activated by the multiple contrasts of GLM2 (see Methods and Results). Columns indicate (L-R): cluster corrected p-value (FDR), cluster size (Voxels), corrected peak p-value for each cluster (Peak p), peak t-value (Peak t), peak z-value (Peak z) and the final column gives the xyz coordinates in MNI space for each peak (Coords).

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	221	0	12.37	6.4	-26;-4;54
		0	10.99	6.09	-22;0;62
		0.023	6.36	4.6	-6;12;46
0	1186	0	10.44	5.95	14;-64;62
		0.001	9.82	5.79	22;-64;58
		0.001	9.68	5.75	-46;-40;42
0	148	0	10.36	5.93	26;4;54
0	67	0.003	7.87	5.19	-42;32;34
		0.056	5.69	4.3	-30;36;42
0.001	44	0.02	6.5	4.66	38;-52;-30
		0.041	5.94	4.41	30;-64;-30
		0.159	4.95	3.92	34;-44;-34
0.004	28	0.026	6.24	4.55	-34;-64;-30
		0.041	5.91	4.4	-34;-48;-34
		0.116	5.17	4.03	-38;-56;-34
0.056	12	0.059	5.63	4.27	-6;-76;-30
0.001	45	0.076	5.46	4.18	46;32;26
		0.187	4.81	3.84	42;40;34
0.042	14	0.133	5.07	3.98	-50;4;34
0.234	5	0.229	4.67	3.77	54;8;30
0.077	10	0.273	4.55	3.7	6;-76;-26

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	5745	0.001	11.27	6.16	46;-36;18
		0.001	11.05	6.1	10;56;26
		0.001	10.65	6.01	2;56;14
0	48	0.001	9.86	5.8	-50;-8;50
		0.184	4.72	3.79	-38;-16;42
0	309	0.006	7.62	5.1	22;-88;30
		0.006	7.5	5.05	-2;-92;30
		0.007	7.22	4.95	-14;-92;30
0	94	0.007	7.29	4.98	10;-44;34
		0.008	7.07	4.89	-6;-40;34
		0.01	6.84	4.8	-2;-48;30
0	100	0.018	6.35	4.6	-46;-84;-10
		0.079	5.28	4.09	-26;-96;-10
		0.079	5.28	4.09	-38;-88;-14
0.235	5	0.138	4.91	3.9	46;-44;-22

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	138	0.203	6.93	4.84	-10;44;50

		0.203	6.7	4.74	-14;28;58
		0.423	5.23	4.07	-6;56;34
0.078	12	0.203	6.74	4.76	38;-16;18
0	74	0.314	6.25	4.55	-6;48;-6
		0.423	4.98	3.94	6;52;-6
		0.439	4.75	3.81	-2;60;2
0.092	11	0.423	5.71	4.3	-38;-16;22
0.024	22	0.423	5.6	4.25	-2;-92;30
		0.468	4.63	3.74	-18;-96;30
0	82	0.423	5.55	4.23	-42;-68;50
		0.423	5.55	4.23	-46;-56;34
		0.423	5.48	4.19	-46;-68;38
0	172	0.423	5.46	4.18	6;-20;62
		0.423	5.33	4.12	18;-20;50
		0.423	5.22	4.06	18;-20;66
0.024	21	0.423	5.22	4.06	-62;-40;2
		0.432	4.78	3.83	-66;-48;2
0.024	21	0.423	5.2	4.05	-18;-48;30
0.004	36	0.423	5.15	4.03	-26;-4;-26
		0.425	4.86	3.87	-30;-16;-18
		0.452	4.7	3.78	-30;-8;2
0.01	29	0.423	5.09	3.99	-58;-16;-14
		0.506	4.53	3.69	-62;-32;-10
		0.598	4.33	3.57	-54;-8;-22
0.043	16	0.423	5.09	3.99	14;68;26
		0.425	4.9	3.89	10;60;30
		0.598	4.33	3.57	2;68;22
0.034	18	0.423	5.05	3.97	-22;0;30
		0.747	4.05	3.39	-26;-8;26
0.026	20	0.425	4.94	3.92	-58;-4;22
		0.563	4.45	3.64	-54;-12;22
		0.598	4.32	3.56	-50;-8;14
0.108	10	0.425	4.83	3.85	14;36;58
0.248	6	0.432	4.78	3.82	-38;16;50
0.078	12	0.452	4.66	3.76	-30;32;-14
		0.838	3.92	3.31	-30;40;-10
0.043	16	0.452	4.66	3.76	34;-96;-2
		0.745	4.08	3.42	30;-92;10
0.078	12	0.581	4.41	3.61	-14;-40;10
0.164	8	0.598	4.36	3.59	26;-4;-18
0.313	5	0.598	4.33	3.57	66;-16;-6
0.248	6	0.72	4.13	3.44	-2;-8;34

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	647	0.001	11.09	6.11	-2;12;46
		0.011	8.52	5.4	-26;-4;54
		0.011	8.42	5.37	-22;4;62
0	962	0.011	8.23	5.31	22;-56;18
		0.011	7.99	5.23	-22;-60;50
		0.011	7.92	5.21	-42;-36;54

0	295	0.011	7.84	5.18	42;20;-2
		0.012	7.61	5.09	54;12;22
		0.012	7.39	5.01	42;8;30
0	156	0.012	7.58	5.08	14;8;10
		0.019	6.79	4.78	14;-4;-2
		0.021	6.72	4.75	6;-24;-6
0	66	0.016	7.08	4.89	-30;24;2
		0.036	6.11	4.49	-42;16;-2
		0.065	5.61	4.26	-34;16;10
0.001	37	0.017	6.95	4.84	-14;4;10
0.233	5	0.347	4.4	3.61	30;44;30

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	730	0.001	10.76	6.04	-38;-32;46
		0.001	10.14	5.88	-38;-44;62
		0.002	9.35	5.66	-54;-16;30
0	334	0.005	8.39	5.36	54;-24;34
		0.007	8.05	5.25	58;-16;34
		0.007	7.8	5.16	46;-24;42
0	47	0.007	7.75	5.14	54;8;34
0.005	25	0.034	6.25	4.55	46;-68;-10
0	41	0.034	6.23	4.54	22;-52;-22
		0.068	5.66	4.28	22;-72;-22
0	47	0.038	6.11	4.49	-54;4;26
		0.166	5.06	3.98	-54;0;38
0.118	7	0.049	5.9	4.4	-42;-4;14
0	68	0.096	5.43	4.17	-14;-20;6
		0.122	5.27	4.09	-26;-12;2
		0.258	4.68	3.77	-22;8;6
0.1	8	0.239	4.78	3.83	-6;4;34
0.004	27	0.239	4.76	3.82	2;0;50
		0.282	4.61	3.73	2;-4;58
0.035	13	0.396	4.39	3.6	-42;-64;2
0.026	15	0.485	4.2	3.49	26;-4;62
		0.593	4.04	3.39	34;-4;58
		0.715	3.84	3.26	30;-4;50
0.19	5	0.63	3.95	3.33	-18;-72;-22

