

# When the ventral visual stream is not enough: A deep learning account of medial temporal lobe involvement in perception

Tyler Bonnen \*<sup>a</sup>, Daniel L.K. Yamins<sup>a,b,c</sup>, and Anthony D. Wagner<sup>a,c</sup>

<sup>a</sup>Department of Psychology, Stanford University

<sup>b</sup>Department of Computer Science, Stanford University

<sup>c</sup>Wu Tsai Neurosciences Institute, Stanford University

The medial temporal lobe (MTL) supports a constellation of memory-related behaviors. Its involvement in perceptual processing, however, has been subject to an enduring debate. This debate centers on perirhinal cortex (PRC), an MTL structure at the apex of the ventral visual stream (VVS). Here we leverage a deep learning approach that approximates visual behaviors supported by the VVS. We first apply this approach retroactively, modeling 29 published concurrent visual discrimination experiments: Excluding misclassified stimuli, there is a striking correspondence between VVS-modeled and PRC-lesioned behavior, while each are outperformed by PRC-intact participants. We corroborate these results using high-throughput psychophysics experiments: PRC-intact participants outperform a linear readout of electrophysiological recordings from the macaque VVS. Finally, *in silico* experiments suggest PRC enables out-of-distribution visual behaviors at rapid timescales. By situating these lesion, electrophysiological, and behavioral results within a shared computational framework, this work resolves decades of seemingly inconsistent experimental findings surrounding PRC involvement in perception.

## 1 Introduction

Animal behavior is informed by previous experience<sup>1</sup>. To understand how the mammalian brain supports this ability, neuroscientific data are often interpreted using two distinct cognitive constructs: ‘perception’ transforms ongoing sensory experience into behaviorally relevant abstractions (e.g. objects), while ‘memory’ enables retrieval of prior task-relevant experience. These informal, descriptive accounts of animal behavior have enabled researchers to characterize the role of the ventral visual stream (VVS) in visual perception<sup>2,3,4</sup>, as well as the role of the medial temporal lobe (MTL) in memory-related behaviors<sup>5,6,7</sup>. Nonetheless, identifying the neuroanatomical—and, by proxy, the computational—distinction between ‘perceptual’ and ‘mnemonic’ processing has been subject to an enduring debate<sup>8,9</sup>.

This debate centers on perirhinal cortex (PRC), an MTL structure situated at the apex of the primate VVS<sup>10,11</sup> (Fig. 1a). Lesion, electrophysiological, and imaging data have documented the role of PRC in memory-related behaviors<sup>11,12,13,14</sup>. This includes early observations that PRC-related memory impairments were modulated by item-level stimulus properties<sup>15,16,17,18</sup>, motivating perceptual experiments in PRC-lesioned primates<sup>18,19,20,21</sup>. A perceptual-mnemonic hypothesis emerged to account for these data, suggesting that PRC jointly supports perceptual and mnemonic behaviors<sup>22,23</sup>. Critically, PRC-related perceptual impairments were only evident in tasks that required sufficiently ‘complex’ representations (original schematic of PRC dependence in Fig. 1b). Methodological concerns were raised with this interpretation of these data<sup>24,25,26</sup>, however, suggesting that PRC-related deficits are a consequence of extra-perceptual task demands (e.g. memory). Additionally, there were concerns that concurrent damage to PRC-adjacent sensory cortices—not to PRC, *per se*—may explain perceptual deficits in lesioned subjects. Together, these concerns reinforced a purely mnemonic interpretation of PRC function.

To resolve these competing interpretations, experimentalists on both sides of the perceptual-mnemonic debate have converged on the use of concurrent visual discrimination (i.e. ‘oddity’) tasks. In each trial, participants freely view a stimulus screen containing multiple objects (Fig. 1d), then choose the item whose identity does not match the others (i.e. the ‘odd one out’). Diagnostic trials are designed to require putatively ‘complex’ perceptual representations while control trials are designed to require perceptual processing that only depends on canonical VVS structures. These studies intend to isolate perceptual and extra-perceptual task demands, as well as evaluate the integrity of PRC-adjacent sensory cortices. Nonetheless, concurrent visual discrimination tasks administered to PRC-lesioned and -intact participants have generated a seemingly inconsistent body of experimental evidence: results from these studies have been used both to support<sup>27,28,29,30,31,32</sup> and refute<sup>33,34,35,36</sup> the perceptual-mnemonic hypothesis (schematized in Fig. 1e left and right,

\*To whom correspondence should be addressed. email: bonnen@stanford.edu

49 respectively). While there is no discernible pattern of PRC-related deficits across these studies,  
50 interpreting these data has been forced to rely on informal, descriptive interpretations of these  
51 diverse stimulus sets.

52 We suggest that these apparent inconsistencies can be resolved by situating experimental behav-  
53 ior in relation to perceptual processing supported by the VVS. Experimental accuracy supported  
54 from a linear readout of the VVS (i.e. ‘VVS-supported performance’ Fig. 1f) offers a direct  
55 assessment of perceptual processing in the absence of extra-perceptual task demands. Stimulus  
56 ‘complexity,’ in this framework, is continuous and inversely related to VVS-supported performance  
57 (Fig. 1f: bottom). A perceptual-mnemonic hypothesis would predict that this approach organizes  
58 the available experimental observations into three distinct distributions. First, PRC-lesioned  
59 behavior is approximated by VVS-supported performance (Fig. 1f: purple). Second, PRC-intact  
60 participants outperform the VVS (Fig. 1f: grey). And third, experiments where VVS-supported  
61 performance is at ceiling. This third distribution may help identify ‘misclassified’ experiments that  
62 have been *described* as ‘complex’ yet are not relevant to the perceptual-mnemonic debate: Because  
63 VVS-supported performance is at ceiling, the perceptual-mnemonic hypothesis predicts no PRC-  
64 lesioned deficits. Any below-ceiling performance can only be due to extra-perceptual task demands  
65 (Fig. 1f: white). Thus, situating human behavior in relationship to VVS-supported performance  
66 may provide a unified account of PRC involvement in visual object perception.

67 Here we evaluate this unified account by situating lesion, electrophysiological, and behavioral  
68 results within a shared computational framework. As neural recordings from the VVS are not  
69 available from human participants in previous studies, we leverage a model class that is able to  
70 predict neural activity throughout the VVS, directly from experimental stimuli: task-optimized  
71 convolutional neural networks<sup>37,38,39</sup>. We use this model as a computational proxy for the VVS,  
72 developing an analytic approach that generates trial-by-trial predictions of VVS-supported perfor-  
73 mance on concurrent visual discrimination tasks (Fig. 1c). We first make use of this approach  
74 retroactively, collecting stimuli and behavioral data from published concurrent visual discrimina-  
75 tion studies administered to PRC-intact and -lesioned participants. In this ‘retrospective dataset’ of  
76 29 experiments, we deploy this modeling approach to estimate mean VVS-supported performance  
77 for each published stimulus set: after excluding misclassified stimulus sets on both sides of the  
78 perceptual-mnemonic debate, we observe a striking correspondence between a computational proxy  
79 for the VVS and PRC-lesioned performance, yet each are outperformed by PRC-intact participants.  
80 Next, we directly compare human behavior with neural responses at multiple levels of the VVS  
81 hierarchy (areas V4 and inferior temporal (IT) cortex) using a novel stimulus set. Results reveal  
82 that PRC-intact human participants outperform a linear readout of electrophysiological recordings  
83 collected from high-level visual cortex in the macaque, validating the computational results from  
84 the retrospective dataset. Finally, given the model’s correspondence with IT and PRC-lesioned  
85 behavior, we conduct experiments *in silico* to evaluate two prominent theories of PRC-dependent  
86 perceptual processing. Taken together, this computational framework enables us to compare the  
87 results from multiple experimental settings—lesion, electrophysiological, and *in silico*—providing a  
88 unified account of PRC involvement in perception.

