Hippocampal and medial prefrontal cortices encode structural task representations following progressive and interleaved training schedules

Sam C. Berens[1,*] and Chris M. Bird[1]
[1]School of Psychology, University of Sussex, UK.
*Correspondence: s.berens@sussex.ac.uk (S.C.B.)

Abstract

Memory generalisations may be underpinned by either encoding- or retrieval-based generalisation mechanisms and different training schedules may bias some learners to favour one of these mechanisms over the other. We used a transitive inference task to investigate whether generalisation is influenced by progressive vs randomly interleaved training, and overnight consolidation. On consecutive days, participants learnt pairwise discriminations from two transitive hierarchies before being tested during fMRI. Inference performance was consistently better following progressive training, and for pairs further apart in the transitive hierarchy. BOLD pattern similarity correlated with hierarchical distances in the left hippocampus (HIP) and medial prefrontal cortex (MPFC) following both training schedules. These results are consistent with the use of structural representations that directly encode hierarchical relationships between task features. However, such effects were only observed in the MPFC for recently learnt relationships and were strongest in participants who were least proficient at the inference task. We conclude that humans preferentially employ encoding-based mechanisms to store map-like relational codes that can be used for memory generalisation. These codes are expressed in the HIP and MPFC following both progressive and interleaved training but are not sufficient for accurate inference.

Introduction

Humans are readily able to generalise information learnt in one situation and apply it in another. For example, if we are told that Abuja is generally hotter than Beirut (A>B), and Beirut is hotter than Carlisle (B>C), then we can infer that Abuja is hotter than Carlisle (A>C), despite never having been given that information directly. This particular type of generalisation is known as transitive inference.

The hippocampal system and medial prefrontal cortices (MPFC) have long been implicated in generalising newly learned information for use in new situations. Broadly speaking, contemporary models pose that these generalisations may be supported in two different ways: *1)* retrieval-based models, and *2)* encoding-based models. Despite these opposing views being present in the literature for many decades [1,2], it is unclear which mechanisms are used to support memory generalisation or, indeed, whether one is favoured over the other in particular situations.

Retrieval-based models suggest that the hippocampus encodes pattern separated representations that express specific relationships between co-presented items [3]. These models argue that generalisation is supported by a recursive neural mechanism to rapidly integrate distinct memories on-the-fly. Such a proposal has received support from fMRI [4] and behavioural studies [5]. These models predict that the brain only needs to store the originally presented information, since generalisation occurs as and when it is necessary, and is achieved by retrieval of the learnt information.

In contrast, encoding-based models suggest that the hippocampal and MPFC systems learn unified representations that directly express inferred structured relationships between task features [6–8]. These 'structural representations' are therefore sufficient to support inference without the need for a specialised inference mechanism. As such, the hallmark of encoding-based models is that the relationships between events have been abstracted and stored, enabling generalisation to occur without the need for online integration. Of course, these knowledge structures may not be created strictly at the point of encoding – it is possible that they emerge after a period of consolidation or after the same information has been experienced several times [9–11].

Consistent with encoding-based models, the hippocampus, entorhinal cortex, and medial prefrontal cortex have been found to encode generalised relationships that were not explicitly trained [6,12–17]. The entorhinal cortex and MPFC also appear to represent similarly structured hierarchies in using a common neural code [18–20]. These neural codes are thought to facilitate inference and knowledge transfer across related tasks, although their relationship to generalisation performance is ambiguous.

Certain training conditions can have a large impact on how information is retained [21] and generalises to new situations [22]. For example, categorisation of previously unseen objects is sometimes improved if exemplars from different categories are presented in an interleaved order, rather than in category-specific blocks, e.g., [23–25]. However, other studies have shown advantages for blocked training schedules, especially when category differences are clearly verbalizable, e.g., [26,27]. Potentially resolving this conflict, interleaving has been shown to aid category generalisation when exemplars are highly similar to one another (both within-

and between-categories), whereas blocking may be best when exemplars are relatively distinct[28]. As such, it has been suggestions that interleaving emphasises between-category differences, whereas blocking emphasises commonalities amongst exemplars [22,28].

Despite this, it remains unclear whether and how interleaved vs blocked training influences the structure of memory representations that are learnt, or the generalisation mechanism that are preferentially employed. Behavioural evidence suggests that blocked training enables the learning of low-dimensional (compressed) stimulus representations that linearly encode task-relevant features [27]. Additionally, recent research shows that training overlapping discriminations in an ordered sequence (e.g., A>B followed by B>C and then C>D, so-called 'chaining') may improve transitive inference by allowing learners to integrate the discriminations into a unified mental model [29,30]. Nevertheless, when this integration takes place, and whether it depends on encoding- or retrieval-based mechanisms has yet to be tested.

We hypothesised that interleaved training would promote the learning of the specific – pattern separated - pairings and consequently bias the use of retrieval-based inference judgements. This follows the proposal that interleaved learning highlights the differences between items and is consistent with the finding that hippocampal pattern separation prevents interference between overlapping relationships learnt in an interleaved order [31]. In contrast, we hypothesised that presenting related pieces of information in ordered blocks (hereafter referred to as 'progressive training'), should facilitate the use of encoding-based inference mechanisms. Specifically, progressive training may enable pattern completion between pairs thereby allowing participants to encode inferred relationships during training [32].

In addition to the training schedule, we also manipulated whether participants experienced an overnight period of consolidation before being tested on their ability to generalise. Sleep-dependent consolidation has long been implicated in abstracting statistical regularities across separate memories, possibly because it allows distinct event representations to be replayed out-of-order [33,34]. In support of this, many studies have shown that memory generalisation improves following a period of sleep, or even wakeful rest [35–39]. We hypothesised that overnight consolidation would allow pattern separated memories of related task contingencies to be re-encoded as structural memory representations, see [40,41]. We therefore predicted that inferences made on items learnt the previous day would depend more on encoding-based mechanisms than inferences on items learnt immediately prior to scanning.

To test these hypotheses, we analysed the effect of training schedule and overnight consolidation on behavioural and fMRI data collected while human participants performed transitive inferences. To do this, we trained a series of 'premise' discriminations via either progressive or interleaved presentations within a reinforcement learning task (see Figure 1). Across consecutive days, 34 participants learnt 2 independent sets of premise discriminations (one set per day), each of which entailed a 1-dimensional transitive hierarchy over 7 visual features (A>B>C>D>E>F>G). Shortly after training on the second day, participants recalled all the premise discriminations, and made inferences whilst being scanned. As such, we were able to investigate progressive/interleaved training, and inferences based on recent/remote memories in a full factorial design.

We first show that progressive training can improve inference performance in simple artificial neural networks (multilayer perceptions; MLPs), particularly when those networks are

constrained to learn low-dimensional (compressed) representations of the task. In particular, we demonstrate that MLPs constrained in this way tend to learn representations that encode the relative value of discriminable stimuli. We also found that progressive training had a large benefit on inference performance in humans. Our analyses show that the behavioural data are better accounted for by encoding-based generalisation mechanisms when directly compared with retrieval-based mechanisms. Surprisingly, this did not depend on the experimental factors of interest, namely, training schedule and overnight consolidation.

Next, we tested neurocognitive predictions of encoding- and retrieval-based models using univariate and multivariate analyses of the imaging data. Contrary to the predictions of retrieval-based models, we did not observe BOLD activations that were uniquely associated with generalisation performance. In contrast, representational similarity analyses (RSAs) were more consistent with encoding-based models. We identified RSA effects in the hippocampus and MPFC suggestive of structural representations for the whole transitive hierarchy. However, these effects we only evident in the MPFC for recently learnt contingencies, suggesting a time-dependent role of this region. Finally, we found that structural representations in the hippocampus and MPFC were associated with different patterns of behavioural performance that may be indicative of different generalisation processes.

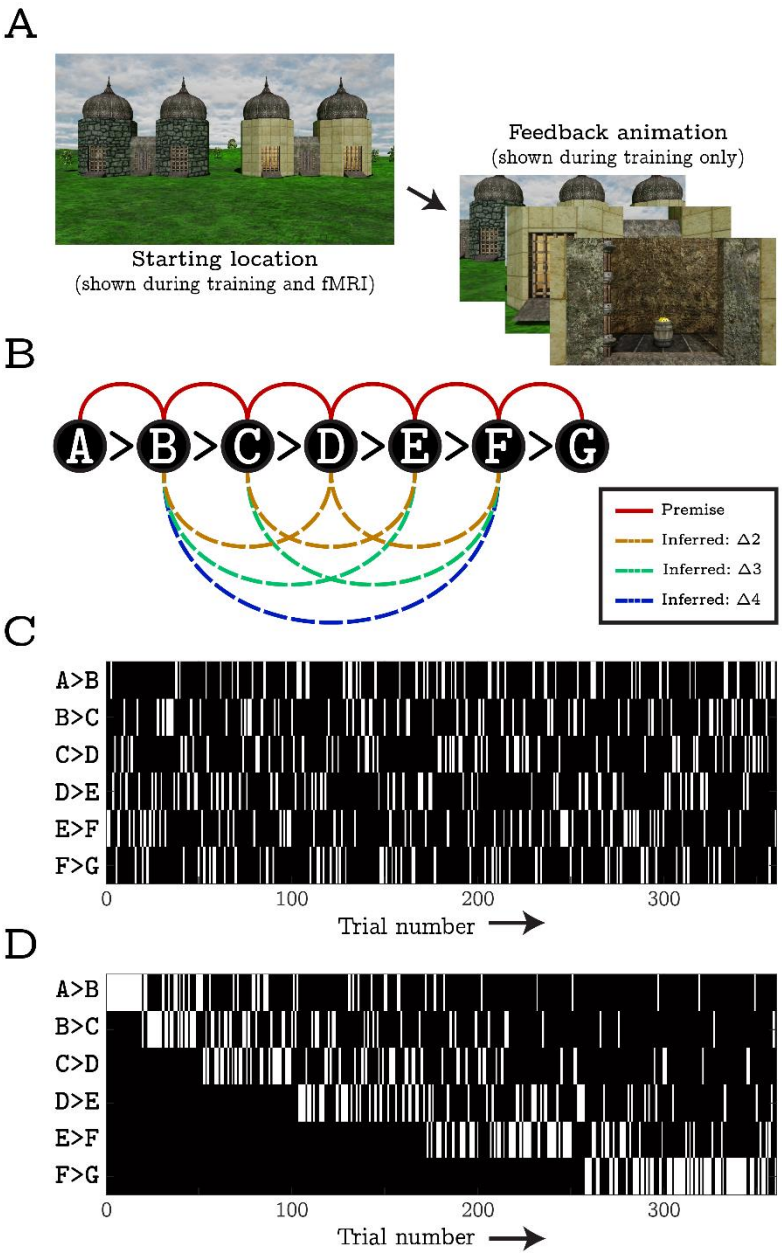


Figure 1. Illustration of the pre-scanner training and in-scanner behavioural tasks. A) Both before and during fMRI, participants saw computer generated images of two buildings with different wall-textures rendered onto their exterior surfaces. One building concealed a pile of virtual gold (reinforcement) and the location of this reward was perfectly determined by the combination of wall-textures shown. In the pre-scanner training phase, participants were tasked with learning the reward contingencies via trial-and-error. A left/right button press was required within 3 seconds of the start of each trail. Following this, a feedback animation was shown indicating whether the response was correct or not. During the in-scanner task, participants were required to respond to still images of the two buildings, yet no feedback was provided. B) A schematic illustration of the reward contingencies trained before scanning (i.e., the premise discriminations, red solid lines) and inferred inside the scanner (i.e., inferred discriminations, dashed lines). Letters denote unique wall textures and the greater than signs indicate the rewarded wall-texture in each premise discrimination. Taken together the 6 premise discriminations implied a 1-dimensional transitive hierarchy. Inferred discriminations did not involve the ends of the hierarchy (i.e., A and G) since such challenges can be solved by retrieving an explicitly trained (featural) contingency (e.g., recalling that A is always rewarded). As such, the set of inferred discriminations included three trials with a 'transitive distance' of Δ2, two trials with a transitive distance of Δ3, and one trial with a transitive distance of Δ4. Note that participants were trained on two independent transitive hierarchies on two separate days: one 24 hours before scanning, one immediately before scanning. While equivalent in structure, the contingencies learnt on each day involved entirely different wall-texture stimuli (counterbalanced across participants) which were never presented in the same trial. C) and D) One each day of training, premise trials were ordered in one of two ways: interleaved training involved presented all 6 premise discriminations in pseudorandom order such that there was a uniform probability (1/6) of encountering any one discrimination on a particular trial (panel C). In contrast, progressive training involved 6 epochs of different lengths that gradually introduced the discriminations whilst ensuring that, once a discrimination had been introduced, it was presented in all subsequent epochs (panel D).

Results

Progressive training in multilayer perceptrons

We first demonstrate that different training schedules can guide even very simple agents to learn fundamentally different task representations which, in turn, affects generalisation. Across multiple independent runs, we trained multilayer perceptrons (MLPs) with 6 different architectures to perform a set of binary discriminations similar in structure to our fMRI task (see Figure 1B). These discriminations were presented either progressively (one after another), or in an interleaved order (see Figures 1C and 1D). We then tested the perceptrons on both explicitly trained (premise) and transitively inferred discriminations.

The goal of this analysis was to characterise effects of each training schedule on learning and inference. As such, the MLPs were designed to be as simple as possible while having the ability to learn a task representation that could support transitive inference. They were not indented to be a faithful model of how humans solve such tasks, nor is this analysis indented to adjudicate between retrieval- and encoding-based models of generalisation. The MLPs consisted of three layers (input, hidden, and output) with varying numbers of hidden units (between 1 and 6) connected to all inputs and output (see Figure 2A). In short, training involved presenting the perceptrons with two input symbols (e.g., 'A' and 'B') coded as a binary vector and updating the network weights via backpropagation to reproduce the "correct" symbol as a one-hot vector on the output layer.

We explicitly trained 6 premise discriminations; 'A>B', 'B>C' ... 'F>G' (correct responses indicated to the left of the greater-than sign) over 3,600 training steps (note: this is 10x the number of training trials given to human participants in preparation for the fMRI task). As such, these contingencies implied a 1D transitive hierarchy (A>B>C>D>E>F>G). For the networks with only one hidden unit, the output layer only received a 1-dimensional latent representation of the inputs and so high-levels of performance depended on non-trivial weights and bias distributed between layers. Nonetheless, many solutions that do yield near-perfect performance on both premise and inferred discriminations do exist (see https://osf.io/ps3ch/ for a scripted example). With more hidden units the networks were capable of learning increasingly sparse representations of each discrimination, see [42,43]. For example, with 6 hidden units, a specific hidden unit could uniquely activate the target output for each trained discrimination.

Interleaved training involved presenting the MLPs with all 6 discriminations in a pseudorandom order such that there was a uniform probability (1/6) of encountering any one on a particular trial (see Figure 1C). As such, training-related prediction error (as measured by a cross-entropy cost function) gradually decreased across interleaved training trials (see Figure 2B). In contrast, progressive training involved 6 epochs of different lengths that gradually introduced discriminations and ensured that, once introduced, they were presented in all subsequent epochs. Given this, progressive training was characterised by sharp increases, and then rapid decreases in prediction error at the start of each epoch. For each MLP architecture, interleaved and progressive training runs were repeated 10,000 times with varied training orders and different network initialisations (i.e., random starting parameters).

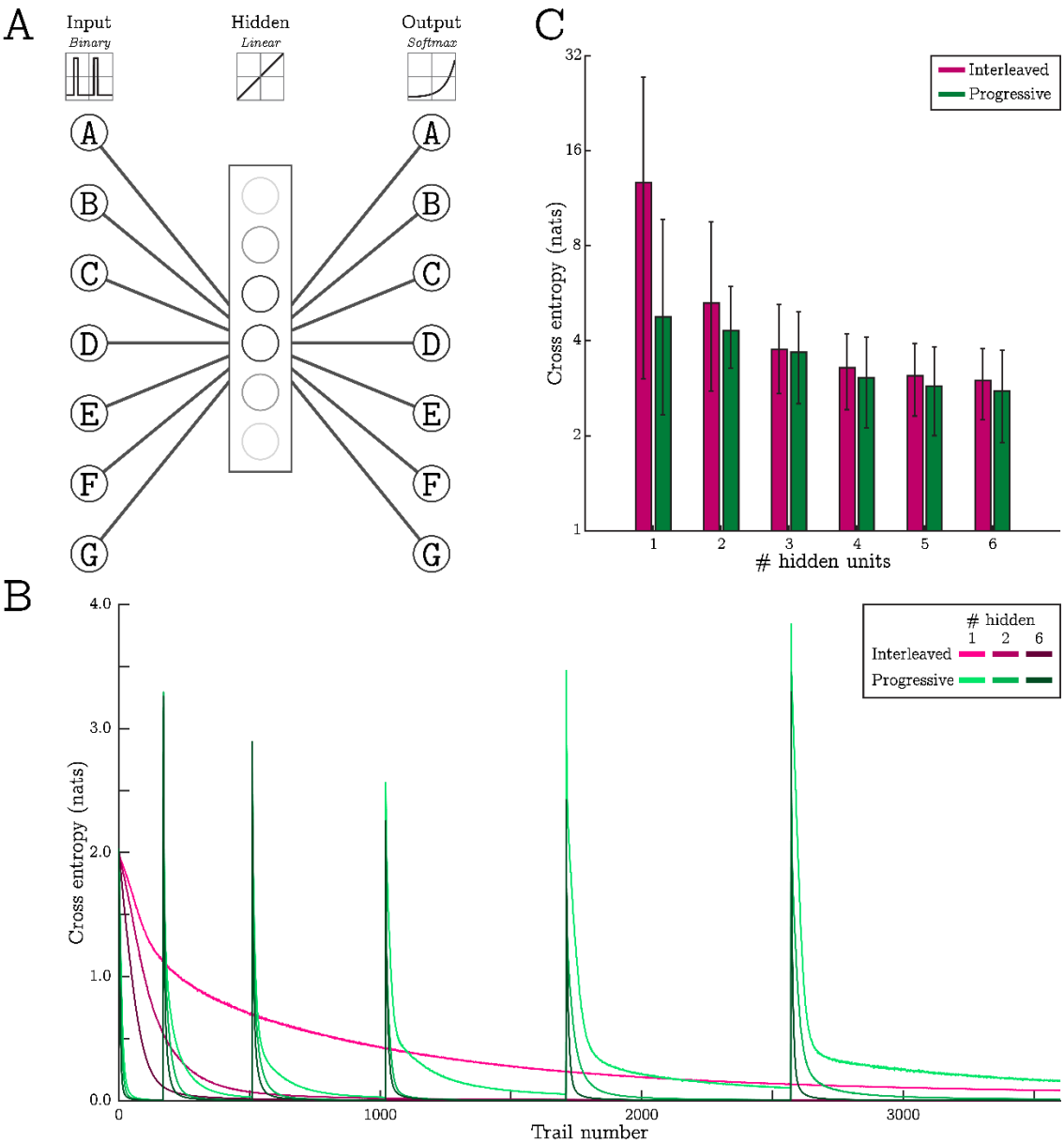


Figure 2. Multi-layer perceptrons (MLPs) show better generalisation in a transitive interference task after progressive training, particularly when the hidden layer only contains a single unit. A) An illustration of the 3-layer MLPs included in the analysis. Seven input units (labelled A-G, one for each discriminable feature) provided weighted binary inputs to between 1 and 6 biased hidden units with a linear activation function. In turn, activity in the hidden units projected along weighted connections to seven biased outputs with a SoftMax activation function. B) We trained the MLPs on the six premise discriminations in Figure 1 via backpropagation. Training was structured in two ways (see Figures 1C and 1D). With interleaved training, a cross entropy loss function decreased exponentially over trials (magenta lines) whereas progressive training resulted in rapid exponential decreases following the start of each epoch (green lines). Note: the learning curves plotted in this figure result from averaging 10,000 independent training runs for networks with either 1, 2 or 6 hidden units. C) After each training run, we tested the performance of MLPs on the full set of inferred discriminations show in Figure 1. Inference performance as measured by the cross-entropy loss improved as more hidden units were added and was invariably better following progressive training. Nonetheless, the effect of progressive training was most pronounced for MLPs with a single hidden unit. Bars indicate mean cross-entropy scores across runs on a log-scale, error bar ranges include 95% of scores.

