A Computational Model of Perceptual and Mnemonic Deficits in Medial Temporal Lobe Amnesia

Patrick S. Sadil[1,*] and Rosemary A. Cowell[1]

[1]Department of Psychological and Brain Sciences, University of Massachusetts, Amherst, MA 01003, USA

*Corresponding author – Contact Information:

Email: psadil@gmail.com

Tel: +1 (413) 545-2383

Address:

Department of Psychological and Brain Sciences

University of Massachusetts

135 Hicks Way

Amherst, MA 01003

USA

Abstract

Damage to the Medial Temporal Lobe (MTL) has long been known to impair declarative memory and recent evidence suggests that it also impairs visual perception. A theory termed the representational-hierarchical account explains such impairments by assuming that MTL stores conjunctive representations of items and events, and that individuals with MTL damage must rely upon representations of simple visual features in posterior visual cortex, which are inadequate to support memory and perception under certain circumstances. One recent study of visual discrimination behavior revealed a surprising anti-perceptual learning effect in MTL-damaged individuals: with exposure to a set of visual stimuli, discrimination performance worsened rather than improved (Barense et al., 2012). We extend the representational-hierarchical account to explain this paradox by assuming that difficult visual discriminations are performed using a familiarity-based differencing rule, in which subjects compare the relative familiarity of the two to-be-discriminated items. Exposure to a set of highly similar stimuli entails repeated presentation of simple visual features, eventually rendering all feature representations equally – maximally – familiar and hence inutile for solving the task. Discrimination performance in patients with MTL lesions is therefore impaired by stimulus exposure. Because the unique conjunctions represented in MTL do not occur repeatedly, healthy individuals are shielded from this perceptual interference. We simulate this mechanism with a neural network previously used to explain recognition memory, thereby providing a model that accounts for both mnemonic and perceptual deficits caused by MTL damage with a unified architecture and mechanism.

INTRODUCTION

The Medial Temporal Lobe (MTL) has classically been associated with declarative memory (Scoville & Milner, 1957; Squire & Wixted, 2011; Squire & Zola-Morgan, 1991). But recent studies have implicated MTL structures in other functions, such as high-level perception, e.g., distinguishing two similar objects or scenes (e.g., Buckley et al., 2001; Lee et al., 2005), decision-making (Wimmer & Shohamy, 2012), statistical learning (Schapiro, Kustner, & Turk-Browne, 2012) and imagination (Maguire, Vargha-Khadem, & Hassabis, 2010; Schacter & Addis, 2007). Thus, while the notion of a 'declarative memory system' in the MTL (Squire and Zola-Morgan, 1991; Squire and Wixted, 2011) has provided critical insight into the organization of memory, it can no longer adequately explain the role of the MTL in cognition. Attention is now turning to the development of theories that explain why MTL structures are implicated in both studies of memory and studies of other cognitive functions.

A theory termed the representational-hierarchical account has been put forward to explain both mnemonic and perceptual deficits caused by damage to different structures within MTL (Bussey & Saksida, 2002; Cowell, Bussey, & Saksida, 2006; Kent, Hvoslef-Eide, Saksida, & Bussey, 2016). The representational-hierarchical account assumes that the ventral visual stream contains a hierarchical organization of representations that continues into the MTL. Early stages of the pathway (e.g., V1, V2, V4) are assumed to represent simple visual features (e.g., color, orientation), whereas more anterior regions are assumed to bring these simple features together into conjunctions of increasing complexity (Fig. 1). The hierarchy culminates in the MTL, where the conjunctions correspond to whole objects, scenes, or complex episodic events. The representational-hierarchical account claims that conjunctive representations in the MTL are important whenever a cognitive task – perceptual or mnemonic – employs stimuli with overlapping features, such that individual feature representations in posterior regions provide ambiguous information (Bussey & Saksida, 2002).

Paradoxical Finding of an Exposure-induced Deficit in Visual Discrimination

A recent study by Barense et al. (2012) documented a new and puzzling way in which MTL lesions impair visual discrimination. MTL amnesics and healthy controls were asked to judge whether pairs of simultaneously presented abstract stimuli were the same or different (Fig. 2a). In the High Ambiguity condition, each pair of to-be-discriminated stimuli shared two out of three explicitly defined features, whereas in the Low Ambiguity condition, the items in a pair shared no explicitly defined features. The task required participants to declare a 'mismatch' if any of the three features differed across the pair, or a 'match' if no difference was observed. Amnesic participants were unimpaired at discriminating Low Ambiguity objects, but in the High Ambiguity condition the performance of MTL patients deteriorated in the second half of trials (Fig. 2b).

Barense et al. explained their data in terms of the representational-hierarchical account: Individuals with MTL damage lack conjunctive representations of objects that are usually stored in perirhinal cortex (PRC), a structure in the MTL. Objects are instead represented only as a collection of simple features in posterior visual cortex. In the task of Barense et al., each stimulus is a unique conjunction of features (since items are trial-unique), but the features comprising the stimuli repeat across trials. After viewing many items, feature-level interference renders the feature representations in posterior visual cortex inadequate for solving difficult (High Ambiguity) discriminations. Control subjects can resolve this interference by utilizing the unique conjunction for each stimulus represented in PRC, but when PRC is damaged, performance is impaired.

These results present a paradox: the *decrease* in MTL amnesics' performance with increasing exposure to task stimuli contrasts with perceptual learning effects. Perceptual learning is often explained by assuming that experience increases the separability of stimuli, either because stimulus representations become less overlapping (Saksida, 1999; Schoups, Vogels, Qian, & Orban, 2001; Yang & Maunsell, 2004), or because the weights via which stimulus representations influence decision-making are optimized (Kumano & Uka, 2013; Liu, Dosher, & Lu, 2015). But if discrimination relies

on the separability of stimulus representations, and exposure differentiates representations, then it is not clear why brain damage should reverse this effect, causing exposure to hurt performance. Put another way, even if individuals with MTL damage possess only feature representations, why should exposure should cause feature representations to become more overlapping, rather than less?

How can this contradiction be resolved? In a previous neural network instantiation of the representational-hierarchical account (Cowell et al., 2006) impairments in *recognition memory* induced by MTL damage were accounted for by considering the familiarity signal evoked by stimulus representations in the brain. Exposure to many items sharing visual features entails frequent repetition of the features. Eventually, the representations of all features in posterior visual cortex appear familiar, causing individuals with MTL damage (who possess only feature representations) to perceive all items as equally familiar, thereby impairing recognition memory. Here, we invoke the same mechanism to explain the *visual discrimination* impairments reported by Barense et al. To apply this account to visual discrimination, we assume that participants used a familiarity-based differencing rule to decide whether two items were identical (Dai, Versfeld, & Green, 1996; Macmillan & Creelman, 2005).

Resolving the Paradox: Visual Discrimination Based on Familiarity Differences

We assume that, for difficult discrimination tasks like that of Barense et al. (2012), participants adopt the strategy of searching for a mismatch between two stimuli (Dai et al., 1996). To do so, they visually scan back and forth between stimuli in a pair; if the second item appears less familiar than the item just examined, this is taken as evidence for a mismatch in identity. That is, the new item appears novel to the extent that it differs from the item just inspected. If switching between items elicits a novelty signal (i.e., a difference in familiarity) that exceeds some criterion, the two stimuli are judged to mismatch (hence 'differencing rule'). In this way, the model's signal for familiarity – the 'tunedness' of the stimulus representations (Cowell et al., 2006) – is used to perform discrimination.

Just as in the memory experiments simulated by Cowell et al. (2006), representations in the model can saturate (i.e., reach a maximum) in terms of tunedness, or familiarity. In a discrimination task in which all stimuli are similar, the stimulus features appear repeatedly, resulting in all feature representations becoming highly tuned. After sufficient repetitions, all the representations for object features are equally, maximally, familiar – rendering the familiarity-based differencing rule inutile. In contrast, because individual objects do not repeat across trials, familiarity for these objects does not saturate and so remains useful for discrimination. In other words, conjunctive representations in PRC shield a person from perceptual interference. When these conjunctive representations are compromised by MTL damage, visual discrimination is impaired. A central assumption of the representational-hierarchical framework is that memory and perception share common neural mechanisms. Our model embodies this by using a mnemonic signal – familiarity – to solve a visual perceptual task.

A REPRESENTATIONAL-HIERARCHICAL ACCOUNT OF VISUAL DISCRIMINATION

Model Architecture

We use the model of Cowell et al. (2006) with minor modifications. The network contains a PRC layer and a layer corresponding to posterior ventral visual stream (Fig. 3). Visual objects are instantiated as 8-dimensional vectors. We assume that posterior regions represent simple conjunctions of two visual dimensions, so the Posterior layer is divided into four grids: posterior grid units receive two input dimensions and combine them into a simple conjunction, termed a ‘feature’. Because PRC is assumed to represent whole objects, all eight input dimensions converge into a single, 8-dimensional conjunction in the PRC layer. Thus, PRC contains unique, conjunctive representations of objects, whereas the Posterior layer represents the four 2-dimensional features separately.

All model layers are constructed from Kohonen grids, which mimic information processing mechanisms of cortex such as Hebbian learning and lateral inhibition (Kohonen, 1984). A Kohonen grid (or self-organizing map) is trained by successively presenting stimulus inputs and incrementally

adapting the weights of the grid's units on each presentation. As the grid learns, its representations for stimuli are sharpened, such that they are more tuned to a particular stimulus. Once a stimulus has been encoded, its representation on the grid is more selective: a smaller region of the grid is active, but the magnitude of that activation is greater (Fig. 4). The selectivity of the activation profile provides an index of familiarity (e.g., Cowell et al., 2006; Norman & O'Reilly, 2003).

Each dimension of an input stimulus takes one of four values in a given stimulus: 0.05, 035, 0.65, or 0.95. This scheme yields $4^8 = 65{,}536$ unique objects, which are represented holistically on the PRC layer, but only $4^2 = 16$ unique, 2-dimensional features on each Posterior grid. This scheme reflects a key assumption of the representational-hierarchical account: that there is a vast number of possible visual objects in the world, which are composed from a small number of visual elements.

Simulating Visual Discrimination Behavior

Fixations. In Barense et al. (2012), participants decided whether two simultaneously presented stimuli were the same or different. On trials declared as a 'match', eye-tracking data from control subjects revealed that participants made approximately 25 and 20 fixations at High and Low Ambiguity, respectively. In addition, participants exhibited a higher ratio of within-stimulus to between-stimulus fixations in High Ambiguity trials (~1.2) than in Low Ambiguity trials (~0.6). Barense et al. conjectured that this reflected a greater tendency to bind features together in the High Ambiguity condition. Accordingly, we hypothesized that differential fixation ratios might contribute to task performance (e.g., sampling stimuli with a higher within:between ratio may enable more reliable conjunctive representations in PRC) and aligned simulation parameters with these empirical data.

Stimuli are 'sampled' by the model via fixations, with each fixation allowing the network to encode the stimulus for 20 cycles (see *Appendix*). A probabilistic rule governs how the network switches back and forth between sampling the two items in a pair, and the probability of a switch is derived from the empirical within:between ratio for each experimental condition (1.2 or 0.6). On any

trial, the maximum number of fixations (i.e., when the search for a mismatch terminates and a 'match' is declared, see below) is 25 and 20 in the High and Low Ambiguity conditions, respectively.

