Medial temporal and prefrontal cortices encode structural task representations at different levels of abstraction

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.), chris.bird@sussex.ac.uk (C.M.B.)

Abstract

Memory generalisations may be underpinned by either encoding- or retrieval-based mechanisms. We used a transitive inference task to investigate whether these generalisation mechanisms are 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 medial temporal lobe (MTL) and medial prefrontal cortex (MPFC). These results are consistent with the use of representations that directly encode structural relationships between different task features. Furthermore, BOLD patterns in MPFC were similar across the two independent hierarchies. We conclude that humans preferentially employ encoding-based mechanisms to store map-like relational codes that can be used for memory generalisation. While both MTL and MPFC support these representations, the MPFC encodes more abstract relational information.

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 (Potts, 1972, 1974), 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 (Kumaran and McClelland, 2012). 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 (Koster et al., 2018) and behavioural studies (Banino et al., 2016). 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 (e.g. Kumaran et al., 2016; Whittington et al., 2020). 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 (Ferreira et al., 2019; Schapiro et al., 2017; Zeithamova et al., 2012).

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 (Kaplan and Friston, 2018; Kumaran et al., 2012, 2016a; Morton et al., 2020; Tavares et al., 2015). The entorhinal cortex and MPFC also appear to represent similarly structured hierarchies in analogous metric spaces (Baram et al., 2021; Park et al., 2020). This coding could facilitate knowledge transfer across related tasks but the relationship between generalised neural codes and task performance is unclear.

The particular learning conditions can have a large impact on how information is retained (Roediger and Karpicke, 2006) and generalises to new situations (Birnbaum et al., 2013). For example, categorisation of previously unseen objects is improved if training exemplars are presented in an interleaved fashion, rather than progressively (i.e., shown within a number of category-specific blocks; e.g., Kang and Pashler, 2012). Birnbaum and colleagues (2013) argued

that interleaved presentation highlights the differences between the categories enabling them to be better discriminated.

Contrary to these previous studies, in the present study we report both modelling results and then human data demonstrating that transitive inference is superior following progressive, rather than interleaved presentation of the training pairs (see Figs. 1-3 and Results). We hypothesised that training condition would affect the mechanism used to make inferential choices. In particular, we predicted that interleaved training would promote 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 (see Favila et al., 2016). By contrast, progressive learning of the training pairs 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 (Schlichting et al., 2015).

In addition to the learning procedure, 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 (Kumaran et al., 2016b; Mcclelland et al., 1995). In support of this, many studies have shown that memory generalisation improves following a period of sleep, or even wakeful rest (Dumay and Gaskell, 2007; Ellenbogen et al., 2007; Javadi et al., 2015; Richards et al., 2014; Schlichtinga and Prestona, 2014). We hypothesised that overnight consolidation would allow pattern separated memories of related task contingencies to be re-encoded as structural memory representations (see Lewis and Durrant, 2011; Liu et al., 2019). 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 procedure 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 boosted inference performance in a simple artificial neural network (a multilayer perception; MLP) by facilitating the learning of representations that encode the relative value of discriminable stimuli. We also found that progressive training had a large positive effect on inference 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 procedure 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 correlated with generalisation performance. In contrast, representational similarity analyses (RSAs) were more consistent with encoding-based models. Specifically, we identified RSA effects in the hippocampus, entorhinal cortex, and medial prefrontal cortex that were suggestive of structural representations for the whole transitive hierarchy. Finally, we found that representations within the MPFC generalised across the different hierarchies learnt on each day of training, suggesting that the region represents the structure of the learnt information, irrespective of the specific information that is held within this structure.

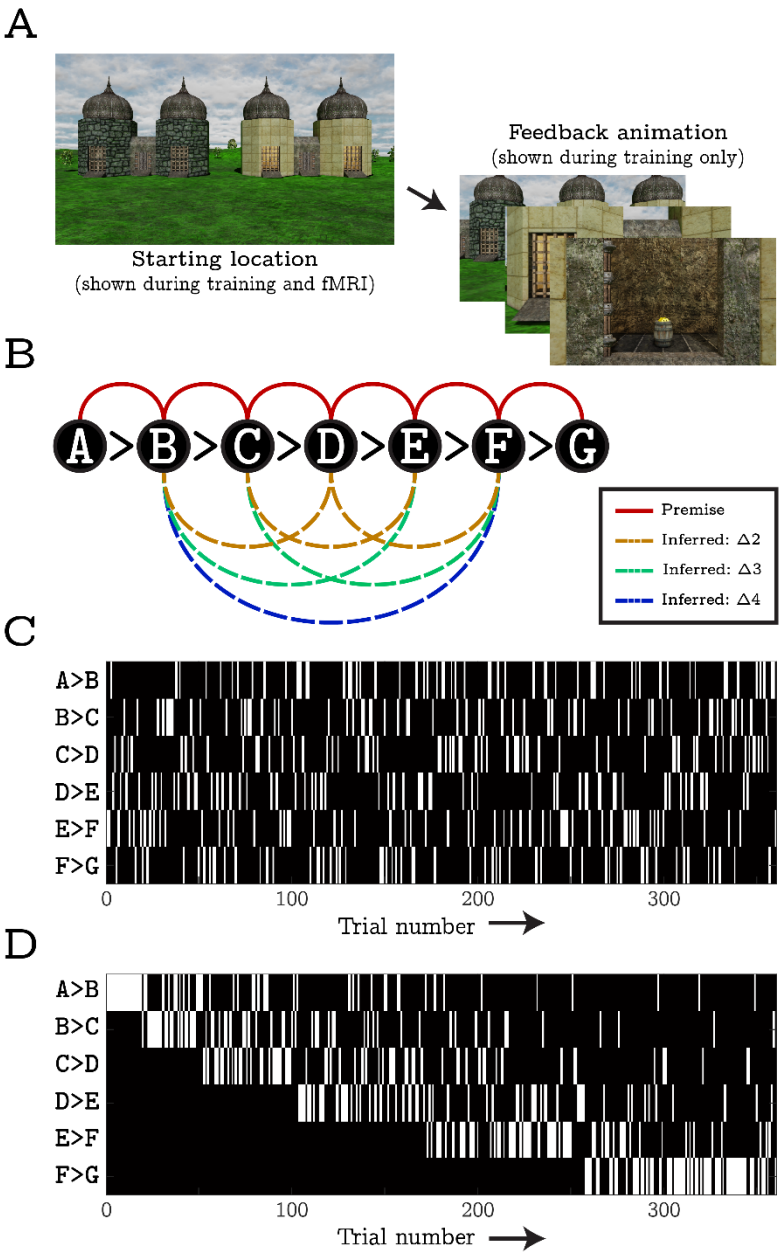


Figure 1. Illustration of the pre-scanner training and in-scanner behavioural tasks. A) Both before and during fMRI, participants saw computer generated images of two buildings with different wall-textures rendered onto their exterior surfaces. One building concealed a pile of virtual gold (reinforcement) 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 a multilayer perceptron

We first demonstrate that different training procedures can guide even very simple agents to learn fundamentally different task representations which, in turn, affects generalisation. Across multiple independent runs, we trained a multilayer perceptron (MLP) 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 perceptron on both explicitly trained (premise) and transitively inferred discriminations.

The goal of this analysis was to characterise general effects of each training procedure on learning and inference. As such, the MLP was designed to be as simple as possible while having the ability to learn a task representation that could support transitive inference. It is not indented to be a faithful model of how humans solve such tasks, nor is it indented to adjudicate between retrieval- and encoding-based models of generalisation. The MLP consists of three layers (input, hidden, and output) with a single hidden neuron connected to all inputs and output (see Figure 2A). In short, training involved presenting the perceptron 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). As the hidden neuron can only provide a 1-dimensional latent representation of the inputs, high-levels of performance depend on non-trivial weights between both layers with solutions being sparsely distributed in the parameter space. Nonetheless, solutions that do yield near-perfect performance on both premise and inferred discriminations do exist (see https://osf.io/ps3ch/ for a scripted example).

Interleaved training involved presenting the perceptron 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). This procedure was repeated 10,000 times with varied training orders and different network initialisations (i.e., random starting parameters). 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 (see Figure 1D). Again, this procedure was repeated 10,000 times with different training orders and network initialisations.

Training-related prediction error (as measured by a cross-entropy cost function) gradually decreased across interleaved training trials (Figure 2B; mean total cross-entropy, $\widehat{h_{sum}} = 1300.5$ nats, 95% range: 1167.2 - 1495.6). In contrast, progressive training was characterised by sharp increases, and then rapid decreases in prediction error after the introduction of each new discrimination ($\widehat{h_{sum}} = 945.6$ nats, 95% range: 859.5 - 1096.7).

After both training procedures, the perceptron was consistently able to produce accurate responses to all 6 premise discriminations (see Figure 2C, final performance measured by a SoftMax function implementing a 2-alternative Luce choice decision rule; see Methods). We then tested the MLP on 6 transitive inference trials that were not explicitly trained (e.g., 'B>D', Figure 2C). Following interleaved training, the mean performance was around chance at 49.7% (CI: +/- 0.364), yet the variability across different training runs was large (95% range: 17.1 - 78.9%). Following progressive training, mean transitive performance was substantially higher at 64.2% (CI: +/- 0.175), but the variability in performance reduced (95% range: 52.0 - 84.1%).

To explore the reasons for this, we computed the linear activation of each output neuron (prior to SoftMax normalisation) that resulted from stimulating the corresponding input neuron with a one-hot vector. We denote this activation, $\Delta\alpha_i^3 \mid \alpha_i^1 = 1$, for the ith input/output neuron, where superscripts denote layer indices. Supplementary Figure 1 plots these values as a function of training trial, averaged across runs. This shows 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 neuron that faithfully represents a generalised value for the corresponding stimulus, as long as that stimulus is presented on the input layer. This pattern does not hold for MLPs trained via an interleaved procedure. Here, activity levels in output neurons are approximately equivalent for all stimuli that are rewarded during training (i.e., excluding stimulus 'G').

Progressive training results in a generalised value representation because learning the first discrimination ('A>B') tunes the MLP parameters 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 yields a very specific error gradient that allows the network to learn a generalised value for each stimulus. In contrast, interleaved training yields (approximately) uniform levels of activity in all output neurons that are rewarded, but network weights that inhibit non-target outputs in response to the presentation of specific premise pairs.

A consequence of this is that progressive training biased the MLP towards learning a limited number of solutions to the discrimination problem (Supplementary Figure 2). We used a gaussian mixture model to characterise the distribution of trained parameters and found that they tended towards 42 distinct 'attractors' after interleaved training, but only 23 attractors after progressive training. Differential entropy statistics also revealed that interleaved training resulted in less clustered solutions relative to progressive training (h = -9.32 and -19.1, respectively).

In sum, we find that progressive training has a significant effect on the type of representations that are learned by an MLP, and this may increase generalisation performance. At the same time, progressive training appears to reduce the diversity of learned representations thereby precluding the development of models that achieve very high levels of performance.