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	239	0.018	8.64	5.44	34;16;50
		0.018	8.53	5.41	22;24;50
		0.026	7.8	5.16	2;32;34
0	89	0.073	6.99	4.86	50;-68;46
		0.13	6.08	4.48	46;-60;54
		0.386	4.77	3.82	58;-48;42
0	111	0.121	6.46	4.64	10;-36;70
		0.123	6.24	4.55	14;-24;54
		0.17	5.71	4.31	10;-16;70
0	59	0.123	6.23	4.54	-34;8;58
		0.17	5.7	4.3	-22;20;46

		0.385	4.84	3.86	-18;20;38
0.005	32	0.13	6.08	4.48	-46;-68;46
		0.17	5.67	4.29	-38;-80;42
0.146	10	0.146	5.96	4.42	-38;-20;-10
		0.386	4.76	3.82	-42;-16;-18
0.17	9	0.199	5.54	4.23	-50;-56;50
0.005	31	0.199	5.51	4.21	-18;-92;30
0.22	7	0.244	5.36	4.13	66;-32;-2
0.067	14	0.292	5.21	4.06	42;8;-18
0.22	6	0.292	5.2	4.05	-66;-48;-2
0.007	28	0.332	5.05	3.97	-42;-4;-14
		0.346	5	3.95	-30;4;-26
		0.814	3.86	3.28	-46;0;-22
0.22	6	0.376	4.89	3.88	18;-8;-22
0.01	25	0.385	4.83	3.86	42;40;-6
		0.826	3.82	3.25	34;32;-14
0.265	5	0.386	4.77	3.82	34;-20;14
0.002	38	0.406	4.71	3.79	-34;44;10
		0.437	4.64	3.74	-26;60;-2
		0.663	4.29	3.54	-38;52;18
0.22	6	0.552	4.46	3.64	-14;-32;26
0.186	8	0.552	4.45	3.64	18;-84;30
0.22	6	0.654	4.33	3.57	-62;-56;34
		0.822	3.83	3.26	-62;-52;26
0.265	5	0.683	4.26	3.52	-22;-52;-10
0.186	8	0.704	4.16	3.47	46;0;-10
		0.89	3.73	3.19	46;-12;-2
0.22	6	0.704	4.15	3.46	-42;40;-2

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	178	0.082	7.55	5.07	-30;-92;-10
		0.134	6.44	4.63	-18;-92;-14
		0.135	6.06	4.47	-38;-80;-14
0	60	0.082	7.24	4.96	-26;8;54
0	137	0.134	6.56	4.69	26;12;54
		0.134	6.35	4.6	30;20;58
		0.164	5.86	4.38	42;4;58
0.004	32	0.135	6.12	4.49	18;-56;22
		0.246	5.36	4.13	10;-44;10
0.001	42	0.164	5.81	4.35	34;-92;-2
0.214	6	0.166	5.74	4.32	-6;-44;14
0.004	31	0.246	5.38	4.14	46;32;22
0.036	16	0.258	5.29	4.1	-14;-56;14
0.077	12	0.731	4.5	3.67	-50;48;10
		0.735	4.4	3.61	-50;52;2
0.111	10	0.731	4.48	3.66	-42;24;30
0.004	29	0.735	4.33	3.57	38;-68;54
		0.735	4.32	3.56	38;-60;46
		0.824	4.11	3.44	38;-72;42
0.033	17	0.735	4.28	3.54	30;24;2

0.214	6	0.735	4.23	3.51	-10;-76;50
		0.979	3.67	3.15	2;-76;50
0.214	6	0.735	4.23	3.51	-2;-48;-2
0.154	8	0.824	4.09	3.42	6;16;46
0.129	9	0.836	4	3.36	-30;-76;50
0.255	5	0.836	3.96	3.34	-50;-52;-18
0.255	5	0.93	3.87	3.28	30;40;18
		0.93	3.81	3.24	34;48;22

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0	152	0.068	7.46	5.04	-38;-24;42
		0.295	5.7	4.3	-42;-20;54
		0.295	5.54	4.22	-46;-28;50
0.1	16	0.295	5.56	4.23	-2;-16;38
0.319	8	0.295	5.55	4.23	2;32;-6
0.319	7	0.446	4.92	3.9	-46;-12;6
0.319	7	0.705	4.54	3.69	54;-12;30
0.462	5	0.926	3.89	3.3	62;-44;22

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0.146	10	0.296	6.55	4.68	2;-80;30
0.022	22	0.576	5.71	4.31	-30;-12;66
0.417	5	0.598	5.2	4.05	22;4;-14
0.002	40	0.598	5.16	4.03	26;-68;-6
		0.598	4.97	3.93	22;-56;-14
		0.86	4.13	3.45	2;-68;-14
0.146	10	0.598	4.96	3.93	2;-4;70
0.022	21	0.769	4.61	3.73	-46;-24;58
		0.86	4.19	3.49	-42;-40;62
0.268	7	0.861	4.05	3.4	-2;-12;18

FDR	Voxels	Peak p	Peak t	Peak z	Coords
0.012	26	0.587	5.7	4.3	50;-60;-6
0.016	21	0.746	4.88	3.88	54;-32;42

Supplemental table ST4.3. Correlation of distances
Mean correlation across the cohort between the multiple measures of distance to goal

	D_L	D_X	D_U
D_S	0.6134	0.6958	0.4228
D_L	-	0.5065	0.4087
D_X	-	-	0.1235

Appendix 4. Supplementary Information for Chapter 5

Text SX5.1. Description of the normative Hierarchical Bayesian (HB) model

We describe here a normative account of behaviour in our task. We chose to use the framework of Bayesian inference because it provided us with a standard and methodological approach to normative inference. Our computational modelling approach (described in the main text) relied heavily on this model in order to characterize how learning deviates from normative behaviour. An illustration of the hierarchical Bayesian model can be found in **Fig 5.2**.

For simplicity's sake, we will denote the probabilities P(T=1) , P(Y=1), P(X=x), P(R=r) as P(t), P(y), P(x), P(r) respectively, where T is a Bernouilli random variable describing if any given trial is target (T=1) or nontarget (T=0); Y is a Bernouilli random variable describing the response of the model; X is a discrete random variable associated with the stimulus presented in a trial; and R is a discrete random variable associated with the rule governing the block. A subscript $P(t_n)$, $P(y_n)$, $P(x_n)$ will be used to refer to these probabilities in trial n. We will use **bold** notation $\mathbf{x_n} = (x_1 \dots x_n)$, $\mathbf{y_n} = (y_1 \dots y_n)$, and $\mathbf{t_n} = (t_1 \dots t_n)$ to refer to the history of trials from 1 to n included. The same notation will be used for conditional probabilities.

The HB model has no free parameters and is built under a few assumptions. Like humans, the model had uniform prior about the space of possible rules and the output (target or nontarget) that each rule predicts for each pair of stimuli presented during a trial. The HB model did not take into account the fact that each block had 50% target trials and 50% nontarget trials, nor the fact that rules with same relevant feature on both sides were excluded (humans were not informed of this). The model probabilities were initialised at the beginning of each block. At any point, the probability for the current trial of being target only depended on the probability of the underlying rule and the current observed stimuli. In familiar blocks, rules associated with the irrelevant dimensions had probability 0, in line with the assumption that the model knew the meaning of the symbolic cues. The HB model is optimal based on the following assumptions: (i) independence of rules across blocks; (ii) independence of the current stimulus presented from the history of previous stimuli; (iii) independence of the stimulus presented with the underlying rule; (iv) knowledge about the space of possible rules; (v) knowledge about the possible cues, as informed by the block cues; (vi) prior equiprobability across possible rules; (vii) uniform distribution across observations. We argue that these assumptions are plausible, as participants had practiced a full session of the task (48 blocks) prior to the main experiment in the scanner.