We tested the performance of MLPs in two ways: *1)* using the same cross-entropy measure employed during training and, *2)* using a 2-alternative Luce choice (2-ALC) decision rule implemented by a SoftMax normaliser (see Methods). Unlike cross-entropy, the 2-ALC measure allowed us to examine whether the activity of the target output unit was greater than the activity of the concurrently presented non-target (which tended to be highly activated by its corresponding input). Both interleaved and progressive training schedules resulted in all MLPs learning the 6 premise discriminations to a high level (Supplementary Table 1). According to both performance measures, premise performance increased as hidden units were added, and were slightly better following interleaved training.

We then tested the MLP on 6 transitive inference trials that were not explicitly trained (e.g., 'B>D'). Cross-entropy statistics indicated that progressively trained MLPs invariably performed better on these inference trials than MLPs trained via an interleaved schedule (Figure 2C). In networks with a single hidden unit, interleaved training did not generally promote greater activity in the target output relative to the non-target output, 2-ALC accuracy: 49.8% (CI: +/- 0.367; see Figure 3A). However, the variability in inference performance across different interleaved training runs was large (95% range: 17.0 - 78.8%). In contrast, progressive training in these single hidden unit networks lead to substantially higher levels of inferential accuracy: 64.1% (CI: +/- 0.174), yet the variability in performance reduced (95% range: 51.6 - 83.7%). This pattern in 2-ALC scores did not hold for networks with two or more hidden units, which tended to activate non-target output units more strongly following both training schedules (Supplementary Table 2).

The above demonstrates that progressive training can yield better inference in some MLPs, particularly those with only a single hidden unit. To explore the reasons for this, we computed the linear activation of each output unit (prior to SoftMax normalisation) that resulted from stimulating the corresponding input unit with a one-hot vector. We denote this activation, $\Delta\alpha_i^3 \mid \alpha_i^1 = 1$, for the ith input/output unit, where superscripts denote layer indices. Figure 3B plots these values as a function of training trial, averaged across runs for MLPs with a single hidden unit. This showed that activity in the output corresponding to stimulus 'A' is generally higher than the activity of all others, activity in 'B' is next highest, and so on. As such, progressive training produces activations in each output unit that faithfully represents a generalised value for the corresponding stimulus as long as that stimulus is presented on the input layer. We note that this is also evident in MLPs with more than one hidden unit yet, the value representation is much weaker, see Supplementary Figure 1.

This generalised value gradient develops because learning the first discrimination ('A>B') tunes the MLP to increase the activation of the 'A' output above all others. When the next discrimination is introduced ('B>C'), activity in the 'B' output starts to increase yet, because the 'A>B' discrimination is still being tested, activation in the 'A' output also must rise to maintain performance. This pattern continues down the transitive hierarchy as long as older discriminations are presented often enough to avoid catastrophic interference. As such, progressive training can yield a very specific error gradient that allows the network to learn a generalised value for each stimulus. The same is not true for MLPs trained via an interleaved schedule. Here, activity levels in output units are approximately uniform for all stimuli that are rewarded during training (i.e., excluding stimulus 'G'). Instead, networks trained in this

way tended to learn the discriminations by inhibiting the activity of non-target outputs in response to the presentation of specific premise pairs.

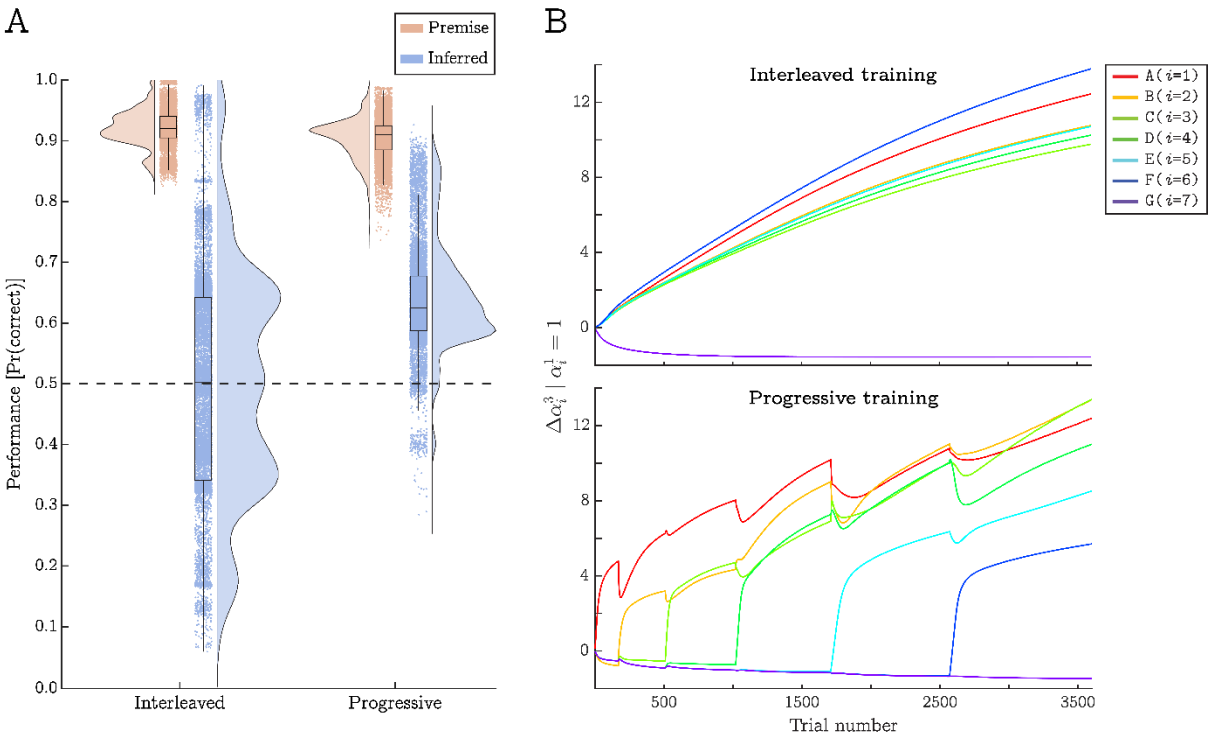


Figure 3. Progressive training promotes transitive inference in MLPs with a single hidden unit due to the learning of a generalised value gradient. A) We tested the performance of the MLP on both trained and inferred discriminations using a SoftMax function to implement a 2-alternative Luce choice decision (see *Eq. 1*). Performance on the premise discriminations was comparable across training methods but average performance on inference trials was much higher following progressive training. Despite this, the variability in inference performance that resulted from interleaved learning was large such that some MLP models trained in this way achieved very high levels of performance. B) When an input stimulus is presented as a one-hot vector, activity in corresponding output unit reflects a generalised value gradient following progressive, but not interleaved, training. Different lines plot activations in different output units (averaged across training runs). With an interleaved training procedure, activity levels are approximately equivalent for all stimuli that are rewarded during training. During progressive training, activity in the output corresponding to stimulus 'A' is generally higher than the activity of all others, activity in 'B' is next highest, and so on. This happens because learning the first discrimination ('A>B') tunes the MLP to increase the activation of the 'A' output above all others. When the next discrimination is introduced ('B>C'), activity in the 'B' output starts to increase yet, because the 'A>B' discrimination is still being tested, activation in the 'A' output also must rise to maintain performance. This pattern continues down the transitive hierarchy as long as older discriminations are presented often enough to avoid catastrophic interference.

A consequence of progressive training is that it biased MLPs with a single hidden unit towards learning a limited number of solutions to the discrimination problem (see Supplementary Figure 2). We used a gaussian mixture model to characterise the distribution of trained parameters and found that they tended towards 53 distinct 'attractors' after interleaved training, but only 28 attractors after progressive training. Summed variance scores across parameters revealed that interleaved training resulted in more diverse solutions relative to progressive training ($\sigma^2 = 328$ and 177, respectively). This patten was also evident for MLPs

with two or more hidden units, although the difference between training schedules was much less pronounced (see Supplementary Table 3).

In sum, we find that progressive training can have a significant effect on the type of representations that are learned by MLPs, especially when they are structured to learn low-dimensional solutions to a particular problem. This can increase generalisation performance, yet progressive training also appears to reduce the diversity of learnt representations thereby precluding the development of models that achieve very high levels of performance.

Inference performance in humans

Over two consecutive days, we trained participants to make 12 binary discriminations in a reinforcement learning task (see Figure 1). Trials presented two buildings that differed in only one respect; the wall textures rendered onto the outside of each building. One building contained a pile of virtual gold (reinforcement) and participants were tasked with learning which wall texture predicted the gold in order to gain as much reinforcement as possible. The contingencies predicting reward were equivalent in structure to those described above. Specifically, we trained 2 independent sets of discriminations, each of which implied a 1-dimensional transitive hierarchy (A>B>C>D>E>F>G). One set of premise discriminations was trained on each day and training sessions were separated by approximately 24 hours. Prior to the first session, participants were randomly assigned to either an interleaved or progressive training condition which determined the type of training they received on both days (see Figures 1C and 1D). After training on the second day, participants underwent fMRI scanning while recalling all the premise discriminations (from both days) as well as 2 sets of inferred discriminations. As such, the experiment involved 3 main experimental factors: *1)* training method (interleaved vs progressive), *2)* session (recent vs remote), *3)* discrimination type (premise vs inferred).

Figure 4A depicts estimates of performance for the in-scanner task in terms of the probability of a correct response. A mixed-effects logistic regression highlighted similar levels of accuracy for the premise discriminations regardless of training method, session, or their interaction; largest effect: $t(804) = 0.765$, $p = .445$. However, there was a large effect of training method on inference performance with progressive learners outperforming interleaved learners; $t(804) = 5.54$, $p < .001$. This effect was also evident as a main effect of training method, $t(804) = 3.83$, $p < .001$, and an interaction between method and discrimination type; $t(804) = 7.39$, $p < .001$. No main effect of session or a session by method/discrimination type interaction was detected; largest effect: $t(804) = 1.44$, $p = .149$.

The logistic regression also examined the effect of 'transitive distance', that is, accuracy differences corresponding to larger or smaller separations between wall textures along the transitive hierarchy (e.g., B>D has a distance of 2, whereas B>F has a distance of 4). We found an overall effect of transitive distance, $t(804) = 2.11$, $p = .035$, indicating that, in general, as the separation between wall textures increased, behavioural accuracy also increased. Additionally, there was a significant 3-way interaction between training method, session and transitive distance, $t(804) = 3.18$, $p = .002$. This suggested that the effect of transitive distance was most consistent for remote discriminations in the progressive condition, $t(804) = 3.07$, $p = .002$, relative to all other conditions, t-values < 2.00 (see Figure 4B).

A generalised linear model of response times (correct responses only) produced a complementary pattern of results (see Figure 4C). Specifically, we detected main effects of training method and discrimination type indicating shorter response times from progressive learners and longer response times to inferred discriminations; $t(5301) = 2.01$, $p = .045$, and $t(5301) = 2.31$, $p = .021$, respectively. These effects were superseded by a training by discrimination type interaction highlighting that longer response times to inferred trials were more pronounced for interleaved learners; $t(5301) = 5.17$, $p < .001$. Unlike the accuracy data, this analysis showed a main effect of session indicative of quicker responses to all remote discriminations; $t(5301) = 3.26$, $p = .001$. No other significant main effects or interactions were detected.

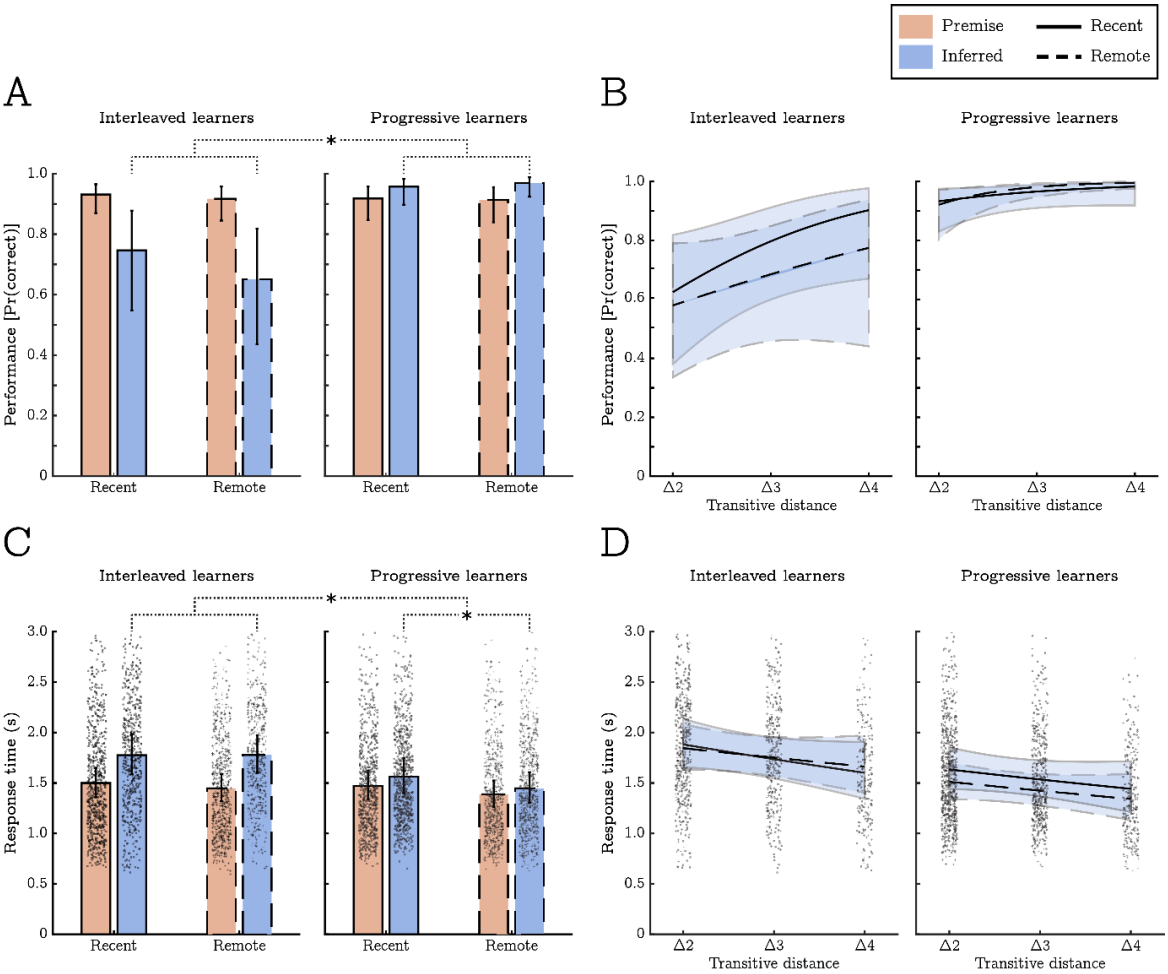


Figure 4. Humans shows better generalisation following progressive training. A) Estimates of the probability of a correct response, $Pr(correct)$, split by trial type (premise vs inferred) and experimental condition (training method and session). While participants showed comparable levels of performance on the premise discriminations across conditions (red bars), inference performance varied by training method with progressive learners showing much higher levels of accuracy (blue bars). B) On inference trials, behavioural performance was positively related to "transitive distance" (the degree of separation between discriminable features along the transitive hierarchy, see Figure 1B). While the correlation between transitive distance and performance was positive in all conditions, the association was most consistent for remote discriminations in the progressive training condition. C) Estimates of the mean response time (in seconds, correct responses only) split by trial type and experimental condition (as in

panel A). Response times closely mirrored the probability of a correct response but showed an additional effect indicating that participants were faster at responding to remote contingencies (overall). D) Response times to inference trails by transitive distance. While not significant, in general, response times decreased as transitive distance increased. All error bars/lines represent 95% confidence intervals.

Computational models of human inference

We predicted that the use of retrieval- and encoding-based generalisation mechanisms would vary by experiment condition. To directly test this, we created two descriptive models that, under similar assumptions, attempted to predict participants' inference performance given their responses on premise trials. Each model was based on general principles of retrieval- vs encoding-based accounts.