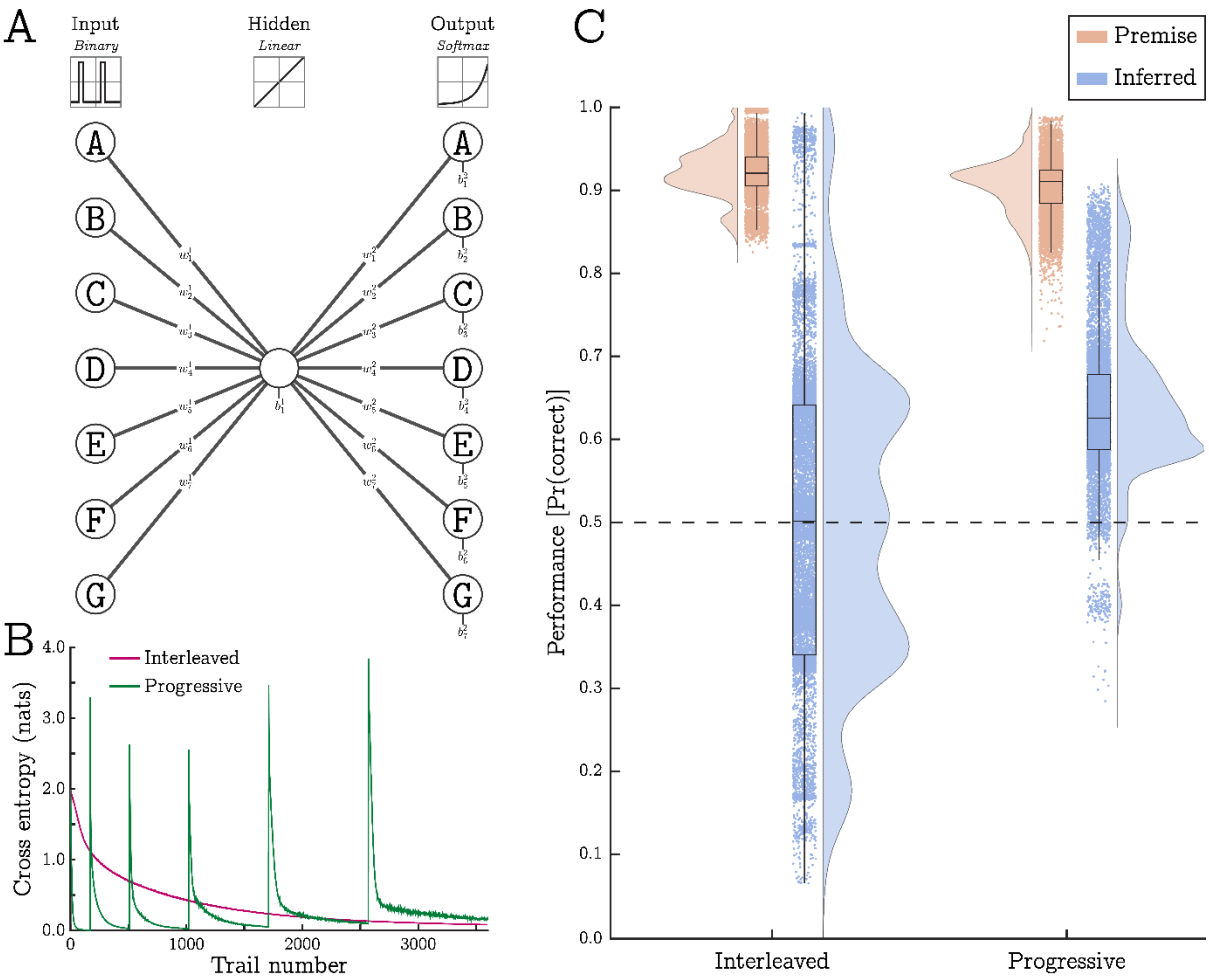


Figure 2. A simple multi-layer perceptron (MLP) shows better generalisation following progressive training. **A)** A schematic illustration of the 3-layer MLP that was tested. Seven input nodes (labelled A-G, one for each discriminable feature in the task set) provided weighted binary inputs to a biased hidden node with a linear activation function. In turn, activity in the hidden node projected along weighted connections to seven biased output nodes with a SoftMax activation function. Weights and biases are indicated as w_j^i and b_j^i (respectively) with superscripts and subscripts denoting layer and node induces (respectively). **B)** We trained this MLP 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 line) whereas progressive training resulted in rapid exponential decreases following the start of each epoch (green line). Note: the learning curves plotted in this figure result from averaging over 10,000 iterations. **C)** After each training iteration, we tested the performance of the MLP on both trained and inferred discriminations using a SoftMax function to implement a 2-alternative forced 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.

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 3A 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 3B).

A generalised linear model of response times (correct responses only) produced a complementary pattern of results (see Figure 3C). 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.

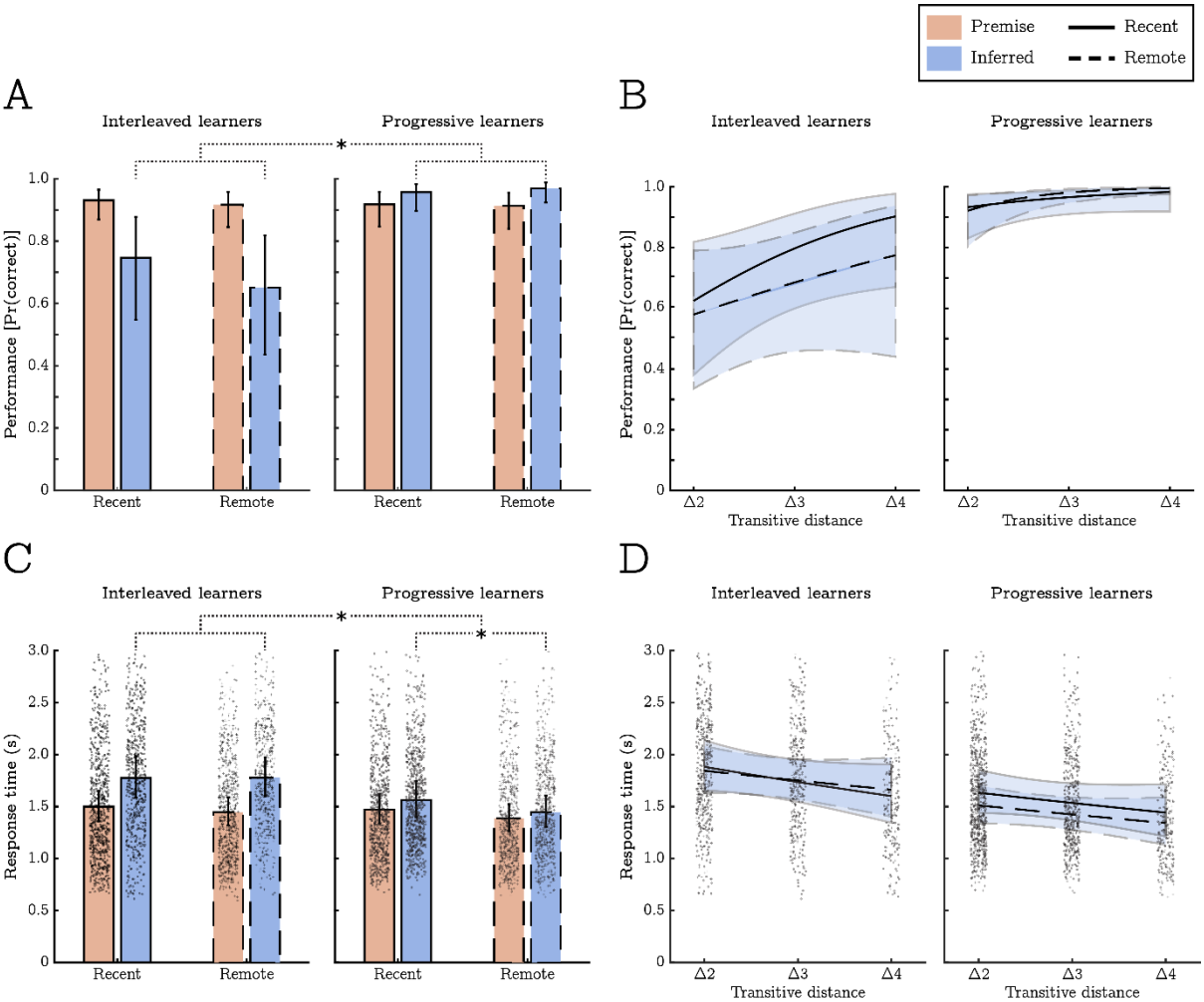


Figure 3. Humans shows better generalisation following progressive training. **A)** Estimates of the probability of a correct response, $Pr(correct)$, split by trial type (premise vs inferred) and experimental condition (training method and session). While participants showed comparable levels of performance 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), and so inferred discriminations are easiest when there is a large 'distance' between discriminable features. As such, these 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 4 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). Contrary to our predictions, these results indicate that inference performance is best accounted for by encoding-based models in all experimental conditions.

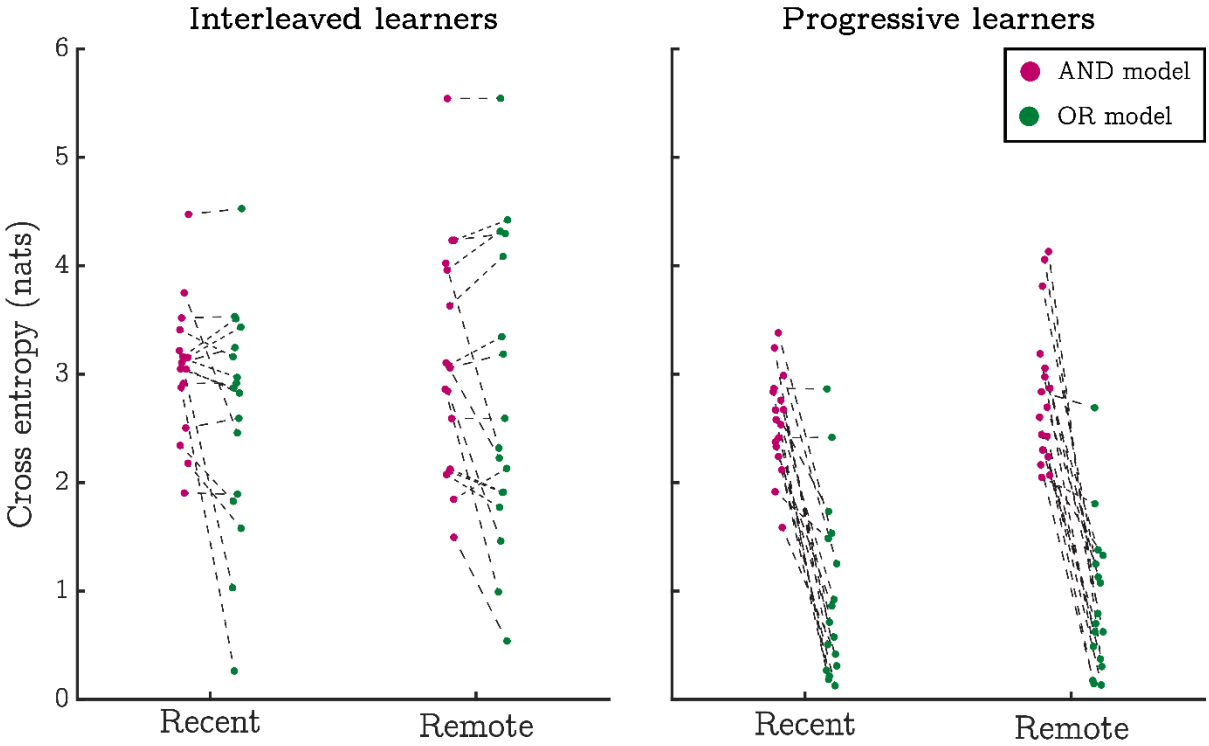


Figure 4. Behavioural performance is better fitted by an encoding-based generalisation mechanism (OR model). Figure shows goodness-of-fit statistics (lower is better) for two models of inference performance. The AND model implements a general assumption of retrieval-based generalisation mechanisms - 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, $t(787) = 2.93$, $p = .003$ (see Figure 5). Similarly, the right superior MPFC produced a significant interaction between trial type and training method $t(787) = 2.76$, $p = .006$, (Figure 5C). 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 Preston et al., 2004; Zalesak and Heckers, 2009).