The HB model uses two layers, which is why we call it *hierarchical*. The two layers are the target layer and the rule layer. The first 'target' layer estimates the probability for a certain trial of being target conditioned on the history of stimuli and feedback on the current block: $P(t_n|\mathbf{x_n}, \mathbf{t_{n-1}})$. This is calculated as the marginal probability across all the possible rules:

$$P(t_n|\mathbf{x_n}, \mathbf{t_{n-1}}) = \sum_r P(t_n|r, x_n) \times P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) \quad \text{(eq. 6)}$$

where $\sum_r$ is the sum across all possible rules (i.e., the 6 × 6 combinations of features between left and right sides). Note that the target probability $P(t_n|r, x_n)$ was only dependent on the underlying rule and the current observation, while the probability of a certain rule $P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}})$ depended on the history of previous trials.

The second 'rule' layer estimates the probability for each rule of being the underlying one from the history of previous trials: $P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}})$. For instance, it can be shown that

$$P(r|\mathbf{x_{n-1}}, \mathbf{t_{n-1}}) \qquad \text{(eq. 7)}$$

Bayes rule

$$= \frac{P(r|\mathbf{x_{n-1}})}{P(\mathbf{t_{n-1}}|\mathbf{x_{n-1}})} \times P(\mathbf{t_{n-1}}|r, \mathbf{x_{n-1}})$$

$$= \frac{P(r|\mathbf{x_{n-1}})}{\sum_r P(\mathbf{t_{n-1}}|r, \mathbf{x_{n-1}}) \times P(r|\mathbf{x_{n-1}})} \times P(\mathbf{t_{n-1}}|r, \mathbf{x_{n-1}})$$

Rules are independent of the stimuli $\mathbf{x_{n-1}}$

$$= \frac{P(r)}{\sum_r P(\mathbf{t_{n-1}}|r, \mathbf{x_{n-1}}) \times P(r)} \times P(\mathbf{t_{n-1}}|r, \mathbf{x_{n-1}})$$

Rules are equiprobable

$$= \frac{P(\mathbf{t_{n-1}}|r, \mathbf{x_{n-1}})}{\sum_r P(\mathbf{t_{n-1}}|r, \mathbf{x_{n-1}})}$$

trial targetness only depends on the current stimulus and the rule

$$= \frac{\prod_{i=1:n-1} P(t_i|r, x_i)}{\sum_r \prod_{i=1:n-1} P(t_i|r, x_i)}$$

where we made use of additional assumptions that $\mathbf{x_n}$ and r are independent,

and x_i is independent across different trials i. Intuitively, the posterior probability will be zero if the rule is inconsistent with the history of previous trials, and otherwise equally distributed across all remaining (consistent) rules. For example, if four possible rules were consistent with the history of trials, the probability for each of these would be of 0.25, while 0 for any other rule. This model is shown to be optimal based on the previous assumptions. An illustration of the model at work is provided in **Fig 5.2**.

Unlike our human participants, we allowed the HB model to respond probabilistically, and thus the optimal policy was to provide a continuous response $P(y_n|\mathbf{x}_n,\mathbf{t}_{n-1})=P(t_n|\mathbf{x}_n,\mathbf{t}_{n-1})$ between 0 and 1 rather than greedily comparing the probability of the current trial being target against 50%.

As can be seen in **Fig 5.1b**, the model largely outperforms the % accuracy achieved by humans, by inferring the probability for each trial of being target in a Bayesian fashion (Familiar cues: HB model 90.1% against humans 77.6%; Novel cues: HB model 84.2% against humans 63.6%). The number of consistent rules explaining the history of previous trials decreases from 9 and 36 candidates (familiar and novel blocks, respectively) to 1 (or equivalently, the rule entropy decreases to zero; see **Fig SF5.6**), when the model can conclude with complete certainty what is the underlying rule for that block. This model thus allows us to estimate an upper boundary on the accuracy of responses given by either human behaviour or any other model predictions.

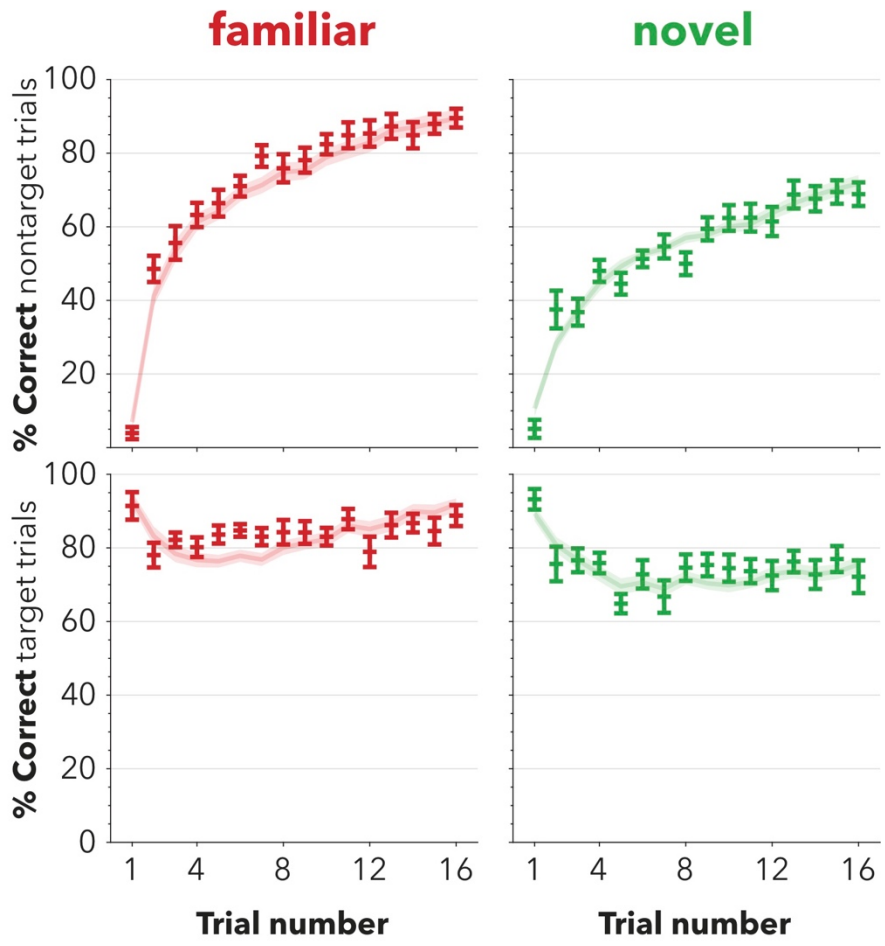

Figure SF5.1. Response and performance dynamics for target and non-target trials
Behavioural data (% accuracy) from human participants and our best-fitting model, independently for target and nontarget trials and blocks with familiar and novel cues; similar to **Fig 5.1b**.