The retrieval-based AND model assumes that correctly inferring a non-trained discrimination (e.g., B>E) involves retrieving all the directly trained response contingencies required to reconstruct the relevant section of the transitive hierarchy (e.g., B>C *and* C>D *and* D>E; so-called mediating contingencies). In contrast, the encoding-based OR model assumes that inferences require the retrieval of a unified representation which may be activated by recalling any one of the mediating contingencies (e.g., B>C *or* C>D *or* D>E). Accordingly, each mediating contingency is sufficient to deduce the hierarchical structure or relative value of stimuli. As such, inferred discriminations are easiest when there is a large 'distance' between discriminable features. The AND and OR models predict different levels of performance across inference trials (see Methods). We measured the fit of these models against participants' performance data using a cross-entropy cost function and analysed these goodness-of-fit statistics using a generalised linear mixed-effects regression with 3 experimental factors: 1) model type (AND vs OR), 2) training method (interleaved vs progressive), and 3) session (recent vs remote).

Figure 5 plots the cross-entropy statistics by all conditions. The mixed-effects regression highlighted main effects of model type, $t(128) = 7.86$, $p < .001$, and training method, $t(128) = 5.60$, $p < .001$, both of which were qualified by a model type by training method interaction: $t(128) = 5.56$, $p < .001$. No other model terms were significant. These results indicated that, relative to the AND model, the OR model provided a better fit to the inference data in general, although it was less predictive in interleaved learners. Nonetheless, the OR model was still preferred over the AND model in interleaved learners, $t(128) = 2.63$, $p = .009$. This was also evident when we used Spearman rank correlations to compare the number of correct responses to each inferred discrimination with the number of correct responses that would be expected under each model. Specifically, the correspondence between model predictions and the observed data tended to be higher across participants for the OR model in both the progressive and interleaved conditions; $t(14) = 6.78$, $p < .001$, and $t(16) = 3.55$, $p = .003$ (respectively, statistics derived from bootstrapped paired-samples t-tests, see Supplementary Table 4). Contrary to our predictions, these results indicate that inference performance is best accounted for by encoding-based models in all experimental conditions.

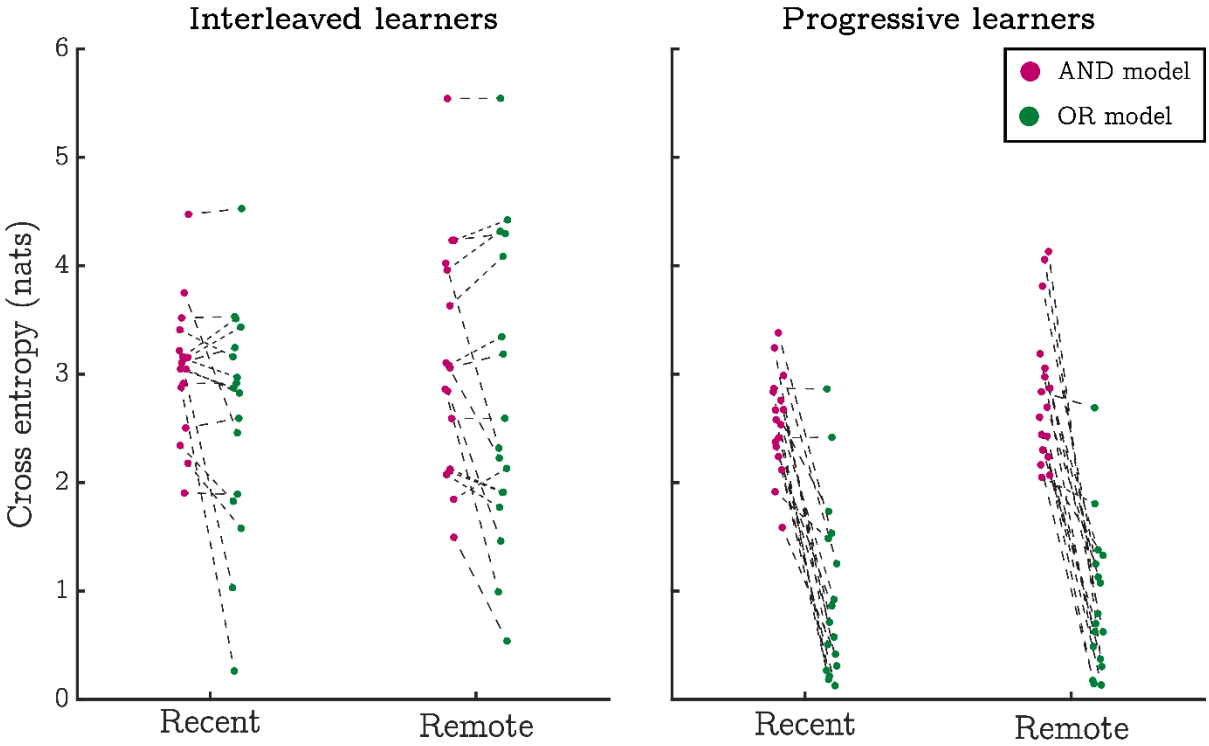


Figure 5. Behavioural performance is better fitted by an encoding-based generalisation mechanism (OR model). Figure shows goodness-of-fit statistics (lower is better) for two models of inference performance. The AND model implements a general assumption of retrieval-based generalisation mechanisms - that inference requires retrieving multiple independent response contingencies in order to evaluate transitive relationships. In contrast, the OR model realises a general assumption of encoding-based generalisation; specifically, that inferences require the retrieval of a unified representation and inferred discriminations are easiest when there is a large distance between features. In all conditions, average goodness-of-fit statistics were lowest for the OR model indicating that it was a better fit to the behavioural data (result qualified by a generalised linear mixed-effects model - see text).

Univariate BOLD effects

Retrieval-based models of generalisation hold that inferences depend on an online mechanism that retrieves multiple premise contingencies from memory and integrates information between them. As such, we used a set linear mixed-effects models to test whether BOLD responses were larger on inferred trials than on premise trials and whether this effect was modulated by 5 factors of interest: *1)* transitive distance, 2*)* training method (interleaved vs progressive), *3)* training session (recent vs remote), *4)* inference accuracy, and *5)* the slope relating transitive distance to inference performance (hereafter referred to as the 'transitive slope'). The rationale for this latter factor follows from considering that encoding- and retrieval-based models predict different transitive slopes (being positive and negative, respectively). Given this, the magnitude of the slope can be used to indicate whether BOLD responses more closely adhere to the predictions of one model or the other.

In comparison to the trained discriminations, inference trials evoked lower levels of BOLD in the right hippocampus (specifically, more deactivation relative to the implicit baseline); $t(787) = 2.79$, $p = .005$ (Supplementary Figure 3). However, this effect was not modulated by training method, $t(787) = 0.31$, $p = .753$, or inference accuracy, $t(787) = 0.04$, $p = .965$, and

so cannot account for variation in inference performance. In contrast, BOLD estimates in the superior MPFC did reflect differences in inference performance. In the left superior MPFC we saw a significant effect of trial type, again indicating more deactivation on inference trials, $t(787) = 3.25$, $p = .001$. This was qualified by a 3-way interaction between trial type, training method, and inference accuracy, $t(787) = 2.93$, $p = .003$ (Figures 6A and 6B). Similarly, the right superior MPFC produced a significant interaction between trial type and training method $t(787) = 2.76$, $p = .006$, (Figures 6C and 6D). Overall, these results indicate that the MPFC produced greater levels of BOLD activity whenever response accuracy was high, regardless of whether participants were responding to premise or inferred discriminations.

In sum, we found no univariate BOLD effects consistent with the use of retrieval-based generalisation mechanisms. While activity in the superior MPFC was associated with behavioural performance, this association was not specific to, or enhanced by novel inferences as would be expected under retrieval-based accounts, see [44,45].

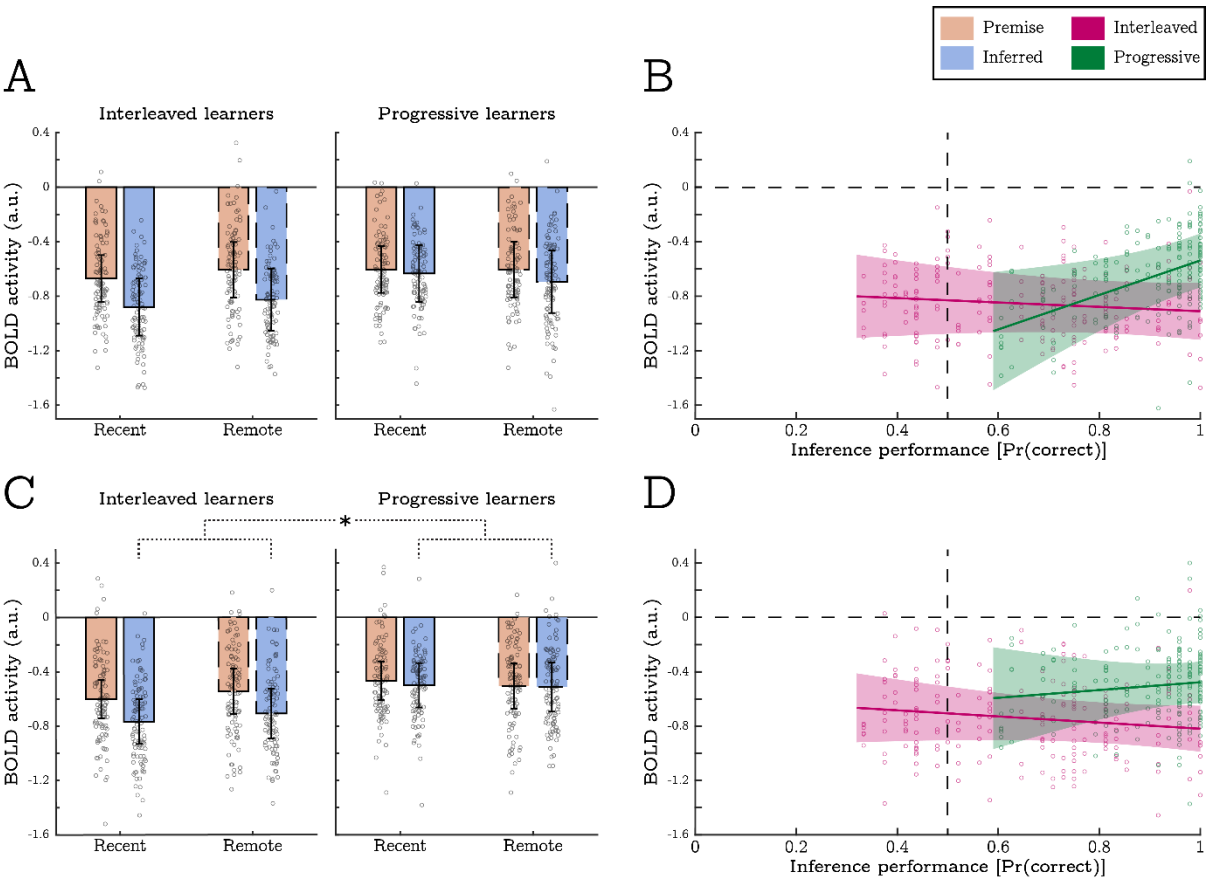


Figure 6. Inference performance within- and across- experimental conditions is associated with univariate BOLD activity in the superior MPFC. Panels A and B show activity in the left superior MPFC. Panels C and D show activity in the right superior MPFC. Bar charts display mean response amplitudes to all in-scanner discriminations split by trial type (premise vs inferred) and experimental condition (training method and session). Scatter plots display mean response amplitudes to all inference trials (both recent and remote) as a function of inference performance, split by training method (interleaved vs progressive). In the left superior MPFC, a main effect of trial type indicated lower levels of BOLD activity on inference trails (panel A). This was superseded by a significant 3-way interaction indicating larger BOLD responses to inference trials in progressive learners who achieved high levels of inference performance (panel B). The right superior MPFC showed a significant 2-way interaction

between trial type and training method. This indicated that BOLD responses in interleaved learners were lower on inference trails (relative to premise trials), but comparable to premise trials in progressive learners (panels C and D). Overall, these data indicate that the MPFC produced greater levels of BOLD activity whenever response accuracy is high. All error-bars indicate 95% confidence intervals.

Representational similarity analysis

We predicted that the training method and the length of the study-test interval would affect how response contingencies were encoded by medial temporal and prefrontal systems. Specifically, we expected that progressive training and longer retention intervals would result in structural representations of the transitive hierarchy and that this would correspond to better inference. To test this, we constructed a series linear mixed-effects models (LMMs) that aimed to *a)* identify neural signatures of structural memory representations, and *b)* reveal whether they are modulated by each experimental factor (and their interactions).

BOLD responses to each discrimination were first used to estimate representations of individual wall-textures via an ordinary least-squares decomposition (see Methods and Figure 7A). The correlational similarity between wall-texture representations was then analysed in the LMMs to identify 'distance effects' within each hierarchy, i.e., where the similarity between wall-textures from the same transitive chain (i.e., trained on the same day) scaled with transitive distance (e.g., corr[B,C] > corr[B,D] > corr[B,E]). Moreover, the LMMs tested whether such distance effects were modulated by 4 factors of interest: *1)* training method (interleaved vs progressive), *2)* training session (recent vs remote), *3)* inference accuracy, and *4)* transitive slope (as above).

Importantly, the LMMs excluded correlations involving wall textures 'A' and 'G' at the extreme ends of each hierarchy. This is because these stimuli were only presented in premise trials and so their estimated voxel representations may differ from all other representations for trivial reasons. Additionally, each LMM included an extensive set of fixed- and random-effect predictors that controlled for nuisance correlations between co-presented wall textures and correlations resulting from the least-squares decomposition (see Methods).

Here, we only report main effects and interactions involving the transitive distance predictor since our hypotheses only concerned these terms. Nonetheless, for completeness, we report all other significant effects in the Supplementary Results. This sections also includes an analysis of hierarchical representations that generalised across the transitive chains learnt on each day of training (i.e., across recent and remote conditions). While no signficnat 'across-hierarchy distance effects' were identified, the supplement details all other main effects and interactions from this analysis.

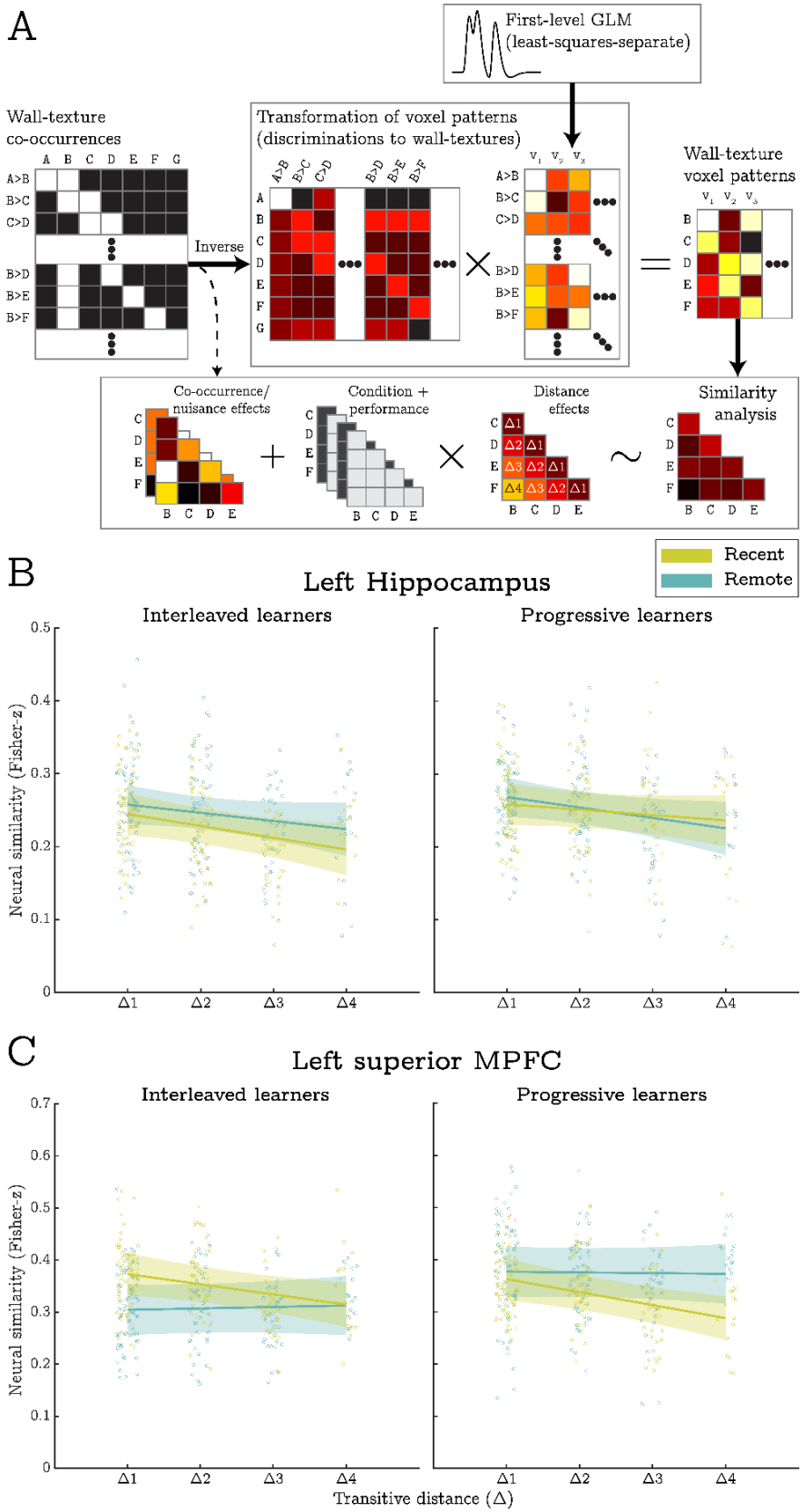


Figure 7. Methods and results for the RSA. A) BOLD responses across voxels (v_1, v_2, etc.) for each in-scanner discrimination (A>B, B>D, etc.) were estimated in a set of 1st level models. These were linearly transformed into representations of specific wall-texture stimuli (A, B, C, etc.) via a least-squares decomposition procedure. Subsequently, BOLD similarity between wall-textures was estimated, Fisher-transformed, and entered into a mixed-effects model that implemented the RSA. Nuisance covariates accounted for trivial correlations between co-presented wall-textures and correlations resulting from the decomposition procedure. Effects of interest modelled the influence of condition, behavioural performance, and transitive distance. B) In the left hippocampus, transitive distance (i.e., the separation between wall-textures) was negatively correlated with BOLD similarity across all conditions. As such, this region appears to encode a structural representation of the transitive hierarchy that is not modulated by training method (i.e., interleaved vs progressive) or session (recent vs remote). C) The left superior MPFC exhibited a distance by session interaction suggesting that structural representations were only expressed for recently learnt contingencies. All shaded error-bars indicate 95% confidence intervals.