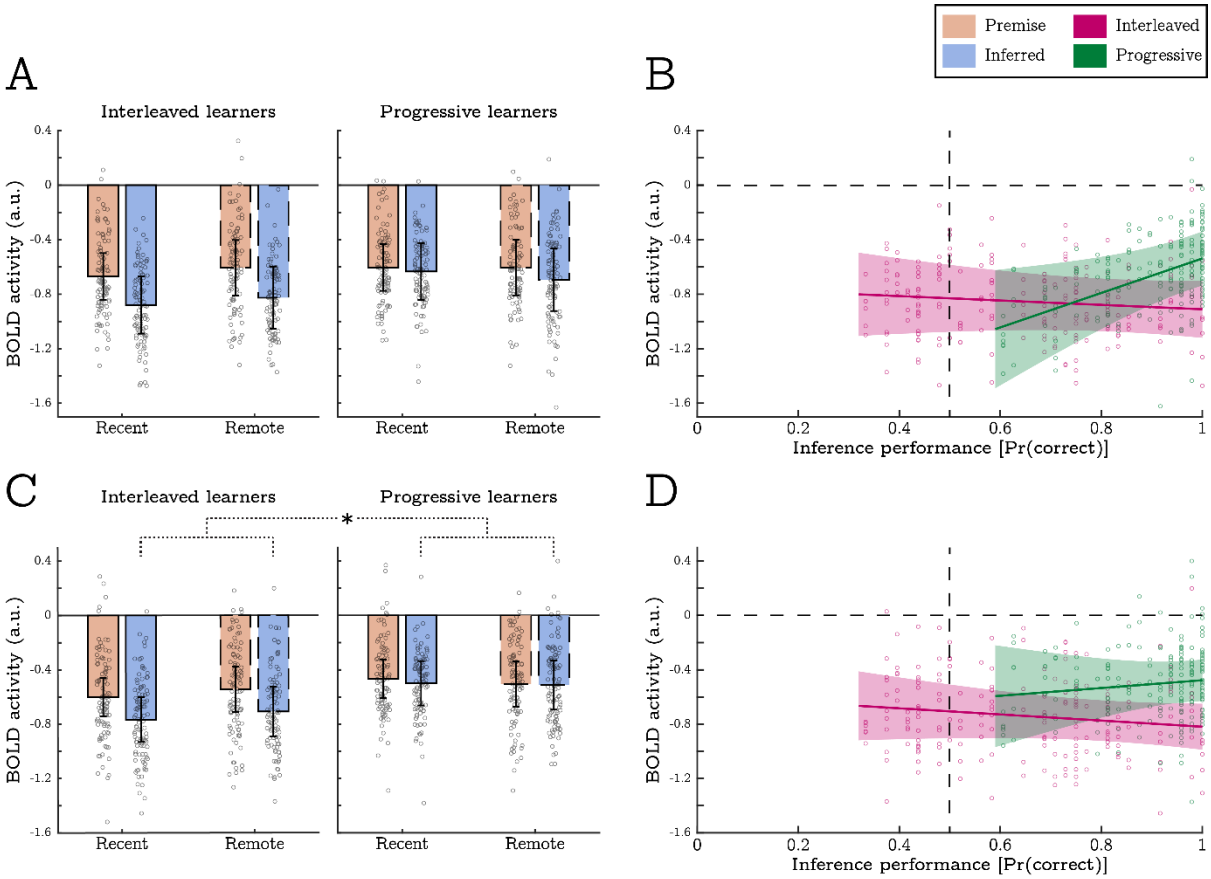


Figure 5. Inference performance within- and across- experimental conditions is associated with univariate BOLD activity in the superior MPFC. Panels **A** and **B** show activity in the left superior MPFC. Panels **C** and **D** show activity in the right superior MPFC. Bar charts display mean response 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

Within-hierarchy RSA

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 interaction).

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 6A). The (correlational) similarity between wall-texture representations was then analysed in the LMMs to identify 'within-hierarchy distance effects', i.e., effects 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[A,B] > corr[A,C] > corr[A,D]). 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).

The LMMs revealed significant distance effects in 3 left-hemisphere ROIs when collapsing across all experimental conditions: the hippocampus, $t(1396) = 6.20$, $p < .001$, the entorhinal cortex, $t(1393) = 4.23$, $p < .001$, and the inferior MPFC, $t(1399) = 2.98$, $p = .003$, see Figure 6. These results indicate that each region encoded a generalised structure of the transitive hierarchy. To illustrate this, we used classical multidimensional scaling (MDS) to reconstruct the hierarchy in 2 dimensions using the fMRI data alone (see inset panels in Figure 6). While wall-textures at the extreme ends of the hierarchy are notably dissimilar to all others, the MDS shows that B-F representations are approximately co-linear, falling along a 1-dimentional path reflective of the transitive hierarchy.

The LMM in the left inferior MPFC revealed a significant interaction between the distance effect and the length of study-test interval (session). This indicated that the strength of the hierarchical representation in the MPFC was greatest for contingencies that had been learnt recently. In fact, a simple effects test showed that while inferior MPFC distance effect was significant in the recent condition, $t(1399) = 4.29$, $p < .001$, it was non-significant in the remote condition, $t(1399) = 0.736$, $p = .462$.

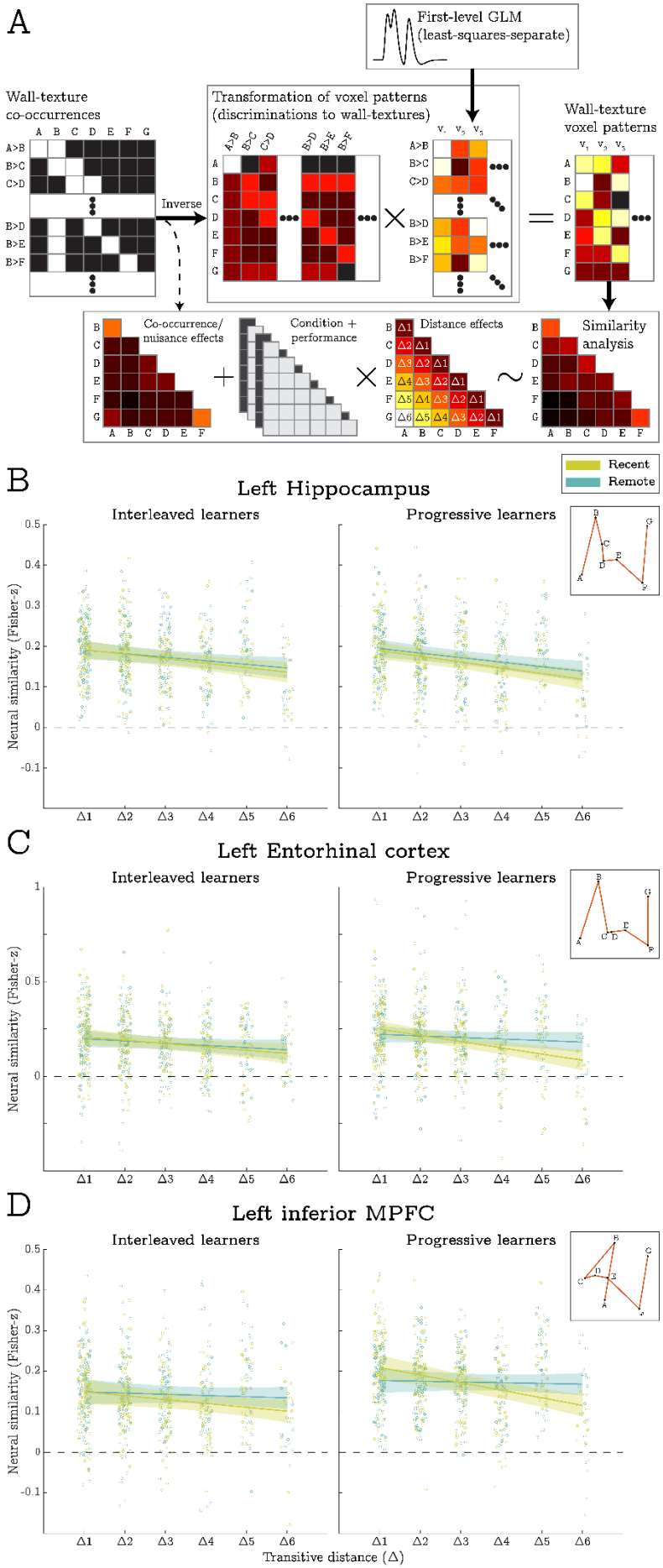


Figure 6. Methods and results for the within-hierarchy RSA. **A)** BOLD responses across voxels (v_1, v_2, etc.) for each in-scanner discrimination (A>B, B>D, etc.) were estimated in a set of 1st level models. These were linearly transformed into representations of specific wall-texture stimuli (A, B, C, etc.) via a least-squares decomposition. 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 while effects of interest modelled the influence of condition, performance, and transitive distance. **B-D)** Transitive distance (i.e., the separation between wall-textures) was negatively correlated with BOLD similarity in 3 ROIs: the left hippocampus, the left entorhinal cortex, and the left inferior MPFC. These effects indicate that each region encoded a generalised structure of the transitive hierarchy that was not modulated by training method (i.e., interleaved vs progressive), or behavioural performance (see text). The distance effect in the left inferior MPFC also showed a distance by session interaction suggesting the strength of hierarchical representations was greatest for contingencies learned most recently. Inset panels show 2D multi-dimensional scaling of the neural similarity data collapsed across training method and session (see Methods). All error-bars indicate 95% confidence intervals.

Importantly, the effect of distance in both the hippocampus and inferior MPFC did not appear to depend on inference ability. Specifically, there was no notable interaction between transitive distance and either *a)* inference performance or *b)* training method in these regions; largest effect: $t(1396) = 1.06$, $p = .290$. A follow-up Bayesian analysis found more evidence in favour of the null hypothesis for the distance by accuracy interaction when considering representational similarity in interleaved learners who performed at chance level on the inference task; BF_{01} statistics: 1432 and 4.11 for the hippocampus and inferior MPFC, respectively (the Bayes factor in the left entorhinal cortex was insensitive: $BF_{01} = 1.02$). Note that these tests were possible because 9 of the 17 interleaved learners scored *below* chance on the inference task in at least one of the recent/remote conditions. Intriguingly, these results suggest that a structured representation of the transitive hierarchy was encoded by participants who were unable to achieve good levels of inference performance.

In the left superior MPFC, we detected two RSA effects associated with inference performance. Here, similarity estimates were generally larger for progressive- relative to interleaved learners (i.e., a main effect of training method), $t(1398) = 3.60$, $p < .001$ (see Figure 7A). Additionally, similarity scores were significantly modulated by an interaction between training session and transitive slope, $t(1398) = 2.89$, $p = .004$. This indicated that representational similarity in the remote condition was highest when behavioural performance closely matched the predictions of encoding-based models, $t(1398) = 2.84$, $p = .005$, yet no such relationship was evident in the recent condition, $t(1398) = -1.33$, $p = .182$ (Supplementary Figure 4). Overall, these results suggest that representations in the left superior MPFC may account for differences in transitive inference across participants. In particular, higher levels of representational similarity were observed in progressive learners (i.e., those who were most proficient at the inference task) and in participants who appeared to make use of encoding-based generalisation mechanisms in the remote condition.

In the right hippocampus, inference performance was negatively related to the similarity between all wall-texture representations, $t(1398) = 2.83$, $p = .005$ (see Figure 7B). As such, *lower* levels of representational similarity in this region were associated with better generalisation. While not hypothesised, this effect may suggest the 'uniqueness' of wall-texture representations in the right hippocampus supports performance. No other significant effects were detected.