## 89 **2 Results**

### 90 **2.1 Retrospective Analysis**

91 Through a comprehensive literature review we identify published, concurrent visual discrimina-  
92 tion studies administered to PRC-intact and -lesioned participants (Methods: Literature Review).  
93 Through correspondence with the original authors we acquired a ‘retrospective dataset’ composed  
94 of stimuli and behavioral data for 29 experiments that have collectively been used as evidence both  
95 for and against the perceptual-mnemonic hypothesis (Methods: Retrospective Dataset). Using one  
96 instance of a task-optimized convolutional neural network, we estimate the model’s cross-validated  
97 fit to previously collected electrophysiological responses<sup>40</sup>, identifying a model layer that best fits  
98 high-level visual cortex (Methods: Model Fit to Electrophysiological Data). We use an unweighted,  
99 linear decoder off of model responses from this layer to solve each trial in the retrospective dataset,  
100 then compute the average performance across trials for a given experiment (Methods: Model Per-  
101 formance on Retrospective Dataset). Thus, for each experiment in the retrospective dataset, we  
102 have a single value corresponding to the averaged performance that would be expected by a linear  
103 readout of high-level visual cortex which we refer to here as ‘model performance.’

#### 104 **2.1.1 Multiple stimulus sets have been misclassified on both side of the debate**

105 We identify 14 experiments in the retrospective dataset that appear to have been misclassified:  
106 Experimentalists have claimed these experiments are diagnostic of PRC involvement in perception,  
107 yet model performance is 100% accurate (as schematized in Fig. 1f: white), suggesting no need

for perceptual processing beyond the VVS. This includes eight experiments in which performance did not differ between PRC-intact and -lesioned participants (1 experiment in Buffalo et al., all 7 experiments in Knutson et al.; Supplemental Figure S2a-b), leading the authors to suggest these experiments provide evidence against perirhinal involvement in perception<sup>26,36</sup>. However, our computational results suggest no perceptual processing beyond the VVS is required for the experiments. In another six experiments performance differed between PRC-lesioned and -intact subjects (all 3 ‘Fribble’ experiments in Barense et al., all 3 ‘Face Morphs’ in Inhoff et al.; Supplemental Figure S2c-d), leading the authors to suggest these experiments provide evidence in support of PRC involvement in perception<sup>31,32</sup>. However, our modeling results suggest the observed divergence is better attributed to extra-perceptual task demands. After excluding all stimulus sets where model performance is at ceiling, including these misclassified experiments, there remain 14 experiments, which were used as evidence on both sides of the perceptual-mnemonic debate. This includes 10 experiments described by the original authors as ‘diagnostic’ and 4 experiments labeled as ‘controls.’

### 2.1.2 PRC-lesioned subjects are impaired on concurrent visual discrimination tasks

To make claims about PRC involvement in concurrent visual discrimination behaviors, we are principally interested in the comparison between PRC-lesioned behavior and their non-lesioned, age and IQ matched controls (i.e. ‘PRC-intact’). However, human PRC lesions are often accompanied by damage to other prominent structures within the MTL, such as the hippocampus (HPC). To ensure that behavioral impairments are a consequence of damage to PRC and not HPC we also compare the behavior of participants with selective hippocampal damage (i.e. ‘HPC-lesioned’) to their non-lesioned, age and IQ matched controls (i.e. ‘HPC-intact’)—where both HPC-lesioned and HPC-intact participants have an intact PRC. This is standard practice in the MTL literature. Across the 14 experiments in the retrospective dataset, PRC-lesioned participants are significantly impaired relative to PRC-intact participants (paired ttest,  $\beta = .14$ ,  $t(13) = 2.68$ ,  $P = .019$ ), while HPC-lesioned participants show no such impairment (paired ttest,  $\beta = .01$ ,  $t(13) = .73$ ,  $P = .479$ ). Directly comparing the difference between PRC-intact/lesioned participants with HPC-intact/lesioned participants, there is a significant difference between lesioned groups (PRC-intact – PRC-lesion vs. HPC-intact – HPC-lesion:  $\beta = .13$ ,  $F(1, 26) = 2.34$ ,  $P = .028$ ). PRC-intact participants perform significantly better than PRC-lesioned participants, while there is no such difference between HPC-intact and -lesioned participants.

### 2.1.3 A computational model of the VVS approximates PRC-lesioned performance

The previous section demonstrates a coarse distinction between PRC-lesioned and -intact performance. A stronger test of the perceptual-mnemonic hypothesis would be to predict the relative impairments observed across different experiments, using our computational proxy for the VVS. To this end, we directly compare model performance with human performance across eligible experiments in the retrospective dataset (Methods: Model Performance on Retrospective Dataset). We observe a striking correspondence between PRC-lesioned behavior and model performance (Fig. 2a, purple;  $\beta = .85$ ,  $F(1, 12) = 5.59$ ,  $P = 1 \times 10^{-4}$ ). Conversely, PRC-intact participants are not predicted by a computational proxy for the VVS (Fig. 2a, grey:  $\beta = .85$ ,  $F(1, 12) = 2.05$ ,  $P = .063$ ); these participants significantly outperform the model ( $\beta = .35$ ,  $t(13) = 7.32$ ,  $P = 6 \times 10^{-6}$ ). Critically, there is a significant interaction between PRC-intact and PRC-lesion groups when predicting human accuracy from model performance ( $\beta = .63$ ,  $F(3, 24) = 2.82$ ,  $P = .010$ ), which is not observed for the hippocampal groups (HPC-lesion/HPC-intact  $F(3, 24) = .28$ ,  $P = .781$ ). To make the correspondence between model performance and PRC-lesioned behavior more explicit, for each experiment we take the difference between PRC-intact and -lesioned participants, resulting in a difference score for each experiment. This difference is predicted by model performance ( $\beta = -.63$ ,  $F(1, 12) = -3.17$ ,  $P = .008$ ) with the sign indicating that as model performance is degraded, the difference between PRC-intact and -lesioned participants increases. These results suggest that as IT-supported performance on a given experiment decreases, the divergence between PRC-lesioned and -intact performance increases. The low sample size in this analysis encourages caution when interpreting these results<sup>41</sup>. Nonetheless, these results offer a stimulus-computable account of why the magnitude of PRC-related deficits might vary across published studies, clarifying PRC contributions to concurrent visual discrimination behaviors.

### 2.1.4 Available experiments do not enable focal claims about VVS dependence

As the final and most stringent test of the perceptual-mnemonic hypothesis, we determine whether high-level visual cortex uniquely explains PRC-lesioned performance. This requires not only that PRC-lesioned behavior reflects a linear readout of high-level visual cortex, but that high-level visual cortex predicts PRC-lesioned behavior significantly better than earlier stages of processing within the VVS. To address this uncertainty, we leverage the differential correspondence between

model layers with early and late stages of processing within the VVS (Methods: VVS Reliance). Borrowing from electrophysiological data previously collected<sup>40</sup>, we first generate a metric for each layer's differential fit to IT cortex ( $\Delta_{IT-V4}$ , Fig. 3a). Next, we estimate model performance on the retrospective dataset from all model layers, not only for the layer that best fits IT cortex (Fig. 3b: PRC-lesion/intact top, HPC-lesion/intact bottom). With these model performance-by-layer estimates we generate a metric for the differential fit to PRC-lesioned ( $\Delta_{prc}$ ) and HPC-lesioned ( $\Delta_{hpc}$ ) performance. We then relate these human behavioral metrics from the retrospective dataset to the electrophysiological metrics from the non-human primate. Across layers, differential correspondence with IT cortex predicts differential fit to PRC-lesion behavior (Fig. 3c, purple, top:  $\beta = .95$ ,  $F(1, 17) = 13.20$ ,  $P = 2 \times 10^{-10}$ ). In addition to these aggregate (i.e. across all layers) analyses, we determine whether there is a significant interaction between lesioned groups at each layer (i.e. PRC-lesioned vs. PRC-intact, repeating previous analyses in Fig. 2a). After correcting for multiple comparisons, there is only a significant interaction in more 'IT like' layers (e.g. fc7:  $P = .00257$ ). Conversely, there are no layers with a significant interaction between HPC-lesioned and -intact participants, even at a liberal (uncorrected) threshold (all  $p > .05$ , e.g. fc7:  $P = .773$ ). Nonetheless, model performance from an IT-like layer is not significantly better at predicting PRC-lesioned behavior than a V4-like layer (conv5\_1 and pool3, respectively:  $\beta = .05$ ,  $F(2, 25) = .86$ ,  $P = .400$ ), which is evident in the similarity across layers in 6b. Taken together, these results suggest that while PRC-lesioned behavior is best fit by later stages of processing within the VVS, the available stimuli do not clearly separate V4 from IT supported behaviors.