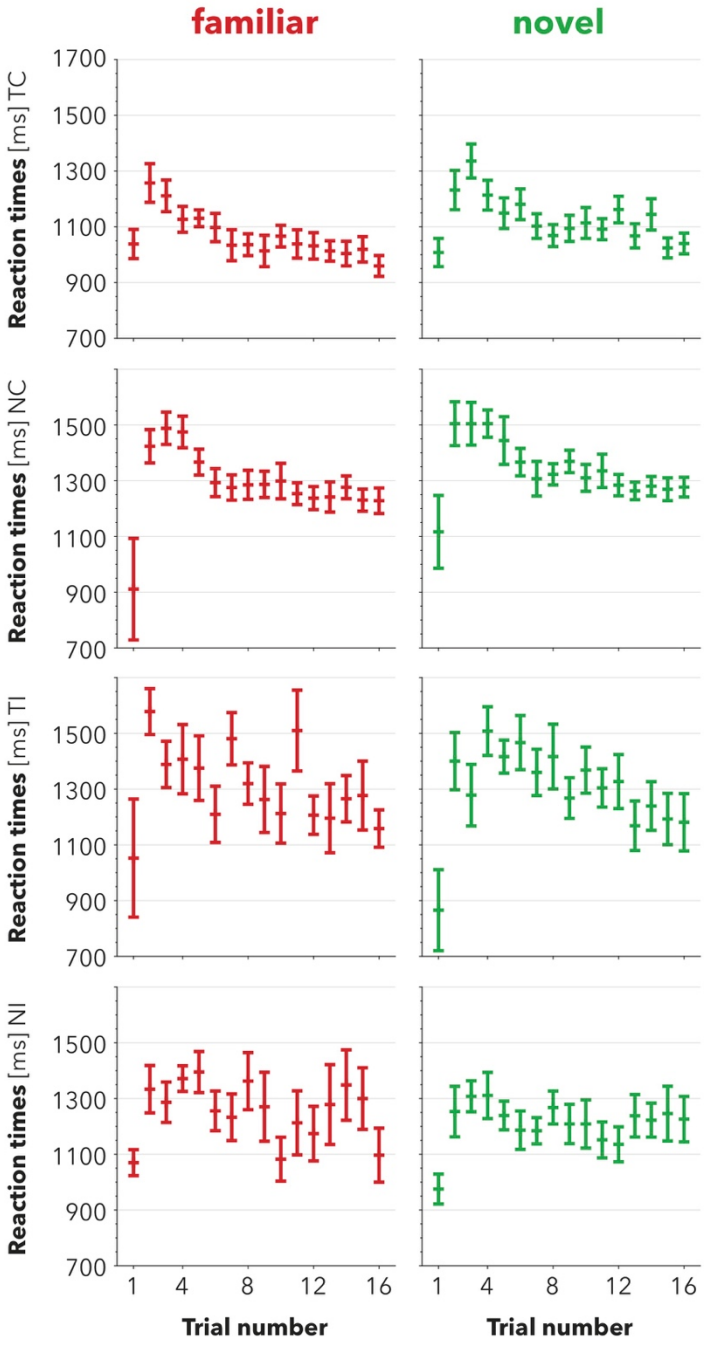

Figure SF5.2. Reaction time dynamics
Reaction times from humans across trials, independently for correct and incorrect responses, target and nontarget trials, and blocks with familiar and novel cues.

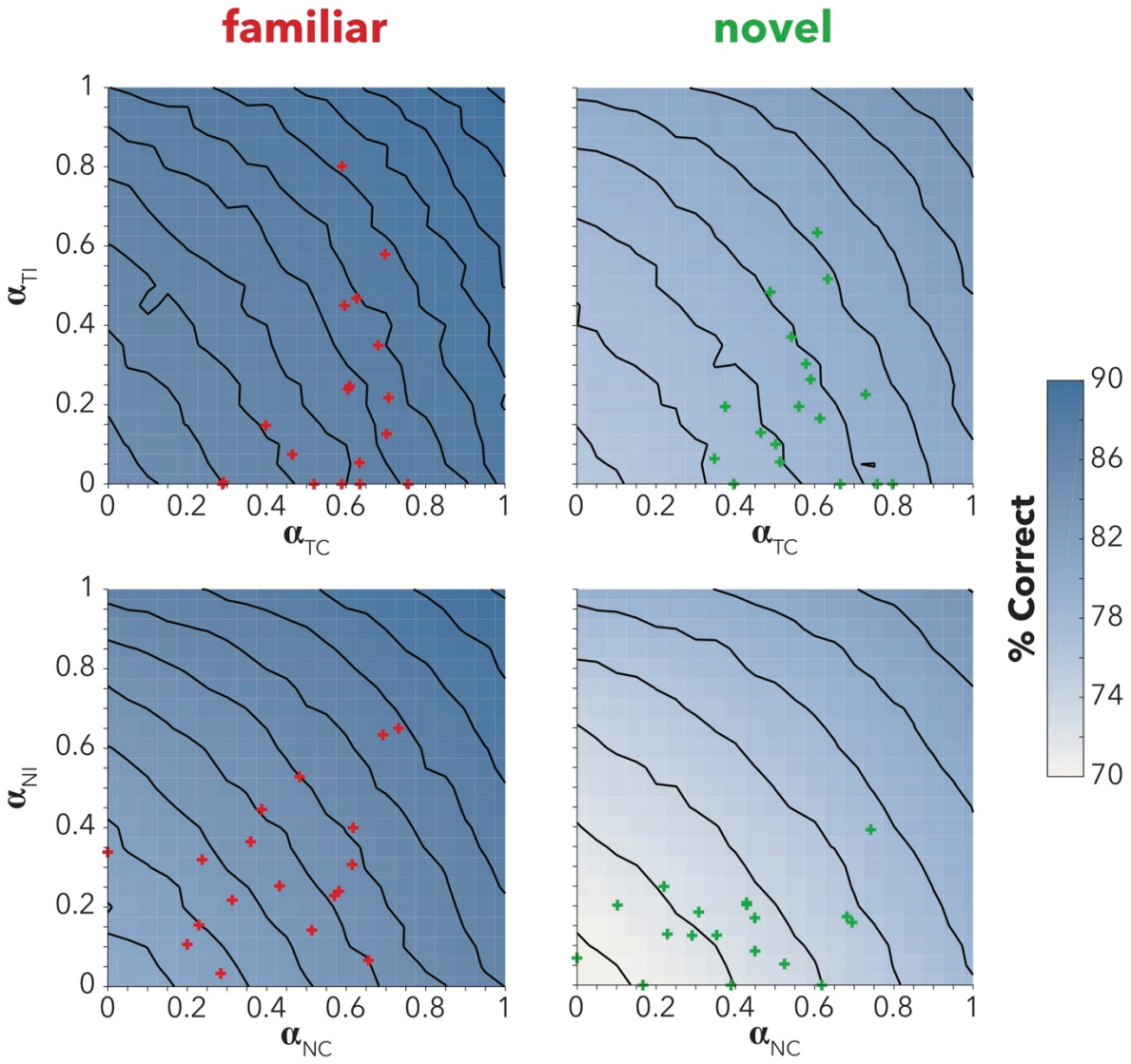

Figure SF5.3. Model performance for correct and incorrect learning rates
Model performance (% correct) as a function of the model learning rates (blue heatmap with black contour lines) on blocks with familiar cues (left panel) and novel cues (right panel) under an optimal policy. Darker blue indicates higher performance (% correct). Red and green crosses show individual humans participants in the novel and familiar cues conditions respectively. This figure complements **Fig 5.3a.** by comparing the effect of suboptimal learning rates for correct and incorrect trials, conditioned on target and nontarget trials.