Discrimination Decisions. On each trial, two stimuli are presented. One is arbitrarily selected first (Item A) and the network samples it via successive fixations until switching to the other stimulus, Item B. Upon switching, the model assesses evidence for a mismatch by computing a novelty score and comparing it to a criterion. The novelty score is a measure of 'familiarity change' obtained by taking the familiarity of Item A and subtracting the familiarity of Item B. When the two items are identical, the novelty score is zero; when they are different, Item B has slightly lower familiarity than Item A (because Item A has just been encoded, whereas Item B has not), yielding a positive novelty score. If the novelty score in any individual grid (any of the 4 Posterior grids or the PRC layer) exceeds the criterion, the items are declared to 'mismatch'. If the network finds no evidence for a mismatch after this switch, fixations continue. In the next comparison, Item B serves as the previously inspected stimulus and Item A as the newly fixated stimulus. Comparisons proceed until either a mismatch is declared or the maximum number of fixations is reached, whereupon a match is declared.

Criterion Shift. We assume that participants adjust their decision rule as their stimulus representations adapt. That is, if participants begin to perceive all items as more similar, they require less evidence to declare that two items are mismatching. To effect this, the criterion value of novelty required to declare a pair of items as mismatching (i.e., the *decision criterion*) is allowed to shift by setting it equal to the average of the novelty ('familiarity change') score on the previous 6 trials. In addition, for each decision, noise drawn from the uniform distribution ($\pm 1e^{-6}$) is added to the criterion. If familiarity change signals are small, this noise can swamp the novelty signal. The starting criterion value (i.e., criterion on Trial 1) is set to twice the maximum noise: $2e^{-6}$.

SIMULATION 1

In this study, we simulated the paradoxical finding that visual discrimination performance worsens with exposure to the task stimuli after damage to MTL. Our target empirical data were those of *Experiment 3* of Barense et al. (2012), in which patients with MTL damage and healthy controls indicated whether two simultaneously presented visual stimuli were a match or a mismatch. Stimuli were trial-unique items composed of three features (Fig. 2a). In High Ambiguity trials the items shared two out of three features, whereas in Low Ambiguity trials the items shared none. Individuals with PRC damage performed similarly to controls at Low Ambiguity, but at High Ambiguity their performance was intact initially, then fell sharply in the second half of the task (Fig. 2b).

Methods

As in *Experiment 3* of Barense et al., all stimuli were trial-unique. Stimulus pairs in the Low Ambiguity condition shared no 2-dimensional features, whereas High Ambiguity pairs shared three out of four features. (Although the stimuli from Barense et al. contained only three explicitly defined features, we used four features per stimulus for consistency with Cowell et al. (2006); the total number of features is arbitrary and does not qualitatively affect simulation results). Because the representational-hierarchical account assumes that all visual objects are composed from a limited pool of visual features (see *Model Architecture*, above) there are many possible unique object-level stimuli, but the features comprising them appear repeatedly. For this simulation, we further constrained the feature set to reflect the high degree of feature-overlap in the stimuli of Barense et al. by constructing stimuli (unique, four-featured objects) using only six out the sixteen possible 2-dimensional features for each Posterior grid (where a 'feature' is a conjunction of 2 input dimensions). This yielded $6^4 = 1296$ total possible unique objects, with high feature-overlap among them. As in Barense et al., each condition contained 36 'match' and 36 'mismatch' trials.

Networks were initialized and pre-trained (see *Appendix*) before performing visual discrimination in two conditions: High and Low Ambiguity. Control networks comprised both

posterior and PRC layers, whereas networks in the 'PRC Lesion' group possessed only a Posterior layer. In order to obtain a reliable estimate of performance we simulated 48 networks in each group, corresponding to 6 networks per human control participant in Barense et al. (2012).

Results

Networks with no PRC layer, like humans with PRC damage, were impaired relative to controls at High but not Low Ambiguity (Fig. 5), and the impairment was worse in the second half of trials. We do not report statistics on simulated data because significance scales arbitrarily with the number of networks run. Instead, we focus on qualitative patterns, which match those of the patient data, including the critical interaction between Lesion Group, Ambiguity and First/Second Half.

Discussion

The simulation of a lesion-induced deficit in discriminating High Ambiguity stimuli in the second half of trials hinges on three assumptions: (1) participants solve the task using a familiarity-based differencing rule; (2) the stimuli contain many low-level features that repeat over trials so that all stimulus features eventually appear familiar; (3) the stimuli are represented in PRC as whole conjunctions but in posterior regions as individual features. Together, these assumptions provide that, following PRC damage, discrimination performance is impaired once all features are maximally tuned.

Control networks possess both a PRC layer, which contains conjunctive representations that bind the object features into a whole, and a Posterior layer, which represents the object features individually. Lesioned networks possess only a Posterior layer, and so their performance relies upon feature-based representations. At the start of the task, individual features are not highly tuned. On each new trial, the network tunes its representations of these features. When the network switches to inspect the second stimulus, if that stimulus is not identical to the first, its features appear novel and a mismatch is correctly declared. However, after many trials, all features have been repeatedly encoded by the network. Now, at the start of a new trial, there can be no increase in tunedness (i.e., familiarity)

of the posterior feature representations when the network inspects the first stimulus. When the network switches to the second stimulus, even if that stimulus differs in identity from the first, these features are now equally tuned. At this point, the familiarity-based differencing rule can no longer reliably detect mismatching stimulus pairs on the Posterior layer.

The feature-level interference has more effect at High than Low Ambiguity for two reasons. First, networks perform more encoding on any given mismatch trial in the High than in the Low Ambiguity condition. This is because the ratio of within:between fixations is higher on High Ambiguity trials (see *Fixations*), so that networks make more fixations of the first inspected item in a pair before switching to the other item and attempting a discrimination decision. Since mismatch trials are often terminated after the first discrimination attempt (because a mismatch is detected), stimulus representations undergo more tuning, on average, at High than at Low Ambiguity. Consequently, the tunedness (familiarity) of features rises more steeply across trials at High Ambiguity. Although, as Barense et al. claimed, the use of a higher within:between fixation ratio might be useful for binding features into a conjunctive representation, that strategy proves detrimental to the lesioned model: in psychological terms, inspecting each stimulus more closely leads to faster build-up of interference. Evidence that humans with brain damage nevertheless adopt this disadvantageous strategy is provided by Erez, Lee, and Barense (2013), in which patients with PRC damage exhibited the same viewing patterns as control subjects. The second reason that lesioned networks' performance deteriorates faster at High Ambiguity is that mismatching High Ambiguity pairs share three out of four features, whereas Low Ambiguity pairs share no features. In seeking a mismatch, the network searches for any pair of features across the two stimuli that differ. In Low Ambiguity pairs, there are four mismatching features therefore a network is has four opportunities to discover a feature that has not yet reached maximum familiarity, which can be used to declare the two items a mismatch. In High Ambiguity pairs, there is only one mismatching feature and so the chance of discovering a mismatch is greatly reduced.

Performance in control networks is maintained throughout the task because of conjunctive representations in the PRC layer. Individual stimuli are trial-unique (i.e., whole conjunctions are never repeated) so whole-object representations in PRC never reach maximum familiarity. At the start of each new trial, the PRC representation for the first stimulus inspected always increases in familiarity during the inspection. When the network switches to inspect the second stimulus, if the second differs from the first, the second will elicit lower familiarity and the pair will be declared to mismatch.

These simulations demonstrate an explicit mechanism by which perceptual interference, along with compromised conjunctive representations, can cause MTL amnesics to suffer a paradoxical worsening of visual discrimination with increasing stimulus exposure. This mechanism aligns with the interpretation of the data offered by Barense et al. (2012). However, in Barense et al.'s *Experiment 3*, perceptual interference was incidental rather than explicitly manipulated – its effects were examined by comparing performance in the first and second halves of the study, which confounded degree of interference with order of presentation. Thus, although interference was hypothesized to account for the patients' greater impairment in the second half of trials, a potential alternative explanation was that patients grew relatively more fatigued than controls as the task wore on, rendering their performance in the harder (High Ambiguity) condition more impaired. To test directly the hypothesis that perceptual interference is central to MTL patients' deficits, Barense et al. conducted another study – *Experiment 4* – in which perceptual interference was explicitly manipulated. Next, we simulate that study.

SIMULATION 2

In *Experiment 4* of Barense et al. (2012), subjects completed three blocks in strict order: Low Interference 1, High Interference, and Low Interference 2. The High Interference block contained 88 trials (44 match, 44 mismatch) identical to the High Ambiguity trials of *Experiment 3*, in which pairs of abstract stimuli shared two out of three features. In Low Interference blocks, High Ambiguity and photographic trials were interleaved: each of 30 High Ambiguity trials (15 match, 15 mismatch) was

followed by two trials containing a pair of color photographs of real-world objects (58 trials in total; 29 match, 29 mismatch). Critically, performance in all blocks was assessed only on every third trial, which was always a High Ambiguity trial. Photographic stimuli shared few low-level features with the abstract stimuli of High Ambiguity trials (Fig. 6). Therefore, Low Interference blocks entailed less feature-level interference than High Interference blocks, and MTL patients were predicted to be less impaired at Low Interference. *Experiment 4* replicated the results of *Experiment 3*: MTL-damaged patients showed impaired discrimination at High but not Low Interference, even in the second Low Interference block (Fig. 7, left). This suggested that the impairment seen in *Experiment 3* was caused by the cumulative effect of perceptual interference, rather than increasing fatigue in the MTL patients.

Methods

As in Simulation 1, we modeled abstract stimuli by using six of the sixteen possible stimulus 'features' (i.e., simple conjunctions of two input dimensions) on each Posterior grid to construct four-featured stimulus wholes, yielding high feature-overlap among stimuli. To reflect the assumption of Barense et al. that abstract stimuli shared few low-level features with photographic stimuli, we used the remaining ten features (i.e., an independent set of features) to construct the photographic stimuli.

Networks performed three discrimination blocks. A block contained 88 trials, in which a network discriminated between a unique pair of stimuli on every trial. Every third trial in all blocks was a critical comparison trial, in which the stimuli were abstract stimuli (15 matching, 15 mismatching), constructed as in Simulation 1; on mismatching critical comparison trials, the stimuli shared two out of three features. In High Interference blocks the remaining 58 trials contained extra pairs of High Ambiguity abstract stimuli. In Low Interference blocks the remaining 58 trials contained pairs of photographic stimuli. As in Barense et al., in both High and Low Interference blocks, performance was judged only on critical comparison trials: the trials occurring at every third position. The difference between High and Low Interference was that, for Low Interference, the trials interposed

between critical trials contained items sharing no features with critical-trial stimuli whereas, for High Interference, interposed trials contained items similar to critical-trial stimuli.

Results

Replicating the results of Barense et al.'s *Experiment 4*, networks with no PRC layer discriminated stimuli at the level of Control networks in Low Interference blocks (the first and third blocks), but their performance was impaired in the High Interference (second) block (Fig. 7, right).

Discussion

The same mechanism that impaired lesioned networks in Simulation 1 drives the impairments in Simulation 2. In the High Interference condition, because all trials contain the same class of stimuli, stimulus features appear repeatedly and posterior feature representations reach maximum familiarity. Once this saturation occurs, the familiarity of posterior, feature-based representations cannot increase substantially when the network inspects a new stimulus at the start of a trial. Consequently, a network with no PRC layer cannot detect novelty (a drop in familiarity) upon switching to the other stimulus in the pair, and the familiarity-based differencing rule no longer discriminates the two stimuli. In contrast, in the Low Interference condition, two-thirds of trials contain photographs composed of different features than the critical-trial stimuli. The critical-trial stimulus features repeat too infrequently for their representations to reach maximum familiarity, and lesioned networks remain unimpaired.