### 2.1.5 Retrospective summary & limitations

Modeling and behavioral results from the retrospective dataset suggest that PRC-lesioned performance reflects a linear readout of the VVS. In contrast, PRC-intact behaviors outperform both PRC-lesioned participants and a computational proxy for the VVS; this includes both PRC-intact participants (i.e. no lesion to the MTL/PRC), and participants with selective damage to the hippocampus that spared PRC. These results suggest that above VVS performance in concurrent visual discrimination tasks is dependent on PRC. While this analysis resolves fundamental questions at the center of the perceptual-mnemonic debate, there are multiple limitations to consider. First, extant stimulus sets do not differentiate V4- from IT-Supported behavior, leaving open what neuroanatomical structures within the VVS PRC-lesioned behavior is reliant on. Second, the available stimulus sets only offer a sparse sampling of the range of VVS-supported behaviors. This is due, in part, to reliance on experimental averages when fitting to human behavior, low experimental Ns (both experiments and participants), and stimulus sets that were designed to result in categorical PRC-related impairments. Finally, there is a considerable amount of hypothesis-orthogonal variability across these studies. For example, the number of stimuli used on each trial varies from 3-9 objects across experiments in the retrospective dataset. Instead of developing a deeper understanding of how these off-hypothesis factors relate to the results presented here, we develop a novel, model-based experimental approach.

## 2.2 Novel Dataset

To address limitations in the retrospective analysis, we design a novel experiment that enables item-level performance estimates, continuously samples the space of stimulus 'complexity,' and clearly disentangles multiple stages of processing across the VVS from PRC-intact behavior. Additionally, these experiments minimize off-hypothesis experimental variance, using the minimum configuration of objects in each trial ( $N = 3$ ) across all levels of stimulus 'complexity.' We leverage our computational approach to generate this stimulus set, then evaluate it using computational, electrophysiological, and behavioral methods.

### 2.2.1 High-throughput human psychophysics experiments

We begin with stimuli that have been previously shown to separate V4- from IT-supported behavior<sup>40</sup>, reconfiguring these images into 3-way, within-category, oddity trials (Methods: Novel Stimulus Set Generation; for examples see Fig. 4a). We develop a novel estimate of 'model performance' on these oddity tasks: a weighted, linear readout from an 'IT-like' model layer, learned via a leave-one-out cross-validated protocol (Methods: Model Performance on Novel Stimuli). We administer these stimuli to PRC-intact human participants ( $N = 297$ ) via high-throughput psychophysics experiments (Methods: High-throughput Psychophysics Experiments). Finally, using the approach developed to estimate a weighted model performance, we determine the performance on these oddity trials that would be supported from a weighted readout of macaque IT and V4. Thus, for the same stimuli, we are able to compare model performance and PRC-intact human behavior, alongside the accuracy supported by a weighted, linear readout of electrophysiological responses collected from macaque IT and V4 (Fig. 4b).

226 **2.2.2 PRC-intact participants outperform electrophysiological recordings from IT** 226  
 227 PRC-intact human behavior outperforms a linear readout of IT on this novel stimulus set (Fig 5c: 227  
 228  $\beta = .24$ ,  $t(31) = 9.50$ ,  $P = 1 \times 10^{-10}$ ) while IT significantly outperforms V4 (Fig. 5a:  $\beta = .18$ , 228  
 229  $t(31) = 6.56$ ,  $P = 2 \times 10^{-7}$ ). This three-way dissociation enables us to disentangle early and late 229  
 230 stage processing within the VVS from PRC-supported behaviors. A computational proxy for IT 230  
 231 demonstrates the same pattern, predicting IT-Supported Performance (Fig 5d, purple:  $\beta = .81$ , 231  
 232  $F(1, 30) = 13.33$ ,  $P = 4 \times 10^{-14}$ ), outperforming V4 (Fig 5d, grey:  $\beta = .26$ ,  $t(31) = 8.02$ ,  $P =$  232  
 233  $5 \times 10^{-9}$ ), and being outperformed by PRC-intact participants (Fig 5d, teal:  $\beta = .16$ ,  $t(31) = 5.38$ , 233  
 234  $P = 7 \times 10^{-6}$ ). Finally, we find that the PRC-intact human reaction time for each item is a reliable 234  
 235 predictor of IT-supported performance, such that greater RTs are observed for items with lower 235  
 236 IT-supported accuracy ( $\beta = -.88$ ,  $t(31) = -10.00$ ,  $P = 4 \times 10^{-11}$ ). Framed more explicitly, for each 236  
 237 item, the difference between IT-supported and PRC-intact performance is predicted by reaction 237  
 238 time (Fig. 5e, purple:  $\beta = .81$ ,  $F(1, 31) = 7.44$ ,  $P = 3 \times 10^{-8}$ ). This relationship is also observed 238  
 239 for model performance ( $\beta = .72$ ,  $F(1, 31) = 5.62$ ,  $P = 4 \times 10^{-6}$ ) but not V4-supported performance 239  
 240 (Fig. 5e, grey:  $\beta = -.08$ ,  $F(1, 31) = -0.41$ ,  $P = .682$ ). These results demonstrate that PRC-intact 240  
 241 human participants require more time to choose among items that are not linearly separable in IT, 241  
 242 in a way that scales inversely with IT-supported performance. 242

243 **2.3 In Silico Experiments** 243  
 244 Here we address two prominent theories around why the VVS—and, by proxy, PRC-lesioned 244  
 245 subjects—fail to perform ‘complex’ discriminations. The first hypothesis posits that PRC provides 245  
 246 another layer of processing within the VVS<sup>22</sup>: Just as IT supports discrimination behaviors 246  
 247 not linearly separable in V4, PRC supports discrimination behaviors not linearly separable in IT. 247  
 248 In this case, PRC is thought to integrate information from neural populations in IT in order to 248  
 249 generate a ‘complex’ or ‘configural’ representation—using computations that are common across the 249  
 250 VVS. More concretely, this implies that adding VVS-like layers “on top” of an IT-like model should 250  
 251 improve performance on concurrent visual discrimination experiments with ‘complex’ stimuli. The 251  
 252 second hypothesis suggests that PRC dependence is not due to stimulus properties, per se (that is, 252  
 253 properties of the stimulus that can be computed directly from the image itself—i.e. pixels) but the 253  
 254 interaction between perceptual properties and task-relevant perceptual experience<sup>42</sup>. This implies 254  
 255 that canonical VVS structures are fully capable of performing ‘complex’ perceptual discriminations, 255  
 256 but that this requires extensive, content-specific training. This suggests that subjecting a VVS-like 256  
 257 model to perceptual training over a putatively ‘complex’ stimulus type should enable these models 257  
 258 to approximate PRC-intact performance on this stimulus class. 258