In the left hippocampus, we saw a main effect of distance that did not survive our correction for multiple comparisons; $t(652) = 2.61$, $p = .009$. While non-significant, this effect is consistent with our prediction that representational similarity between wall textures should be inversely related to their hierarchical distance (see Figure 7B). Nonetheless, the left hippocampus did produce a significant distance by transitive slope interaction, $t(652) = 3.31$, $p = .001$. Contrary to predictions, this indicated that distance effects were most strongly expressed in participants who had a relatively low (or negative) transitive slope (see Figure 9A). As such, the left hippocampus appears to encode structural task representations when behavioural performance is less indicative of encoding-based generalisation.

In the left superior MPFC, transitive distance was negatively correlated with representational similarity in the recent, but not the remote, conditions; $t(650) = 3.59$, $p < .001$, and $t(650) = 0.113$, $p = .910$ (respectively). This resulted in a significant interaction between transitive distance and training session suggesting the presence of structural representations for recently learnt stimuli alone, $t(650) = 3.00$, $p = .003$ (see Figure 7C). Moreover, this effect was superseded by a 3-way interaction between distance, session and inference accuracy indicating that the strength of structural representations was greatest for participants who did not achieve a high level of inferential performance, $t(650) = 4.24$, $p < .001$ (see Figure 8B).

In the right superior MPFC, we also saw a 3-way interaction between distance, session, and accuracy, $t(652) = 2.90$, $p = .004$. Again, this was suggestive of structural representations in recent condition, but only when generalisation performance was relatively low, and only for learners in the interleaved training condition. Furthermore, the right superior MPFC produced a distance by transitive slope interaction that was superseded by a 3-way interaction between distance, session, and transitive slope, $t(652) = 2.79$, $p =.006$, and $t(652) = 2.96$, $p = .003$ (respectively, see Figure 8C). This highlighted that the expected distance effects were only expressed in the remote condition when participants' behavioural data was heavily indicative of encoding-based generalisations. Importantly, this contrasts with the distance effects identified in the left hippocampus which were strongest when participants behavioural data was *less* indicative of encoding-based generalisation.

Finally, we report a significant 3-way interaction between distance, session, and training method in the left inferior MPFC, $t(648) = 2.99$, $p = .003$. This was mainly driven by distance effects in the progressive training condition. Specifically, we observed the predicted negative correlation between distance and pattern similarity in the recent condition ($t = 1.95$), but the opposing (positive) relationship for progressive learners in the remote condition ($t = 2.40$). Post-hoc tests indicated that the positive distance effect was principally attributable to high levels of similarity between responses to wall textures 'B' and 'F' (i.e., those with the largest transitive distance in the analysis). While such an effect may reflect how the stimuli were being encoded, this pattern of data was entirely unpredicted and so we do not draw any inferences based on this result.

Taken together these findings suggest that the hippocampus may contribute to transitive inference by representing structural memory codes when participants behavioural performance is less typical of encoding-based generalisation performance. In contrast, the superior MPFC appears to perform the same function when participants do appear to rely on encoding-based generalisation mechanisms, especially following short retention intervals.

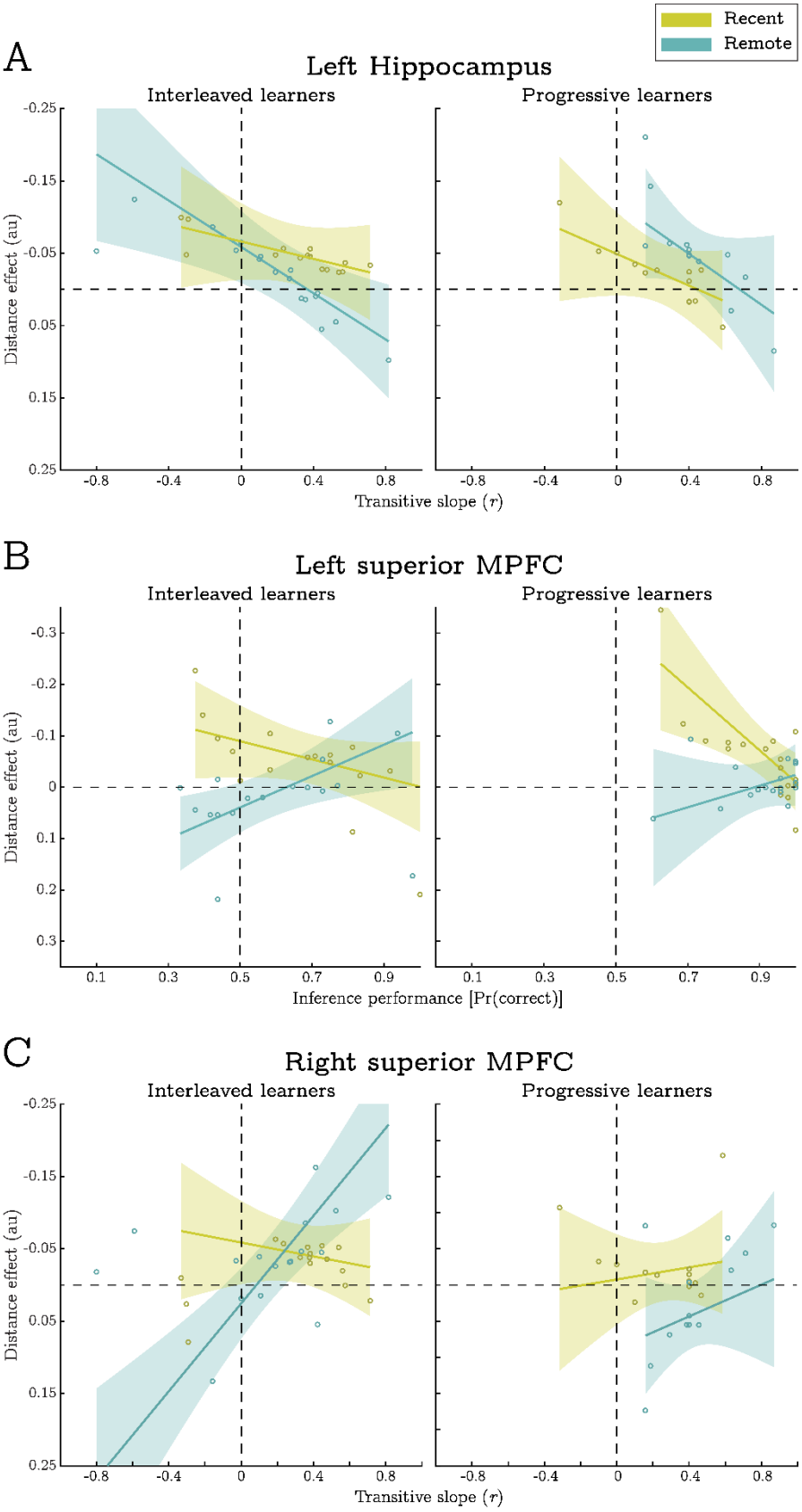


Figure 8. Associations between measures of behavioural performance and the magnitude of distance effects in the RSA. Solid tend lines depict the fitted fixed-effect relationship, while shaded error-bars indicate 95% confidence intervals. Note that negative distances effects (plotted above the dashed horizontal) represent the predicted association between transitive distance and BOLD pattern similarity. Individual datapoints depict participant-specific random slopes for each association. A) Distance effects in the left hippocampus were strongest when participants produced relatively low transitive slopes (less indicative of encoding-based generalisation). B) Distance effects in the left superior MPFC were only significant in the recent condition and were strongest when participants did not achieve high levels of inference performance. C) In contrast to the left hippocampus, distance effects in the right superior MPFC were most evident when participants produced relatively large transitive slopes (most indicative of encoding-based generalisation), yet this association was limited to the remote condition.

Discussion

We sought to determine whether the use of encoding- and retrieval-based generalisation mechanisms are influenced by two factors: 1) the order in which task contingencies are learnt (i.e., interleaved vs progressive training), and 2) whether there has been a period of overnight consolidation.

Our first key finding is that both humans and artificial neural networks with a single hidden unit are better able to make transitive inferences after learning premise pairs progressively. Next, contrary to our hypotheses, model-based analyses of the behavioural data revealed that encoding-based mechanisms were preferentially used across all experimental conditions. Representational similarity analyses of the fMRI data were also suggestive of the use of encoding-based mechanisms. BOLD pattern similarity correlated with the hierarchical distances between stimuli in the hippocampal and medial prefrontal cortices. These effects imply the presence of map-like structural representations that directly express inferred relationships between stimuli in an abstract task space.

Notably, we only saw consistent evidence of these structural codes in the MPFC for recently learnt stimuli and they tended to be most strongly expressed in participants who did not achieve high levels of inferential performance. Furthermore, structural codes in hippocampus and MPFC were differentially related to participants' transitive slopes - a behavioural measure indexing the degree to which performance was suggestive of encoding-based generalisations. Structural codes in the right superior MPFC were strongest when behaviour was most indicative of encoding-based generalisations. The reverse was true in the left hippocampus which exhibited structural codes most strongly when behaviour was less typical of encoding-based mechanisms.

Our behavioural analyses clearly demonstrate that progressive training substantially increases generalisation performance compared with randomly interleaving contingencies. This happens despite comparable accuracy in remembering the directly trained (premise) contingencies that generalisations were based upon. We observed similar effects in an MLP that was specifically structured to allow the learning of compressed representations of the transitive hierarchy. Here, progressive training in MLPs with a single hidden unit boosted successful inference. This happens because progressively ordering task contingences induces an error gradient that favours learning the relative value of all stimuli (A>B>C...; see Results). Progressive training did not, however, confer the same advantage for MLPs with two or more hidden units. This is because these networks were biased toward learning localist representations of the premise pairs that did not encode the higher-order transitive structure.

Using similar MLPs, Flesch et al [42] found that networks initialised with high-variance parameters tended to learn 'lazy' representations of stimuli in a categorisation task that indiscriminately preserved irrelevant features. On the other hand, when the networks were initialised with low-variance parameters, they were biased towards learning low-dimensional (compressed) representations of the input space that linearly encoded task-relevant features. These highly structured, 'rich' representations were learnt slower than lazy ones but were more resilient noise in the input layer. Our findings extend these results by highlighting that progressive training may selectively promote generalisation when representations using rich coding schemes are acquired during training.

As noted in the introduction, previous studies have shown that some generalisations are enhanced by interleaved training when compared to blocked (or progressive) schedules. Zhou et al [43] report that both humans and connectionist model of the hippocampal system show elevated levels of memory integration and property generalisation following interleaved training. It is likely that the specific type of generalisation under test is a major determinant in whether interleaved or blocked/progressive schedules will confer the greatest benefit and further research is needed to account for this.

When different pieces of information are presented in specific blocks, there is a potential for 'catastrophic interference' [27,46]. This is when the introduction of (e.g.) premise pairs or category exemplars towards the end of training results in forgetting for those that were presented earlier. To date, state-of-the-art machine learning applications have circumvented this issue by relying on interleaved training [47]. Our progressive schedule largely avoided interference by ensuring that once a discrimination had been introduced, it was presented in all subsequent epochs. Furthermore, while there is evidence of some interference in MLPs (see Figure 3B), processes that operate in the human brain likely limit the effect of interference [27]. Kirkpatrick et al [47] introduce an updated loss function for gradient decent that protects previously acquired expertise from catastrophic interference. This cost function incorporates a Fisher information matrix describing how critical each parameter is to maintaining performance on previously trained objectives. It is possible that such protections against interference may allow artificial networks to learn task representations that better support generalisation.

Although progressive training improved inference performance in humans, it appeared to have very little effect on the mechanisms used to make inferential judgements. Analyses of both the behavioural and fMRI data suggested that participants principally used encoding-based mechanisms in all experimental conditions. Moreover, the results of our univariate analyses did not meet the predictions of retrieval-based mechanisms. The use of encoding-based strategies in this experiment may be due to how well the premise pairs were learnt. Retrieval-based mechanisms are known to explain inference performance when memories have been acquired in a single episode (i.e., one-shot learning; [5]). Given this, our findings are consistent with proposals that encoding-based mechanisms may be preferentially engaged when generalisations are based on well-leant information [33].

As noted, in line with the use of encoding-based mechanisms, we found evidence that transitive hierarchies were represented in the hippocampus and MPFC in the form of structural codes. Various models of memory generalisation are inspired by how representations of physical space may be applied to support inferences in abstract, non-spatial tasks [7,48,49]. Our findings support these models: whilst training only involved pairwise presentations of stimuli, participants encoded generalised relationships between all stimuli in a map-like representation. Importantly, this was the case following both interleaved and progressive training, with no detectable differences between conditions. In contrast to the MLPs therefore, it appears that humans have a robust means of learning hierarchical structures regardless of the training schedule.

Despite the presence of structural representations in both progressive and interleaved learners, we observed large differences in inference performance between these groups. Furthermore,

structural representations in the left and right superior MPFC were most strongly expressed in participants who did not achieve high levels of inference performance – perhaps because those who found the inferences most difficult tended to activate MPFC representations for longer and/or less efficiently. Indeed, BOLD patterns in the MPFC exhibited significant distance effects when participants performed at chance level on the inference task (see Figure 8B). It therefore appears that merely having structural representations is not sufficient for good inference performance. We speculate that progressive exposure may facilitate inferences whenever they depend on structural representations. This may explain why some types of generalisation benefit from blocked/progressive exposure, while forms of generalisation (not reliant on structural codes), do not, e.g., [23–25,43].

Our findings of structural representations in the absence of above-chance inference performance is incompatible with models that propose knowing the relative value of stimuli is all that is needed to build a hierarchical task representation and make novel inferences (e.g., [6], and the MLPs studied above). Nevertheless, other models can account for this observation. Neural codes postulated by both the Tolman-Eichenbaum Machine (TEM) and the successor representation (SR) model dissociate the learned values of stimuli from structural relationships between them [7,48,49]. In the TEM, structural information derived from previous experience is bound to sensory codes in the hippocampus via a fast Hebbian learning rule. However, the ability to use these representations for transitive inference depends on additional path-integration steps that may bottleneck performance. Similarly, SRs can encode the distance between all stimuli in a transitive hierarchy based on knowledge of which stimuli were presented in the same premise pairs (see Supplementary Figure 4). However, in order to support transitive inference, SRs must be combined with a representation encoding the average reward returned by each stimulus.

We also found evidence of a subtle dissociation between structural representations in the left hippocampus and right superior MPFC. Encoding- and retrieval-based generalisation mechanisms make divergent predictions regarding the correlation between transitive distance (i.e., the separation between items in a hierarchy) and inference performance, what we refer to as the 'transitive slope'. Whereas encoding-based models predict a positive transitive slope (performance increases with transitive distance), retrieval-based models predict a negative transitive slope. Overall, participants tended to produce positive transitive slopes strongly suggestive of encoding-based generalisations; a finding that was backed up by our computational modelling results. However, we also tested whether individual difference in transitive slopes covaried with the strength of structural representations in each ROI. We predicted that structural representations (themselves suggestive of encoding-based mechanisms) would be principally expressed in participants with positive transitive slopes. This was indeed the case in the right superior MPFC yet, the left hippocampus showed an opposing relationship; a significant distance effect that was strongest for participants who produced negative transitive slopes (see Figure 8).

We are unable to fully interpret these results as they were not predicted a priori and because the association between transitive slope and the MPFC distance effect was not consistent across conditions (only being evident in the interleaved, remote condition). However, our results do suggest that certain participants preferentially expressed structural representations in distinct brain regions and that this predicts behavioural performance. As such, the

hippocampal and medial prefrontal effects may be associated with contrasting generalisation processes. Further work will be needed to explicitly test this.

Aside from a slight decrease in response latencies on both premise and inferred trials, we did not observe any benefit of overnight consolidation on inference performance (in either training condition). This finding is in contrast to a similar test by Ellenbogen et al [38] who found that inference performance increased after a period of sleep. We also did not support our hypothesis that consolidation would bias the use of encoding-based inference mechanisms. Indeed, contrary to this, we found that structural representations in the superior MPFC became less evident for transitive hierarchies that were learnt 24 hours before testing. While it is possible that this effect reflects consolidation of structural representations outside of the MPFC, or more efficient processing within the region, additional research will be needed to clarify the role of consolidation in memory generalisations.

In summary, we show that progressive training dramatically improves transitive inference and that humans tend to use encoding-based mechanisms to inform inferential judgements based on well-learnt contingencies. Both the hippocampus and MPFC encode structural representations of transitive hierarchies, yet the presence of these representations appears to be insufficient for successful inference. Furthermore, we found that representations in the hippocampus and MPFC are associated with different patterns of behaviour suggesting that they may contributed to generalisation in different ways. Taken together, these findings provide strong support for encoding-based models that predict map-like structural representations underpin spatial and non-spatial generalisations.

Methods

Participants

Right-handed participants were recruited from the University of Sussex, UK. All gave written informed consent and were reimbursed for their time. Participants had either normal or corrected-to-normal vision and reported no history of neurological or psychiatric illness. During the study, they were randomly assigned to one of the two between-subject conditions (i.e., the interleaved or progressive training conditions) such that there were an equal number of useable datasets in each. Data from 5 participants could not be included in the final sample because of problems with fMRI data acquisition (one participant), excess of motion-related artifacts in the imaging data (three participants), and a failure to respond during the in-scanner task (one participant). After these exclusions, the final sample included 34 participants (16 females) with a mean age of 25.9 years (SD = 4.60 years). The study was approved by the Brighton and Sussex Medical School's Research Governance and Ethics Committee.