INTERIM SUMMARY: ACCOUNTING FOR VISUAL DISCRIMINATION

Barense et al. (2012) reported a striking perceptual deficit in patients with MTL damage: the accumulation of perceptual experience impairs visual discrimination. This result is paradoxical because perceptual discrimination typically improves with exposure to the stimuli. Barense et al. argued that MTL-lesioned patients suffer from accumulated feature-level interference, which − in the absence of conjunctive MTL representations − cannot be overcome by feature representations in posterior visual cortex. Although we concur with this explanation, we suggest that it is incomplete.

Standard theories of perceptual learning claim that experience improves discrimination performance by reducing the overlap between stimulus representations (i.e., training increases representational separation). In such theories, the assumed mechanism for visual discrimination is that discriminability is proportional to the overlap between representations (Saksida, 1999; Schoups et al., 2001). But this mechanism does not appear to account for the data of Barense et al.: if exposure separates representations, even feature-based discrimination should improve with exposure, because even feature representations should become less overlapping with exposure. To explain why the performance of MTL patients in Barense et al. worsened after exposure to the stimuli, a theory based on representational overlap would require the counter-intuitive assumption that – although cortical representations of stimuli underlying perceptual learning typically become less overlapping with exposure (Jenkins, Merzenich, Ochs, Allard, & Guíc-Robles, 1990) – posterior feature representations in this task become more so.

In the account provided by Simulations 1 and 2, we eschew representational overlap as the mechanism for visual discrimination. Instead, we take the explanation offered by Barense et al. – that amnesics suffer from compromised conjunctive MTL representations – and combine it with a less commonly invoked discrimination mechanism: a familiarity-based differencing rule, which capitalizes on differences in familiarity caused by moment-to-moment encoding. Under this account – as in prior instances of the representational-hierarchical account applied to memory (Cowell et al., 2006; McTighe, Cowell, Winters, Bussey, & Saksida, 2010) – representations of features, but not conjunctions, reach maximum familiarity as interference accrues. Thus, after MTL damage, perceptual experience impairs visual discrimination.

Two aspects of the proposed mechanism for visual discrimination require clarification. First, although we simulate only two layers in the ventral pathway, our model comprises only a subset of the full hierarchy of representations in the brain, which includes simpler layers (containing lower-

dimensional representations) prior to the model's posterior layer and higher-dimensional layers, such as hippocampus, after PRC. Other tasks with different representational requirements may require other layers (Cowell, Bussey, & Saksida, 2010). For example, a discrimination task involving whole objects that repeat would require hippocampal representations capable of combining objects with context or with temporal information, to shield participants from object-level interference that could not be resolved by PRC representations alone. Second, we do not suggest that a familiarity heuristic must be used in all discrimination tasks. Dai et al. (1996) point out that a differencing rule is optimal (and adopted) when the familiarity signals for the to-be-discriminated stimuli are highly correlated. Familiarity signals would be less highly correlated during easy discrimination tasks in which the stimuli differ on the basis of distinct, salient features such as color. Such tasks could be solved by a more standard discrimination mechanism that assesses representational overlap. Our model suggests only that a familiarity heuristic is used for difficult discriminations between highly similar stimuli.

In Simulations 1 and 2, the model of Cowell et al. (2006) – originally developed to explain recognition memory – accounted for the deficits in perceptual discrimination observed in patients with PRC damage by Barense et al. (2012). Critically, the mechanism by which lesioned networks are impaired at visual discrimination is the same mechanism that causes deficits on recognition memory tasks. In both cases, lesioning the PRC layer removes the conjunctive representations that are required to shield the network from feature-level perceptual interference. In both cases, the remaining posterior feature-based representations reach an asymptotic level of tunedness, rendering them incapable of supporting the judgment of familiarity, or novelty detection process, required to solve the task.

A UNIFIED MODEL OF MNEMONIC AND PERCEPTUAL DEFICITS

Barring some minor modifications, the neural network that accounted for visual discrimination in Simulations 1 and 2 retained the architecture and parameters of the model that simulated recognition memory in Cowell et al. (2006). Nonetheless, to verify that the modifications did not qualitatively

change the predictions for recognition memory, and to provide a fully unified account of recognition memory and visual discrimination, we resimulated the three key findings explained by Cowell et al. (2006) using the network presented in Simulations 1 and 2. The three findings concerned deficits in recognition memory caused by PRC lesions, as follows: (1) the deficit in recognition memory is delay-dependent: it worsens as the retention interval increases; (2) the deficit is exacerbated by increasing the length of the list of to-be-remembered items; and (3) recognition memory for repeatedly-presented stimuli is not impaired by PRC lesions.

Simulating Recognition Memory

Following Cowell et al. (2006), these deficits will be accounted for by simulating an object recognition task that aligns with the spontaneous object recognition (SOR) and delayed non-match-to-sample tasks (DNMS) often used in animals (Ennaceur & Delacour, 1988; Mishkin & Delacour, 1975). In such tasks, subjects are presented with a list of items (the 'study' or 'sample' phase). After a variable retention interval, a copy of each studied item is presented, paired with a novel item. In the SOR task, healthy animals spontaneously spend more time exploring the novel item, yielding a 'recognition score' that reflects the difference in exploration time between the novel and familiar items. In the DNMS task, if the animal chooses the item that was not previously encountered, it receives a food reward. In both tasks, good performance depends upon discerning a difference in familiarity between the novel and familiar items. Our model simulations recapitulate this process.

In the recognition memory simulations below we follow closely the protocol used in Cowell et al. (2006). Accordingly, some aspects of the protocol for Simulations 1 and 2 that were intrinsic to the visual discrimination behavior (e.g., fixations, criterion shifts) are not included. We present a brief overview of procedures here; more details are provided in the Appendix and in Cowell et al. (2006). A pre-trained network encodes a list of stimuli during the study (or sample) phase, and provides a judgment of familiarity for both studied and novel objects during the test (or choice) phase. In tasks

involving a delay (*Delay-dependent Deficits* and *Trial-Unique Versus Repeated Stimuli*), we simulate interference between the study and test phases by presenting the network with stimuli sampled randomly from the set of all possible stimuli. (This implements the assumption that forgetting over a delay is caused by visual interference, see Cowell et al., 2006). In the final test phase, the network is presented with the list of sample stimuli, each paired with a novel item. An index of familiarity, termed 'Selectivity', is calculated at each grid in the network, for both sample and novel items. Selectivity values are averaged across grids first within a layer, then across layers, and a recognition score is then calculated by combining the selectivity of sample and novel items in a normalized difference score (see Appendix, Equation 5). A higher recognition score indicates greater familiarity of the sample than the novel item, i.e., better recognition memory performance.

Empirical SOR and DNMS tasks typically employ a diverse sample of everyday objects as stimuli; accordingly, we constructed input stimuli using all possible visual features (16 per posterior grid; $16 * 4 = 64$ in total) to reflect the greater variation among stimuli than in Barense et al. (2012). This stimulus set nonetheless retains the assumption of the representational-hierarchical account that real-world visual objects are composed from a limited pool of repeatedly occurring features.

As in Simulations 1 and 2, any assessment of the reliability of networks' performance with statistical tests would depend on the arbitrary choice of the number of simulated networks. We therefore assess model performance by examining the qualitative trends.

SIMULATION 3: DELAY-DEPENDENT DEFICITS IN RECOGNITION MEMORY

Establishing the involvement of a brain region in recognition memory requires a demonstration that, following damage to that region, the recognition memory deficit is delay-dependent. Increasing the delay between study and test is assumed to increase the load on memory, and so manipulating the length of the delay is akin to manipulating the extent to which the memory system is taxed (Cowell et al., 2006; Gaffan, 1974). In a number of studies of object recognition, animals with PRC lesions have

exhibited worsening recognition memory deficits with increasing delay (Buffalo, Ramus, Squire, & Zola, 2000; Buffalo, Reber, & Squire, 1998; Eacott, Gaffan, & Murray, 1994; Malkova, Bachevalier, Mortimer, & Saunders, 2001; Meunier, Bachevalier, Mishkin, & Murray, 1993; Mumby & Pinel, 1994). Accordingly, Cowell et al. (2006) began by simulating this delay-dependent deficit.

Methods

As in Cowell et al. (2006), a delay was implemented by presenting the network with interfering stimuli between the study and test phases. There were four trials. On each trial, a sample stimulus was presented to the network for 500 encoding cycles. Next, interfering stimuli were randomly selected, with replacement, from the pool of all possible stimuli (65,536 objects), with each being presented to the network for one encoding cycle. Finally, both the sample and a novel stimulus were presented to the network, enabling a judgment of relative familiarity (see Appendix). More interfering stimuli corresponded to a longer delay; we used delays of 0, 2000, 4000, 6000, and 8000 stimuli.

Results and Discussion

Replicating the findings of Cowell et al. (2006), removing the PRC layer of the network caused deficits in recognition memory that worsened as the length of the delay increased (Fig. 8). The mechanism underlying the deficit is shared with Simulations 1 and 2 of the Barense et al. data. The interfering items encountered during a delay are composed of a limited number of features: individual features appear repeatedly, whereas the unique conjunctions of features corresponding to whole objects do not. Consequently, feature-based representations on the Posterior layer are rendered familiar by interference, whereas conjunctive PRC representations are not. The longer the delay, the more interfering items, and the more closely the features approach maximum familiarity. In the test phase, after a delay, the Posterior representations of both the sample and the novel object appear familiar, because all features have appeared repeatedly. Lesioned networks, which possess only the Posterior layer, can no longer adequately discriminate the items on the basis of familiarity.

SIMULATION 4: THE EFFECT OF LIST-LENGTH

Increasing list length, like increasing delay, is assumed to increase the load on memory in a recognition memory task (Gaffan, 1974). Therefore a second key piece of evidence that the PRC is critical to recognition memory is that the use of longer sample lists exacerbates the deficit caused by PRC lesions (Eacott et al., 1994; Malkova et al., 2001; Meunier et al., 1993). The effect of list-length was simulated in Cowell et al. (2006), and we replicate that result here.

Methods

We used lists of items that contained either 1, 6, 12, or 18 pairs of sample stimuli. All stimuli were unique, and the sample and novel items in each pair shared no features, but features were allowed to appear repeatedly across items within a list. As in the animal studies that generated the target data, networks were required to encode all items in the list before proceeding to the test phase. To simulate simultaneous retention of all list items in memory, network weights were not reset between encoding of one list item and the next, and memory for all items was tested after the encoding phase was completed. No delay was simulated between study and test.

Results and Discussion

The recognition memory impairment in lesioned networks increased as a function of list length (Fig. 9), in line with empirical data (e.g., Eacott et al., 1994; Malkova et al., 2001; Meunier et al., 1993) and replicating Cowell et al. (2006). As in the delay-dependent effect of Simulation 3, lesioned networks in this list length simulation are impaired by the accumulation of feature-level interference that renders feature representations in the Posterior layer inadequate for solving the task. With longer lists, networks are more likely to repeatedly encounter all possible visual features during the study phase. After encoding a list of sufficient length, the representations of all objects – including novel test items – appear familiar on the Posterior layer, making the sample and novel items indiscriminable.