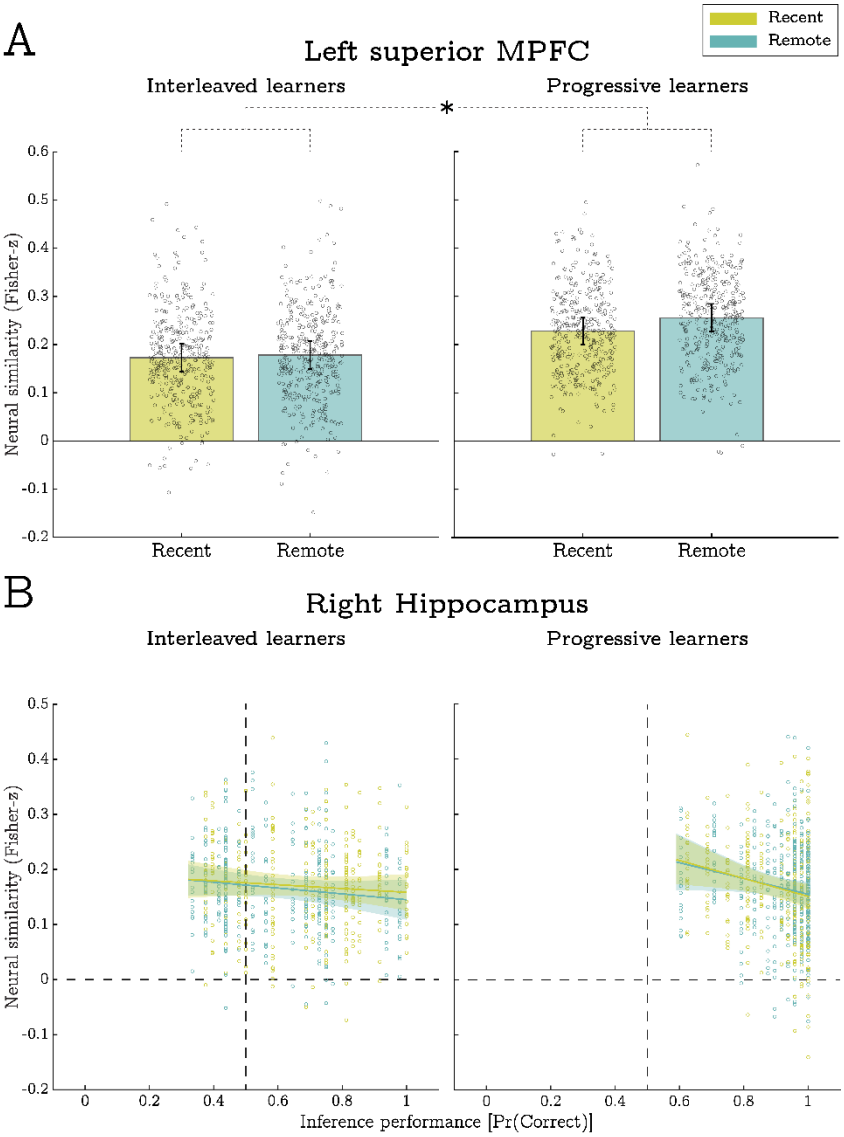


Figure 7. Inference performance within- and across- experimental conditions is associated with representational similarity between wall-textures from the same transitive hierarchy. **A)** In the left superior MPFC, we detected a main effect of training method indicating that similarity estimates were largest for progressive learners (who achieved high levels of inference performance: mean Pr(Correct) = .900), relative to interleaved learners (who performed just above chance: mean Pr(Correct) = .639). **B)** In the right hippocampus, individual differences in inference performance within experimental conditions were negatively correlated with similarity scores. As such, high levels of hippocampal similarity were associated with poorer inference. All error-bars indicate 95% confidence intervals.

Across-hierarchy RSA

In a final mixed-effects regression model, 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 (Whittington et al., 2020). 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.

Consistent with the presence of generalised hierarchical representations, we identified across-hierarchy distance effects in the left and right, superior, and inferior MPFC, smallest effect: $t(1651) = 3.07$, $p = .002$ (see Figure 8). As above, we found no evidence that these effects were modulated by inference accuracy. In each ROI, a Bayesian test found more evidence in favour of the null hypothesis for the distance by accuracy interaction when considering representational similarity in interleaved learners who performed at chance level on the

inference task; smallest BF_{01}: 3.57. Thus, generalised hierarchical representations in each region appeared to be present in participants who were unable to achieve good levels of inference performance.

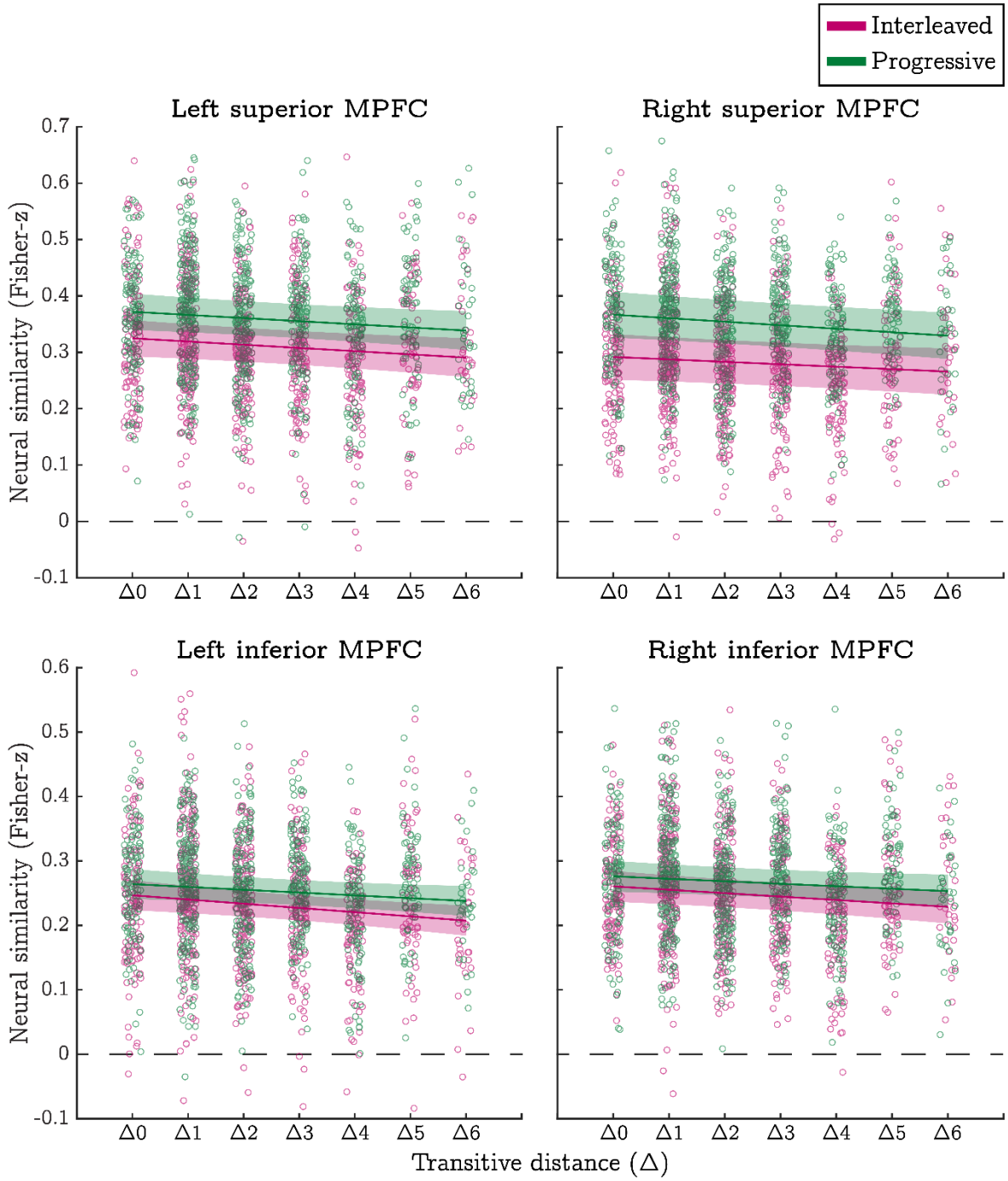


Figure 8. The superior and inferior medial prefrontal cortices (bilaterally) encode generalised representations that abstract the structure of a transitive hierarchy across different stimuli. Specifically, we found that each region exhibited across-hierarchy distance effects such that neural similarity between wall-texture representations learned in different sessions (i.e., remote vs recent) scaled with transitive distance. The effects were similar in both interleaved and progressive learners and did not depend on individual differences in inference performance. Note: unlike the previous analyses, the across-hierarchy distance effects test the similarity between wall-textures at the same hierarchical level (e.g., A[recent] vs A[remote]). All error-bars indicate 95% confidence intervals.

In line with the previous analyses, the left superior MPFC showed a main effect of training method indicating that representational similarity was generally higher in progressive- relative to interleaved-learners, $t(1651) = 2.89$, $p = .004$. This analysis also revealed the same effect in the right superior MPFC, $t(1650) = 5.38$, $p < .001$ (see Figure 8). Additionally, both the left and right superior MPFC exhibited an interaction between training method and transitive slope, smallest effect: $t(1651) = 3.08$, $p < .001$. This highlighted that the effect of training method on representational similarity was most pronounced in participants who applied encoding-based generalisations (i.e., those with a large transitive slope, Supplementary Figure 5). Finally, the left inferior MPFC produced a significant interaction between training method and inference performance, $t(1644) = 3.29$, $p = .001$. This suggested that inferential ability was negatively related to similarity estimates in the interleaved, $t(1644) = -2.94$, $p = .003$), but not progressive, $t(1644) = 1.70$, $p = .090$, training conditions (Supplementary Figure 6).

Discussion

In this study, 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 was that both humans and a simple computational model made transitive inferences better after learning the original premise pairs progressively. Next, contrary to our hypotheses, model-based analyses of the behavioural data revealed that encoding mechanisms were preferentially used across all experimental conditions. Considering the fMRI data, representational similarity analyses were also suggestive of the use of encoding-based mechanisms. BOLD pattern similarity correlated with the distances between stimuli in both of the learnt hierarchies in the hippocampal, entorhinal, 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. Finally, in the MPFC, pattern similarity effects were driven by the structure of the learnt hierarchies (A>B>C...), irrespective of which hierarchy the stimuli came from (stimulus "B" from the hierarchy learnt on day 1 was treated the same as stimulus "B" from the hierarchy leant on day 2). This suggests that MPFC encodes structural information at a higher level of abstraction than the MTL, where fMRI distance effects were specific to each hierarchy.

Our results 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 also observed similar effects in a multilayer perceptron. Here, progressive training in a simple neural network boosted successful inference. This happens because progressively ordering task contingences induces an error gradient that favours learning representations encoding the relative value of all stimuli (A>B>C...; see Results). When information is presented progressively, there is a potential for 'catastrophic interference' where the introduction of premise pairs towards the end of training results in forgetting of premise pairs that were presented earlier (McCloskey and Cohen, 1989). Our progressive training procedure largely avoided this by ensuring that once a discrimination had been introduced, it was presented in all subsequent epochs. Furthermore, while there is

evidence of some catastrophic interference in the MLP (see Supplementary Figure 1), processes that operate in the human brain likely limit the effect of catastrophic interference (e.g., by dynamically adjusting learning rates during training).

Although progressive training resulted in better inference performance, it had very little effect on the mechanisms used to make inference judgements. In fact, analyses of both the behavioural and fMRI data suggested that participants encoded structured representations of the relationships between the stimuli in all experimental conditions. Moreover, our univariate analyses of fMRI activity yielded results that were not consistent with the use of retrieval-based inferences. The lack of evidence for retrieval-based inferences may be due to how well the premise pairs were learnt in our study. Retrieval-based generalisation mechanisms are known to explain inference performance when memories have been acquired in a single episode (i.e., one-shot learning; Banino et al., 2016). As such, our findings are consistent recent proposals that encoding-based mechanisms may be preferentially engaged when generalisations are based on well-leant information (Kumaran et al., 2016b).