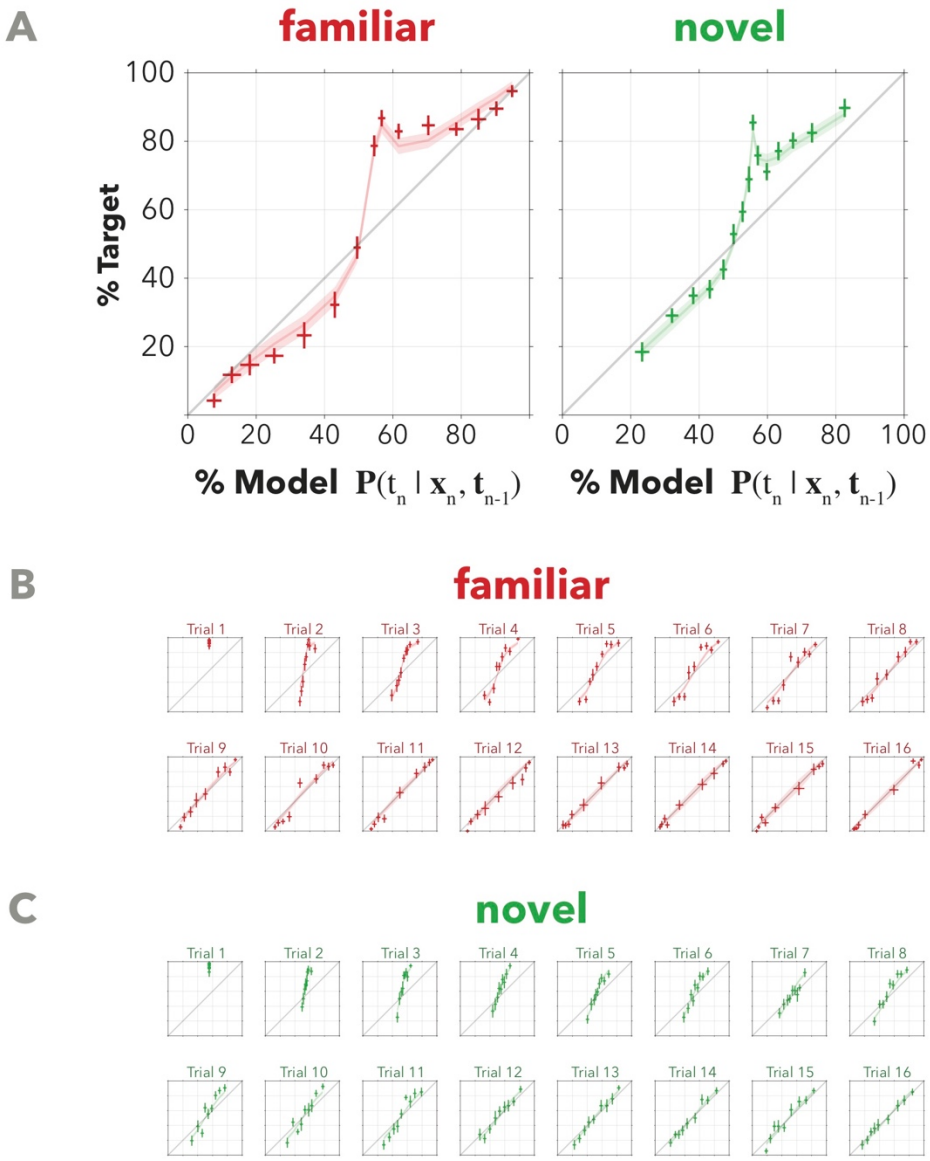

Figure SF5.4. Model policy across block trials
Psychometric curves illustrate the policy of the best-fitting model. **A.** Overall profile of responses, with proportion of target responses (y-axis) as a function of the probability of the current trial being target as estimated by the model (x-axis), for blocks with familiar (red, left) and novel cues (green, right). Trials were uniformly grouped into 15 bins. Crosses and their height/width correspond to human mean and SEM. The continuous line and its shade correspond to mean and SEM of the best-fitting model. The gray line corresponds to the optimal policy, e.g. adopted by the HB model. Note the probability of a target response assumed by the model in the first trial is 55.5% both in familiar and novel cues, corresponding to the 'bump' slightly above 50%. **B.** Psychometric curves obtained independently for each trial in blocks with familiar cues. **C.** Psychometric curves obtained independently for each trial in blocks with novel cues.

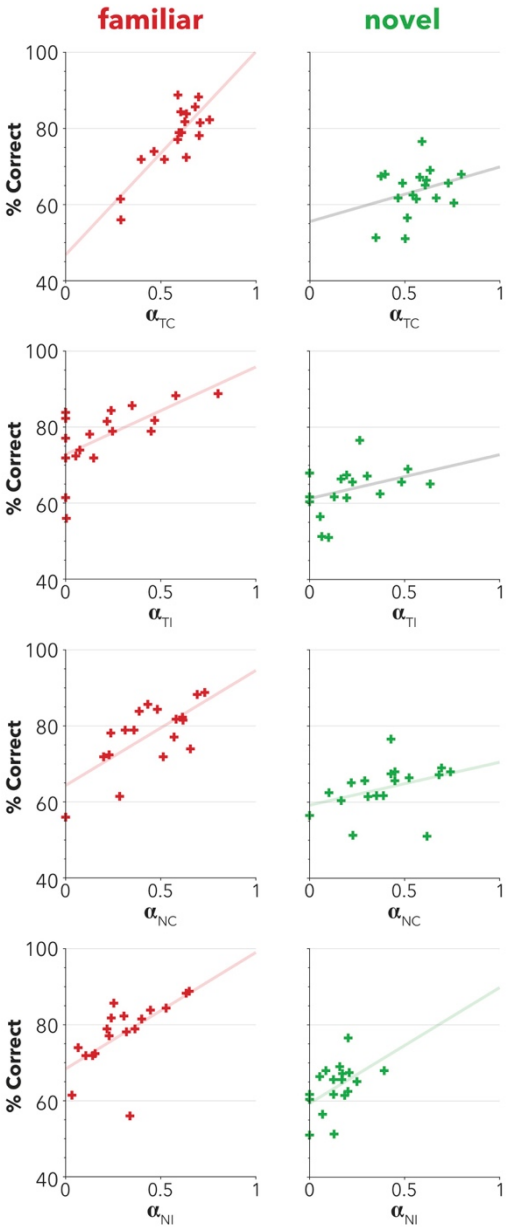

Figure SF5.5. Relationship between learning rate estimates and performance
Scatter plots describe the relationship between the four learning rate parameter estimates (x-axis) and the accuracy (%correct; y-axis) independently for blocks with familiar and novel cues. Straight correspond to the best-fitting linear approximation, coloured for significant spearman correlation and gray otherwise.