SIMULATION 5: TRIAL-UNIQUE VERSUS REPEATED STIMULI

The final recognition memory finding simulated by Cowell et al. (2006) was that PRC lesions cause impairments only when trial-unique items are used as stimuli; when items are presented repeatedly, animals with PRC damage are unimpaired (Eacott et al., 1994). In the most extreme version of repeated items object recognition, the stimulus set comprises only two stimuli. Both stimuli appear on every trial, with the designation of sample and novel assigned randomly within the pair. Under these conditions, the deficit of PRC-lesioned animals vanishes and, in addition, neither lesioned nor control animals perform well at a relatively short study-test delay of 30 seconds (Eacott et al., 1994).

Methods

The goal of this simulation was to demonstrate no deficit in lesioned networks for repeated-items recognition, alongside a deficit for trial-unique recognition. Because deficits in trial-unique recognition are revealed more reliably in animal studies when a delay is interposed between study and test, we simulated a short delay of 200 items in both conditions. To simulate this task, two sets of 30 pairs of stimuli were created. In the 'Trial-unique' set, no item appeared more than once. In the 'Repeating' set, the same pair of items appeared 30 times, with the sample and novel status assigned randomly on each trial. In both conditions, networks completed 30 trials, with each trial comprising a sample presentation (500 encoding cycles), a brief delay (200 interfering items, 1 encoding cycle each) and a choice phase (presentation of sample and novel items for assessment of relative familiarity). As in Cowell et al. (2006), to allow the cumulative effects of stimulus type (trial-unique versus repeated) to influence recognition performance, network weights were not reset between successive trials.

Results and Discussion

Replicating the simulations of Cowell et al., lesioned networks performed worse than control networks when stimuli were trial-unique, but when stimulus items were repeated the two groups' recognition scores were not different (Fig. 10). The poor performance even in control networks for repeated-items recognition illustrates an important tenet of the representational-hierarchical account of cognition: that

different tasks require conjunctive representations at different levels of the hierarchy. In the delay-dependent and list length simulations above, repeatedly occurring features created feature-level interference that was resolved by conjunctions in PRC. In repeated-items recognition, whole objects themselves occur repeatedly: now, even the PRC layer suffers from interference, in that all objects in the task appear familiar. The resolution of interference now requires an even higher-dimensional layer (not simulated here), in which objects are combined with time or context to form more complex conjunctions that enable recency judgments. A candidate brain region □ well-placed anatomically to provide such representations and known for its role in associative learning or binding (i.e., forming conjunctions of elements) □ is the hippocampus (Cowell et al., 2006). In line with this suggestion, empirical data indicate that animals with intact hippocampi can perform recency judgments, at least at zero delay (Eacott et al., 1994), but that lesions to the hippocampal formation impair this ability (Charles, Gaffan, & Buckley, 2004; Rawlins, Lyford, Seferiades, Deacon, & Cassady, 1993).

GENERAL DISCUSSION

The goal of this study was to investigate whether deficits in both memory and perception caused by MTL damage can be accounted for by a single model. Using the model of recognition memory presented by Cowell et al. (2006), we simulated the impaired performance of MTL amnesics on two visual discrimination tasks (Barense et al., 2012). To demonstrate that the account is truly unified, we then replicated the original recognition memory simulations of Cowell et al. (2006) using model parameters identical to those of the discrimination simulations. We thereby report, to our knowledge, the first computational model to explicitly simulate both mnemonic *and* perceptual deficits caused by MTL damage using a unified architecture and mechanism.

There exist numerous detailed and successful models of the role of MTL structures in memory (Bogacz, Brown, & Giraud-Carrier, 2001; Cowell et al., 2006; Greve, Donaldson, & Van Rossum, 2010; Linster & Hasselmo, 1997; Marr, 1971; McClelland, McNaughton, & O'Reilly, 1995; Norman

& O'Reilly, 2003; Treves & Rolls, 1994). More recently, at least two theoretical accounts of how MTL structures contribute to perception have been proposed (Aly, Ranganath, & Yonelinas, 2013; Bussey & Saksida, 2002; Cowell et al., 2010). However, no model has accounted for both mnemonic and perceptual contributions of the MTL with a single system. By doing so, the present model provides a parsimonious account of the data and raises a deeper question concerning how to characterize the mechanisms of cognition in the brain. The question arises from a central claim of the model: that more than one cognitive function – in this case, recognition memory and high-level perception – can depend on the same brain structure. When this occurs, should the function of that structure be characterized as dichotomous or should we reconsider how its function is defined? We argue for a reconsideration.

Consider the present study: we simulated two tasks, one traditionally defined as a perceptual task (the discrimination of simultaneously presented stimuli, without the need to retain information while stimuli are absent), and the other as a memory task (in which good performance requires the retention of information over a delay). One might interpret the simulation results as suggesting that the PRC can support two distinct cognitive functions. On the other hand, the mechanism that accounted for both tasks involved a judgment of familiarity, or novelty detection process, which is traditionally associated with memory (Bogacz et al., 2001; Mandler, 1980). So one might instead interpret the model as suggesting that the visual discrimination task of Barense et al. was in fact a memory task, and that perirhinal cortex specializes in making familiarity judgments. We reject both of these interpretations, advocating a very different alternative: The essence of the model is that cognition in the brain is best explained in terms of representations and representational changes. Accordingly, traditionally intuitive labels for cognitive processes – such as *memory* and *perception*, or *familiarity* and *recollection* – should be eliminated from the account of cognitive function. Although the term 'familiarity' provides an intuitive description of how stimulus representations in the model change with experience, a more accurate description is 'representational tunedness'. As a network is exposed

to a stimulus, the representation of the stimulus is tuned; the relative tunedness of two representations provides an index by which the items can be discriminated. Just as we can describe this as a 'novelty detection' mechanism in the visual discrimination task (proposing a mnemonic mechanism for a perceptual task), we could call it a 'perceptual learning' mechanism in the recognition memory task (a perceptual mechanism for a memory task). The most accurate characterization of the model avoids process-based labels altogether, replacing those notions with a parsimonious, single-system account in terms of representations.

In sum, we present a unified account of the mnemonic and perceptual deficits caused by MTL damage. Under this account, the contribution of a brain region to cognition is determined by the representations that the region contains. Cognitive functions are realized through operations upon those representations, and are influenced by changes to those representations. Representational changes can be critical to the performance of a cognitive task, or they can be disadvantageous. In the visual discrimination and recognition memory tasks simulated here, representational changes in PRC critically support performance in healthy participants, but representational changes in posterior visual cortex produce interference, disrupting performance when a person has MTL damage. Although a complete understanding of MTL function – to include decision-making, imagination, and spatial navigation – requires much future investigation, we suggest that a representational approach to building unifying theories may prove fruitful in this endeavor.

ACKNOWLEDGEMENTS

We thank Morgan Barense and Lok-kin Yeung for sharing their data, and Morgan Barense for comments on an earlier draft of the manuscript. This work was supported by NSF grant 1554871 to R.A.C.

APPENDIX

Initialization and Pretraining. Control networks comprise four Posterior grids and one PRC grid. To simulate PRC lesions we remove the PRC layer of the network, so that Lesioned networks comprise only the four Posterior grids. Every grid contains $200x200$ nodes whose weights are initialized with random values between 0 and 1. For all simulations, networks are pretrained for 500 cycles according to the standard Kohonen learning rule,

$$w_i(t+1) = w_i(t) + f(r,t) * \big(stim - w_i(t)\big) \quad (1)$$

in which,

$$f(r,t) = \eta(t) * v(r,t), \quad (2)$$

where w_i refers to the weights of node i, t is the current cycle, $stim$ is stimulus input, $\eta(t)$ is the learning rate, r is the city-block distance of node i from the most strongly active (winning) node, and $v(r,t)$ is a neighborhood function that scales the learning rate.

In the pre-training phase, both η and v decrease with each cycle. The neighborhood function $v(r,t)$ is defined by a Gaussian function:

$$v(r,t) = \exp\left(-\left(\frac{r}{G(t)}\right)^2\right) \quad (3)$$

where, $G(t) = 0.5 + 10t^{-B}$, and B is a constant determining the rate of shrinkage of the neighborhood function. The learning rate decreases as $\eta(t) = t^{-A}$, where the constant A determines the rate of decrease. In each pre-training cycle, the network is exposed to a different, unique stimulus.

In all simulations, networks begin each task condition (e.g., Low Ambiguity versus High Ambiguity in Simulation 1, or 0 delay versus 2000 delay in Simulation 3) by assuming the weight state reached at the end of pre-training, unless otherwise indicated. That is, the effects of encoding in the first simulated condition do not influence performance in the second simulated condition.

Encoding a Stimulus. In Simulations 1 and 2, the stimuli are encoded during the course of each trial, as the network stochastically fixates on one stimulus at a time. Each fixation of a stimulus entails encoding it for 20 cycles, following equations 1, 2 and 3, except that η and v are constants fixed at the values of the final pre-training cycle ($\eta = \eta(500)$; $G = G(500)$). After the first between-stimulus fixation, the network calculates a novelty score, which is the selectivity of current stimulus subtracted from the selectivity of the other stimulus. If the novelty score exceeds a criterion (see *Criterion Shift*, main text), the stimuli are declared to mismatch and a new trial begins. If a mismatch is not declared, encoding proceeds on the current stimulus, and fixations continue until either a mismatch is declared or the maximum number of fixations for the condition is exceeded.

In Simulations 3, 4 and 5, a stimulus is encoded during the sample phase following Equations 1, 2 and 3, with fixed η and v, for 500 encoding cycles. This reflects the encoding that would be achieved through multiple fixations (although fixations are not an important part of the mechanism and are not modeled). Because trials in Simulations 1 and 2 include multiple fixations, these parameters produce similar numbers of encoding cycles per trial across all simulations. However, the qualitative trends in the results do not depend on the exact number of encoding cycles, and could be achieved with other parameters. When simulating a delay, each interfering item is encoded for one cycle.

Measuring Familiarity (or Selectivity, or Tunedness). In Simulations 1 and 2, the familiarity of both stimuli is assessed whenever the network switches from fixating one stimulus to the other. In Simulations 3, 4 and 5, the familiarity of both stimuli is assessed during the test (choice) phase.

Activation a of node i is determined by the sigmoid function:

$$a_i = 1 \Big/ \left(1 + \exp\left(-k * \ln\left(\frac{1}{dist}\right)\right)\right) \quad (4)$$

where k is a constant that determines the steepness of the sigmoid function, $dist$ is the mean squared error between a node's weights w_i and the stimulus input vector.

The familiarity of a stimulus is given by the selectivity, or tunedness, of its activation pattern, calculated separately in each grid. Selectivity, *S*, is the activation of the peak (the summed activation of the winning node and its nearest 4 neighbors) divided by the summed total activation of the grid. Thus, via normalization, higher familiarity corresponds to more selective, or 'tuned', representations. Selectivity is measured separately in each grid in the network, yielding a single, object-level selectivity score from the PRC layer and four separate feature-level selectivity scores in the posterior layer. When comparing two stimuli in Simulations 1 and 2, the posterior selectivity scores are compared separately for each pair of features. Stimulus representations are not updated during the choice phase, in recognition memory simulations (Simulations 3, 4 and 5).

Recognition Score. For Simulations 3, 4 and 5, the recognition score, *R*, is given by,

$$R = \frac{S_{samp} - S_{nov}}{S_{samp} + S_{nov}} \quad (5)$$

where S_{samp} is the selectivity (tunedness, or familiarity) of the sample stimulus representation, and S_{nov} is the selectivity of the novel stimulus. R is calculated using S_{samp} and S_{nov} values that are averaged over the separate grids of the network. In Control networks, which possess a PRC grid, each S is computed by first averaging across all four posterior S values, then taking the mean of the Posterior and PRC S values. In Lesioned networks, S is given by the average of the Posterior S values.