Behavioural tasks

Pre-scanner training

We developed a reinforcement learning task designed to train participants on pairwise discriminations before scanning. Two different versions of the task were produced so that each participant could be trained on two occasions; once immediately prior to scanning (recent condition), and once 24 hours before scanning (remote condition).

Unreal Development Kit (Epic Games) was used to generate a number of unique scenes within a first-person virtual environment (see Figure 1A for examples). On each trial, a scene depicted two buildings positioned equidistantly from a start location. One building concealed a pile of virtual gold (reinforcement), yet the only features that predicted the rewarded location were the wall textures rendered onto the towers of each building. Participants were tasked with learning which wall-textures predicted reward in each scene and selecting them in order to gain as much reward as possible.

In total, seven unique wall textures were used in each version of the task. During training, these were combined to generate 6 binary discriminations (e.g., A>B, B>C, etc.) that implied a 1-dimensional transitive hierarchy (A>B>C>D>E>F>G, where each letter denotes a unique wall texture; see Figure 1B). As such, every wall texture could be assigned a scalar value representing its utility in predicting reward. Importantly, each wall texture was rendered onto the left and right buildings an equal number of times to ensure that non-target strategies (e.g., always selecting the building on the left) would not result in above chance performance.

Figure 1 presents a schematic of the training schedule for participants in either the interleaved or progressive learning conditions. All trials initially depicted the participant at the start location, in front of two buildings, for up to 3 seconds. During this time participants were required to select the building they believed contained the gold via a left/right button press (decision period). Immediately following a response, a 4-second animation was played showing the participant approaching their chosen building and opening its central door to reveal whether or not it contained gold (feedback period). If no response was made within the 3-

second response window, a 4-second red fixation cross was shown in place of the feedback video.

For participants in the interleaved learning condition, all discriminations were presented in a pseudorandom order such that there was a uniform probability (1/6) of encountering any one discrimination on any particular trial (see Figure 1C). For participants in the progressive learning condition, the task was composed of 6 sequentially presented epochs of different lengths which gradually introduced each discrimination one-by-one. The first epoch exclusively trained the discrimination at the top of the transitive hierarchy (A>B) across 17 trials. The second epoch involved an additional 14 trials of the A>B discrimination but also introduced the next-highest discrimination (B>C) across 20 trial (~59%). This pattern continued down the hierarchy such that, after a discrimination had been introduced, the number of times it was tested in subsequent epochs linearly decreased but remained above zero so that all discriminations were tested in the final epoch (see Figure 1D). Full details of this training procedure are provided on the Open Science Framework (https://osf.io/uzyb7/).

Regardless of the learning condition participants were assigned to, all pairwise discriminations were tested 60 times each by the end of the training procedure (i.e., 360 trials in total, ~37 minutes). Before the first training session, participants were briefed on the experimental procedure and told that the wall textures were the only features that predicted reward. They were not given any other details regarding the number or type of discriminations.

In-scanner task

Following the second training session, participants were tested on the 6 directly trained (premise) discriminations, and a set of 6 transitive inferences (e.g., B>D), whilst being scanned (see Figure 1B). This tapped knowledge acquired during both of the preceding training sessions. Note that the inferred discriminations did not involve wall-texture stimuli from the ends of each hierarchy (i.e., A and G). This is because discriminations involving these terminal stimuli may be made by applying simple feature-based response policies (i.e., "Always select A", "Always avoid G"), without the need to use a generalised value function.

Similar to the training task, all in-scanner trials initially depicted the participant at a start location in front of two buildings. Participants were instructed to select the building that they believed contained virtual gold based on what they had learned during training. Guesses were strongly encouraged if the participant was not confident. Unlike the previous training sessions, the image of the start location persisted on-screen throughout the 3-second response window regardless of when/whether a response was made. Importantly, no feedback videos were shown during the in-scanner task meaning that participants could not (re-)learn the contingencies via external feedback. Following the response window, a fixation cross was displayed centrally for 3.5 seconds before the next trial commenced.

The in-scanner task tested each premise/inferred discrimination 8 times (the higher value wall-texture appeared on the left-hand building in exactly 50% of trials). As such, the task involved a total of 192 trials: 2 trial types (premise vs inferred) x 2 training sessions (recent vs remote) x 6 unique discriminations x 8 repetitions. Additionally, we included 16 null events (lasting 6.5 seconds each) in order to facilitate the estimation of a resting baseline. All trials were presented in a single run and progressed in a pseudorandom order that was determined by an

optimization procedure to enable maximally efficient decoding of trial-specific BOLD responses (https://osf.io/eczjf/).

Refresher task

As noted, participants were trained on 2 independent sets of premise discriminations in pre-scanner training sessions that occurred approximately 24 hours apart. The wall-textures used in each session were counterbalanced across participants. Just before entering the scanner, participants practised each of the directly trained discriminations in a short refresher task. This ran identically to the training tasks but only included 12 trails of each discrimination (lasting approximately 15 minutes). The refresher was not intended to act as an additional training phase but served to remind participants of the appearance of all wall textures so that they were easily identifiable.

Multilayer perceptron analyses

To explore whether the order in which multiple discriminations are trained can have an effect on the representations that are learned, we examined interleaved and progressive training in 6 different multilayer perceptrons (MLPs) with similar architectures. Each MLP consisted of 3 fully connected layers: 1) an input layer with 7 binary inputs, 2) a linearly activated hidden layer with between 1 and 6 biased hidden units, and 3) a SoftMax output layer with 7 biased outputs (see Figure 2). These architectures were chosen to be as simple as possible while still having the ability to learn parameters that would support near-perfect performance on both premise and inferred discriminations (see https://osf.io/ps3ch/).

Before training, the network was initialised with uniformly distributed random weights and biases sampled from the interval [-0.5, 0.5]. During training, we presented the network with 6 premise discriminations that had the same structure as those in the main behavioural task (i.e., A>B, B>C, etc.). This was done by setting the activity of two input units (e.g., A & B) to a value of 1 and setting all other input units (e.g., C, D, E, F, & G) to a value of 0. After each forward pass, a cross-entropy cost-function was used to quantify the error between the actual and target activity patterns on the output layer. The target patterns respected the contingencies described previously such that, when (e.g.) 'F>G' was presented on the input, output 'F' should have been maximally active relative to all other outputs. Backpropagation was then used to adjust all network weights and biases with a learning rate of 0.1. This rate parameter was chosen to allow for sufficient levels of learning following 3,600 training trials (10 times the number used in the pre-scanner task, see https://osf.io/uzyb7/). However, to examine the sensitivity of our results to this rate, we ran a supplementary analysis that involved training the MLPs with a learning rate of 0.01 over 36,000 trials. This yielded an identical pattern of results to those presented above (see Supplementary Results).

The MLPs were trained using both interleaved and progressive schedules across many independent iterations. As before, interleaved training involved presenting all 6 discriminations in a pseudorandom order such that there was a uniform probability of each discrimination occurring on any particular trial (10,000 iterations per MLP architecture). Progressive training involved 6 sequential epochs of different lengths which gradually introduced each discrimination one-by-one (10,000 iterations per architecture). The only difference between the training schedules used in the behavioural task (described above) and those used to train the MLP was the number of trials.

After training, we tested the MLP on the premise and inferred discriminations detailed in Figure 1B, quantifying performance using both the cross-entropy cost function and 2-alternative Luce choice (2-ALC) decision rule:

$$\Pr(correct) = \frac{\exp\left(\frac{a_+}{\tau}\right)}{\exp\left(\frac{a_+}{\tau}\right) + \exp\left(\frac{a_-}{\tau}\right)} \qquad Eq.\ 1$$

Where, a_+ denotes the activity level of the target, a_- denotes the activity level of the non-target alternative, and τ is a constant temperature parameter fixed at a value of 2. Unlike the cross-entropy measure, 2-ALC performance allowed us to examine whether the activity of the target output unit was greater than the activity of the concurrently presented non-target (which tended to be activated by its corresponding input).

Analysis of in-scanner performance

We used a generalised-linear mixed-effects model (GLMM) to characterise the pattern of correct vs incorrect responses during the in-scanner task. Specifically, this tested the relationship between response accuracy and 3 binary-coded fixed-effect predictors: *1)* trial type (premise vs inferred), *2)* training method (interleaved vs progressive), and *3)* training session (recent vs remote). Additionally, a continuous (mean-centered) fixed-effect predictor accounted for the effect of transitive distance on inference trials. All possible interactions between these variables were included meaning that the model consisted of 12 fixed-effects coefficients in total (including the intercept term). We also included random intercepts and slopes for each within-subject variable (grouped by participant), and random intercepts for each unique wall-texture discrimination (to account for any stimulus specific effects). Covariance components between random effects were fully estimated from the data.

The outcome variable was the number of correct responses to the 8 repeated trials for each in-scanner discrimination. This outcome was modelled as a binomial process such that parameter estimates encoded the probability of a correct response on a single trial, $Pr(correct)$. To avoid any biases resulting from failures to respond (1.81% of trials on average), we resampled missing responses as random guesses with a 50% probability of success. The model used a logit link-function and was estimated via maximum pseudo-likelihood using the Statistics and Machine Learning toolbox in MATLAB R2020a (The MathWorks Inc.).

In addition to the model of response accuracy, we estimated a similar GLMM that characterised behavioural patterns in response latencies (correct trials only). This GLMM used exactly the same fixed- and random-effect predictors as above. Response times were modelled using a log link-function and the distribution of observations was parameterised by the gamma distribution. As before, the model was fit via maximum pseudo-likelihood in MATLAB.

Models of human inference performance

We predicted that inference performance would vary by experimental condition due to differences in the way inferences were made, but not because of any differences in performance for the directly trained discriminations. To test this, we produced two competing models of the behavioural data referred to as the AND and OR models. Both of these attempted to

predict participants' inference performance given responses to the directly trained discriminations alone.

The AND model assumes that correctly inferring a non-trained discrimination (e.g., 'B>E') involves retrieving all the directly learnt response contingencies required to reconstruct the relevant transitive hierarchy (e.g., 'B>C' and 'C>D' and 'D>E'). We refer to these directly trained discriminations as "mediating contingencies". As such, this model captures a common assumption of retrieval-based models of generalisation.

In contrast, the OR model assumes that correctly inferring a non-trained discrimination requires retrieving a unified representation which may be activated by recalling any one of the mediating contingencies. Accordingly, each mediating contingency is sufficient to deduce the hierarchical structure or relative value of stimuli and so inferences are easiest when there is a large 'distance' between features. This model therefore captures a common assumption of encoding-based models of generalisation, that contingencies are retrieved non-independently.

Note, these models are not intended to be process models of how humans solve the task. They are merely intended to describe the data and test whether the behaviour in each condition better accord with general predictions of encoding and retrieval models.

To formalise both models, we first computed a likelihood function describing plausible values for the probability of correctly retrieving each premise discrimination (r_p, where the index p denotes a specific premise discrimination). To do this we assume the probability of observing k_p correct responses, to the $n = 8$ test trials, depends on a joint binomial process involving r_p and, if retrieval is not successful, a random guess that yields a correct response with a probability of 0.5:

$$\Pr(k_p \mid n, r_p) = \sum_{m=0}^{k_p} \frac{n!}{m!\,(k_p - m)!\,(n - k_p)!}\, r_p^m (1 - r_p)^{n-m} (1/2)^{n-m} \qquad \textit{Eq. 2}$$

From this, the likelihood function for the parameter r_p (denoted $L(r_p|k_p, n)$) is given by dividing out a normalising constant, $c(k_p|n)$, computed by numerical integration:

$$L(r_p \mid k_p, n) = \frac{\Pr(k_p \mid n, r_p)}{c(k_p \mid n)} \qquad \textit{Eq. 3}$$

Where:

$$c(k_p \mid n) = \int_0^1 \Pr(k_p \mid n, r)\, dr \qquad \textit{Eq. 4}$$

Supplementary Figure 5A displays the likelihood function for r_p under different values of k_p. Based on these likelihoods, we then sampled random values of r_p for each premise discrimination that mediated the generalisation trails. To do this, we used an inverse transform sampling method where a value of r_p was selected such that the cumulative likelihood up to that value (i.e., $\int_0^{r_p} L(x|k_p, n)\, dx$) was equal to a unique, uniformly distributed random number in the range [0,1] (see Supplementary Figure 5B). We denote one set of sampled r_p values that

mediate an inferred discrimination, R_i, where the index i denotes a specific non-trained discrimination, and the number of elements in R_i is equal to the transitive distance.

As noted, the AND model assumes that a non-trained discrimination depends on successfully retrieving all the mediating contingencies. Given the sampled values in R_i, we therefore computed the probability this for each non-trained discrimination (denoted g_i^{AND}):

$$g_i^{AND} = \kappa \prod_{r_p \in R_i} r_p \qquad \textit{Eq. 5}$$

The constant term κ is a scalar value in the range [0,1] that determines the probability of engaging in memory-guided generalisations rather than simply guessing. This parameter was fit to the inference data by a bounded nonlinear optimiser ("fmincon", MATLAB Optimization Toolbox, R2020a). Specifically, given the sampled values in R_i, the optimiser was tasked with finding a single value of κ across all inference trials from a particular participant/condition that minimised a cross-entropy term (H) relating model predictions to the observed inference data (see https://osf.io/mcyuv/). H was based on equation 8 and describes the mean log-probability of observing k_i correct responses to all inference trials.

The OR model assumes that performance on the non-trained discriminations depend on successfully retrieving a unified representation that can be activated by recalling any one of the mediating contingencies. Given the sampled values in R_i, we computed the probability of successful inference under the OR model (g_i^{OR}) as follows:

$$g_i^{OR} = \kappa \left(1 - \prod_{r_p \in R_i} (1 - r_p) \right) \qquad \textit{Eq. 6}$$

Note that the value of κ was estimated as above, but independently for each model. We then computed model-derived probabilities for the number of correct inference responses k_i to the $n = 8$ inference trials (similar to *Eq. 1*):

$$\Pr(k_i \mid n, g_i) = \sum_{m=0}^{k_i} \frac{n!}{m!\,(k_i - m)!\,(n - k_i)!}\, g_i^m (1 - g_i)^{n-m} (1/2)^{n-m} \qquad \textit{Eq. 7}$$

In order to estimate the expected distribution of $Pr\,(k_i|n, g_i)$ for each type of inference, we repeatedly sampled sets of R_i over 10000 iterations using the aforementioned likelihood functions (*Eq. 2*). The cross entropy H of each model was then taken as the mean negative log probability over all I inferences in a particular condition, from a particular participant:

$$H = -\frac{1}{I} \sum_{i=1}^{I} \log\big(\Pr(k_i \mid n, g_i)\big) \qquad \textit{Eq. 8}$$

To analyse condition-dependent differences in the cross-entropy statistics, we entered them into a GLMM with 3 binary-coded fixed effect predictors: *1)* inference model (AND vs OR), *2)* training method (interleaved vs progressive), and *3)* training session (recent vs remote). All possible interactions between these predictors were also included. The GLMM further

contained random intercepts and slopes of each fixed effect (grouped by participant), with a covariance pattern that was fully estimated form the data. Cross-entropy was modelled using a log link-function and the distribution of observations was parameterised by the gamma distribution. The model was fitted via maximum pseudo-likelihood in MATLAB.

MRI acquisition

All functional and structural volumes were acquired on a 1.5 Tesla Siemens Avanto scanner equipped with a 32-channel phased-array head coil. T2∗-weighted scans were acquired with echo-planar imaging (EPI), 34 axial slices (approximately 30° to AC-PC line; interleaved) and the following parameters: repetition time = 2520 ms, echo time = 43 ms, flip angle = 90°, slice thickness = 3 mm, inter-slice gap = 0.6 mm, in-plane resolution = 3 × 3 mm. The number of volumes acquired during the in-scanner task was 537. To allow for T1 equilibrium, the first 3 EPI volumes were acquired prior to the task starting and then discarded. Subsequently, a field map was captured to allow the correction of geometric distortions caused by field inhomogeneity (see the MRI pre-processing section below). Finally, for purposes of co-registration and image normalization, a whole-brain T1-weighted structural scan was acquired with a $1mm^3$ resolution using a magnetization-prepared rapid gradient echo pulse sequence.

MRI pre-processing

Image pre-processing was performed in SPM12 (www.fil.ion.ucl.ac.uk/spm). This involved spatially realigning all EPI volumes to the first image in the time series. At the same time, images were corrected for geometric distortions caused by field inhomogeneities (as well as the interaction between motion and such distortions) using the Realign and Unwarp algorithms in SPM [50,51]. All BOLD effects of interest were derived from a set of first-level general linear models (GLM) of the unsmoothed EPI data in native space. Here, we estimated univariate responses to the 24 discriminations (i.e., 6 premise + 6 inferred, from each day) using the least-squares-separate method [52]. To do this, a unique GLM was constructed for each discrimination such that one event regressor modelled the effect of that discrimination while a second regressor accounted for all other discriminations. As such, one beta estimate from each model encoded the BOLD response for a particular discrimination. These models also included the following nuisance regressors: 6 affine motion parameters, their first-order derivatives, squared values of the motion parameters and derivatives, and a Fourier basis set implementing a 1/128 Hz high-pass filter.

For the analysis of univariate BOLD activity, the 24 beta estimates related to each discrimination were averaged within regions of interest and entered into a linear mixed-effects regression model (see 'Analysis of univariate BOLD' below). For the RSA, these beta estimates were linearly decomposed into voxel-wise representations of each wall texture in the reinforcement learning task (n=7 per transitive chain). This decomposition involved multiplying the 24 beta values from a given voxel with a 15*24 transformation matrix that encoded the occurrence of each wall texture across discriminations (see Figure 7A). The first 7 outputs of this transformation related to the recently learnt wall textures, the second 7 outputs related to remotely learnt wall textures, and the final output encoded overall BOLD differences between premise and inferred trials (a nuisance term that was not included in any further analysis). Importantly, BOLD representations for wall textures 'A' and 'G' may have trivially differed from all other patterns since these stimuli were only presented in premise

trails and so were only shown alongside one other wall texture ('B' and 'F', respectively). As such, similarity scores involving the 'A' and 'G' patterns were excluded from the RSA leaving only scores related to 'B', 'C', 'D', 'E', and 'F'.