Irrespective of whether information was learnt via an interleaved or progressive procedure, we found that inferences were made on the basis of structural representations of the whole transitive hierarchy. Many models of memory generalisation are inspired by how representations of physical space may be applied to support inferences in abstract, non-spatial tasks (Momennejad, 2020; Stachenfeld et al., 2017; Whittington et al., 2020). This appears to be the case in our study: although participants learnt the information through simple pairwise presentations of the stimuli, our RSA results suggest that they represent the relationships between all the stimuli in a map-like manner, as if they occupied positions in a physical space (see Figure 6). Progressive presentation of the premise pairs may aid the construction of these structural representations and it is noteworthy that physical spaces are almost always explored progressively along connected paths. We speculate that progressive exposure may facilitate inference whenever generalisations must be made on the basis of structural representations of the information.

A critical feature of our study was that participants learnt two separate 7-item transitive hierarchies which comprised independent stimuli. This enabled us to test a core prediction of a recently proposed model of memory generalisation, the Tolman Eichenbaum Machine (TEM; Whittington et al., 2020). The TEM posits that the brain encodes structural representations of transitive hierarchies that are not directly tied to specific stimuli. Our results corroborate this prediction as the MPFC coded hierarchical positions in similar ways across the two sets of stimuli encountered on each day of training (see Figure 8). The TEM specifically predicts the existence of these representations within the medial entorhinal cortex. While we did not observe such effects in our entorhinal ROI, we suggest that the MPFC may be involved in abstracting relationships between similarly structured tasks at a very high level. Indeed, human brains likely host schematic codes that abstract relationships across a large number of hierarchical levels to allow the efficient transfer of learning across domains (Kemp and Tenenbaum, 2008, 2009; Kumaran, 2012).

Surprisingly, we found evidence of structural representations in participants who performed at chance level on the transitive inference task. Furthermore, the strength of structural representations was similar in both progressive and interleaved learners, even though these

groups exhibited very different levels of inference performance. It therefore appears that merely having structural task representations is not sufficient for good inference. This finding is incompatible with models that propose knowing the relative value of stimuli is instrumental in building a hierarchical task representation (e.g., Kumaran et al., 2016a). Nevertheless, other models can account for this observation. Neural codes postulated by both the TEM and the successor representation (SR) model dissociate the learned values of specific stimuli from structural relationships between them (Momennejad, 2020; Stachenfeld et al., 2017; Whittington et al., 2020). 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 7). However, in order to support transitive inference, SRs must be combined with a representation encoding the average reward returned by each stimulus.

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 somewhat similar study by Ellenbogen and colleagues (2007). We also did not support our hypothesis that consolidation would bias the use of encoding-based inference mechanisms; the few effects of consolidation on the fMRI data were not predicted and difficult to interpret. Further work will be needed to clarify the role of sleep consolidation in making memory generalisations.

In summary, we show that progressive learning results in a dramatic improvement in performance on a transitive inference task and that humans use encoding-based mechanisms to inform their inference judgements. Both MTL and MPFC support structed representations of the transitive hierarchies which likely support transitive inference but are not in themselves sufficient for successful generalisation. Information represented within the MPFC is encoded at a higher level of abstraction compared with MTL regions, being less bound to the specific stimuli that were experienced. Taken together, these findings provide strong support for encoding-based models such as the TEM, which explicitly predicts that generalisations are based on learning map-like structural representations during spatial and non-spatial tasks.

Methods

Participants

Thirty-four, right-handed participants were recruited from the University of Sussex, UK (16 females, mean age = 25.9 years, SD = 4.596). 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. Participants 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 in each. 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 procedure 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 of these trials were presented in a pseudorandom order that was selected 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 pieces of information are trained can have a fundamental effect on the representations that are learned, we examined interleaved and progressive training in a multilayer perceptron (MLP). The MLP was 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/). Specifically, the MLP consisted of 3 fully connected layers: 1) an input layer with 7 binary inputs, 2) a linearly activated hidden layer with a single biased neuron, and 3) a SoftMax output layer with 7 biased outputs (see Figure 1).

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 neurons (e.g., A & B) to a value of 1 and setting all other input neurons (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.

We trained the MLP using both interleaved and progressive procedures 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). Progressive training involved 6 sequential epochs of different lengths which gradually introduced each discrimination one-by-one (10,000 iterations). The only difference between the training procedures used in the behavioural task (described above) and those used to train the MLP was the number of trials. Specifically, we increased the number of trials by a factor of 10 when training the MLP to allow for sufficient learning (see https://osf.io/uzyb7/).

After training, we tested the MLP on the premise and inferred discriminations detailed in Figure 1B using a 2-alternative Luce choice decision rule:

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

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.

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: *1)* correctly inferring a non-trained discrimination requires retrieving a unified representation which may be activated by recalling any one of the mediating contingencies, and that *2)* inferences are easiest when there is a large 'distance' between discriminable features. As such, this model captures a common assumption of encoding-based models of generalisation; that learned 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:

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

Supplementary Figure 8A 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 8B). 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 of that discrimination.

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 nonlinear optimiser ("fmincon", MATLAB Optimization Toolbox, R2020a) tasked with maximised the likelihood of the observed data for a given R_i.

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 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 1000 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 (Andersson et al., 2001; Hutton et al., 2002). 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 (Mumford et al., 2012). 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, beta estimates for 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, beta estimates were linearly decomposed into voxel-wise representations of each wall texture in the reinforcement learning task. As most wall textures were present in multiple discriminations, this decomposition involved multiplying the 24 beta values from a given voxel with a 14*24 transformation matrix that encoded the occurrence of each wall texture across discriminations (see Figure 6A). Importantly, this decomposition left trivial correlations between the resulting texture representations since it combined measurements between different observations. Given this, we regressed out these trivial correlations in all RSA analyses (see 'Representational similarity analyses' below).

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. (2015). 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 (Schaefer et al., 2018). Normalised group averages of each ROI are shown in Supplementary Figure 9 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 (see 'MRI pre-processing' above). 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.

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 models also included a fixed-effect predictor of no-interest that accounted for trivial correlations between the wall-texture representations associated with the task structure (see 'MRI pre-processing'). This effect was also included as a random slope within each model (grouped by participant). Finally, the LMMs included an extensive set of random intercepts and slopes that accounted for all dependencies across wall-textures and participants.

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.

Bayesian analyses

We directly tested whether the effect of transitive distance was not dependent on inference performance in brain regions that appeared to encode a hierarchical task representation. Specifically, we tested the hypothesis that the effect of transitive distance was absent in participants in the interleaved training condition who performed at chance level on the inferred discriminations. To do this, we constructed a Bayesian prior for the distance by performance interaction that assumed the effect of distance on representational similarity would be abolished when inference performance was at $Pr(correct) = 0.5$. This prior was taken to be a normal distribution with a mean of $-\mu_\pi$ and standard deviation of σ_π, where μ_π and σ_π are the estimated raw effect size for the interleaved distance effect and its associated standard error (respectively). To compute a Bayes factor in favour of the null, this prior was combined with a model derived likelihood for the distance by performance interaction which had a mean and standard error of μ_θ and σ_θ:

$$BF_{01} = \frac{f(0 \mid \mu_\theta, \sigma_\theta)}{\int_{-\infty}^{\infty} f(x \mid -\mu_\pi, \sigma_\pi) \cdot f(x \mid \mu_\theta, \sigma_\theta) dx} \qquad Eq.\ 9$$

Where, $f(x|\mu, \sigma)$ denotes the probability density of a normal distribution at location x, with parameters μ and σ. Importantly, the likelihood represented by μ_θ and $\sigma_\theta$corresponded to a raw effect size evaluated at chance levels of performance in the interleaved condition.

Plots of RSA effects

All plots of the similarity data show similarity scores after regressing out between-subject random effects and nuisance effects of no interest (i.e., trivial correlations between co-presented wall-textures). For the multidimensional scaling plots presented in Figure 6, adjusted Fisher-z scores encoding the similarity between wall-textures of the same hierarchy were averaged across recent/remote conditions and participants. They were then transformed into mean cosine distance scores ($\hat{d}$) by the function $\widehat{d_{i,j}} = 1 - tanh(\widehat{z_{i,j}})$, where $tanh$ denotes the hyperbolic tangent function, $\hat{z}$ denotes mean adjusted Fisher-z scores, and the subscripts i and j represent different wall-textures indices. These cosine distance scores were then entered into the classical multidimensional scaling algorithm implemented in MATLAB R2020a.

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):

$$z = \Phi^{-1}\left(\frac{1}{2}\left(1 - 2^{-\frac{1}{n}}\right)\right) \qquad \textit{Eq. 10}$$

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 0.720%; see https://osf.io/dvmr2/). Additionally, visual inspection of scatter plots showing residual versus predicted scores indicated no evidence of heteroscedasticity or non-linearity (see https://osf.io/d2jky/).

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

Andersson, J.L.R., Hutton, C., Ashburner, J., Turner, R., and Friston, K. (2001). Modeling geometric deformations in EPI time series. Neuroimage *13*, 903–919.

Banino, A., Koster, R., Hassabis, D., and Kumaran, D. (2016). Retrieval-Based Model Accounts for Striking Profile of Episodic Memory and Generalization. Sci. Rep. *6*, 1–15.

Baram, A.B., Muller, T.H., Nili, H., Garvert, M.M., and Behrens, T.E.J. (2021). Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. Neuron *109*, 713-723.e7.

Birnbaum, M.S., Kornell, N., Bjork, E.L., and Bjork, R.A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. Mem. Cogn. *41*, 392–402.

Dumay, N., and Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words: Research report. Psychol. Sci. *18*, 35–39.

Ellenbogen, J.M., Hu, P.T., Payne, J.D., Titone, D., and Walker, M.P. (2007). Human relational memory requires time and sleep. Proc. Natl. Acad. Sci. U. S. A. *104*, 7723–7728.

Favila, S.E., Chanales, A.J.H., and Kuhl, B.A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. Nat. Commun. *7*.

Ferreira, C.S., Charest, I., and Wimber, M. (2019). Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. Neuroimage *201*, 115996.

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., and Turner, R. (2002). Image distortion correction in fMRI: A quantitative evaluation. Neuroimage *16*, 217–240.

Javadi, A.H., Tolat, A., and Spiers, H.J. (2015). Sleep enhances a spatially mediated generalization of learned values. Learn. Mem. *22*, 532–536.

Kang, S.H.K., and Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. Appl. Cogn. Psychol. *26*, 97–103.

Kaplan, R., and Friston, K.J. (2018). Hippocampal-entorhinal transformations in abstract frames of reference. Entorhinal Transform. Abstr. Fram. Ref. 414524.

Kemp, C., and Tenenbaum, J.B. (2008). The discovery of structural form. Proc. Natl. Acad. Sci. U. S. A. *105*, 10687–10692.

Kemp, C., and Tenenbaum, J.B. (2009). Structured Statistical Models of Inductive Reasoning. Psychol. Rev. *116*, 20–58.

Koster, R., Chadwick, M.J., Chen, Y., Berron, D., Banino, A., Düzel, E., Hassabis, D., and Kumaran, D. (2018). Big-Loop Recurrence within the Hippocampal System Supports Integration of Information across Episodes. Neuron *99*, 1342-1354.e6.

Kumaran, D. (2012). What representations and computations underpin the contribution of the hippocampus to generalization and inference? Front. Hum. Neurosci. *6*, 1–11.

Kumaran, D., and McClelland, J.L. (2012). Generalization through the recurrent interaction

of episodic memories: A model of the hippocampal system. Psychol. Rev. *119*, 573–616.

Kumaran, D., Melo, H.L., and Duzel, E. (2012). The Emergence and Representation of Knowledge about Social and Nonsocial Hierarchies. Neuron *76*, 653–666.

Kumaran, D., Banino, A., Blundell, C., Hassabis, D., and Dayan, P. (2016a). Computations Underlying Social Hierarchy Learning: Distinct Neural Mechanisms for Updating and Representing Self-Relevant Information. Neuron *92*, 1135–1147.

Kumaran, D., Hassabis, D., and McClelland, J.L. (2016b). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. Trends Cogn. Sci. *20*, 512–534.