Parameters. In all simulations, $d = 0.08$, $B = 0.3$, and $A = 0.6$. For consistency with Cowell et al. (2006), we simulated 6 networks per group in the recognition memory simulations (Simulations 3, 4 and 5). In the simulations of Barense et al., the stochastic nature of the decision rule rendered simulation results more variable, so to account for this we simulated 6 control networks for each of the 8 control subjects tested in the empirical study, and the same number of lesioned networks, giving 48 networks per group.

REFERENCES

Aly, M., Ranganath, C., & Yonelinas, A. (2013). Detecting Changes in Scenes: The Hippocampus Is Critical for Strength-Based Perception. *Neuron*, *78*(6), 1127–1137. doi:10.1016/j.neuron.2013.04.018

Barense, M. D., Groen, I. I. A., Lee, A. C. H., Yeung, L. K., Brady, S. M., Gregori, M., … Henson, R. N. A. (2012). Intact memory for irrelevant information impairs perception in amnesia. *Neuron*, *75*(1), 157–167. doi:10.1016/j.neuron.2012.05.014

Bogacz, R., Brown, M. W., & Giraud-Carrier, C. (2001). Model of familiarity discrimination in the perirhinal cortex. *Journal of Computational Neuroscience*, *10*(1), 5–23. doi:10.1023/A:1008925909305

Buffalo, E. A., Ramus, S. J., Squire, L. R., & Zola, S. M. (2000). Perception and Recognition Memory in Monkeys Following Lesions of Area TE and Perirhinal Cortex. *Learning & Memory*, *7*(6), 375–382. doi:10.1101/lm.32100

Buffalo, E. A., Reber, P. J., & Squire, L. R. (1998). The Human Perirhinal Cortex and Recognition Memory. *Hippocampus*, *8*, 330–339.

Bussey, T. J., & Saksida, L. M. (2002). The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex. *The European Journal of Neuroscience*, *15*(2), 355–364. doi:10.1046/j.0953-816x.2001.01850.x

Charles, D. P., Gaffan, D., & Buckley, M. J. (2004). Impaired Recency Judgments and Intact Novelty Judgments after Fornix Transection in Monkeys. *Journal of Neuroscience*, *24*(8), 2037–2044. doi:10.1523/JNEUROSCI.3796-03.2004

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2006). Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. *The Journal of Neuroscience*, *26*(47), 12186–12197. doi:10.1523/JNEUROSCI.2818-06.2006

Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Functional dissociations within the ventral object processing pathway: cognitive modules or a hierarchical continuum? *Journal of Cognitive Neuroscience*, *22*(11), 2460–2479. doi:10.1162/jocn.2009.21373

Dai, H., Versfeld, N. J., & Green, D. M. (1996). The optimum decision rules in the same-different paradigm. *Perception & Psychophysics*, *58*(1), 1–9. doi:10.3758/bf03205469

Eacott, M. J., Gaffan, D., & Murray, E. A. (1994). Preserved recognition memory for small sets, and impaired stimulus identification for large sets, following rhinal cortex ablations in monkeys. *Eur J Neurosci*, *6*(9), 1466–1478. doi:10.1111/j.1460-9568.1994.tb01008.x

Ennaceur, A., & Delacour, J. (1988). A new one-trial test for neurobiological studies of memory in rats. 1: Behavioral data. *Behavioural Brain Research*, *31*(1), 47–59. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3228475

Erez, J., Lee, A. C. H., & Barense, M. D. (2013). It does not look odd to me: Perceptual impairments and eye movements in amnesic patients with medial temporal lobe damage. *Neuropsychologia*, *51*(1), 168–180. doi:10.1016/j.neuropsychologia.2012.11.003

Gaffan, D. (1974). Recognition impaired and association intact in the memory of monkeys after transection of the fornix. *Journal of Comparative and Physiological Psychology*, *86*(6), 1100–1109. doi:10.1037/h0037649

Greve, A., Donaldson, D. I., & Van Rossum, M. C. W. (2010). A single-trace dual-process model of episodic memory: A novel computational account of familiarity and recollection. *Hippocampus*, *20*(2), 235–251. doi:10.1002/hipo.20606

Jenkins, W. M., Merzenich, M. M., Ochs, M. T., Allard, T., & Guíc-Robles, E. (1990). Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *Journal of Neurophysiology*, *63*(1), 82–104. doi:2299388

Kent, B. A., Hvoslef-Eide, M., Saksida, L. M., & Bussey, T. J. (2016). The representational–

hierarchical view of pattern separation: Not just hippocampus, not just space, not just memory? *Neurobiology of Learning and Memory*, *129*, 99–106. doi:10.1016/j.nlm.2016.01.006

Kohonen, T. (1984). *Self Organizing Maps* (Third Edit). Berlin-Verlag: SPRINGER.

Kumano, H., & Uka, T. (2013). Neuronal mechanisms of visual perceptual learning. *Behavioural Brain Research*, *249*, 75–80. doi:10.1016/j.bbr.2013.04.034

Linster, C., & Hasselmo, M. (1997). Modulation of inhibition in an olfactory bulb model reduces overlap in the neural representation of oflactory stimuli. *Behav. Brain Res*, *84*, 117–127.

Liu, J., Dosher, B. A., & Lu, Z. (2015). Augmented Hebbian reweighting accounts for accuracy and induced bias in perceptual learning with reverse feedback. *Journal of Vision*, *15*(10), 1–21. doi:10.1167/15.10.10.doi

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. doi:10.1017/CBO9781107415324.004

Maguire, E. A., Vargha-Khadem, F., & Hassabis, D. (2010). Imagining fictitious and future experiences: Evidence from developmental amnesia. *Neuropsychologia*, *48*(11), 3187–3192. doi:10.1016/j.neuropsychologia.2010.06.037

Malkova, L., Bachevalier, J., Mortimer, M., & Saunders, R. (2001). Neurotoxic lesions of Perirhinal cortex impair visual recognition memory in rhesus monkeys. *Learning and Memory*, *12*(9), 1913–1917.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252–271. doi:10.1037/0033-295X.87.3.252

Marr, D. (1971). Simple Memory□: A Theory for Archicortex Author (s): D . Marr Source□: Philosophical Transactions of the Royal Society of London . Series B , Biological Published by□: Royal Society Stable URL□: http://www.jstor.org/stable/2417171 Accessed□: 03-06-2016. *Philosopical Transactions of the Royal Society of London. Series B, Biological Sciences*,

262(841), 23–81. Retrieved from http://www.jstor.org/stable/2417171

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457. doi:10.1037/0033-295X.102.3.419

McTighe, S. M., Cowell, R. A., Winters, B. D., Bussey, T. J., & Saksida, L. M. (2010). Paradoxical False Memory for Objects After Brain Damage. *Science*, *330*(6009), 1408–1410. doi:10.1126/science.1194780

Meunier, M., Bachevalier, J., Mishkin, M., & Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *The Journal of Neuroscience*, *13*(12), 5418–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8254384

Mishkin, M., & Delacour, J. (1975). An analysis of short-term visual memory in the monkey. *Journal of Experimental Psychology. Animal Behavior Processes*, *1*(4), 326–334. doi:10.1037/0097-7403.1.4.326

Mumby, D. G., & Pinel, J. P. J. (1994). Rhinal Cortex Leions and Object Recognition in Rats. *Behavioral Neuroscience*, *108*(1), 11–18.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review*, *110*(4), 611–646. doi:10.1037/0033-295X.110.4.611

Rawlins, J. N. P., Lyford, G. L., Seferiades, A., Deacon, R. M. J., & Cassady, H. J. (1993). Critical determinants of nonspatial working memory deficits in rats with conventional lesions of the hippocampus or fornix. *Behavioral Neuroscience*, *107*(3), 420–433.

Saksida, L. M. (1999). Effects of similarity and experience on discrimination learning: a

nonassociative connectionist model of perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*(3), 308–323. doi:10.1037/0097-7403.25.3.308

Schacter, D., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773–786. doi:10.1098/rstb.2007.2087

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622–1627. doi:10.1016/j.cub.2012.06.056

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*(6846), 549–553. doi:10.1038/35087601

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *20*(11), 11–21. doi:10.1136/jnnp.20.1.11

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience*, *34*, 259–288. doi:10.1146/annurev-neuro-061010-113720

Squire, L. R., & Zola-Morgan, S. (1991). The Medial Temporal Lobe Memory System Human Memory. *Science*, *253*(5026), 1380–1386.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391. doi:10.1002/hipo.450040319

Wimmer, G. E., & Shohamy, D. (2012). Preference by Association: How Memory Mechanismsin the Hippocampus Bias Decisions. *Science*, *338*(October), 270–3. doi:10.1126/science.1223252

Yang, T., & Maunsell, J. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *Journal of Neuroscience*, *24*(7), 1617–1626. doi:10.1523/JNEUROSCI.4442-03.2004

Figure 1. The organization of representations according to the representational-hierarchical account. Simple visual features, such as an oriented line or a color, are represented in early visual regions. Features are combined into increasingly complex conjunctions going from posterior to anterior regions. The MTL lies at the apex of the hierarchy, where conjunctions correspond to whole objects (in perirhinal cortex) or to spatial scenes or episodic events comprising multiple objects and their context (in hippocampus).

Figure 2. Left Panel: Stimuli from Experiment 3 of Barense et al. (2012). Pairs of stimuli were presented simultaneously. Each stimulus was defined by 3 features: inner shape, outer shape, and fill pattern. High Ambiguity mismatching pairs shared 2 of these features, but Low Ambiguity mismatching pairs share 0 features. Right Panel: Empirical data from Experiment 3 of Barense et al. (2012). Subjects with perirhinal (PRC) lesions were impaired at discriminating High Ambiguity stimuli in the second half of trials. Significance was assessed via Crawford's t-test for each Lesion participant separately (Control n=8; Lesion n=2, Error Bars = SEM).

Figure 3. Model architecture. An object stimulus has eight input dimensions, paired into four 2-dimensional 'features'.The PRC layer is a single Kohonen grid, representing an object as a unique conjunction. The Posterior layer is composed of four Kohonen grids, which each represent one visual feature. Encoding and retrieval are identical on Posterior and PRC layers; the layers differ only in the complexity of representations.

Figure 4. Schematic illustration of activations in the network. Left Column: Activation due to a novel stimulus (upper panel shows PRC layer, lower panel shows Posterior layer). Right Column: activation due to a stimulus that has been encoded, i.e., sampled many times. The novel stimulus elicits a broadly distributed pattern of activity across the grid, whereas the encoded stimulus elicits a highly selective activation pattern with a peak over a subset of grid units.

Figure 5. Simulated data for Experiment 3 of Barense et al. (2012) (Simulation1). As in the empirical data (see Fig. 2), networks with PRC lesions were impaired at discriminating High Ambiguity stimulus pairs in the second half of trials. Error Bars = SEM.