259 **2.3.1 Changing model architecture does not enable PRC-intact performance** 259  
 260 We first identify experiments in the retrospective dataset for which model performance increases 260  
 261 with the ‘depth’ that model responses are extracted from (Methods: Model Depth & Architecture 261  
 262 Analyses). We observe depth-dependent performance enhancements for some experiments (e.g. 262  
 263 ‘Low Snow’ stimuli in Stark et al. 2000,  $\beta = 0.78$ ,  $F(1, 19) = 2.62$ ,  $P = .017$ ) but not others 263  
 264 (‘Family High Ambiguity’ stimuli in Barense et al. 2007,  $\beta = -0.00$ ,  $F(1, 19) = -0.05$ ,  $P = .959$ ). 264  
 265 PRC-lesioned participants performed significantly better ( $t(6) = 5.17$ ,  $P = 4 \times 10^{-4}$ ) on experiments 265  
 266 that exhibited depth improvements ( $n = 7$ ,  $\mu = .88$ ) than those that did not ( $n = 7$ ,  $\mu = .52$ ); 266  
 267 experiments that did not exhibit depth-dependent improvements are those with the most substantial 267  
 268 PRC-related deficits. Can changing the model architecture—in this case, adding layers ‘on top’ 268  
 269 of IT-like layers—increase performance on these experiments? To test this, we repeat previous 269  
 270 analyses (Methods: Model Performance on Retrospective Dataset) but estimate model performance 270  
 271 from numerous models, each of which has an increasingly deep architecture (from 18 to 152 layers, 271  
 272 Methods: Model Depth & Architecture Analyses). These architectural modifications do not lead to 272  
 273 increased correspondence with PRC-intact behavior ( $\beta = -.55$ ,  $t(4) = -1.33$ ,  $P = .255$ ). Moreover, 273  
 274 just as in the original model, for each of these architectures we observe an interaction between PRC- 274  
 275 lesioned and -intact participants (Fig. 6, e.g. 152 layers:  $\beta = -.54$ ,  $F(3, 24) = -3.97$ ,  $P = 6 \times 10^{-4}$ ). 275  
 276 Increasing the number of VVS-like computations over a given stimulus does not better predict PRC- 276  
 277 supported behaviors. Finally, we repeat this analysis for numerous convolutional architectures (e.g. 277  
 278 inception-v3, squeezeNet, alexnet, densenets, etc.), taking responses from a penultimate model 278  
 279 layer to estimate model performance on the retrospective dataset. The interaction between PRC- 279  
 280 lesioned and -intact participants is consistent across all competitive architectures evaluated; no 280  
 281 model better approximates PRC-intact performance, suggesting that our findings are robust across 281  
 282 instances within the convolutional neural network model class. 282

### 283 2.3.2 Content-specific training enables VVS models to achieve PRC-intact accuracy 283

284 Faces are an example of a putatively ‘complex’ stimulus category. In the retrospective analysis, faces 284  
285 are an object class in which PRC-intact participants outperform both PRC-lesioned participants 285  
286 ( $\beta = .20$ ,  $t(3) = 4.25$ ,  $P = .024$ ) and model performance ( $\beta = .46$ ,  $t(3) = 8.47$ ,  $P = .003$ ). 286  
287 Similarly, for face stimuli in the novel high-throughput experiment, PRC-intact participants reliably 287  
288 outperform both IT-supported performance ( $\beta = .41$ ,  $t(41) = 15.36$ ,  $P = 1 \times 10^{-18}$ ) and model 288  
289 performance ( $\beta = .35$ ,  $t(41) = 14.04$ ,  $P = 3 \times 10^{-17}$ ). Thus, faces are an example of putatively 289  
290 ‘complex’ experimental stimuli where computational models, PRC-lesioned participants, and IT- 290  
291 supported performance fail to approximate PRC-intact behavior. We optimize a computational 291  
292 proxy for the VVS to perform these putatively ‘complex’ stimuli by changing the distribution of 292  
293 its training data (Methods: Content-Specific Optimization Procedure): in short, we train this 293  
294 model to perform face discrimination, instead of object classification. On the retrospective dataset, 294  
295 this content-specific optimization procedure leads to an increase in model performance on these 295  
296 putatively ‘complex’ stimuli (Fig. 7a, left red; paired  $t(3) = 4.75$ ,  $P = .018$ ). Moreover, this 296  
297 optimization procedure results in performance on face experiments that is not significantly different 297  
298 from PRC-intact participants (Fig. 7a, bottom right, red;  $\beta = .16$ ,  $t(3) = 1.91$ ,  $P = .153$ ). 298  
299 However performance is degraded on all other (i.e. non-face) stimuli ( $\beta = -.14$ ,  $t(9) = -2.67$ , 299  
300  $P = .026$ ). This reveals a significant interaction between training data and testing performance, 300  
301 as a function of stimulus type (Fig. 7a, left greys;  $\beta = .44$ ,  $F(3, 24) = 2.53$ ,  $P = .018$ ). We can 301  
302 state the results more generally: content-specific optimization leads to increased model performance 302  
303 on ‘within distribution’ stimuli, while not demonstrating these same levels of performance ‘out of 303  
304 distribution.’ Given the low sample size, these results should be interpreted with caution. To 304  
305 address this shortcoming, we conduct the same analysis as above using the novel experimental 305  
306 dataset (Methods: Content-Specific Optimization Procedure). When comparing between models, 306  
307 performance is significantly better on within distribution stimuli at the single-trial level for both 307  
308 the face- (Fig. 7b, red:  $\beta = 0.33$ ,  $t(46) = 11.24$ ,  $P = 9 \times 10^{-15}$ ) and object-trained models (Fig. 308  
309 7b, greys:  $\beta = 0.15$ ,  $t(167) = 8.04$ ,  $P = 2 \times 10^{-13}$ ). Critically, content-specific optimization leads to 309  
310 model performance on these putatively ‘complex’ stimuli that is now statistically indistinguishable 310  
311 from item-level performance of PRC-intact human participants (Fig. 7b, bottom right, red;  $\beta = .03$ , 311  
312  $t(46) = .91$ ,  $P = .369$ ). Nonetheless, model performance is degraded on ‘out of distribution’ stimuli, 312  
313 with a significant interaction between training data and testing performance, as a function of 313  
314 stimulus type (Fig. 7b, left greys:  $\beta = -.48$ ,  $F(3, 426) = -9.51$ ,  $P = 6 \times 10^{-20}$ ). Interestingly, unlike 314  
315 models optimized for object classification, which predict IT-supported performance (Fig 7c, top left: 315  
316  $\beta = .86$ ,  $F(1, 30) = 9.13$ ,  $P = 4 \times 10^{-10}$ ) as well as reaction times inherent in supra-IT performance 316  
317 (Fig 7c, top right;  $\beta = -.80$ ,  $F(1, 30) = -7.37$ ,  $P = 3 \times 10^{-8}$ ), there is no correspondence between 317  
318 face-optimized model performance and IT-supported performance (Fig 7c, bottom left;  $\beta = -.08$ , 318  
319  $F(1, 30) = -.42$ ,  $P = .679$ ) or reaction time (Fig 7c, bottom right;  $\beta = -.16$ ,  $F(1, 30) = -.88$ , 319  
320  $P = .386$ ). These results demonstrate that a content-specific optimization procedure enables VVS- 320  
321 like architectures to perform perceptual discriminations on putatively ‘complex’ stimuli. However, 321  
322 VVS-like architectures achieve this level of performance in a manner that is a biologically and 322  
323 behaviorally implausible approximation of PRC-intact performance. 323

## 324 3 Discussion 324

325 We have provided a unified account of PRC involvement in concurrent visual discrimination tasks. 325  
326 We began this work by developing a computational proxy for VVS-supported performance on 326  
327 concurrent visual discrimination tasks; this approach enables us to formalize perceptual demands 327  
328 in these experiments, directly from their stimuli. We first deploy this approach on a ‘retrospective 328  
329 dataset’ composed of 29 published, concurrent visual discrimination experiments administered to 329  
330 PRC-lesioned and -intact participants. We find a number of experiments that appear to have been 330  
331 misclassified: while they have been described as ‘complex,’ the model performs them at ceiling, 331  
332 suggesting that there is no need for perceptual processing beyond the VVS. Across the remaining 332  
333 experiments we observe a striking correspondence between this computational proxy for the VVS 333  
334 and PRC-lesioned human behavior. Critically, PRC-intact behavior outperforms this VVS-like 334  
335 model and PRC-lesioned participants; this is true for PRC-intact participants with an entirely 335  
336 intact MTL and those with selective damage to the hippocampus that spared PRC. Accordingly, 336  
337 PRC-lesioned behavior only diverges from PRC-intact performance to the degree that the model 337  
338 fails to perform these tasks. Together, these results suggest that PRC-lesioned human behavior 338  
339 reflects a linear readout of the VVS, PRC-intact human behaviors on these tasks outperform the 339  
340 VVS, and this behavior is dependent on PRC. 340

341 To address limitations inherent in the retrospective analysis, we next generate a novel con- 341  
342 current visual discrimination stimulus set. We evaluate these stimuli using data collected via 342  
343 high-throughput psychophysics experiments administered to PRC-intact human participants, elec- 343