Regions of interest

Our a priori ROIs included the hippocampus, entorhinal cortex, and medial prefrontal cortices (bilaterally). For each participant, we generated 8 binary masks to represent these ROIs in native space (separately in each hemisphere). This was done by transforming group-level masks in MNI space using the inverse warp utility in SPM12. For the hippocampus, we used an MNI mask provided by Ritchey et al [53]. The entorhinal masks were derived from the maximum probability tissue labels provided by Neuromorphometrics Inc. Finally, 4 separate masks corresponding to the left and right inferior and superior MPFC were defined from a parcellation that divided the cortex into 100 clusters based on a large set of resting-state and task-based functional imaging data [54]. Normalised group averages of each ROI are shown in Supplementary Figure 6 and are available at https://osf.io/tvk43/.

Analysis of univariate BOLD

Univariate BOLD effects were investigated within a set of linear mixed-effects models (LMMs). These characterised condition-dependent differences in ROI-averaged beta estimates that derived from a first level GLM of the in-scanner task (see 'MRI pre-processing' above). The LMMs included 3 binary-coded fixed-effect predictor variables: *1)* trial type (premise vs inferred), *2)* training method (interleaved vs progressive), and *3)* training session (recent vs remote). Additionally, 3 mean-centred continuous fixed-effects were included: *i)* inference accuracy (averaged across discriminations, per participant, per session), *ii)* 'transitive slope' (the simple correlation between transitive distance and accuracy, per participant, per session), and *iii)* transitive distance per se (applied to inference trials only). All interactions between these variables were also included (excluding interactions between the continuous predictors) meaning that the model consisted of 28 fixed-effects coefficients in total (including the intercept term). We also included random intercepts and slopes for each within-subject fixed-effect (group by participant), as well as random intercepts for each unique wall-texture discrimination (both grouped and ungrouped by participant). Covariance components between random effects were fully estimated from the data. The model used an identity link-function and was estimated via maximum likelihood in MATLAB.

Representational similarity analysis

Condition-dependent differences in the similarity between wall-texture representations were also investigated using LMMs. To generate these models, we first estimated BOLD similarity in each ROI by producing a pattern-by-pattern correlation matrix from the decomposed wall-texture representations (including wall textures 'B' to 'F' only, see 'MRI pre-processing'). The resulting correlation coefficients were then Fisher-transformed before being entered into each LMM as an outcome variable. These models were structured to predict the Fisher-transformed similarity scores as a function of various predictors of interest. As above, covariance components between random effects were fully estimated from the data. The models used an identity link-function and were estimated via maximum likelihood in MATLAB.

Critically, the temporal structure of the in-scanner task and least-squared decomposition procedure introduced nuisance correlations between wall texture representations. To account for these, we derived two predictor variables of no-interest and used them to model nuisance effects in each LMM described below. The first predictor accounted for trivial correlations resulting from shared sources of noise across co-presented wall textures. This was taken as the Fisher-transformed correlation between binary vectors encoding whether each pair of wall textures were presented in same the in-scanner trails. Across analyses, this predictor invariably accounted for a significant amount of variance in the similarity scores, $r \in [0.105,\ 0.277]$.

The second predictor of no-interest modelled nuisance correlations attributable to the temporal proximity of trials during the in-scanner task and the pattern decomposition procedure itself. To estimate these correlations, we simulated independent, normally distributed voxel patterns for all wall textures across a large number of iterations, mixed them together in accordance with the trial timings for each subject, and re-estimated the voxel patterns using the least-squares-separate decomposition procedure outlined above. The predictor of no-interest was then taken as the mean Fisher-transformed correlation between simulated wall-textures across iterations. Using this procedure, we aimed to model the effect of fMRI repetition suppression by parametrically modulating the simulated BOLD responses such that repeated presentations evoked an attenuated response. This adjustment was based on Fritsche et al [55] who report that repetition suppression effects in the parahippocampal cortex yield a BOLD attenuation of approximately 23% following a 100ms delay, and 10% following a 1 second delay. Given this, we applied repetition suppression effects assuming an exponential recovery from adaptation over time. Our simulations showed that the presence of such effects did not notably bias BOLD pattern recovery. Furthermore, the temporal signal-to-noise ratio of the fMRI signal had little effect on the correlational structure of the recovered BOLD patterns. Nonetheless, the simulations did reveal some minor nuisance correlations that tended to account for a significant amount of variance in the pattern similarity scores, $r \in [0.001,\ 0.183]$.

Within-hierarchy RSA

The first set of similarity analyses tested for differences between wall-texture representations from the same transitive hierarchy. Similar to the models described previously, these LMMs included 5 fixed-effect predictors of interest: *1)* training method, *2)* training session, *3)* transitive distance, *4)* inference accuracy, and *5)* transitive slope. All interactions between these variables were also included (excluding interactions between inference accuracy and transitive slope). The LMMs also included random intercepts and slopes for each effect derived from a repeated measure variable. Finally, the models comprised an extensive set of random intercepts and slopes (grouped by participant) that accounted for all dependencies between pattern correlations (see https://osf.io/w5pvh/)

Across-hierarchy RSA

The second set of similarity analyses tested for differences between wall-texture representations from different transitive hierarchies (i.e., those learnt in different training sessions). These LMMs included 4 fixed-effect predictors of interest: *1)* training method, *2)* transitive distance, *3)* inference accuracy, and *4)* transitive slope. Note that the effect of training session was not included as did not apply when examining the similarity between representations learnt in different sessions. As before, the effect of transitive distance accounted for comparisons between wall-textures at different levels of the hierarchy. However, in this set of models, the

distance predictor included an additional level ($\Delta 0$), corresponding to comparisons between wall-textures at the same hierarchical level. The across-hierarchy LMMs included the same nuisance variables and random-effects as in the within-hierarchy RSA (https://osf.io/kgcz2/).

Statistical validation and inference

To ensure that each linear mixed-effects regression model was not unduly influenced by outlying data points, we systematically excluded observations that produced unexpectedly large residual values above or below model estimates. The threshold for excluding data points was based on the number of observations in each model rather than a fixed threshold heuristic. We chose to do this because the expected range of normally distributed residual values depends on the sample size which varied between models. Across all linear models, we excluded data points that produced an absolute standardised residual larger than the following cut-off threshold (z):

$$z = \Phi^{-1}\left(\frac{1}{2}\left(1 - 2^{-\frac{1}{n}}\right)\right) \qquad \textit{Eq. 9}$$

Where, Φ^{-1} is the Probit function, and n is the sample size. This threshold was chosen as it represents the bounds of a standard normal distribution that will contain all n normally distributed data points of a random sample, 50% of the time. The value of z is approximately 2.7 when $n = 100$ and 3.4 when $n = 1000$. After excluding outliers, Kolmogorov–Smirnov tests indicated that the residuals were normally distributed across all the linear mixed-effects models (across analyses, the proportion of excluded outliers ranged between 0 and 0.941%; see https://osf.io/akt3g/). Additionally, visual inspection of scatter plots showing residual versus predicted scores indicated no evidence of heteroscedasticity, non-linearity or overly influential datapoints (see https://osf.io/jbuh5/).

All p-values are reported as two-tailed statistics. Unless otherwise stated, we only report significant effects from the fMRI analyses that survive a Bonferroni correction for multiple comparisons across our 8 regions of interest.

Acknowledgments

This work was supported by a European Research Council Consolidator Grant awarded to C.M.B. (Project: EVENTS). We are grateful to Prof Neil Burgess and Prof Anil Seth for feedback on early drafts of the article manuscript.

References

1. Potts GR. Information processing strategies used in the encoding of linear orderings. J Verbal Learning Verbal Behav. 1972;11: 727–740. doi:10.1016/S0022-5371(72)80007-0

2. Potts GR. Storing and retrieving information about ordered relationships. J Exp Psychol. 1974;103: 431–439. doi:10.1037/h0037408

3. Kumaran D, McClelland JL. Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. Psychol Rev. 2012;119: 573–616. doi:10.1037/a0028681

4. Koster R, Chadwick MJ, Chen Y, Berron D, Banino A, Düzel E, et al. Big-Loop Recurrence within the Hippocampal System Supports Integration of Information across Episodes. Neuron. 2018;99: 1342-1354.e6. doi:10.1016/j.neuron.2018.08.009

5. Banino A, Koster R, Hassabis D, Kumaran D. Retrieval-Based Model Accounts for Striking Profile of Episodic Memory and Generalization. Sci Rep. 2016;6: 1–15. doi:10.1038/srep31330

6. Kumaran D, Banino A, Blundell C, Hassabis D, Dayan P. Computations Underlying Social Hierarchy Learning: Distinct Neural Mechanisms for Updating and Representing Self-Relevant Information. Neuron. 2016;92: 1135–1147. doi:10.1016/j.neuron.2016.10.052

7. Whittington JCR, Muller TH, Mark S, Chen G, Barry C, Burgess N, et al. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. Cell. 2020;183: 1249-1263.e23. doi:10.1016/j.cell.2020.10.024

8. Ciranka S, Linde-Domingo J, Padezhki I, Wicharz C, Wu CM, Spitzer B. Asymmetric reinforcement learning facilitates human inference of transitive relations. Nat Hum Behav. 2022; 1–36. doi:10.1038/s41562-021-01263-w

9. Zeithamova D, Dominick A, Preston AR. Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. 2012;75: 168–179. doi:10.1016/j.neuron.2012.05.010

10. Schapiro AC, Turk-Browne NB, Botvinick MM, Norman KA. Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. Philos Trans R Soc B Biol Sci. 2017;372. doi:10.1098/rstb.2016.0049

11. Ferreira CS, Charest I, Wimber M. Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. Neuroimage. 2019;201: 115996. doi:10.1016/j.neuroimage.2019.07.009

12. Morton NW, Schlichting ML, Preston AR. Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. Proc Natl Acad Sci U S A. 2020;117: 29338–29345. doi:10.1073/pnas.1912338117

13. Kaplan R, Friston KJ. Hippocampal-entorhinal transformations in abstract frames of reference. Entorhinal Transform Abstr Fram Ref. 2018; 414524. doi:10.1101/414524

14. Tavares RM, Mendelsohn A, Grossman Y, Williams CH, Shapiro M, Trope Y, et al. A

Map for Social Navigation in the Human Brain. Neuron. 2015;87: 231–243. doi:10.1016/j.neuron.2015.06.011

15. Kumaran D, Melo HL, Duzel E. The Emergence and Representation of Knowledge about Social and Nonsocial Hierarchies. Neuron. 2012;76: 653–666. doi:10.1016/j.neuron.2012.09.035

16. Zhou J, Gardner MPH, Stalnaker TA, Ramus SJ, Wikenheiser AM, Niv Y, et al. Rat Orbitofrontal Ensemble Activity Contains Multiplexed but Dissociable Representations of Value and Task Structure in an Odor Sequence Task. Curr Biol. 2019;29: 897-907.e3. doi:10.1016/j.cub.2019.01.048

17. Kaefer K, Nardin M, Blahna K, Csicsvari J. Replay of Behavioral Sequences in the Medial Prefrontal Cortex during Rule Switching. Neuron. 2020;106: 154-165.e6. doi:10.1016/j.neuron.2020.01.015

18. Baram AB, Muller TH, Nili H, Garvert MM, Behrens TEJ. Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. Neuron. 2021;109: 713-723.e7. doi:10.1016/j.neuron.2020.11.024

19. Park SA, Miller DS, Nili H, Ranganath C, Boorman ED. Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. Neuron. 2020;107: 1226-1238.e8. doi:10.1016/j.neuron.2020.06.030

20. Luyckx F, Nili H, Spitzer B, Summerfield C. Neural structure mapping in human probabilistic reward learning. Elife. 2019;8: 1–19. doi:10.7554/eLife.42816

21. Roediger HL, Karpicke JD. Test-enhanced learning: Taking memory tests improves long-term retention. Psychol Sci. 2006;17: 249–255. doi:10.1111/j.1467-9280.2006.01693.x

22. Birnbaum MS, Kornell N, Bjork EL, Bjork RA. Why interleaving enhances inductive learning: The roles of discrimination and retrieval. Mem Cogn. 2013;41: 392–402. doi:10.3758/s13421-012-0272-7

23. Kornell N, Bjork RA. Learning concepts and categories: Is spacing the "enemy of induction"? Psychol Sci. 2008;19: 585–592. doi:10.1111/j.1467-9280.2008.02127.x

24. Kang SHK, Pashler H. Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. Appl Cogn Psychol. 2012;26: 97–103. doi:10.1002/acp.1801

25. Carvalho PF, Goldstone RL. The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. Psychon Bull Rev. 2014;22: 281–288. doi:10.3758/s13423-014-0676-4

26. Noh SM, Yan VX, Bjork RA, Maddox WT. Optimal sequencing during category learning: Testing a dual-learning systems perspective. Cognition. 2016;155: 23–29. doi:10.1016/j.cognition.2016.06.007

27. Flesch T, Balaguer J, Dekker R, Nili H, Summerfield C. Comparing continual task learning in minds and machines. Proc Natl Acad Sci U S A. 2018;115: E10313–E10322. doi:10.1073/pnas.1800755115

28. Carvalho PF, Goldstone RL. Putting category learning in order: Category structure and

temporal arrangement affect the benefit of interleaved over blocked study. Mem Cogn. 2013;42: 481–495. doi:10.3758/s13421-013-0371-0

29. Markant DB. Active transitive inference: When learner control facilitates integrative encoding. Cognition. 2020;200: 104188. doi:10.1016/j.cognition.2020.104188

30. Markant DB. Chained study and the discovery of relational structure. Mem Cogn. 2021;50: 95–111. doi:10.3758/s13421-021-01201-1

31. Favila SE, Chanales AJH, Kuhl BA. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. Nat Commun. 2016;7. doi:10.1038/ncomms11066

32. Schlichting ML, Mumford JA, Preston AR. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat Commun. 2015;6: 1–10. doi:10.1038/ncomms9151

33. Kumaran D, Hassabis D, McClelland JL. What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. Trends Cogn Sci. 2016;20: 512–534. doi:10.1016/j.tics.2016.05.004

34. Mcclelland JL, Mcnaughton BL, Reilly RCO. Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionist Models of Learning and Memory. Psychol Rev. 1995;102: 419–457. doi:10.1037/0033-295X.102.3.419

35. Javadi AH, Tolat A, Spiers HJ. Sleep enhances a spatially mediated generalization of learned values. Learn Mem. 2015;22: 532–536. doi:10.1101/lm.038828.115

36. Richards BA, Xia F, Santoro A, Husse J, Woodin MA, Josselyn SA, et al. Patterns across multiple memories are identified over time. Nat Neurosci. 2014;17: 981–986. doi:10.1038/nn.3736

37. Schlichtinga ML, Preston AR. Memory reactivation during rest supports upcoming earning of related content. Proc Natl Acad Sci U S A. 2014;111: 15845–15850. doi:10.1073/pnas.1404396111

38. Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker MP. Human relational memory requires time and sleep. Proc Natl Acad Sci U S A. 2007;104: 7723–7728. doi:10.1073/pnas.0700094104

39. Dumay N, Gaskell MG. Sleep-associated changes in the mental representation of spoken words: Research report. Psychol Sci. 2007;18: 35–39. doi:10.1111/j.1467-9280.2007.01845.x

40. Lewis PA, Durrant SJ. Overlapping memory replay during sleep builds cognitive schemata. Trends Cogn Sci. 2011;15: 343–351. doi:10.1016/j.tics.2011.06.004

41. Liu Y, Dolan RJ, Kurth-Nelson Z, Behrens TEJ. Human Replay Spontaneously Reorganizes Experience. Cell. 2019;178: 640-652.e14. doi:10.1016/j.cell.2019.06.012

42. Flesch T, Juechems K, Dumbalska T, Saxe A. Rich and lazy learning of task representations in brains and neural networks. Bioarxiv. 2021; 2021.04.23.441128. Available: https://www.biorxiv.org/content/10.1101/2021.04.23.441128v1%0Ahttps://www.biorx

iv.org/content/10.1101/2021.04.23.441128v1.abstract

43. Zhou Z, Singh D, Tandoc MC, Schapiro AC. Distributed representations for human inference. bioRxiv. 2021.

44. Preston AR, Shrager Y, Dudukovic NM, Gabrieli JDE. Hippocampal contribution to the novel use of relational information in declarative memory. Hippocampus. 2004;14: 148–152. doi:10.1002/hipo.20009

45. Zalesak M, Heckers S. The role of the hippocampus in transitive inference. Psychiatry Res - Neuroimaging. 2009;172: 24–30. doi:10.1016/j.pscychresns.2008.09.008

46. McCloskey M, Cohen N. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. Psychol Learn Motiv. 1989;24: 109–165.

47. Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, et al. Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci U S A. 2017;114: 3521–3526. doi:10.1073/pnas.1611835114

48. Momennejad I. Learning Structures: Predictive Representations, Replay, and Generalization. Curr Opin Behav Sci. 2020;32: 155–166. doi:10.1016/j.cobeha.2020.02.017

49. Stachenfeld KL, Botvinick MM, Gershman SJ. The hippocampus as a predictive map. Nat Neurosci. 2017;20: 1643–1653. doi:10.1038/nn.4650

50. Andersson JLR, Hutton C, Ashburner J, Turner R, Friston K. Modeling geometric deformations in EPI time series. Neuroimage. 2001;13: 903–919. doi:10.1006/nimg.2001.0746

51. Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R. Image distortion correction in fMRI: A quantitative evaluation. Neuroimage. 2002;16: 217–240. doi:10.1006/nimg.2001.1054

52. Mumford JA, Turner BO, Ashby FG, Poldrack RA. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage. 2012;59: 2636–2643. doi:10.1016/j.neuroimage.2011.08.076

53. Ritchey M, Montchal ME, Yonelinas AP, Ranganath C. Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. Elife. 2015;2015: 1–19. doi:10.7554/eLife.05025

54. Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, et al. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb Cortex. 2018;28: 3095–3114. doi:10.1093/cercor/bhx179

55. Fritsche M, Lawrence SJD, de Lange FP. Temporal tuning of repetition suppression across the visual cortex. J Neurophysiol. 2020;123: 224–233. doi:10.1152/jn.00582.2019

Supplementary Tables

Supplementary Table 1

Performance of the MLPs on premise trials as reported in the main text. Cross-entropy (x-entropy) statistics are presented in natural units (nats), and 2-ALC accuracy scores are presented as proportions between 0 and 1. Rounded parentheses represent 95% confidence intervals and square braces indicate ranges that include 95% of scores.