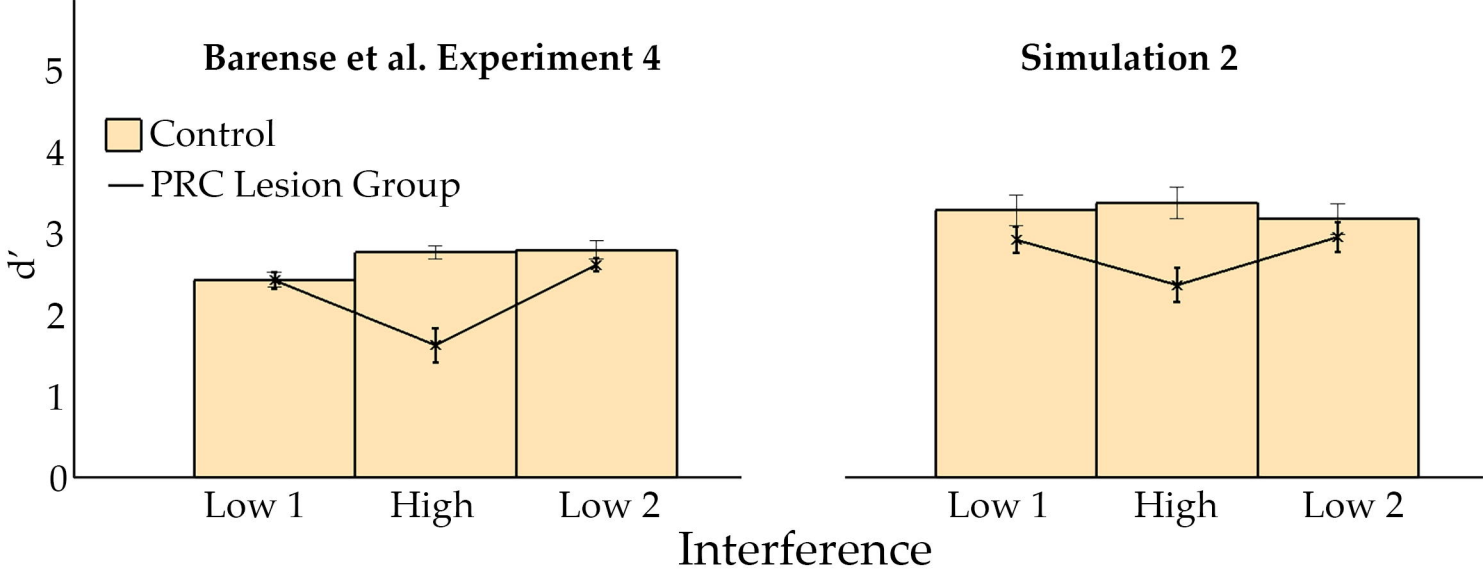
Figure 6. Stimuli from Experiment 4 of Barense et al. (2012). Low Interference blocks used photo stimuli in 2/3 of trials, which shared few features with the abstract stimuli used on critical comparison trials. High Interference blocks used abstract stimuli on every trial.

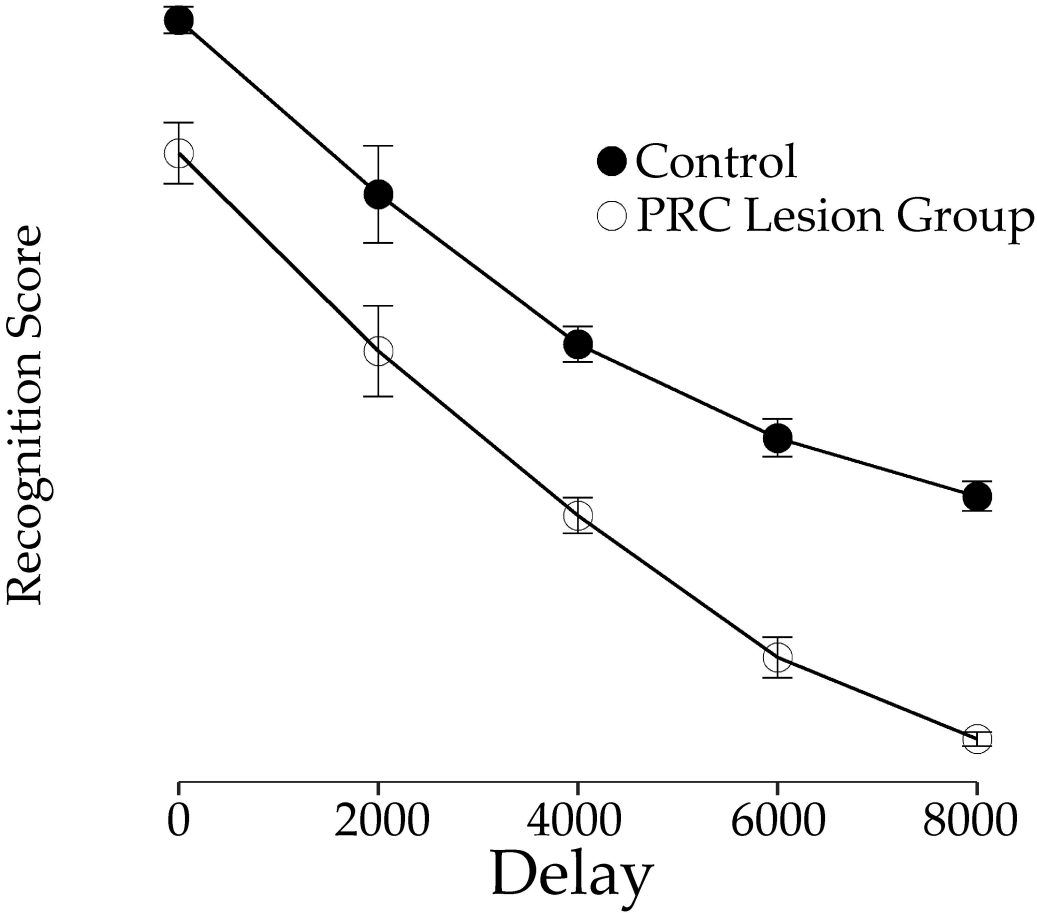
Figure 7. Left: Empirical data from Experiment 4 of Barense et al. (2012). Right: Model simulations for Experiment 4 (Simulation 2). Error Bars = SEM.

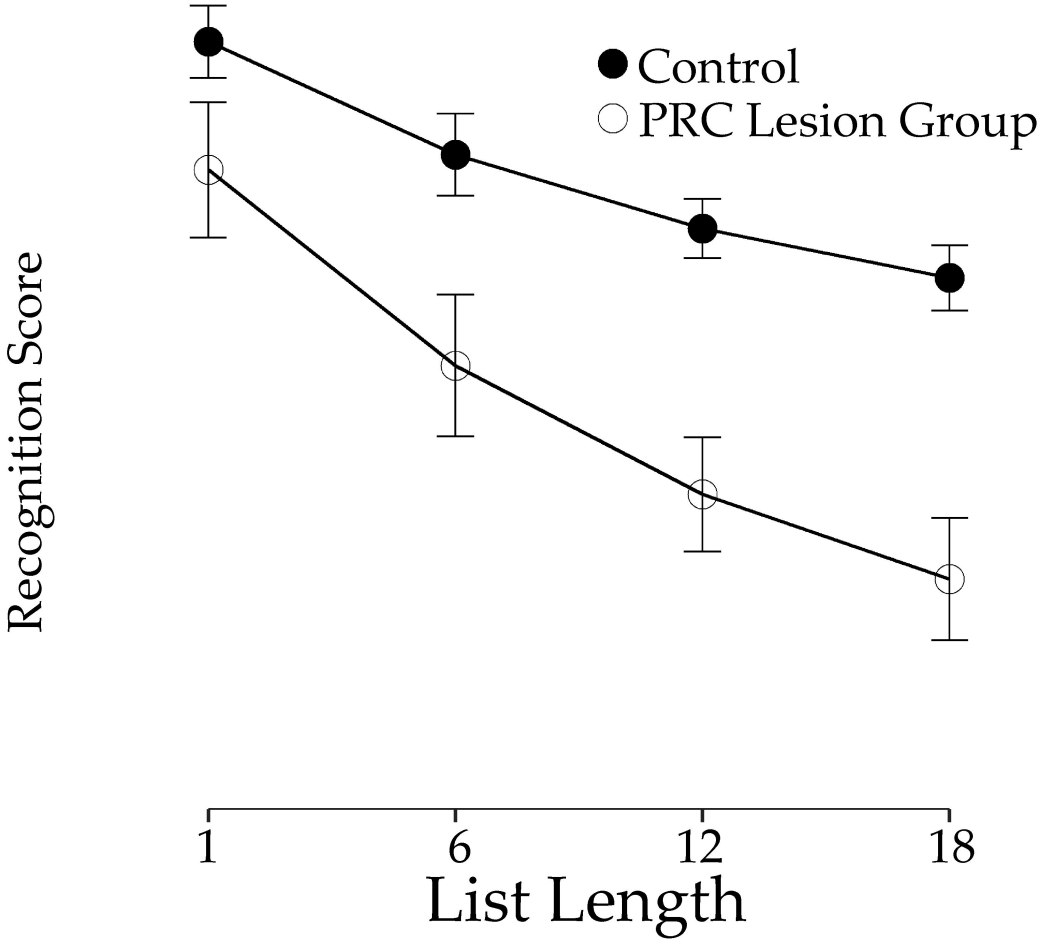
Figure 8. Simulation of a delay-dependent deficit in recognition memory (Simulation 3). The impairment in recognition memory is evident in the lower recognition scores for Lesioned networks, and this impairment increases as the delay between study and test increases. Abscissa indicates number of interfering stimuli sampled during the delay. These data replicate Cowell et al. (2006). Error Bars = 95% CIs.

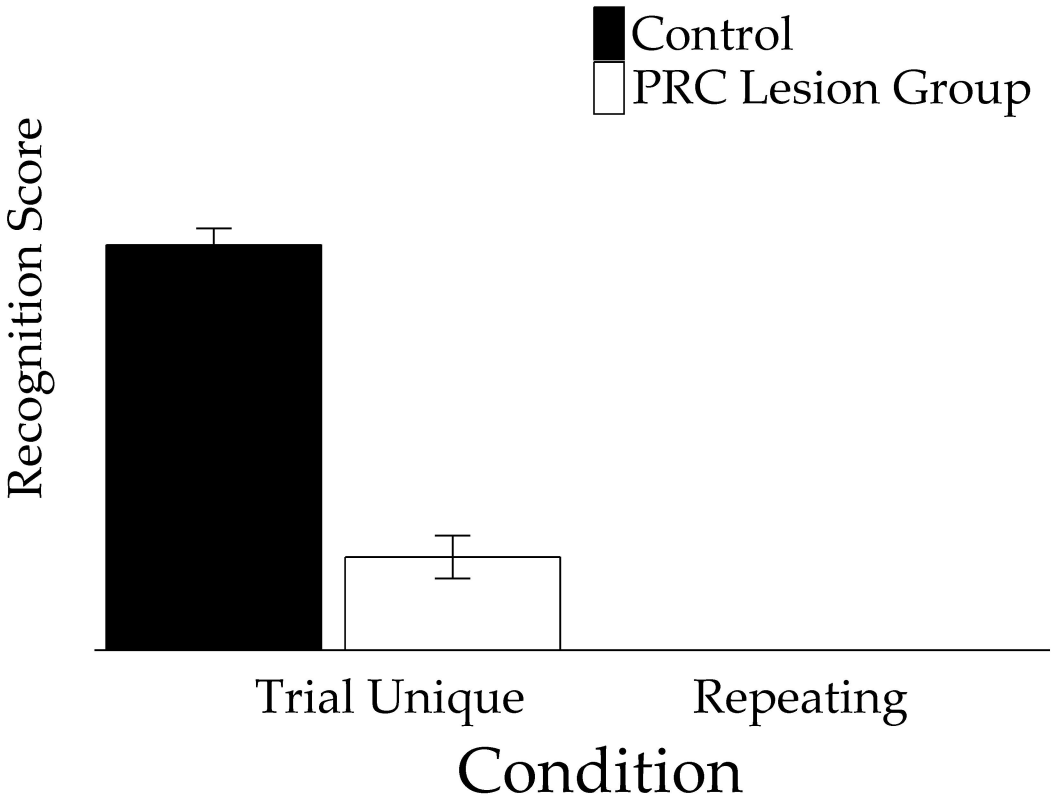
Figure 9. Simulation of the effects of list length on recognition memory (Simulation 4). Recognition scores decreases as the sample list length increases, and the rate of decrease is faster rate for networks without a PRC layer. These data replicate Cowell et al. (2006). Error Bars = 95% CI.