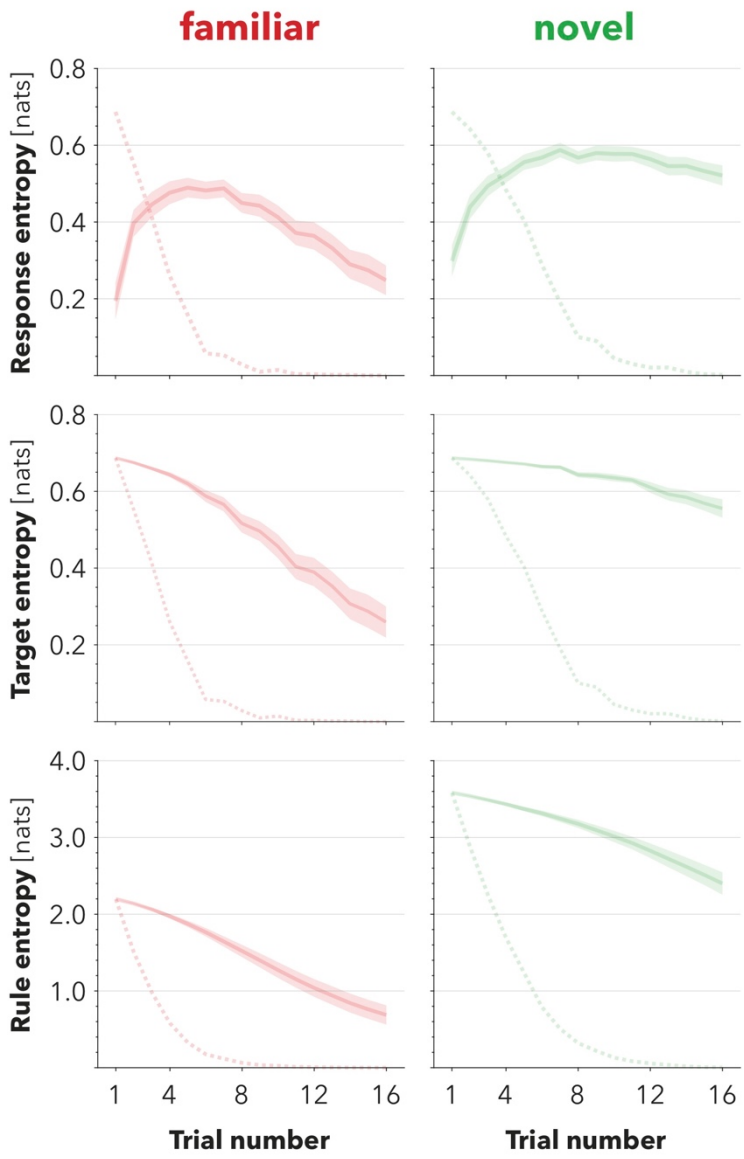

Figure SF5.6. Rule entropy across trials
A. The entropy of the response provided by our best-fitting model (continuous line with shading; mean and SEM respectively) and by the normative HB model (dashed lines) captured the uncertainty about the correct response in any given trial. **B.** The entropy of the stimulus being target as estimated by the best-fitting model (continuous lines) and the optimal HB model (dashed lines). In the case of the optimal model, this matches the entropy of the response. **C.** The entropy of the rule provided by our best-fitting model (continuous line with shading; mean and SEM respectively) and by the normative HB model (dashed lines) captured the uncertainty about the underlying rule governing the feedback at any point during the block. The rule entropy was lower for familiar blocks (log(9), corresponding to a uniform prior among 9 candidate rules) than for novel blocks (log(36), i.e. a uniform prior among all possible rules). Null entropies (meaning that the one correct rule had been discovered with total certainty) were as expected achieved faster by the HB model than the model fitted to the human data.

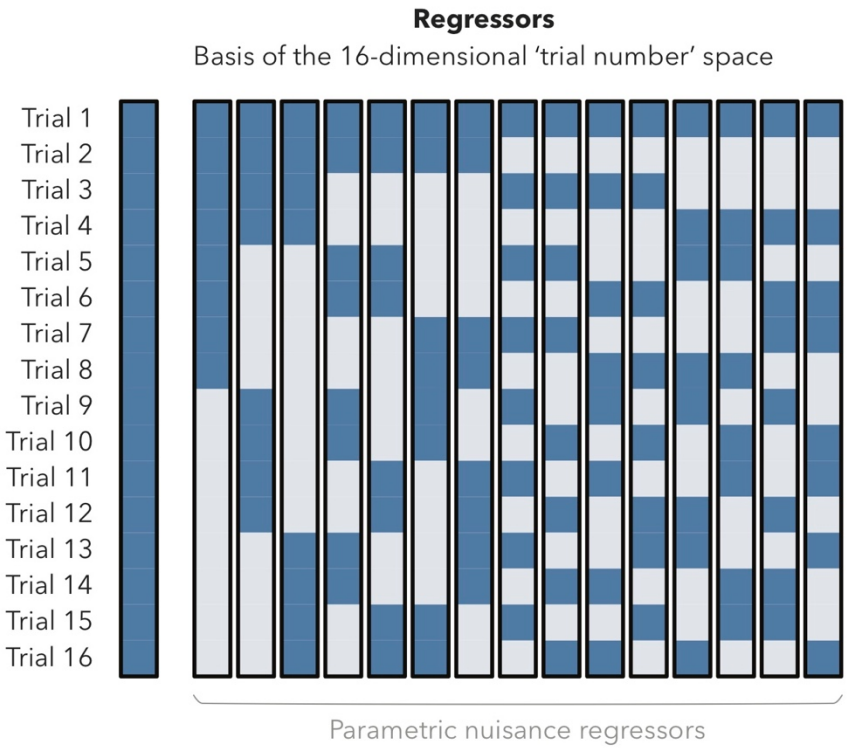

Figure SF5.7. fMRI basis set for trial number
We defined a set of regressors such that they would cover any general temporal tendency as determined by the trial number within the block. One example of such a set would be 16 regressors, with values being 1 for each given trial number and zero everywhere else. The columns shown in this figure display an alternative basis set, with values +1 and -1 for blue and white respectively, that respected some additional constraints: i) all regressors are orthogonal, ii) all regressors have same norm; iii) it includes a main effect of a trial event, i.e. a regressor with +1 everywhere. This basis set allowed us to model our GLM as a main effect of trial, and 15 parametric nuisance regressors. In GLM5.1, we applied this independently for blocks with familiar and novel cues (setting the regressors with non-zero values only for the relevant blocks) and thus had a total of 30 (15+15) nuisance regressors, plus a parametric regressor with the main effect of cue type.

Table ST5.1. Model parameter estimates

Mean and standard error of the mean (SEM) values for the best-fitting (maximum likelihood) parameters of the model separately for familiar and novel blocks. Note that gradient-descent was initialized with all parameters set to zero. The learning rates were bound to be between 0 and 1. For stability's sake, we also constrained the policy parameters to values between 0 and 100. However, policy parameters higher than 50 were only achieved for β_A in three participants, all in the familiar condition.