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

Liu, Y., Dolan, R.J., Kurth-Nelson, Z., and Behrens, T.E.J. (2019). Human Replay Spontaneously Reorganizes Experience. Cell *178*, 640-652.e14.

Mcclelland, J.L., Mcnaughton, B.L., and Reilly, R.C.O. (1995). Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionist Models of Learning and Memory. Psychol. Rev. *102*, 419–457.

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

Momennejad, I. (2020). Learning Structures: Predictive Representations, Replay, and Generalization. Curr. Opin. Behav. Sci. *32*, 155–166.

Morton, N.W., Schlichting, M.L., and Preston, A.R. (2020). Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. Proc. Natl. Acad. Sci. U. S. A. *117*, 29338–29345.

Mumford, J.A., Turner, B.O., Ashby, F.G., and Poldrack, R.A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage *59*, 2636–2643.

Park, S.A., Miller, D.S., Nili, H., Ranganath, C., and Boorman, E.D. (2020). Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. Neuron *107*, 1226-1238.e8.

Potts, G.R. (1972). Information processing strategies used in the encoding of linear orderings. J. Verbal Learning Verbal Behav. *11*, 727–740.

Potts, G.R. (1974). Storing and retrieving information about ordered relationships. J. Exp. Psychol. *103*, 431–439.

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

Richards, B.A., Xia, F., Santoro, A., Husse, J., Woodin, M.A., Josselyn, S.A., and Frankland, P.W. (2014). Patterns across multiple memories are identified over time. Nat. Neurosci. *17*, 981–986.

Ritchey, M., Montchal, M.E., Yonelinas, A.P., and Ranganath, C. (2015). Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. Elife *2015*, 1–19.

Roediger, H.L., and Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychol. Sci. *17*, 249–255.

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., and Yeo, B.T.T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb. Cortex *28*, 3095–3114.

Schapiro, A.C., Turk-Browne, N.B., Botvinick, M.M., and Norman, K.A. (2017). 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. *372*.

Schlichting, M.L., Mumford, J.A., and Preston, A.R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat. Commun. *6*, 1–10.

Schlichtinga, M.L., and Prestona, A.R. (2014). Memory reactivation during rest supports upcoming earning of related content. Proc. Natl. Acad. Sci. U. S. A. *111*, 15845–15850.

Stachenfeld, K.L., Botvinick, M.M., and Gershman, S.J. (2017). The hippocampus as a predictive map. Nat. Neurosci. *20*, 1643–1653.

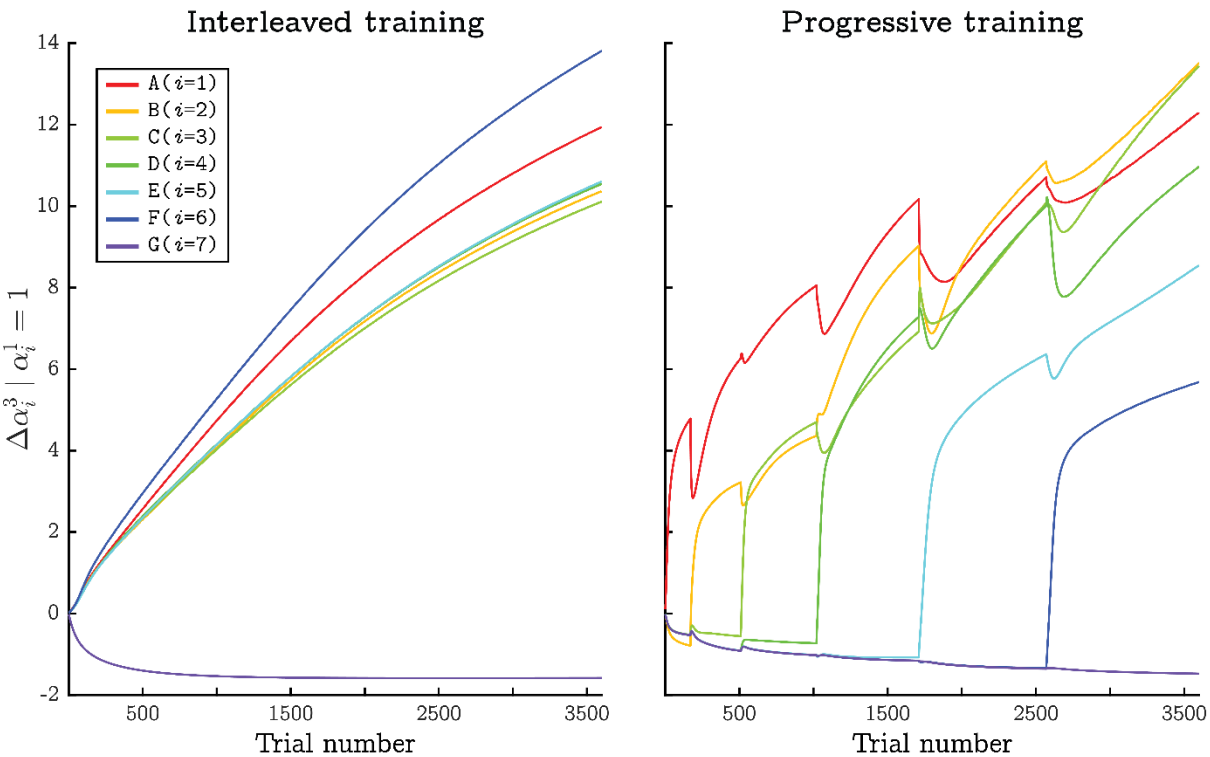
Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., and Schiller, D. (2015). A Map for Social Navigation in the Human Brain. Neuron *87*, 231–243.

Whittington, J.C.R., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T.E.J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. Cell *183*, 1249-1263.e23.

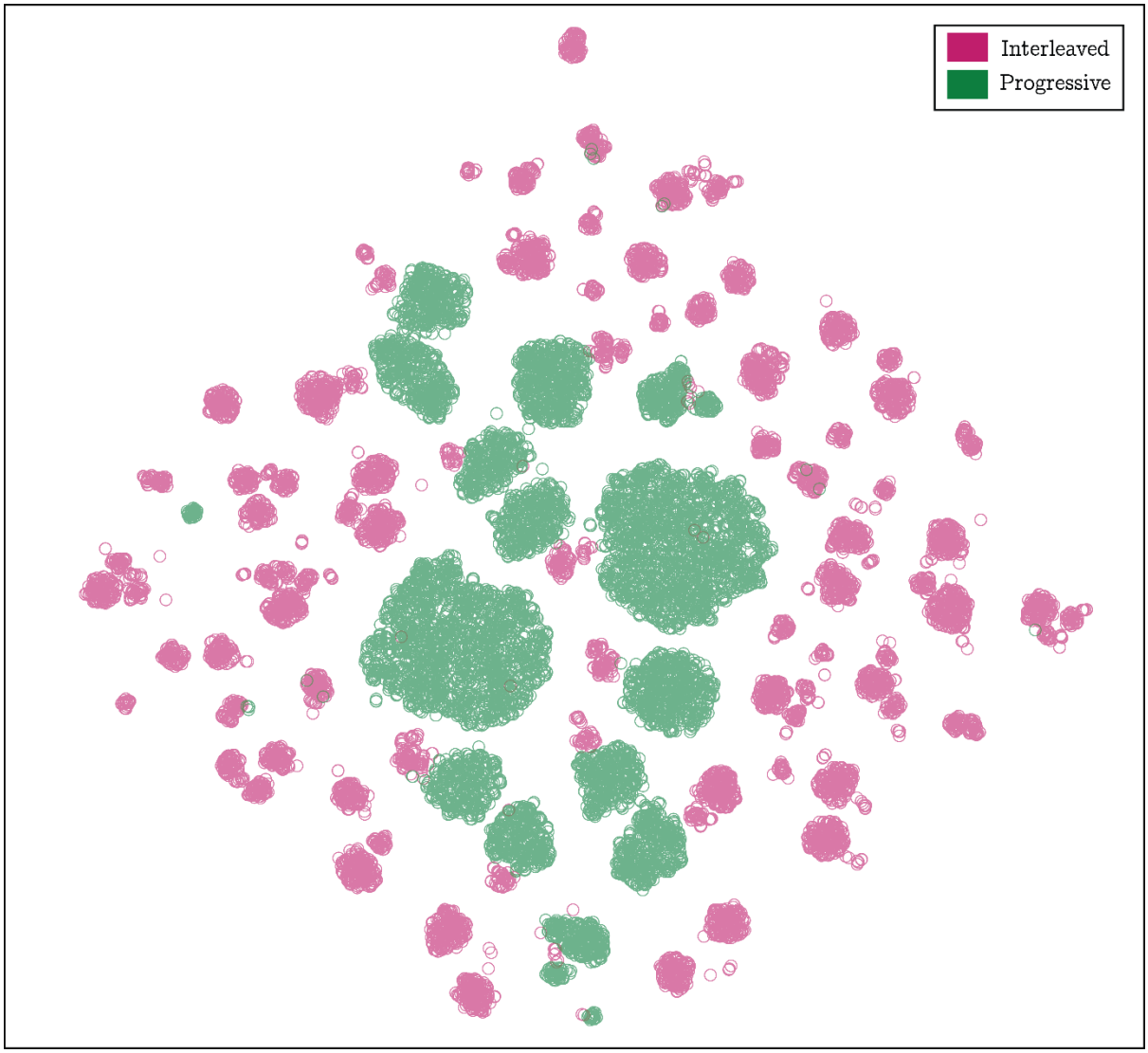
Zalesak, M., and Heckers, S. (2009). The role of the hippocampus in transitive inference. Psychiatry Res. - Neuroimaging *172*, 24–30.

Zeithamova, D., Dominick, A., and Preston, A.R. (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *75*, 168–179.