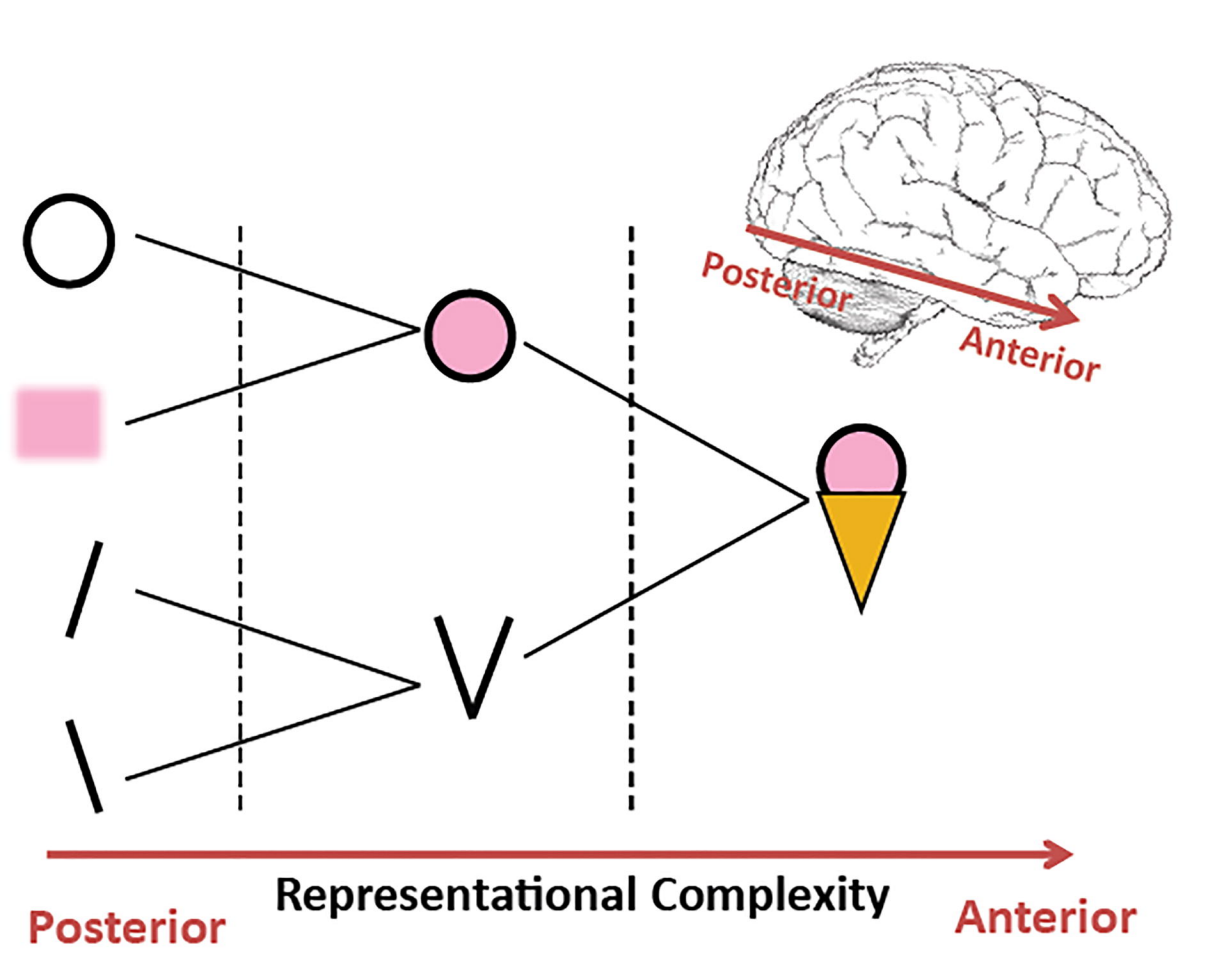
Figure 10. Simulation of the effects of trial-unique versus repeated stimuli (Simulation 5). Recognition scores are lower in lesioned networks when sample and novel stimuli are trial-unique. There is no group difference in recognition memory when stimuli are repeated because scores in both groups are equally poor. These data replicate Cowell et al. (2006). Error Bars = 95% CI.

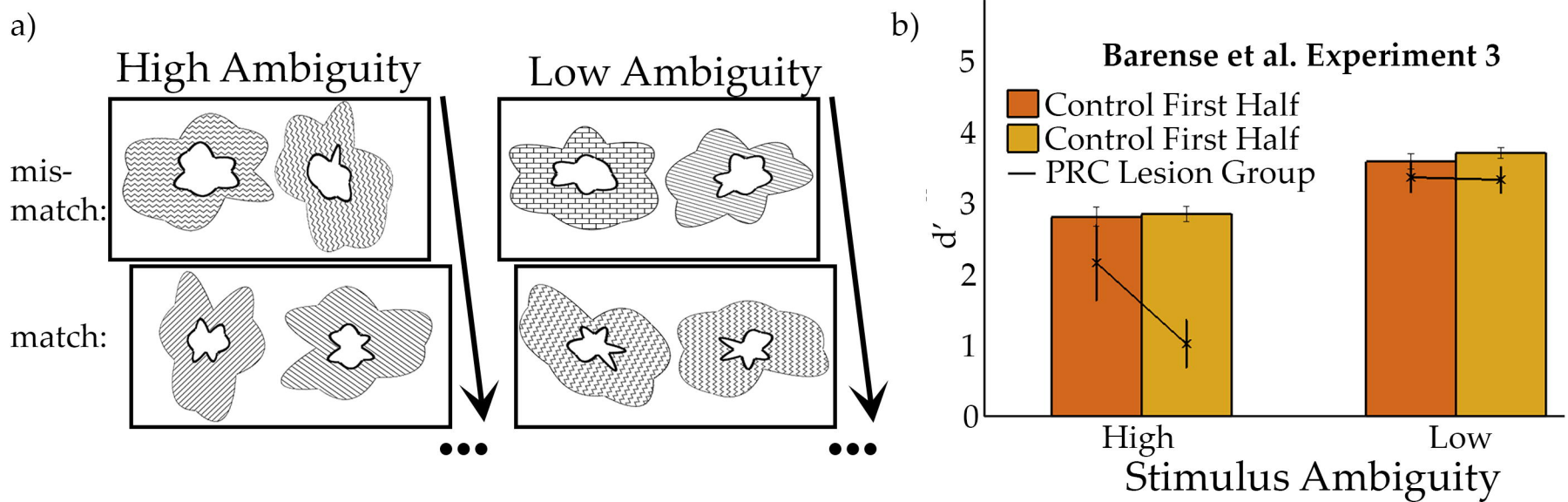

Barense et al. Experiment 4
Simulation 2
Control
PRC Lesion Group
d′
0
1
2
3
4
5
Low 1
High
Low 2
Low 1
High
Low 2
Interference