344 trophysiological data previously collected from the non-human primate, as well as our own compu- 344  
345 putational approaches. We find that PRC-intact behavior diverges from IT-Supported Performance 345  
346 in this novel stimulus set, validating the main finding from the retrospective analysis. Additionally, 346  
347 there is a clear separation between multiple structures throughout the VVS: not only do PRC-intact 347  
348 participants outperform a weighted readout of electrophysiological responses from IT, but IT out- 348  
349 performs V4. Moreover, model performance provides a close approximation for a linear readout 349  
350 of IT on these concurrent visual discrimination tasks—a well validated example of how computa- 350  
351 tional proxies for the VVS can be integrated in future experimental work that aims to formalize 351  
352 the involvement of MTL structures in perceptual processes. Interestingly, in this well-controlled 352  
353 experimental setting, reaction time is a reliable predictor of the divergence between PRC-intact 353  
354 and IT-supported behavior: supra-IT performance in the human scales, parametrically, with time. 354

355 Using *in silico* experiments we address two prominent theories surrounding why the VVS (and, 355  
356 by proxy, PRC-lesioned behavior) fails to support discrimination of increasingly ‘complex’ visual 356  
357 stimuli. The first hypothesis posits that PRC provides *another* layer of processing within the VVS. 357  
358 However, we observe that increasing model depth (i.e. increasing the number of VVS-like layers) 358  
359 does not enable better correspondence with PRC-intact behaviors. To the contrary, all instances 359  
360 of this model class exhibited the same pattern of differential fit to PRC-lesioned behaviors. A 360  
361 second hypothesis suggests that PRC dependence emerges through the interaction between stimulus 361  
362 properties and task-relevant experience. To address this claim, we subject VVS-like models to 362  
363 ‘perceptual training’ (i.e. content-specific optimization) over a putatively ‘complex’ stimulus type: 363  
364 faces. This optimization procedure leads to PRC-intact performance levels on ‘within distribution’ 364  
365 stimuli, while model performance degrades for out-of-distribution (i.e. not face) stimuli. These 365  
366 computational results suggests that PRC-dependence on ‘complex’ stimuli is not about stimulus 366  
367 properties, *per se* (i.e. VVS-like architectures can perform these tasks with training), but the 367  
368 interaction between stimulus properties and stimulus-relevant experience. 368

369 Given these behavioral, neural, and computational results, how might we characterize PRC 369  
370 involvement in concurrent visual discrimination tasks? We must first acknowledge that the VVS 370  
371 provides a basis space for visual perception, generating linearly separable representations that sup- 371  
372 port many downstream behaviors<sup>43</sup>. However, not all visual inputs are linearly separable in this 372  
373 space—they remain ‘entangled,’ even in high-level visual regions. Achieving accuracy above what 373  
374 is linearly separable within the VVS requires time. Extensive training can slowly disentangle these 374  
375 representations within the VVS itself; our *in silico* experiments corroborate a rich literature on 375  
376 perceptual learning<sup>44,45</sup> and make explicit the temporal dynamics/advantages in consolidating per- 376  
377 ceptual information within the VVS<sup>46</sup>. However, PRC can disentangle task-relevant information 377  
378 from VVS responses within a single trial, enabling out-of-distribution visual behaviors at rapid 378  
379 timescales. Interestingly, the degree to which stimuli are not linearly separable within the VVS 379  
380 scales with the amount of time required for supra-VVS performance. We do not interpret these 380  
381 PRC-dependent temporal dynamics as either ‘perceptual’ or ‘mnemonic;’ neither of these terms 381  
382 elucidates the computations that enable this behavior. Instead, what we offer is a tractable, exten- 382  
383 sible framework to formalize how experimental variables relate to PRC-dependent behaviors. We 383  
384 believe this biologically plausible computational approach will continue to offer novel insights into 384  
385 how the MTL supports such enchanting—indeed, at times, indescribable—behaviors. 385

## 386 4 Methods 386

### 387 4.1 Literature Review 387

388 Criteria for inclusion in the retrospective analysis was threefold. First, behavioral data from PRC- 388  
389 lesioned and PRC-intact participants must have been collected. Second, the experiment must have 389  
390 been administered to either human or non-human primate participants. Third, participants must 390  
391 have performed concurrent visual discrimination tasks. The initial Google Scholar search terms used 391  
392 were “perirhinal lesion oddity” resulting in 425 results. The terms “human” or “primate” were not 392  
393 included in this search as experimental participants in human primate research are often referred 393  
394 to simply “subjects.” Instead of “concurrent visual discrimination task” we used “oddity” as it is 394  
395 a more commonly used shorthand in the literature, and the extended task description is applied 395  
396 irregularly. After candidate experiments were identified from these 425 results, the references cited 396  
397 in each of these candidate papers were used as a source of candidate papers missed in the initial 397  
398 search. An additional exclusion criterion was incorporated, as one concurrent visual discrimination 398  
399 experiment (Lee & Rudebeck 2010) required that participants reference real-world shape properties 399  
400 of objects not presented on the stimulus screen alongside the stimuli. This experiment was not 400  
401 included in further analysis. The corresponding authors in each experiment were contacted via 401  
402 email and asked to provide experimental materials necessary to the current computational approach. 402  
403 This included, first, behavioral data from PRC-lesioned and -intact participants with the finest 403  
404 granularity that could be collected (e.g. trial, subject, or group level data). When available, this also 404

405 included behavioral data from hippocampal-lesioned and hippocampal-intact participants. Second, 406 the stimuli corresponding to these behavioral data; ideally, the exact stimuli presented in each 407 experiment conducted. For all studies, the corresponding authors (or their associates) responded 408 promptly and were eager to provide the data requested. The complete list of experiments identified 409 through this search is presented below.

## 410 Studies Requested

- 411 – Buffalo, E. A., Reber, P. J., & Squire, L. R. (1998). The human perirhinal cortex and 412 recognition memory. *Hippocampus*, 8(4), 330-339.
- 413 – Stark, C. E., & Squire, L. R. (2000). Intact visual perceptual discrimination in humans in 414 the absence of perirhinal cortex. *Learning & Memory*, 7(5), 273-278.
- 415 – Buckley, M. J., Booth, M. C., Rolls, E. T., & Gaffan, D. (2001). Selective perceptual impairments 416 after perirhinal cortex ablation. *Journal of Neuroscience*, 21(24), 9824-9836.
- 417 – Levy, D. A., Shrager, Y., & Squire, L. R. (2005). Intact visual discrimination of complex and 418 feature-ambiguous stimuli in the absence of perirhinal cortex. *Learning & memory*, 12(1), 419 61-66.
- 420 – Lee, A. C., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., ... & Graham, 421 K. S. (2005). Specialization in the medial temporal lobe for processing of objects and scenes. 422 *Hippocampus*, 15(6), 782-797.
- 423 – Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., ... & 424 Graham, K. S. (2005). Perceptual deficits in amnesia: challenging the medial temporal lobe 425 'mnemonic' view. *Neuropsychologia*, 43(1), 1-11.
- 426 – Lee, A. C., Buckley, M. J., Gaffan, D., Emery, T., Hodges, J. R., & Graham, K. S. (2006). 427 Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond long- 428 term declarative memory: a double dissociation in dementia. *Journal of Neuroscience*, 26(19), 429 5198-5203.
- 430 – Shrager, Y., Gold, J. J., Hopkins, R. O., & Squire, L. R. (2006). Intact visual perception in 431 memory-impaired patients with medial temporal lobe lesions. *Journal of Neuroscience*, 26(8), 432 2235-2240.
- 433 – Barene, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe 434 processes online representations of complex objects. *Neuropsychologia*, 45(13), 2963-2974.
- 435 – Knutson, A. R., Hopkins, R. O., & Squire, L. R. (2012). Visual discrimination performance, 436 memory, and medial temporal lobe function. *Proceedings of the National Academy of Sciences*, 437 109(32), 13106-13111.
- 438 – Inhoff, M. C., Heusser, A. C., Tambini, A., Martin, C. B., O'Neil, E. B., Köhler, S., ... 439 & Davachi, L. (2019). Understanding perirhinal contributions to perception and memory: 440 Evidence through the lens of selective perirhinal damage. *Neuropsychologia*, 124, 9-18.