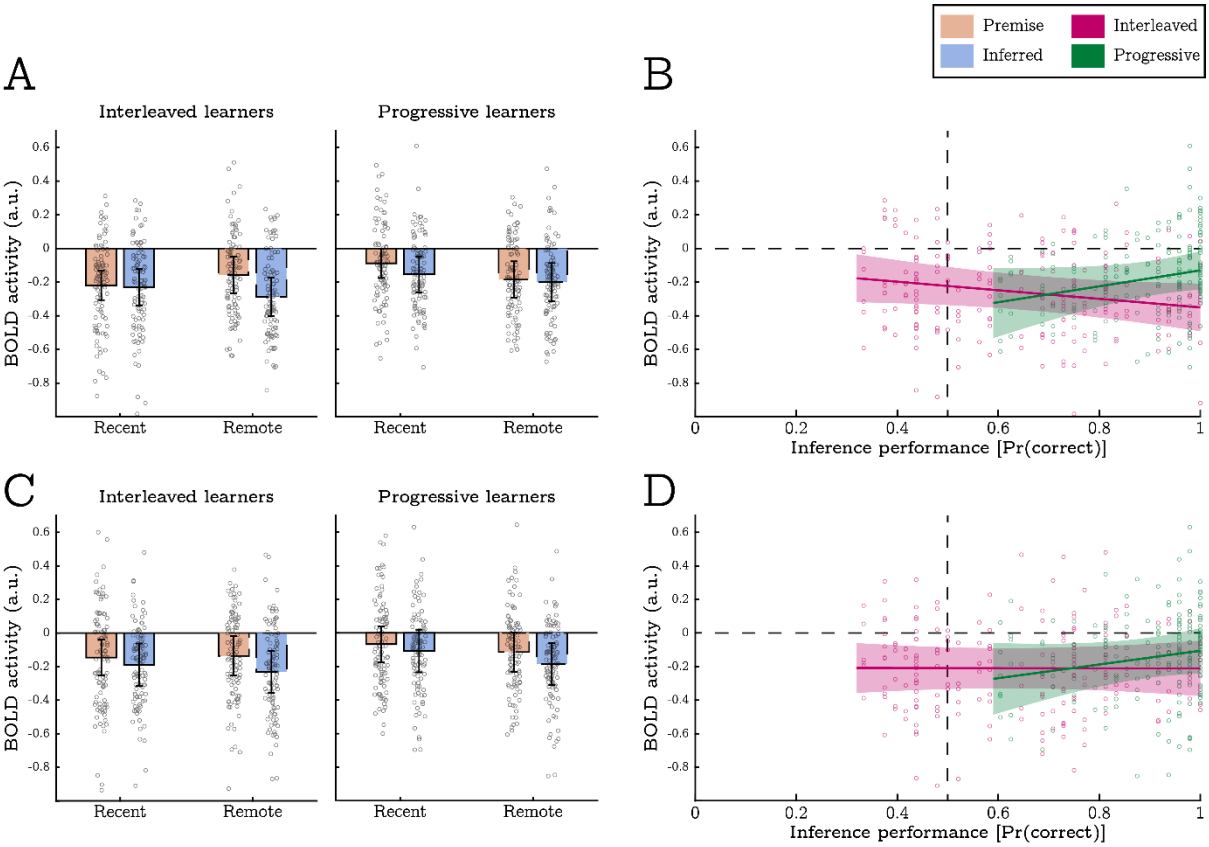
Supplementary Figures



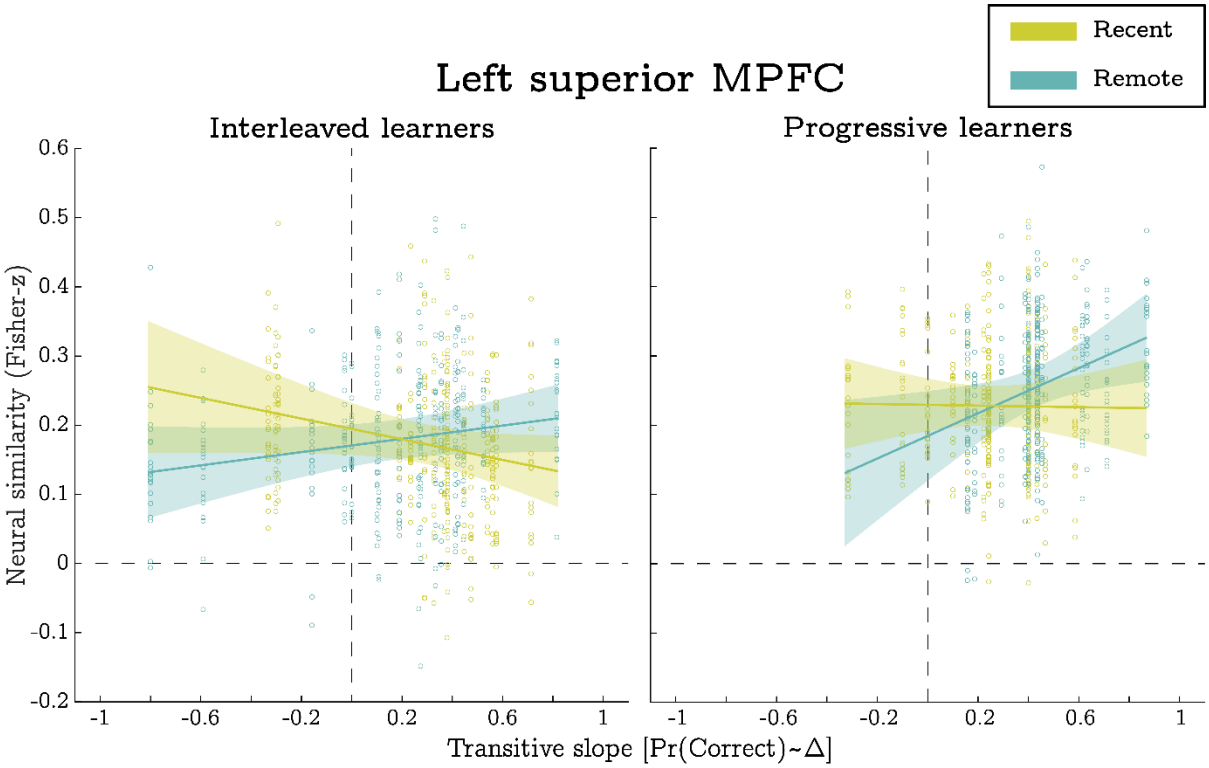
Supplementary Figure 1. Progressive training of configural discriminations allows a multilayer layer perceptron (MLP) to learn generalised values for each stimulus in the task. The plots indicate how the activity of an output neuron changes over training trials when the corresponding input stimulus is presented as a one-hot vector on the input layer (averaged across training runs). Different lines plot activations in different output neurons. 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.



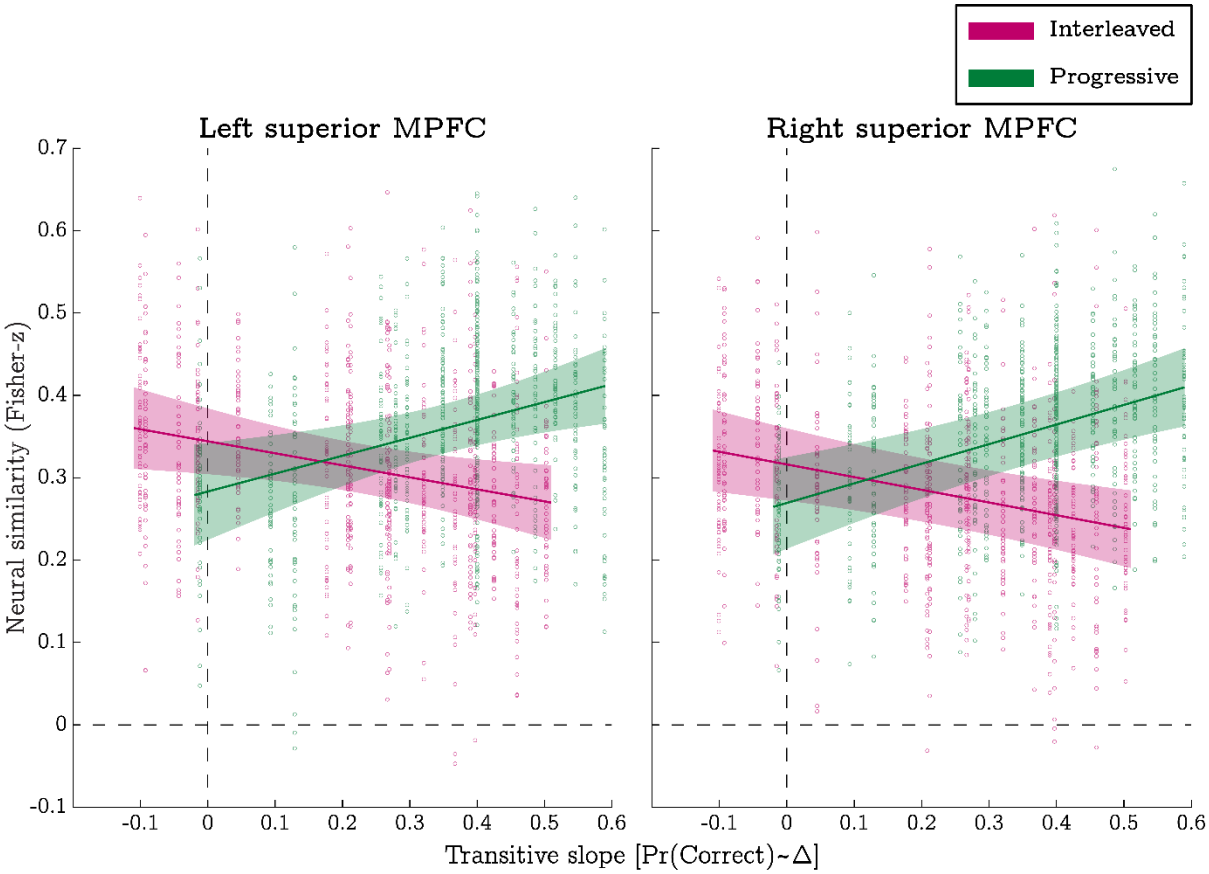
Supplementary Figure 2. Progressive training of configural discrimination contingencies in a multilayer layer perceptron (MLP) biases the network to learning a limited number of solutions. The plot shows a t-distributed stochastic neighbour embedding (tSNE) of the 22 network parameters generated in each MLP training iteration (N=10,000 per training method). 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 42 clusters following interleaved training, but only 23 clusters following progressive training. Differential entropy statistics also showed that interleaved training resulted in less clustered solutions relative to progressive training (h = -9.32 and -19.1, respectively).



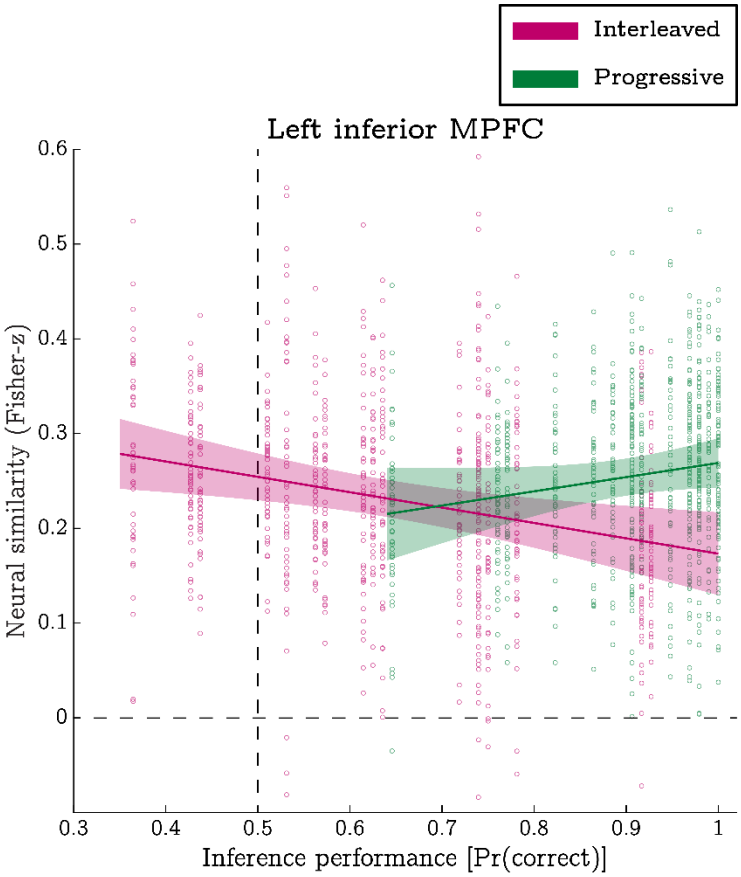
Supplementary Figure 3. Univariate BOLD activity in the hippocampus did not significantly predict inference performance. Panels **A** and **B** show activity in the left hippocampus. Panels **C** and **D** show activity in the right hippocampus. Bar charts display mean response amplitudes to all in-scanner 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.