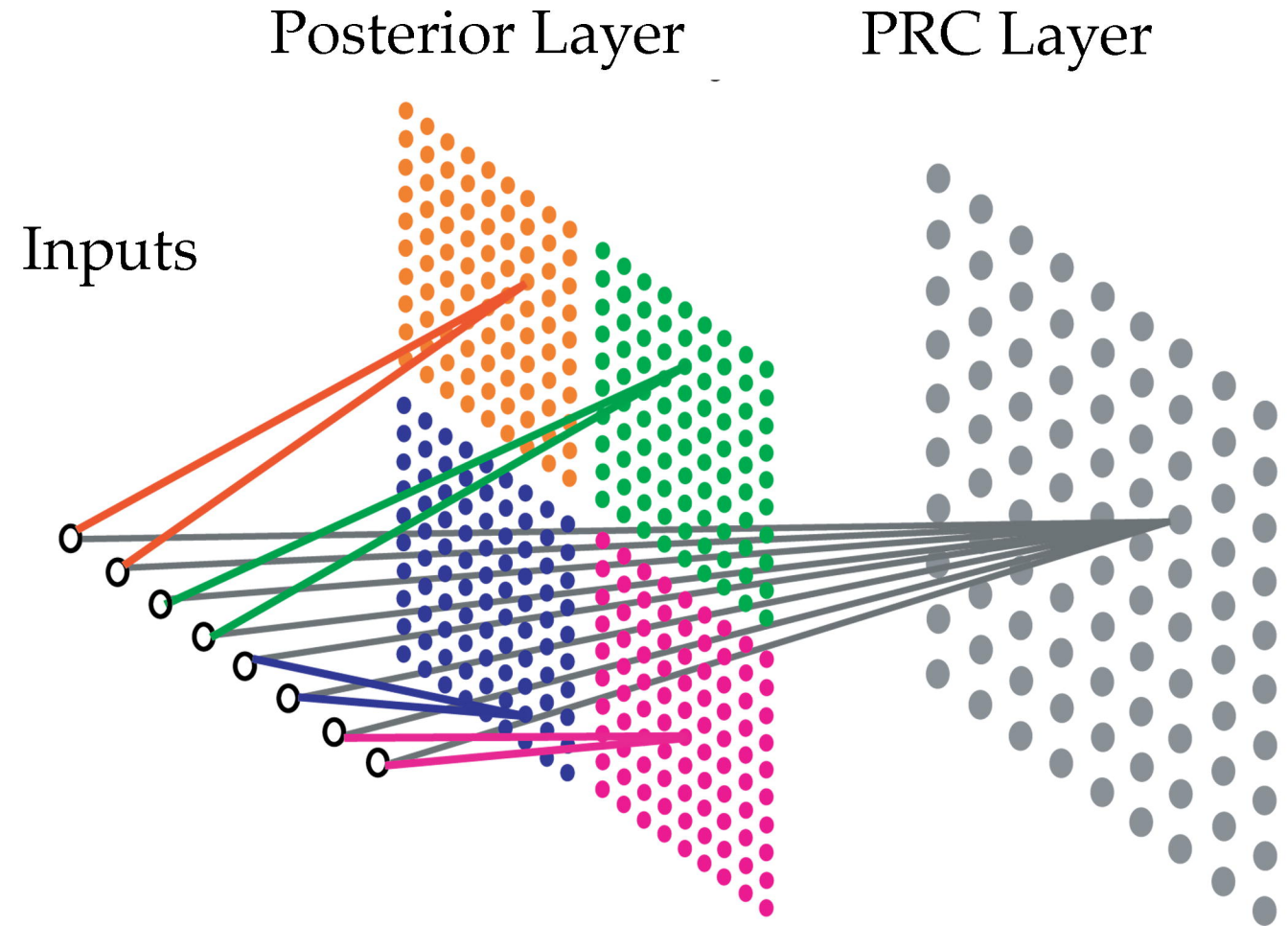

Recognition Score
Control
PRC Lesion Group
0
2000
4000
6000
8000
Delay

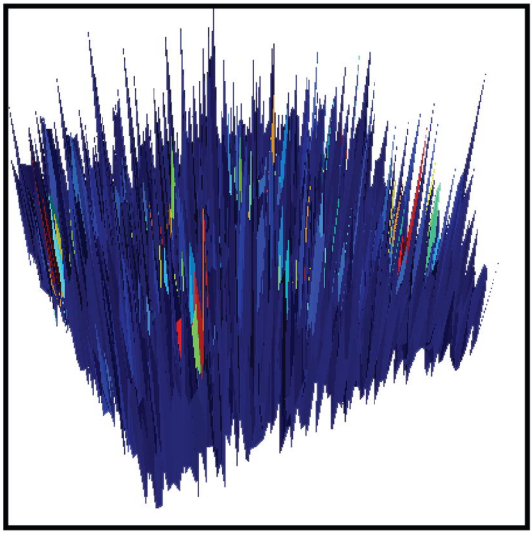

Control
PRC Lesion Group
Recognition Score
1
6
12
18
List Length

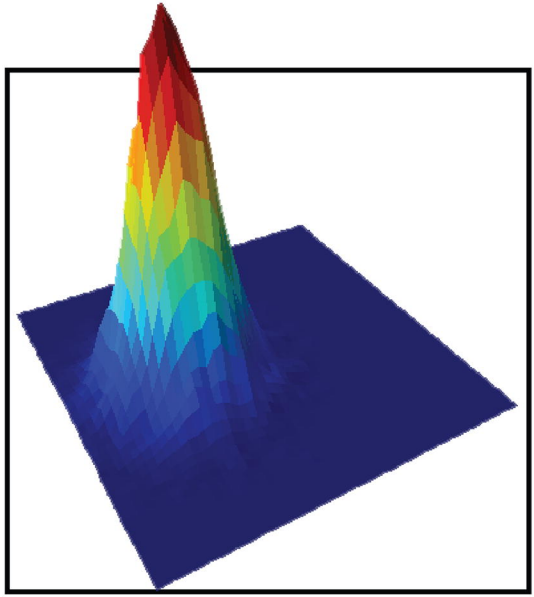

Control
PRC Lesion Group
Recognition Score
Trial Unique
Repeating
Condition

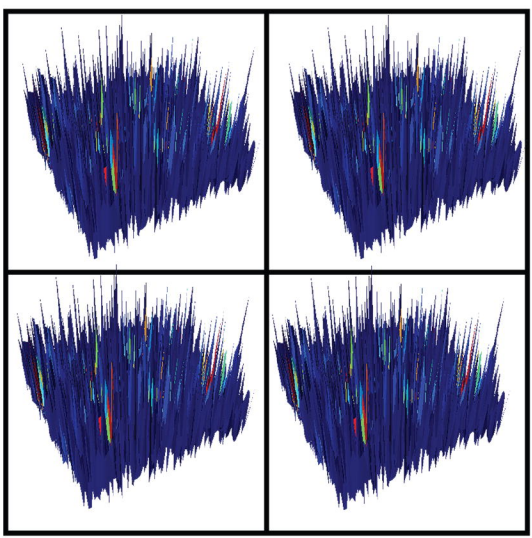

Posterior
Anterior
Representational Complexity
Posterior
Anterior

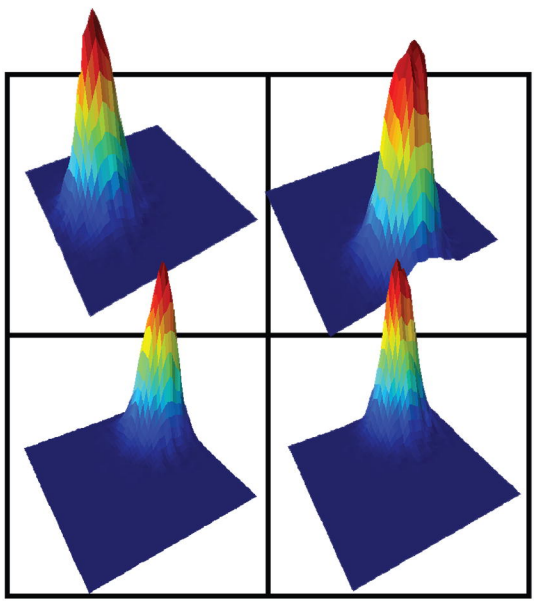

a)
High Ambiguity
Low Ambiguity
mis-
match:
match:
b)
Barense et al. Experiment 3
Control First Half
Control First Half
PRC Lesion Group
d'
0
1
2
3
4
5
High
Low
Stimulus Ambiguity

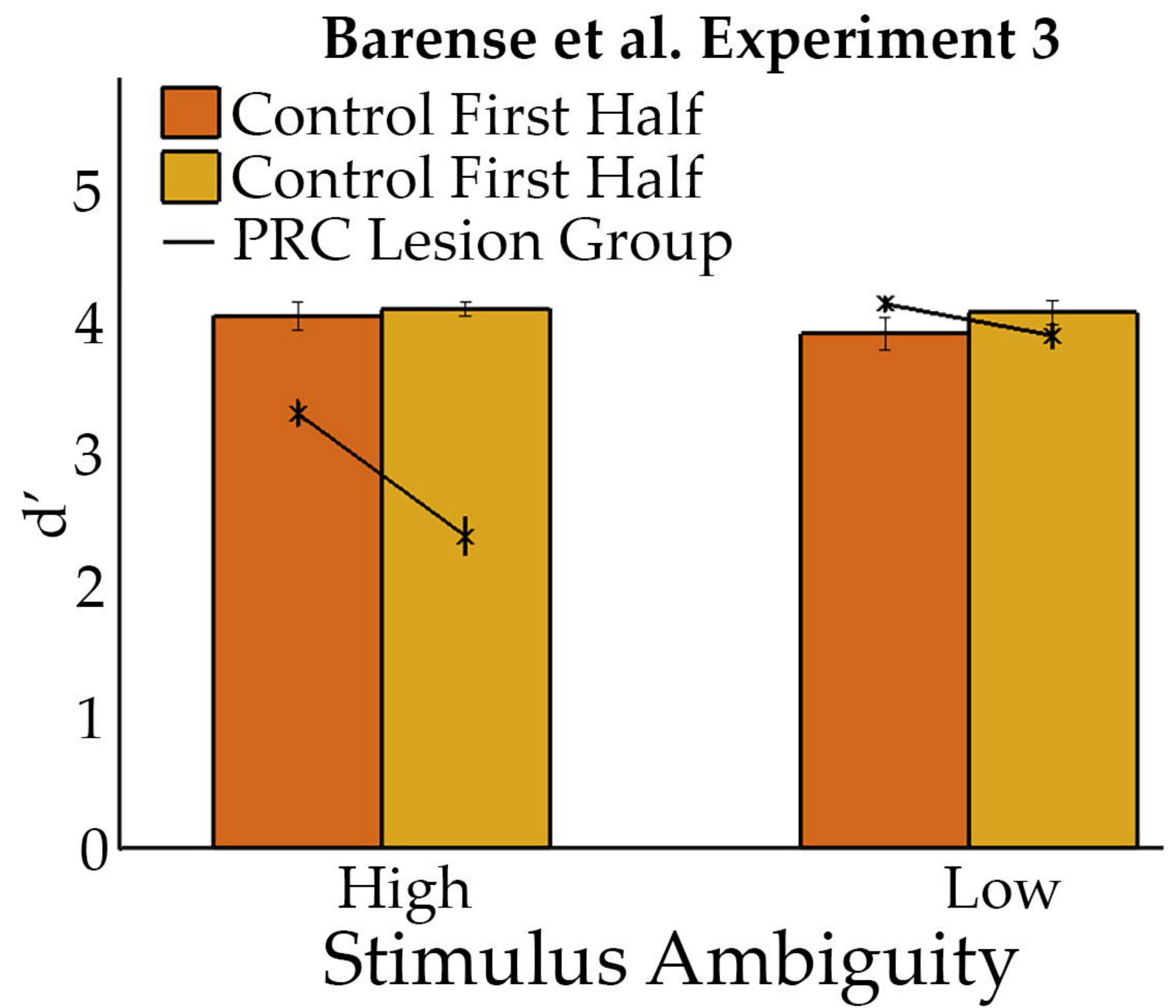

Posterior Layer
PRC Layer
Inputs

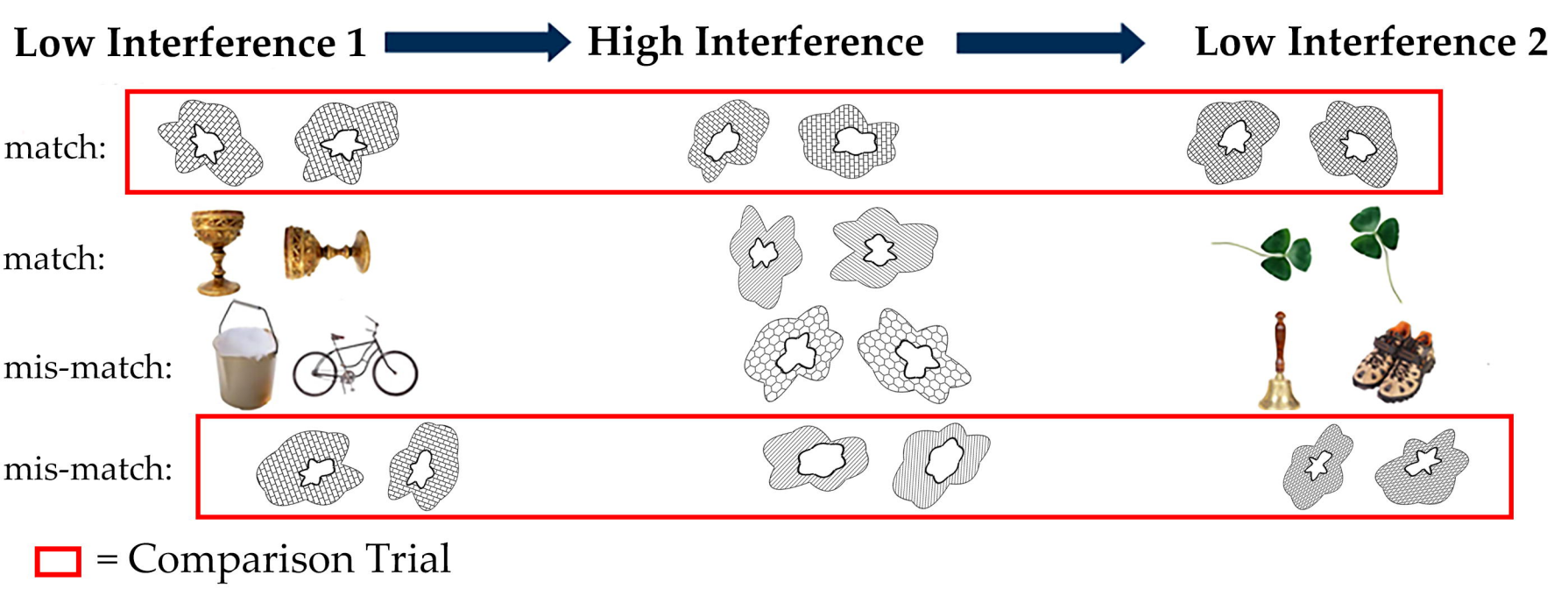









Barense et al. Experiment 3
Control First Half
Control First Half
PRC Lesion Group
5
4
3
2
1
0
d'
High
Low
Stimulus Ambiguity


Low Interference 1
High Interference
Low Interference 2
match:
match:
mis-match:
mis-match:
= Comparison Trial