Number of hidden units	Interleaved training			Progressive training	
	x-entropy	*2-ALC*		*x-entropy*	*2-ALC*
1	0.081 ($\pm 1.67\times10^{-4}$) [0.072, 0.097]	.921 ($\pm 5.70\times10^{-4}$) [.864, .967]		0.335 ($\pm 1.97\times10^{-3}$) [0.224, 0.535]	.904 ($\pm 6.49\times10^{-4}$) [.844, .954]
2	0.003 ($\pm 2.31\times10^{-5}$) [0.002, 0.005]	.977 ($\pm 6.65\times10^{-5}$) [.973, .984]		0.014 ($\pm 8.42\times10^{-5}$) [0.008, 0.021]	.964 ($\pm 1.30\times10^{-4}$) [.954, .976]
3	0.001 ($\pm 3.15\times10^{-6}$) [0.001, 0.002]	.980 ($\pm 4.36\times10^{-5}$) [.977, .984]		0.005 ($\pm 2.32\times10^{-5}$) [0.003, 0.007]	.966 ($\pm 7.40\times10^{-5}$) [.961, .974]
4	0.001 ($\pm 1.92\times10^{-6}$) [<0.001, 0.001]	.980 ($\pm 2.56\times10^{-5}$) [.978, .982]		0.003 ($\pm 1.54\times10^{-5}$) [0.002, 0.005]	.965 ($\pm 5.06\times10^{-5}$) [.961, .970]
5	0.001 ($\pm 1.48\times10^{-6}$) [<0.001, 0.001]	.980 ($\pm 1.97\times10^{-5}$) [.978, .982]		0.003 ($\pm 1.21\times10^{-5}$) [0.002, 0.004]	.965 ($\pm 4.25\times10^{-5}$) [.961, .968]
6	0.001 ($\pm 1.22\times10^{-6}$) [<0.001, 0.001]	.980 ($\pm 1.65\times10^{-5}$) [.979, .981]		0.003 ($\pm 1.03\times10^{-5}$) [0.002, 0.004]	.964 ($\pm 3.84\times10^{-5}$) [.961, .968]

Supplementary Table 2

Performance of the MLPs on inferred trials as reported in the main text. Cross-entropy (x-entropy) statistics are presented in natural units (nats), and 2-ALC accuracy scores are presented as proportions between 0 and 1. Rounded parentheses represent 95% confidence intervals and square braces indicate ranges that include 95% of scores.

Number of hidden units	Interleaved training			Progressive training	
	x-entropy	*2-ALC*		*x-entropy*	*2-ALC*
1	12.7 (±0.142) [3.03, 27.4]	.498 (±.004) [.170, .788]		4.75 (±0.045) [2.33, 9.69]	.641 (±.002) [.516, .837]
2	5.26 (±0.042) [2.77, 9.54]	.513 (±.003) [.270, .696]		4.30 (±0.018) [2.28, 5.95]	.474 (±.001) [.356, .572]
3	3.760 (±0.016) [2.72, 5.22]	.432 (±.002) [.293, .607]		3.68 (±0.015) [2.53, 4.95]	.356 (±.002) [.244, .519]
4	3.28 (±0.011) [2.42, 4.21]	.398 (±.002) [.283, .537]		3.05 (±0.012) [2.12, 4.11]	.347 (±.001) [.244, .465]
5	3.09 (±0.010) [2.31, 3.93]	.381 (±.001) [.277, .502]		2.87 (±0.011) [2.00, 3.82]	.342 (±.001) [.241, .453]
6	2.99 (±0.009) [2.25, 3.78]	.372 (±.001) [.275, .482]		2.78 (±0.011) [1.90, 3.74]	.339 (±.001) [.238, .449]

Supplementary Table 3

Variance in MLP parameters over 10,000 independent training runs, summed across all the parameters in each network (note: MLPs with more hidden units had more parameters). Progressive training results in less variable parameters, yet this effect is minimal for MLPs with three or more hidden units.

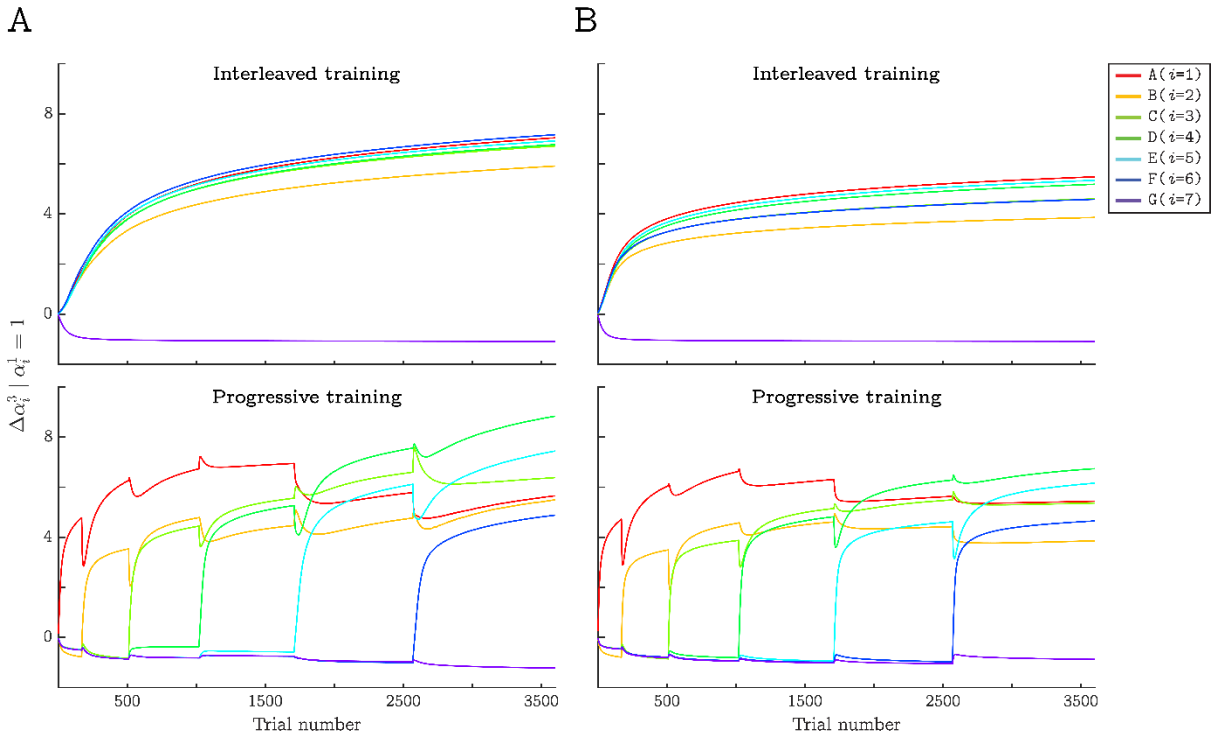
Number of hidden units	Interleaved training	Progressive training
1	328	177
2	119	100
3	89.4	88.5
4	85.1	83.8
5	84.0	82.7
6	83.6	82.3

Supplementary Table 4

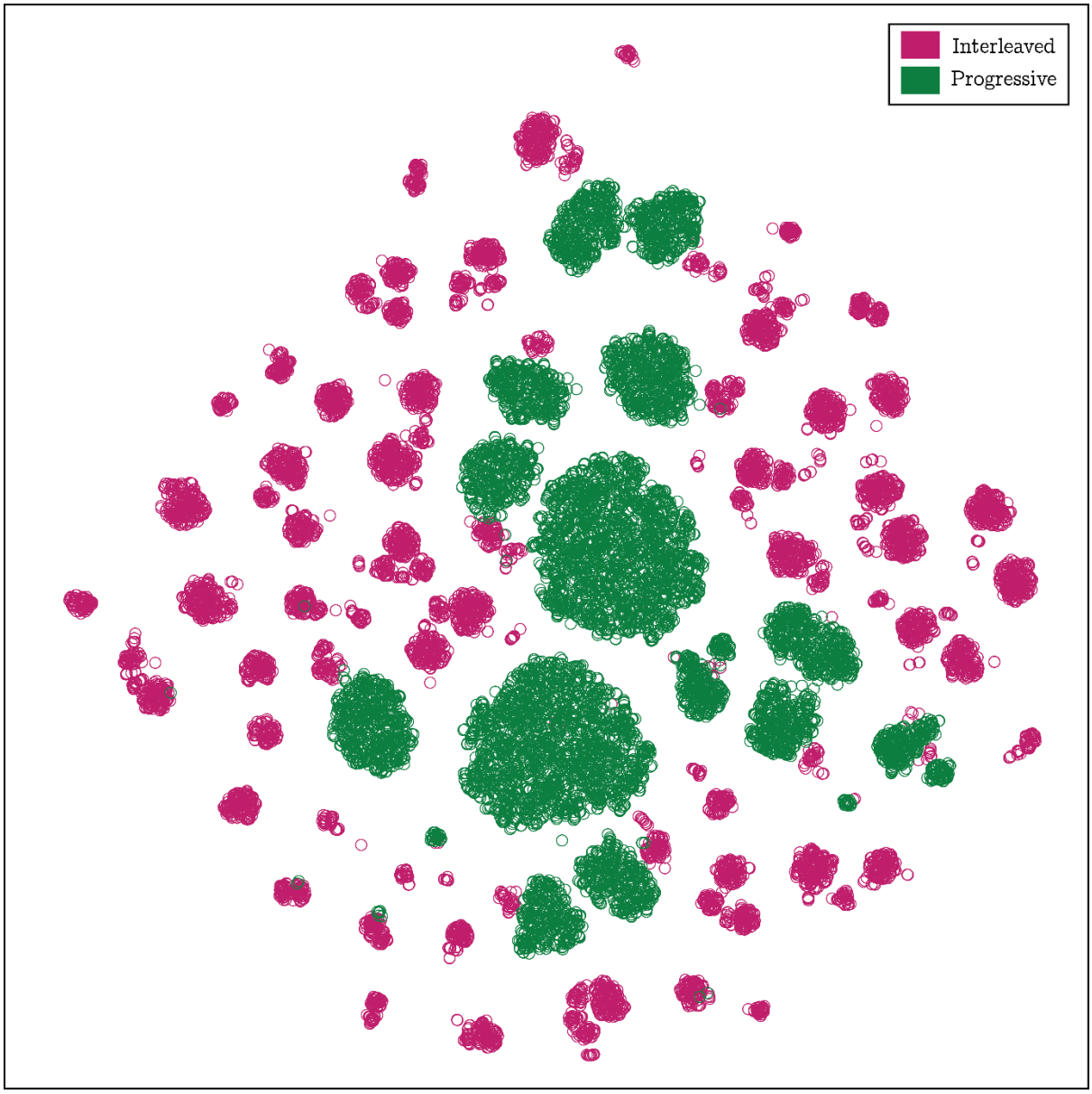
Mean Spearman rank correlations between the number of correct responses to each inferred discrimination and the number of correct responses expected under the AND and OR models. Rounded parentheses represent bootstrapped standard errors and square braces indicate bootstrapped 95% confidence intervals.

	Interleaved training			Progressive training	
	Recent	*Remote*		*Recent*	*Remote*
AND model	-.015 (.104) [-.198, .210]	.051 (.098) [-.154, .234]		.080 (.115) [-.179, .276]	-.295 (.081) [-.430, -.108]
OR model	.334 (.103) [.116, .524]	.304 (.117) [.049, .514]		.385 (.071) [.236, .515]	.473 (.062) [.352, .593]

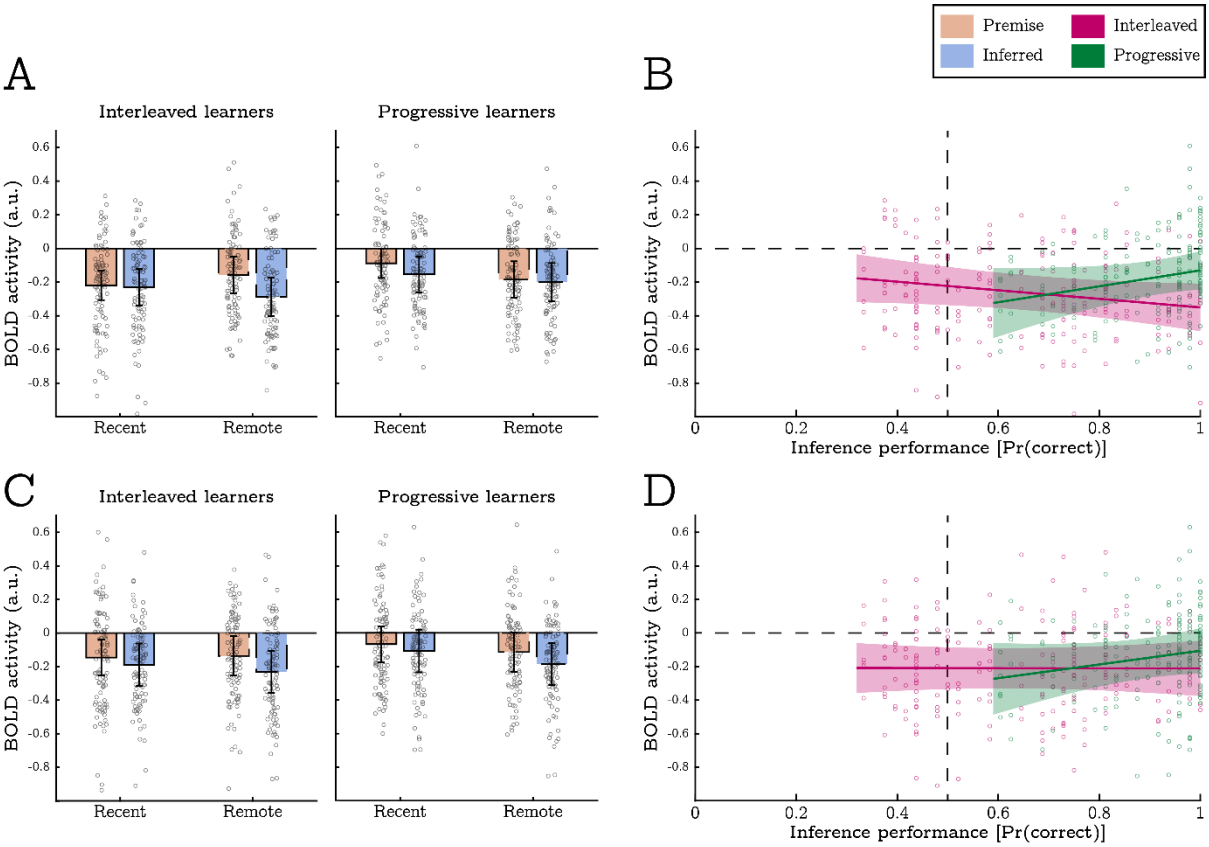
Supplementary Figures



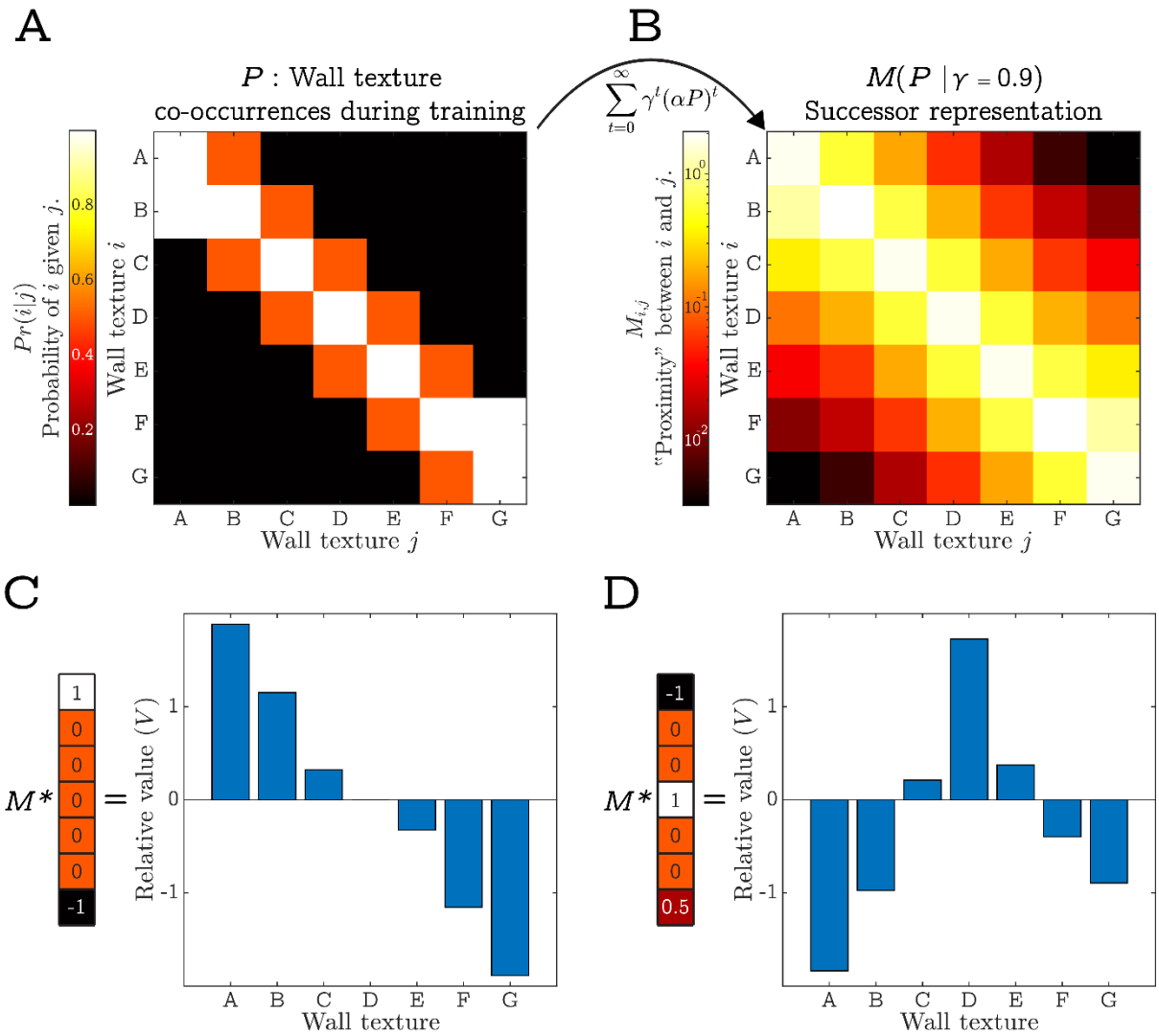
Supplementary Figure 1. Progressive training promotes the learning of a generalised value gradient in MLPs with a single hidden unit (see Figure 3B), yet this effect is substantially reduced in MLPs with two or more hidden units. These plots show $\Delta\alpha_i^3 \mid \alpha_i^1 = 1$ values indicating how the activity of an output unit changes over training trials when the corresponding input stimulus is presented as a one-hot vector on the input layer (averaged across training runs). A) $\Delta\alpha_i^3 \mid \alpha_i^1 = 1$ values for a MLPs with 2 hidden units. B) $\Delta\alpha_i^3 \mid \alpha_i^1 = 1$ values for a MLPs with 6 hidden units.