	Familiar cues	***Novel cues***
α_{TC}	0.58 ± 0.03	0.56 ± 0.03
α_{TI}	0.21 ± 0.05	0.21 ± 0.04
α_{NC}	0.44 ± 0.05	0.39 ± 0.05
α_{NI}	0.30 ± 0.04	0.14 ± 0.02
β_A	25.4 ± 6.4	10.8 ± 1.43
β_T	4.99 ± 3.0	0.50 ± 0.08

Table ST5.2. Goodness of fit

Our best-fitting model achieved lowest (cross-validated) leave-one-block-out negative log-likelihoods (NLL) obtained independently for blocks with familiar and novel cues, compared to two alternative models (see Methods section). The NLL scores correspond to the sum over the 16 trials of the block left out within the cross-validation fold. Lower scores are interpreted as a better fit to the data. We averaged the score across all folds within a participant. The deviation (±) corresponds to the standard error of the mean (SEM) across participants.\

	Familiar cues	**Novel cues**
Model	6.70 ± 0.43	8.84 ± 0.33
Simple model	6.83 ± 0.44	9. 02 ± 0.31
Descriptive model	8.07 ± 0.45	10.19 ± 0.25
HB model	∞	∞

Table ST5.3. Regressors for GLM5.1

This is an exhaustive list of the task regressors included in GLM5.1. See **Fig SF5.7** for more details on the parametric regressors for 'trial number basis'.

Onset	Parametric Regressor	Number of Regressors
Cue	(main effect)	1
	cue (novel – familiar)	1
Trial	rule entropy	1
	target entropy	1
	feedback (correct – incorrect)	1
	response (target - nontarget)	1
	cue (novel - familiar)	1
	trial number basis (familiar)	15
	trial number basis (novel)	15

Table ST5.4. Regressors for GLM5.2

This is an exhaustive list of the task regressors included in GLM5.2.

Onset	Parametric Regressor	Number of Regressors
Cue	(main effect)	1
	cue (novel – familiar)	1
Trial	rule entropy 'mean'	1
	rule entropy 'error'	1
	feedback (correct – incorrect)	1
	feedback (target - nontarget)	1
	response (target - nontarget)	1
	rule entropy 'mean' × feedback (correct – incorrect)	1
	rule entropy 'mean' × feedback (target - nontarget)	1
	rule entropy 'mean' × response (target - nontarget)	1
	rule entropy 'error' × feedback (correct – incorrect)	1
	rule entropy 'error' × feedback (target - nontarget)	1
	rule entropy 'error' × response (target - nontarget)	1

Table ST5.5. Regressors for GLM5.3

This is an exhaustive list of the task regressors included in GLM5.3.

Onset	Parametric Regressor	Number of Regressors
Cue	(main effect)	1
	cue (novel – familiar)	1
Trial	rule entropy	1
	feedback (correct – incorrect)	1
	feedback (target - nontarget)	1
	response (target - nontarget)	1
	cue (novel – familiar)	1
	rule entropy × cue (novel – familiar) × TC	1
	rule entropy × cue (novel – familiar) × TI	1
	rule entropy × cue (novel – familiar) × NC	1
	rule entropy × cue (novel – familiar) × NI	1
	rule entropy × feedback (correct – incorrect)	1
	rule entropy × feedback (target - nontarget)	1
	rule entropy × response (target - nontarget)	1

Table ST5.6. Regressors for GLM5.4

This is an exhaustive list of the task regressors included in GLM5.4.

Onset	Parametric Regressor	Number of Regressors
Cue	(main effect)	1
	cue (novel – familiar)	1
Trial	(main effects)	64

Table ST5.7. Activation clusters for GLM5.1

Tables show voxel clusters (larger than 5 voxels; threshold at 0.005 uncorrected) activated by the multiple contrasts of GLM5.1 (see Methods and Results). Columns indicate (L-R): cluster corrected p-value (FDR), cluster size (Voxels), uncorrected peak t-value (Peak t), peak z-value (Peak z) and the final column gives the xyz coordinates in MNI space for each peak (Coords).

Target entropy > 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0.003	238	7.4	4.88	0	-42;0;54
		5.18	3.96	0	-54;16;26
		4.11	3.38	0	-50;8;42
0.037	112	4.95	3.84	0	-26;24;-10
		3.4	2.93	0.002	-30;44;-2
		3.39	2.92	0.002	-10;16;-2
0.232	54	4.73	3.73	0	-30;-56;46
0.031	131	4.68	3.7	0	10;4;-2
		4.22	3.44	0	30;28;-10
		3.51	3	0.001	46;24;2
0.593	22	4.39	3.54	0	-14;-20;18
0.774	12	4.12	3.39	0	14;-20;18
		3.73	3.14	0.001	10;-24;10
0.774	6	3.99	3.31	0	38;-40;38
0.538	27	3.97	3.29	0	10;40;34
		3.41	2.93	0.002	-6;36;38
		3.13	2.74	0.003	14;36;22
0.593	20	3.91	3.26	0.001	30;4;42
		3.03	2.67	0.004	30;8;34
0.342	40	3.91	3.26	0.001	-2;12;54
		3.87	3.23	0.001	-2;8;62
0.774	6	3.54	3.02	0.001	-18;0;-10
0.774	6	3.5	2.99	0.001	-42;-40;38
0.774	6	3.45	2.96	0.002	34;-68;50
0.774	8	3.45	2.96	0.002	50;-32;-2
0.774	6	3.26	2.83	0.002	50;-20;-6
0.774	7	3.17	2.77	0.003	6;-60;46

Rule entropy > 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0.148	70	5.65	4.18	0	-38;60;18
0.001	408	5.64	4.18	0	42;36;38
		4.83	3.78	0	46;40;30
		4.3	3.49	0	26;60;-6
0.007	257	4.69	3.71	0	54;-48;50
		4.63	3.67	0	62;-52;46
		3.95	3.28	0.001	38;-40;42
0.026	150	4.46	3.58	0	-46;24;38
		4.24	3.46	0	-46;32;30
		3.93	3.27	0.001	-50;40;22
0.026	147	4.16	3.41	0	-42;-60;-30
		3.84	3.21	0.001	-10;-80;-30
		3.11	2.73	0.003	-10;-56;-30
0.023	176	4.14	3.4	0	-34;-52;42
		4.11	3.38	0	-46;-48;54
		3.97	3.29	0	-50;-60;54
0.323	33	3.95	3.28	0.001	10;-28;26
		3.56	3.03	0.001	-10;-32;26
		3.38	2.92	0.002	-2;-28;26
0.302	42	3.73	3.14	0.001	10;-12;2
		3.23	2.81	0.002	10;-16;14
0.365	27	3.62	3.07	0.001	30;-64;-30
0.576	11	3.51	3	0.001	38;4;62
0.425	21	3.5	3	0.001	34;24;-2
0.323	35	3.44	2.95	0.002	6;24;42
		3.43	2.95	0.002	-6;32;38
0.576	12	3.34	2.89	0.002	-26;56;-10
0.627	7	3.31	2.86	0.002	50;16;6
0.627	7	3.16	2.76	0.003	-10;-12;-2
		3.15	2.75	0.003	-14;-4;2

Rule entropy < 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0.341	78	7.91	5.06	0	-30;-12;70
		4.78	3.76	0	-34;-24;74
		4.55	3.63	0	-18;-12;74
0.501	55	6.7	4.63	0	-22;-20;-18
		6.56	4.57	0	-30;-32;-14
0.263	132	5.64	4.18	0	-58;-8;-14
		4.4	3.54	0	-54;0;-26
		4.37	3.53	0	-46;12;-30
0.838	5	4.48	3.59	0	14;48;54
0.838	5	4.13	3.39	0	62;-4;-18
0.317	96	3.96	3.29	0.001	2;56;-14
		3.72	3.14	0.001	-6;56;2
		3.64	3.08	0.001	-6;52;-6
0.838	7	3.91	3.26	0.001	18;-32;74
0.838	11	3.67	3.11	0.001	-6;-52;22
		3.36	2.9	0.002	-14;-52;14
0.838	6	3.45	2.96	0.002	38;-96;-6

Table ST5.8. Activation clusters for GLM5.2

Tables show voxel clusters (larger than 5 voxels; threshold at 0.005 uncorrected) activated by the multiple contrasts of GLM5.2 (see Methods and Results). Columns indicate (L-R): cluster corrected p-value (FDR), cluster size (Voxels), uncorrected peak t-value (Peak t), peak z-value (Peak z) and the final column gives the xyz coordinates in MNI space for each peak (Coords).