## 441 4.2 Retrospective Dataset

442 Across all of the obtained experiments, we were able to reliably secure experiment-level behavioral 443 data (i.e. averaged across trials) for each group within a given study (e.g. the performance of PRC- 444 lesioned participants performing condition A, B, etc., within a given study). In order to compare 445 model and human behaviors, we compare behavior at the level of the experiment (i.e. averaged 446 across trials). For most of the obtained experiments, the exact trial-level stimuli presented to 447 participants were used in the modeling approach. However, there were two experiments (Stark et 448 al. 2000 and Lee et al. 2006) where the distribution of all stimuli was obtained, but the specific 449 trials shown to each subject had to be approximated. For Stark et al. 2000, the authors randomly 450 selected stimuli to be used in each trial, from a set of all possible stimuli. They could not recover the 451 exact trial-by-trial stimuli shown to experimental participants. Instead, the corresponding authors 452 provided all stimuli used across faces and "snow" (partially occluded object) experiments, as well 453 as the pseudo-random protocol used to generate each experiment: For each "typical" item, five 454 different viewpoints were drawn from all available stimuli of this item. Faces had a total of six 455 items, each corresponded to different (but common across faces) viewpoints. For each object, there 456 were a total of five viewpoints, such that all viewpoints of this item were used in each trial. In 457 'snow' conditions, for each trial, the typical object was selected at random, and all of its exemplars 458 are used; the oddity object is selected at random, and one of its exemplars is selected at random to 459 be that trial oddity. For faces, after selecting a typical face, and a subset of 5 of its exemplars, the 460 oddity identity was sampled randomly, with a viewpoint distinct from that present in the typical 461 faces. Consequently, each face trial included an oddity that was always from a different viewpoint 462 from all typical faces. For Lee et al. 2006, the corresponding authors were able to provide all

463 stimuli. However, as with Stark et al. 2000, in experiment two only a subset of the stimuli were  
464 presented to participants. Across participants, the number of trials in this subset was constant  
465 (31/40), but the exact items presented to each subject was drawn randomly from all available  
466 stimuli. For both the Stark et al. 2000 and Lee et al. 2006 we approximate the stimuli presented  
467 to participants by generating a population of experiments (N=100) that adhered to the protocols  
468 outlined above. We then compare the model performance across this population of experiments (i.e.  
469 averaged performance across all N iterations generated by this sampling protocol) to the obtained  
470 human behavior for each experiment.

### 471 4.3 Model Fit to Electrophysiological Data

472 We use one instance of a task-optimized convolutional neural network (VGG16<sup>47</sup>), implemented  
473 in tensorflow and pre-trained to perform object classification on a large-scale object classification  
474 dataset<sup>48</sup>. To identify a model layer that best fits IT cortex, we utilize previously collected<sup>40</sup>  
475 electrophysiological responses from macaque V4 and IT cortex, along with the stimuli that elicited  
476 these responses. Using ‘medium’ and ‘high’ variation images from this data set, we convert each  
477 image from greyscale to RGB then resize it to accommodate model input dimensions (224x224x3).  
478 We pass each image to the model and extract responses from all layers (e.g. convolutional, pooling,  
479 and fully connected layers), vectorize each layer’s output. We randomly segment these model  
480 responses to each image into training and testing data using a 3/4th split. Thus, we use multi-  
481 electrode responses from macaque V4 and IT to a set of image, and model responses to those same  
482 images. For each layer, we learn a linear mapping between vectorized model responses and a single  
483 electrode’s responses to the training images, using sklearn’s implementation of PLS regression (with  
484 five components). We evaluate this mapping between model and neural responses by computing the  
485 Pearson’s correlation between model-predicted responses and observed responses for each electrode  
486 across all test images. For each layer, this results in a single correlation value for each electrode,  
487 which we repeat over all electrodes. This results in a distribution corresponding to that layer’s  
488 cross-validated fits to population-level neural responses, both for electrodes in IT and V4. We  
489 compute the split half reliability for V4 ( $r = .63 \pm .22\text{STD}$ ) and IT ( $r = .73 \pm .24\text{STD}$ ) across  
490 neurons in each region. We then divide the distribution of cross-validated fits to IT and V4 by the  
491 reliability in each region—as a noise-corrected adjustment. This results in a single score—the noise-  
492 corrected, median cross-validated fit to both IT and V4—which we repeat across all layers (Fig.  
493 3a: black and dotted lines for IT and V4 fits across layers, respectively). We determine also each  
494 layer’s differential fit with primate IT,  $\Delta_{IT-V4}$ , by taking the difference between the model’s fit to  
495 IT and V4 (Fig. 3a: hollow line). Early model layers (i.e. first half of model layers) better predict  
496 neural responses in early (V4) regions of the visual system ( $t(8) = 2.70, P = .015$ ), with peak V4  
497 fits occurring in pool3 (noise-corrected  $r = .95 \pm .30\text{STD}$ ) while later layers (e.g. second half of  
498 model layers) better predict neural responses in more anterior (IT) regions ( $t(8) = 3.70, P = .002$ ),  
499 with peak IT fits occurring in con5\_1 (noise-corrected  $r = .88 \pm .16\text{STD}$ ). We use model responses  
500 at this layer, con5\_1, as an ‘IT-like’ model layer in subsequent analyses.

### 501 4.4 Model Performance on Retrospective Dataset

502 For each trial, in each available experiment, the stimulus screen containing N objects was segmented  
503 into N object-centered images, using one of three protocols. For some experiments (e.g. Stark et al.  
504 2000) stimuli were already segmented, requiring no additional processing. For other experiments  
505 (e.g. Lee et al. 2006) the stimulus screen was segmented using a kmeans clustering approach that  
506 automatically identified the centroid of each object, defined a bounding box around each of these  
507 centroids, and extracted each object from the coordinates of each bounding box. There were a  
508 final class of experiments with more irregular dimensions (e.g. “familiar” objects in Barense et  
509 al. 2007); these stimuli were segmented by splitting the original stimulus screen into quadrants of  
510 equal size. We passed these object-centered N images to the model, then extracted model responses  
511 from an ‘IT like’ layer. These layer responses were flattened into length F vectors, resulting in an  
512 FxN response matrix for each trial. To identify the item-by item similarity between objects in this  
513 trial, were used Pearson’s correlation between items in this FxN response matrix, generating an  
514 NxN (item-by-item) correlation matrix. The item with the lowest mean off-diagonal correlation  
515 was the model-selected oddity (i.e. the item least like the others) which we labeled as either correct  
516 or incorrect, depending on its correspondence with ground truth. After repeating this protocol (for  
517 visualization see Supplement: Fig. S1) for each trial in the experiment, we computed the average  
518 accuracy across all trials. This single value, “model performance”, represents the performance that  
519 would be expected from a uniform readout of IT.

520 **4.5 Misclassified Experiments**

521 By definition, experiments that are fully supported by canonical VVS regions are not informative  
522 as to PRC involvement in perception; if the VVS enables 100% accuracy on a given experiment, no  
523 further perceptual processing is necessary. This does not, however, imply that human performance  
524 on these VVS-supported tasks will also be at ceiling: While a *lossless* readout of the VVS should  
525 perform these tasks at ceiling, a *lossy* readout—due to, for example, attentional or memory-related  
526 demands of maintaining those perceptual representations—will be systematically below ceiling. In  
527 this way, below-VVS performance on these trials can be attributed to extra-perceptual task de-  
528 demands that are orthogonal to the perceptual-mnemonic hypothesis. As a validation, we observe that  
529 all color experiments in the retrospective dataset adhere to this logic: model performance achieves  
530 100% accuracy on all trials (both 'Easy' and 'Difficult' experiments) and PRC-lesioned performance  
531 on these conditions is statistically indistinguishable from PRC-intact behavior<sup>31</sup>. Nonetheless, hu-  
532 man performance on 'difficult' trials is significantly lower than 'easy' trials. These results corrobo-  
533 rate researchers' expectations that these control stimuli are not diagnostic of PRC function, while  
534 the difficulty manipulation imposes extra-perceptual task demands.