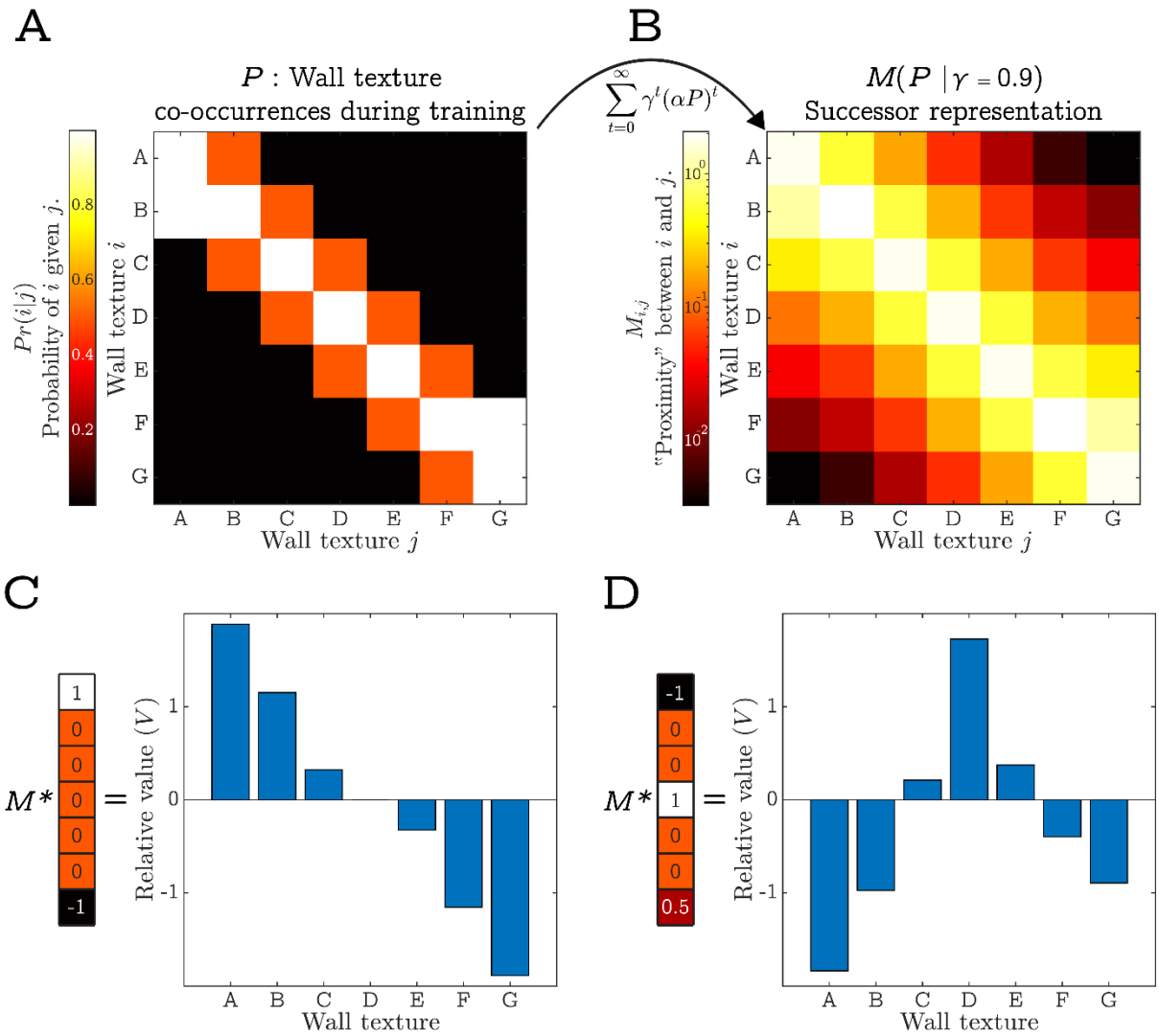
Supplementary Figure 4. The role of the left superior MPFC in supporting transitive inference may be modulated by systems consolidation. Each panel plots the association between transitive slope (i.e., the transitive distance vs accuracy correlation) and the neural similarity amongst wall-textures from the same transitive hierarchy (split by training session and method). The within-hierarchy RSA of the left superior MPFC highlighted a significant interaction between training session (recent vs remote) and transitive slope. This indicates that representational similarity in the remote condition was elevated over the recent condition when participants behavioural performance closely matched the predictions of encoding-based generalisations mechanisms (large, positive slopes imply the use of such mechanisms). All error-bars indicate 95% confidence intervals.



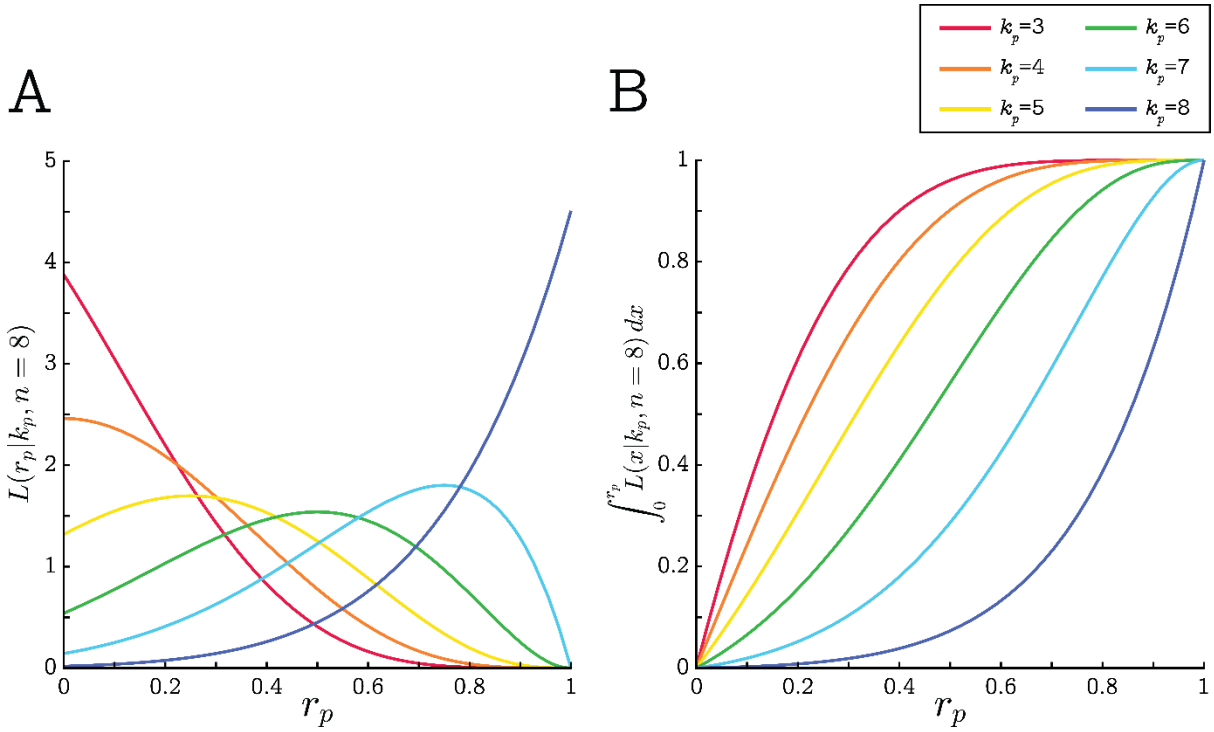
Supplementary Figure 5. Training method has a significant effect on representations in the superior MPFC, but this is most pronounced in participants who apply encoding-based generalisations. Each panel plots the association between mean transitive slope (the transitive distance vs accuracy correlation, averaged across sessions) and neural similarity amongst wall-texture stimuli from different transitive hierarchies (split by training method). The across-hierarchy RSA identified superior MPFC representations that abstracted across different stimuli regardless of overall inference performance, yet the similarity between these representations also varied as a function of training method and transitive slope. In addition to a main effect of training method, representations in the left and right superior MPFC showed a positive correlation between transitive slope and neural similarity for progressive learners (min $t = 2.95$), but a negative correlation for interleaved learners (min $t = 2.41$). This implies that training method most strongly affects MPFC representations when participants were heavily reliant on encoding-based generalisation mechanisms (since large, positive slopes imply the use of such mechanisms). All error-bars indicate 95% confidence intervals.



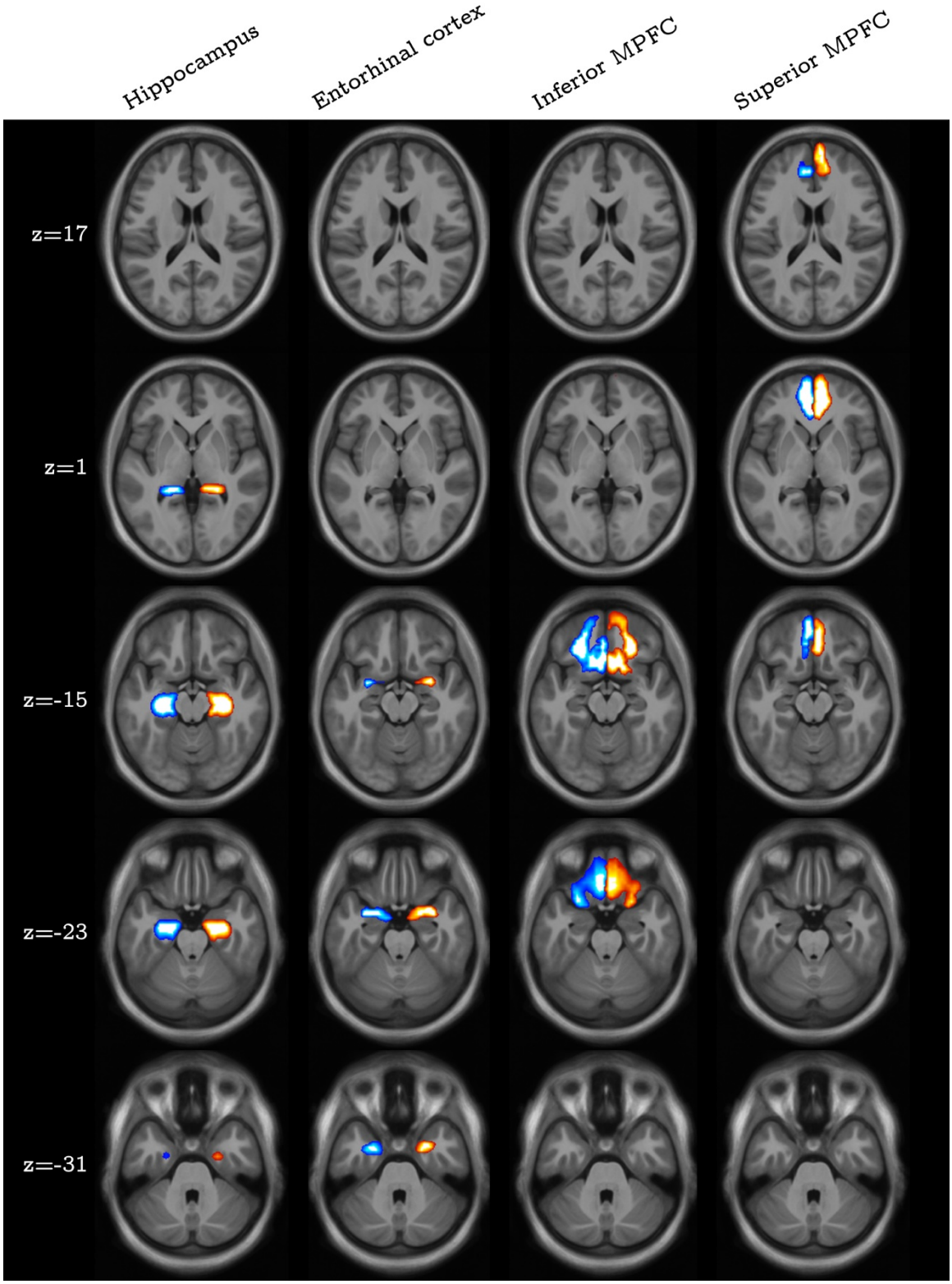
Supplementary Figure 6. Representational similarity across different transitive hierarchies is negatively related to inference performance in the left inferior MPFC, but only after interleaved training. The plot shows the association between mean inference performance (averaged across sessions) and neural similarity amongst wall-texture stimuli from different transitive hierarchies (split by training method). The across-hierarchy RSA in the left inferior MPFC identified representations that abstracted across different stimuli regardless of inference performance, yet the similarity between these representations also varied as a function inference performance in interleaved learners. All error-bars indicate 95% confidence intervals.



Supplementary Figure 7. 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 8. 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 9. 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.