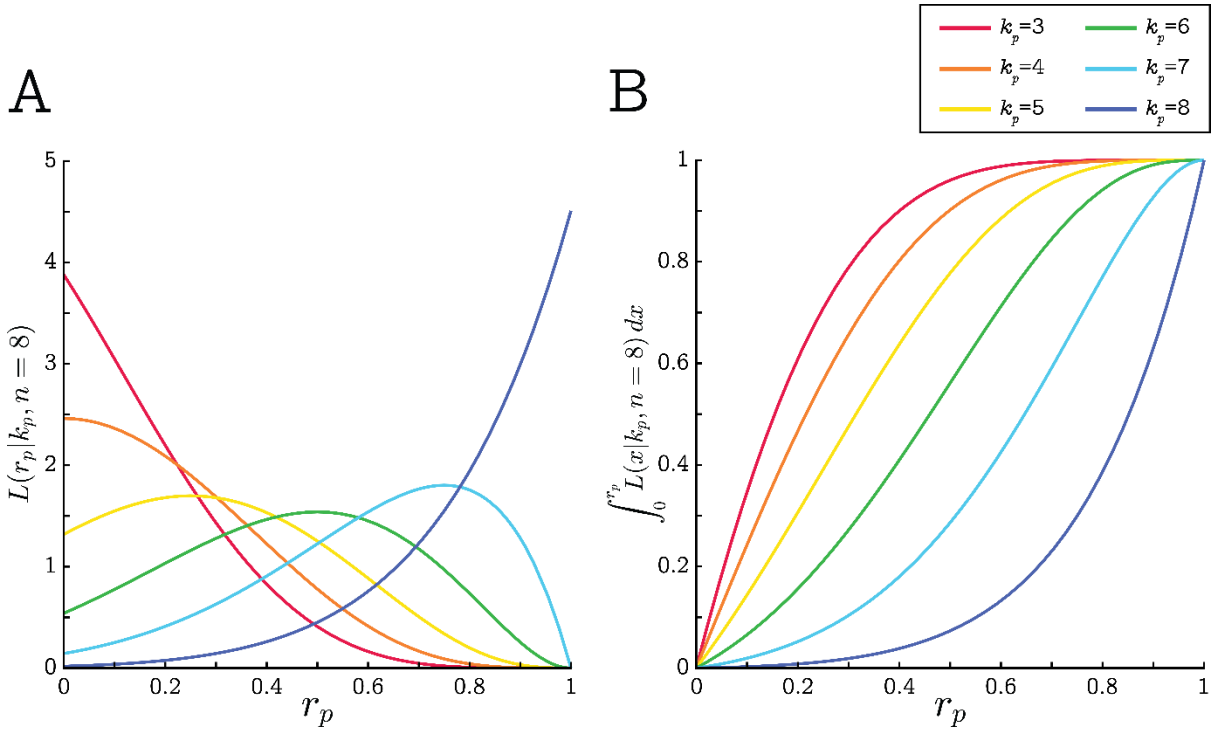
Supplementary Figure 2. Progressive training in MLPs with a single hidden unit biases the networks towards learning a limited number of solutions. The plot shows a t-distributed stochastic neighbour embedding (tSNE) of the 22 network parameters generated in each training iteration (N=10,000 per training schedule). tSNE allows the visualisation many parameter-sets in a 2-dimentaional space by reducing dimensionality in a way that preserves the local structure within clusters of similar parameters. Independent of this visualisation, a gaussian mixture model detected the presence of 53 clusters following interleaved training, but only 28 clusters following progressive training. Summed variance scores across parameters revealed that interleaved training resulted in less clustered solutions relative to progressive training (σ^2 = 328 and 177, respectively).



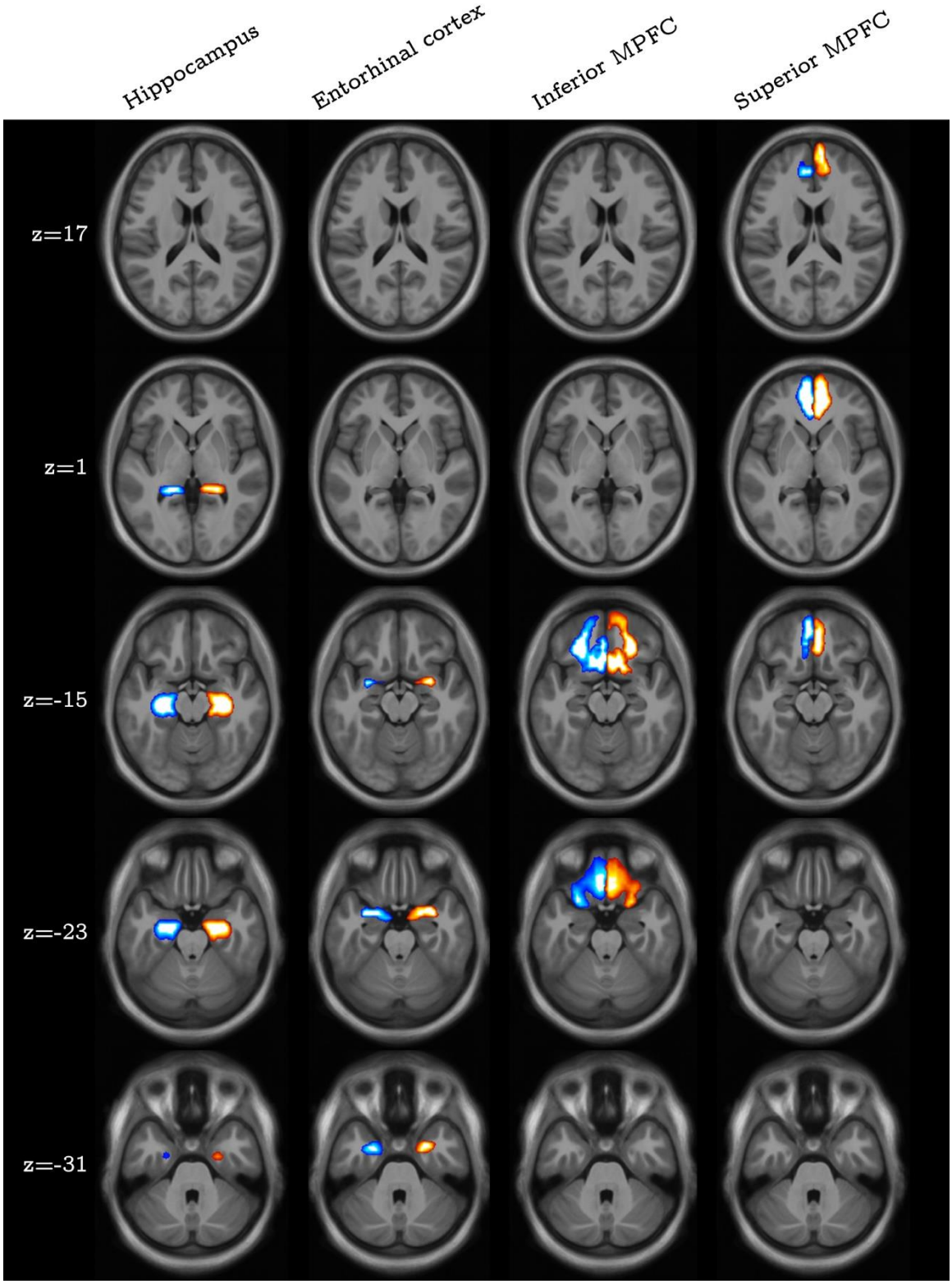
Supplementary Figure 3. Univariate BOLD activity in the hippocampus did not significantly predict inference performance. Panels A and B show activity in the left hippocampus. Panels C and D show activity in the right hippocampus. Bar charts display mean response amplitudes to all in-scanner discriminations split by trial type (premise vs inferred) and experimental condition (training method and session). Scatter plots display mean response amplitudes to all inference trials (both recent and remote) as a function of inference performance, split by training method (interleaved vs progressive). The only effect that reached statistical significance in these regions was detected in the right hippocampus. Here, a main effect of trial type indicated lower levels of BOLD activity on inference trails (panel C), yet this effect was not modulated by training method or inference performance. All error-bars indicate 95% confidence intervals.



Supplementary Figure 4. Using a successor representation (SR) to support transitive inference. A) Co-occurrence probabilities of the wall-texture stimuli during the pre-scanner training task (denoted P). B) A successor representation (denoted M) that results from transforming the co-occurrence probabilities via a computation hypothesised to be carried out by the hippocampal system (see Stachenfeld et al., 2017). Note: the scalar value α is chosen to ensure the sum of matrix rows after multiplication with Pdoes not exceed a value of 1. C) The SR can be combined (left multiplied) with a representation of the average reward provided by each wall texture (R) to compute a generalised value function (V) that may support transitive inference. The quantities in R are zero for wall textures that are rewarded and unrewarded an equal number of times, whereas values of 1 and -1 indicate that wall textures that are always/never predictive of reward (respectively). A value function over wall textures accounts for the observation that inference performance is positively related to transitive distance since this distance scales with differences in value. D) One benefit of the SR is that, if the reward contingencies change suddenly, a new value function can be computed quickly, without the need to relearn the entire task structure.



Supplementary Figure 5. In constructing the AND/OR models of human inference performance, memory for the premise discriminations was parametrised by computing a likelihood function of plausible values for the probability of correct retrieval ($L(r_p)$, where p denotes a specific premise discrimination). To do this we assume the probability of observing k_p correct responses to the $n = 8$ test trials depends on a joint binomial process involving r_p and, if retrieval is not successful, a random guess that yields a correct response with a probability of 0.5 (see Methods). Panel A presents the likelihood function for r_p under different values of k_p. To approximate this distribution for the analysis, we randomly sampled values of r_p using inverse transform sampling. This involved generating uniformly distributed random numbers in the range [0, 1] and selecting values of r_p that returned cumulative likelihoods matching those random values (panel B).



Supplementary Figure 6. Normalised group averages of each ROI in MNI space. Each column relates to a different brain region and the blue-/orange- coloured overlays depict left-/right- hemisphere ROIs (respectively). Overlay lightness represents ROI coverage across participants.

Supplementary Results

Progressive training in multilayer perceptrons

In our main analyses, all MLPs were trained with a learning rate of 0.1 such that sufficient levels of performance on the premise pairs were achieved after 3,600 trials (10 times the number in the pre-scanner task). However, this learning rate is relatively large comped to those used in most machine learning applications. As such, we ran a supplementary analysis that involved training the MLPs with a learning rate of 0.01 over 36,000 trials. In contrast to the main analyses, this only involved 1,000 independent runs per training schedule in order to limit computer memory usage.

As in the main analyses, both interleaved and progressive schedules resulted in all MLPs learning the 6 premise discriminations to a high level, with interleaved training resulting in slightly better performance; mean cross-entropy (h): <0.081 and <0.284, mean 2-ALC accuracy: >92.0% and >89.8% (for interleaved and progressive schedules respectively). When tested on the 6 transitive inference trials, cross-entropy statistics indicated that progressively trained MLPs invariably performed better than MLPs trained via interleaved schedules; mean interleaved $h = 12.1/4.64$, mean progressive $h = 3.00/2.87$ (for MLPs with 1/6 hidden units).

In networks with a single hidden unit, interleaved training did not generally promote greater activity in the target output relative to the non-target output, 2-ALC accuracy: 49.8% (CI: +/- 1.10). However, the variability in inference performance across different interleaved training runs was large (95% range: 17.0 - 76.9). In contrast, progressive training in these single hidden unit networks lead to substantially higher levels of inferential accuracy: 62.5% (CI: +/- 0.542), yet the variability in performance reduced (95% range: 48.9 - 81.6). This pattern in 2-ALC scores did not hold for networks with two or more hidden units, which tended to activate non-target output units more strongly following both training schedules; mean interleaved accuracy = 50.9/37.4%, mean progressive accuracy = 47.2/33.8% (for MLPs with 1/6 hidden units).

Plotting $\Delta\alpha_i^3 \mid \alpha_i^1 = 1$ values also revealed the same pattern as in the main analyses. As before, progressive training in the MLP with a single hidden unit produces activations in each output that faithfully represents a generalised value for the corresponding input stimulus. This is also evident in MLPs with more than one hidden unit yet, the representation is much weaker.

Representational similarity analysis

Here we report significant effects from the representational similarity analyses that did not involve interactions with transitive distance and so are not directly related to our a priori hypotheses. These results therefore reflect changes in representational similarity across all wall textures that only vary by inferential accuracy, transitive slope, or experimental condition. Such effects may suggest that a brain region is involved in transitive inference, yet their precise contributions cannot be inferred from these results alone. This is because there is no unambiguous relationship between uniform changes in pattern similarity across all wall textures, and theoretical accounts of how representational structure may facilitate inference.

Within-hierarchy RSA

Correlations with inferential accuracy

In both the left entorhinal cortex and left superior MPFC, we detected a main effect of inference accuracy indicating greater pattern similarity between stimulus representations in participants who were more proficient at the inference task, $t(651) = 3.85$, $p < .001$, and $t(650) = 2.95$, $p = .003$ (respectively). These effects were principally driven by accuracy-similarity correlations in the interleaved training condition ($t = 4.51$, & $t = 3.14$) since the effect was non-signficnat for progressive learners ($t = 1.04$, & $t = 1.34$). Nonetheless, the accuracy by group interaction did not reach our statistical threshold in either ROI; $t(651) = 2.37$, $p = .018$, and $t(650) = 0.89$, $p = .379$, for the left entorhinal cortex and MPFC, respectively. These results suggest that representations in the left entorhinal and medial prefrontal cortices may be involved in successful inference following interleaved training.

Correlations with transitive slope

The right inferior MPFC exhibited a main effect of transitive slope such that pattern similarity between wall texture representations was lowest when participants' behavioural performance was most indicative of encoding-based generalisations (i.e., a negative association), $t(654) = 3.29$, $p = .001$. The right superior MPFC also showed a negative correlation between transitive slope and pattern similarity, yet this effect was limited to the interleaved learners, $t = 3.82$. Progressive learners showed the opposing relationship such that behavioural responses strongly indicative of encoding-based generalisations corresponded to higher levels of pattern similarity, $t = 2.48$. This resulted in a training method by transitive slope interaction, $t(652) = 4.50$, $p < .001$. In the left superior MPFC, we detected a significant interaction between training session and transitive slope, $t(650) = 4.38$, $p < .001$. This indicated that pattern similarity was highest in the remote condition when participant's behavioural responses were most consistent with encoding-based generalisations.

Categorical effects of condition

Both the left and right superior MPFC showed a significant training method by session interaction, $t(650) = 4.29$, $p < .001$, and $t(652) = 6.34$, $p < .001$ (respectively). These effects indicated that pattern similarity estimates were comparable across all learners in the recent condition yet differed from this level in the remote condition. In the interleaved learners, BOLD responses between remotely trained stimuli were less similar compared with the responses between the recently learnt stimuli. However, the reverse was true for the progressive learners; relative to the recent condition, pattern similarity estimates increased in the remote condition.

Across-hierarchy RSA

In a final set of mixed-effects regression models, we examined whether any ROIs exhibited BOLD representations that encoded the structure of the transitive hierarchy in a way that generalised across the transitive chains learnt on each day of training (i.e., across recent and remote conditions). Such representations are predicted by encoding-based models of generalisation, notably, the Tolman Eichenbaum Machine [7]. This analysis involved estimating the similarity between wall-texture representations from different days of training and identifying across-hierarchy distance effects. Similar to above, we tested whether such

distance effects were modulated by 3 factors of interest: *1)* training method (interleaved vs progressive), *2)* inference accuracy, and *3)* transitive slope.

None of our regions of interest produced significant across-hierarchy distance effects. However, the left hippocampus, and left and right superior MPFC showed strong correlations between across-hierarchy pattern similarity and our behavioural measures of performance. In the left hippocampus, we observed a significant interaction between training method and transitive slope such that behavioural performance strongly suggestive of encoding-based generalisations was positively associated with increased similarity estimates in the progressive learning condition ($t = 2.63$), but negatively associated with similarity in the interleaved condition ($t = -2.03$); interaction effect: $t(835) = 3.31$, $p < .001$.

Results in the left and right superior MPFC were very similar. Both regions produced a training method by inference accuracy interaction, $t(835) = 4.29$, $p < .001$ and $t(835) = 3.22$, $p = .001$ (respectively). These effects indicated that accuracy scores were positively correlated with representational similarity in the progressive learners, but negatively correlated with similarity scores in the interleaved learners. Both regions also produced a main effect of transitive slope such that participants who produced behavioural responses most indicative of encoding-based generalisations had the highest levels of representational similarity, $t(835) = 3.91$, $p < .001$ and $t(835) = 4.29$, $p < .001$ (respectively).