535 We estimate model performance for all experiments in the retrospective dataset and, using  
536 the logic outline above, we exclude all stimulus sets where model performance is 100% accurate.  
537 As expected, this eliminates control experiments (e.g. color experiments in Barense et al. 2007).  
538 But it also eliminates many experiments that the original authors *described* as 'complex' stimu-  
539 lus sets, used to evaluate the role of PRC in perception. These 'misclassified' experiments be-  
540 long to two groups. The first group contains experiments that were argued as evidence *against*  
541 perirhinal involvement in perception<sup>25,36</sup> because performance did not significantly differ between  
542 PRC-intact and -lesioned participants. However, model performance suggests that canonical VVS  
543 regions should be sufficient for ceiling performance (Supplemental Figure S2a-b); consequently, the  
544 matched PRC-lesion/intact performance is expected, and entirely consistent with predictions from  
545 the perceptual-mnemonic hypothesis. The second group contains experiments that were argued  
546 to reveal evidence *in support* of perirhinal involvement in perception<sup>31,32</sup> because PRC-lesioned  
547 subject behavior was impaired relative to PRC-intact controls. However, the model suggests that  
548 canonical VVS regions should be entirely sufficient for performance on these tasks (Supplemental  
549 Figure S2c-d); consequently, the observed divergence may not be due to perceptual demands in  
550 these tasks. After excluding these experiments, we find 14 experiments that are able to adjudicate  
551 the involvement of PRC in concurrent visual discrimination tasks. This includes 10 experiments  
552 the original authors identified as diagnostic (all 'snow' experiments in Stark et al. 2000, 'high am-  
553 biguity' experiments, both 'novel' and 'familiar' experiments in Barense et al. 2007, 'novel objects'  
554 and 'faces' experiments in Lee et al. 2005, and 'different faces' experiments in Lee et al 2006).  
555 Additionally, this includes 4 experiments that were designated as 'control trials' by the original  
556 authors ('low ambiguity novel objects' and 'low ambiguity familiar objects' in Barense et al. 2007,  
557 'familiar objects' in Lee et al. 2005, and 'different scenes' in Lee et al. 2006). Note that the only  
558 criteria for this analysis is that model performance is not at ceiling: This selection procedure makes  
559 no claim about whether each individual experiment will exhibit PRC-related deficits.

560 **4.6 VVS Reliance**

561 Using electrophysiological data from prior work<sup>40</sup>, we estimate the cross-validated fit to neural  
562 data in macaque IT and V4, for each layer (Fig. 3a: solid black and dashed lines for IT and V4,  
563 respectively; Methods: Model Fit to Electrophysiological Data). We then compute each layer's  
564 differential fit to IT by computing the difference between noise-corrected IT and V4 neural fits  
565 (Fig. 3a:  $\Delta_{IT-V4}$ , hollow). The differential fit to IT cortex increased in 'deeper' layers ( $\beta = .98$ ,  
566  $F(1, 17) = 21.75, P = 10^{-13}$ ). Using the retrospective stimulus set (Fig. 3b top and bottom panels  
567 for PRC- and HPC-lesioned groups, respectively), we determine each layer's fit to human behavior,  
568 across all subject groups, using the mean squared prediction error (MSPE) between subject and  
569 model behavior:  $MSPE_s = \frac{1}{n} \sum_{i=1}^n (g_s(x_i) - \hat{g}_\ell(x_i))^2$  where  $x_i$  is each experiment,  $\ell$  is a single layer  
570 within the model,  $\hat{g}_\ell$  is the function (Methods: Model Performance on Retrospective Dataset) that  
571 operates over all trials in  $x_i$ , resulting in model performance on this experiment, for this layer of the  
572 model, while  $g_s(x_i)$  is the performance of participants in group  $s$  on experiment  $x_i$ , averaged across  
573 trials. We compute the average of the difference between Model ( $\hat{g}$ ) and Human ( $g$ ) Performance  
574 across all experiments, resulting in a single value for the fit to each subject group  $s$ , for each layer  
575 (e.g.  $MSPE_{prc.lesion}$ ). We then compute the difference between lesioned and intact model fits at  
576 each layer ( $\Delta_{group} = MSPE_{intact} - MSPE_{lesion}$ ) for both PRC- and HPC-lesioned groups (e.g.  
577  $\Delta_{prc} = MSPE_{prc.intact} - MSPE_{prc.lesion}$ ). Additionally, we determine whether the interaction  
578 between lesioned and intact subject behavior is significant, repeating previous analyses across all  
579 layers, for each patient group. To assess whether PRC-lesioned behavior is better fit by late-stage  
580 processing within the VVS we relate the model's differential fit with lesioned performance (for both

581  $\Delta_{prc}$  and  $\Delta_{hpc}$ ) to the model's differential fit to IT cortex ( $\Delta_{IT-V4}$ ). Model layers that better fit IT  
582 cortex ( $\Delta_{IT-V4}$ ) are better predictors of differential fit with PRC-lesioned behavior ( $\Delta_{prc}$ , Fig. 3c:  
583 top). Moreover, only 'IT-like' layers demonstrate significant interactions between subject groups  
584 (e.g. PRC-lesioned vs PRC-intact) after correcting for multiple comparisons across layers (Fig. 3c:  
585 black outlines). There is no correspondence with HPC-lesioned behavior ( $\Delta_{hpc}$ , Fig. 3c: bottom).  
586

## 586 4.7 Novel Stimulus Set Generation

587 We utilize stimuli and electrophysiological data from a previous experiment<sup>40</sup> consisting of 5760  
588 unique images, each with population-level electrophysiological responses recorded from primate  
589 V4 and IT. Every black and white image contains one of 64 objects, each belonging to one of  
590 eight categories, rendered in different orientations and projected onto random backgrounds—for a  
591 total of 90 images per object. We reconfigure these stimuli into within-category concurrent visual  
592 discrimination tasks. Each trial is designed to have the minimal configuration of objects ( $n = 3$ )  
593 required to be an oddity task: two of the three objects share an identity (two images of the 'typical'  
594 object<sub>i</sub>, presented from two different viewpoints and projected onto different random backgrounds)  
595 and the other is of a different identity (one image of the 'oddity', object<sub>j</sub>, e.g. two animals, where  
596 'elephant' and 'hedghog' are object<sub>i</sub> and object<sub>j</sub>, respectively). We generate a sample trial<sub>ij</sub> for the  
597 pair<sub>ij</sub> of objects *i* and *j* by randomly sampling two different objects from the same category, then  
598 sampling two images of object<sub>i</sub> (without replacement) and one image of the oddity of object<sub>j</sub>, all  
599 with random orientations and backgrounds. These three images comprise sample<sub>ij</sub> of the pair<sub>ij</sub>.  
600