'Residual RE' × feedback(correct - incorrect) > 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0	370	7.48	4.91	0	-50;-40;38
		5.94	4.31	0	-38;-52;54
		4.29	3.48	0	-10;-72;58
0	222	5.95	4.32	0	46;-40;46
		5.27	4	0	38;-56;42
		4.42	3.56	0	14;-64;46
0	284	5.58	4.15	0	-34;0;58
		5.14	3.94	0	-42;20;22
		4.99	3.86	0	-42;0;34
0	271	5.3	4.02	0	54;28;30
		5.16	3.95	0	30;8;58
		5.06	3.9	0	34;28;26
0.252	26	4.83	3.78	0	-54;-48;6
0.26	21	4.8	3.76	0	6;-76;-26
		3.81	3.2	0.001	-6;-76;-26
0.012	92	4.56	3.63	0	6;20;46
		4.52	3.61	0	-2;16;46
		3.55	3.03	0.001	-6;8;58
0.295	17	4.15	3.4	0	34;20;2
0.26	22	4.14	3.39	0	26;56;6
		4.04	3.33	0	30;56;-10
0.295	16	4.06	3.35	0	-26;52;10
0.311	14	3.94	3.28	0.001	-66;-32;-2
		3.16	2.76	0.003	-50;-28;-2
0.362	11	3.4	2.93	0.002	-34;20;2
0.584	5	3.38	2.92	0.002	38;-64;-26

‘Residual RE’ × feedback(correct - incorrect) < 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0	274	9.13	5.42	0	-2;36;-2
		5.42	4.08	0	-10;48;2
		5.39	4.06	0	-2;32;10
0.007	144	5.5	4.11	0	-18;-20;70
		5.37	4.05	0	-26;-12;70
		5.23	3.98	0	-26;-24;66
0.009	126	5.02	3.88	0	10;-20;50
		4.45	3.57	0	-6;-20;38
		4.21	3.44	0	-2;-28;50
0.43	23	4.89	3.81	0	18;-44;70
0.166	50	4.52	3.62	0	-38;4;10
		4.03	3.33	0	-30;4;-18
		3.92	3.26	0.001	-38;-4;-10
0.759	5	4.23	3.45	0	-18;-40;-22
0.43	28	4.15	3.4	0	30;-20;66
		3.29	2.85	0.002	34;-28;58
0.581	15	3.94	3.28	0.001	-30;-20;18
		3.46	2.97	0.001	-46;-24;22
0.43	25	3.77	3.17	0.001	38;-12;14
		3.31	2.87	0.002	50;-12;14
0.565	17	3.76	3.17	0.001	-10;-56;14
0.759	8	3.62	3.07	0.001	-22;28;46
0.66	12	3.44	2.96	0.002	62;-32;26
0.759	5	3.18	2.78	0.003	-46;-76;30

'Residual RE' × feedback(target - nontarget) > 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0	297	5.99	4.33	0	46;44;6
		5.12	3.93	0	38;36;22
		4.43	3.56	0	46;32;14
0	467	4.8	3.76	0	-54;8;38
		4.68	3.7	0	-42;36;22
		4.62	3.67	0	-38;0;42
0.784	17	4.28	3.48	0	26;24;-2
		3.3	2.86	0.002	18;24;-10
0.198	58	4.1	3.37	0	-18;8;6
		4.06	3.35	0	-10;8;26
		3.72	3.14	0.001	-22;0;-6
0.784	16	3.82	3.2	0.001	-46;-44;46
0.275	44	3.77	3.17	0.001	22;0;14
		3.57	3.04	0.001	18;-8;14
		3.39	2.92	0.002	10;8;26
0.784	6	3.67	3.11	0.001	18;-60;-26
0.784	13	3.54	3.02	0.001	-50;-36;2
0.784	9	3.38	2.91	0.002	-30;-68;42
0.784	8	3.36	2.9	0.002	-34;20;2
0.784	5	3.13	2.74	0.003	-34;-48;34

'Mean RE' × feedback(target - nontarget) > 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0	2180	10.32	5.74	0	34;24;-10
		8.29	5.18	0	2;28;46
		7.93	5.06	0	2;16;46
0	931	8.98	5.38	0	42;-48;42
		7.66	4.97	0	38;-60;46
		7.05	4.76	0	50;-36;42
0.287	41	4.93	3.83	0	50;-32;-6
		3.66	3.1	0.001	66;-40;-2
0.499	23	4.83	3.78	0	-10;4;2
0.287	43	4.47	3.59	0	10;4;2
		3.87	3.23	0.001	10;-12;6
0.718	8	3.94	3.27	0.001	6;-32;26
0.718	8	3.77	3.17	0.001	-6;-80;-26
0.533	18	3.61	3.07	0.001	-34;-60;-34

Table ST5.9. Activation clusters for GLM5.3

Tables show voxel clusters (larger than 5 voxels; threshold at 0.01 uncorrected) activated by the multiple contrasts of GLM5.3 (see Methods and Results). Columns indicate (L-R): cluster corrected p-value (FDR), cluster size (Voxels), uncorrected peak t-value (Peak t), peak z-value (Peak z) and the final column gives the xyz coordinates in MNI space for each peak (Coords).

'Residual RE' × cue (novel - familiar) × feedback(NI - TC, TI, NT) > 0

FDR	Voxels	Peak t	Peak z	Uncor P	Coords
0.001	480	5.02	3.88	0	-50;24;22
		4.8	3.77	0	-38;24;42
		4.41	3.55	0	-30;24;-10
0.001	530	4.63	3.67	0	54;20;26
		4.28	3.48	0	34;52;6
		3.79	3.18	0.001	34;16;46
0.097	117	4.23	3.45	0	2;32;38
0.075	152	4.02	3.33	0	-50;-64;42
		3.53	3.02	0.001	-30;-72;50
		3.52	3.01	0.001	-30;-64;34
0.257	65	3.92	3.26	0.001	6;-64;46
0.095	127	3.82	3.2	0.001	-18;56;6
		3.64	3.09	0.001	-34;44;-10
		3.15	2.76	0.003	-18;48;22
0.075	166	3.59	3.06	0.001	30;-64;54
		3.48	2.99	0.001	38;-40;38
		3.31	2.87	0.002	34;-40;30
0.6	29	3.26	2.83	0.002	-14;-76;-34
		3.01	2.66	0.004	-26;-72;-30
0.864	8	3.16	2.76	0.003	-42;-72;-34
		2.86	2.55	0.005	-42;-64;-30
0.805	17	3.15	2.76	0.003	50;-52;-14
		3.04	2.68	0.004	42;-60;-14
0.815	14	3.06	2.69	0.004	6;-8;14