## 600 4.8 Model Performance on Novel Stimuli

601 For each ( $N = 448$ ) unique within-category object pairing in the novel stimulus set we estimate  
602 model performance in two ways. First, we use a modified leave-one-out cross validation strategy.  
603 For a given sample<sub>ij</sub> trial we construct a random combination of three-way oddity tasks to be  
604 used as training data; we sample without replacement from the pool of all images of object<sub>i</sub> and  
605 object<sub>j</sub>, excluding only those three stimuli that were present in sample<sub>ij</sub>. This yields 'pseudo  
606 oddity experiments' where each trial contains two typical objects and one oddity that have the  
607 same identity as the objects in sample<sub>ij</sub> and are randomly configured (different viewpoints, different  
608 backgrounds, different orders). These 'pseudo oddity experiments' are used as training data. We  
609 reshape all images, present them to the model independently, and extract model responses from  
610 an 'IT-like' model layer (in this case, we use fc6 which has a similar fit to IT as conv5\_1 but fewer  
611 parameters to fit in subsequent steps). From these model responses, we train an L2 regularized  
612 linear classifier to identify the oddity across all ( $N = 52$ ) trials in this permutation of pseudo oddity  
613 experiments generated for sample<sub>ij</sub>. After learning this weighted, linear readout, we evaluate the  
614 classifier on the model responses to sample<sub>ij</sub>. This results in a prediction which is binarized into  
615 a single outcome {0 | 1}, either correct or incorrect. We repeat this protocol across 100 random  
616 sample<sub>ij</sub>s, for each pair<sub>ij</sub>. Second, we determine model performance using a uniform, linear (i.e.  
617 the distance metric used in the retrospective analyses) readout of model responses: For each pair<sub>ij</sub>,  
618 we generate 100 random sample<sub>ij</sub>s, determine the item with the lowest off-diagonal correlation  
619 as the model-selected oddity, which is binarized into a single outcome {0 | 1}, either correct or  
620 incorrect. Thus, we have 100 binarized outcomes for each randomly generated sample<sub>ij</sub> for both  
621 the uniform and non-uniform readouts for each pair<sub>ij</sub>. We average across sample<sub>ij</sub>s to estimate  
622 the expected performance on pair<sub>ij</sub> as our measures of uniform (model performance<sub>uniform</sub>) and  
623 weighted (model performance<sub>weighted</sub>) readouts. As expected, the more expressive weighted readout  
624 of model responses outperforms a uniform distance metric (paired ttest,  $t(447) = 33.55, P = 10^{-123}$ ;  
625 Fig. S3a: points on the y axis consistently above the diagonal). For both uniform and weighted  
626 readouts we order each pair<sub>ij</sub> according to accuracy, then compute the difference between each  
627 adjacent pair<sub>ij</sub> ( $\Delta_{pair}$ ); together, these 448 unique pairs (Fig. S3a: black) densely and continuously  
628 span the range of model performance (averaged uniform  $\bar{\Delta}_{pair} = .0018$ , averaged weighted  $\bar{\Delta}_{pair} =$   
629  $.0017$ ). Additionally, we learn a linear transform ( $\beta = 1.01, F(1, 446) = 23.28, P = 10^{-79}$ ) that  
630 projects model performance<sub>uniform</sub> to the expected value for (model performance<sub>weighted</sub> Fig. S3a:  
631 green). We can use this transform to project model performance<sub>uniform</sub> in the retrospective analysis  
632 into the performance that would be expected from model performance<sub>weighted</sub>. This transformed  
633 model performance<sub>transformed</sub> does significantly better at predicting PRC-lesioned behavior than  
634 the original model performance<sub>uniform</sub> ( $\beta = -.20, F(2, 25) = -4.26, P = 2 \times 10^{-4}$ ; Fig. S3b),  
635 motivating the need for novel experimental designs that enable model performance to be estimated  
636 with learned, weighted readouts of model responses. We select 4 categories that continuously span  
637 the space of model performance<sub>weighted</sub> ( $\min = .26, \max = 1.0, \bar{\Delta}_{pair} = .003$  Fig. S3b: Faces,  
638 Chairs, Planes, and Animals), which contains a total of 224 unique typical-oddity pairs. We use  
639 these 224 objects in subsequent analyses.

## 640 4.9 High-throughput Psychophysics Experiments 640

641 We create concurrent visual discrimination tasks composed of stimuli containing these 224 objects 641  
642 identified in the preceding analyses. To create each trial, we adopt the same the protocol used to 642  
643 generate each sample $_{ij}$ . We use this protocol for each of the 224 pair $_{ij}$ s: we generate 5 random 643  
644 combination of trials from each pair $_{ij}$  and fix these trials across all experiments (i.e. trial $_{ij_1}$ , trial $_{ij_2}$ , 644  
645 ..., trial $_{ij_5}$ ), resulting in (224 x 5) 1120 unique trials. We administer a randomized subset ( $N = 100$ ) 645  
646 of these concurrent visual discrimination trials to 297 human participants. In each trial, one of 1120 646  
647 oddity stimuli is presented for 10 seconds. participants are free to respond with a button press at any 647  
648 point to indicate the location of the oddity (right, left, bottom). If participants respond before 10 648  
649 seconds, their responses are recorded and the trial terminated. If participants fail to respond within 649  
650 10 seconds, the trial is marked as incorrect and terminated. After an initial trial phase (5 trials) to 650  
651 acclimate participants to the task, no further feedback is given at any point during the experiment. 651  
652 Each trial is self paced, such that participants initiated the beginning of the next trial with a 652  
653 button press (spacebar). All participants are compensated with a initial base rate for participating 653  
654 in this study. Additionally, each subject is given a monetary bonus for each correct answer, and 654  
655 receives a monetary penalty for each incorrect answer. This monetary incentive structure was 655  
656 titrated to ensure that participants are encouraged to attempt even the most difficult perceptual 656  
657 trials, while ensuring that all participants are compensated fairly (at least earning California's 657  
658 minimum wage for the time they participate in the experiment) given average performance. At 658  
659 the end of each experiment, participants are informed of their performance, alongside their total 659  
660 bonus; participants complete these tasks and received compensation through Amazon's Mechanical 660  
661 Turk. Given the truly random experimental generation procedure—and, subsequently, the highly 661  
662 variable nature of the stimuli used to compose each trial—there is no guarantee that one given 662  
663 trial $_{ij_n}$  will contain the information sufficient to complete the task. All of the faces, for example, 663  
664 may be rotated out of view in a given trial, such that the correct oddity can not be determined. 664  
665 To address this, of the 5 stimuli presented, for each of the 224 pair $_{ij}$ s, we restrict our analysis to 665  
666 1 trial $_{ij}$ . We select this exemplar for each pair $_{ij}$  using a single criterion: the item whose average 666  
667 accuracy (across participants) is closest to the average accuracy measured across all trials (across 667  
668 participants) belonging to other categories. This procedure enables us to exclude outliers (due to, 668  
669 for example, the objects not being fully visible on the viewing screen) while not biasing the results 669  
670 in future analyses. For all analyses, performance estimates are computed across the population 670  
671 of human participants. In this pooled population behavior, accuracy was reliable at the category 671  
672 ( $r = .97 \pm .03$ ), object ( $r = .69 \pm .07$ ), and image level ( $r = .24 \pm .05$ ) when estimated using the 672  
673 averaged correlation over 1000 split halves. This effect was even more prominent in the estimates of 673  
674 reaction time at the category ( $r = .99 \pm .01$ ), object ( $r = .91 \pm .02$ ), and image level ( $r = .76 \pm .02$ ). 674  
675 In order to relate human performance on these oddity tasks with model performance<sub>weighted</sub>, we 675  
676 employ the same pseudo experimental leave-one-out cross-validation strategy as outlined above, 676  
677 but now perform 100 train-test splits for each trial $_{ij}$ , across all ( $N = 224$ ) unique typical-oddity 677  
678 pairings. In order to relate human and model performance with the electrophysiological data, we 678  
679 repeat the leave-one-out cross-validation strategy developed for determining model performance, 679  
680 but in place of the fc6 model representations, we run the same protocol on the population level 680  
681 neural responses from IT and V4 cortex to those same images. We perform all analyses comparing 681  
682 human, electrophysiological, and model performance at the object level: for each object $_i$ , we average 682  
683 the performance on this object across all oddities (i.e. object $_j$ , object $_k$ , ...) resulting in a single 683  
684 estimate of performance on this item across all oddity tasks ( $N = 32$ ). Results from this analysis 684  
685 are plotted in Fig. 5. 685

## 686 4.10 Model Depth & Architecture Analyses 686

687 To examine the effect of model depth, we first ask whether model performance on the retrospective 687  
688 dataset varies depending on the *readout layer* used within the original architecture. For each exper- 688  
689 iment, we determine whether there is a significant positive relationship between model performance 689  
690 and model depth using ordinary least squares linear regression. Model performance increases with 690  
691 depth for some experiments in the retrospective dataset ('Low Snow' stimuli in Stark et al. 2000, 691  
692  $\beta = .01$ ,  $F(1, 19) = 6.17$ ,  $P = 10^{-5}$ ; 'Medium Snow' stimuli in Stark et al. 2000,  $\beta = .01$ ,  $F(1, 19)$  692  
693 = 7.37,  $P = 10^{-6}$ , 'Low Ambiguity Familiar' stimuli in Barense et al. 2007,  $\beta = .02$ ,  $F(1, 19)$  693  
694 = 13.61,  $P = 10^{-10}$ , 'Low Ambiguity Novel' stimuli in Barense et al. 2007,  $\beta = .02$ ,  $F(1, 19)$  694  
695 = 5.84,  $P = 10^{-4}$ , 'Novel Objects' in Lee et al. 2006,  $\beta = .01$ ,  $F(1, 19) = 3.91$ ,  $P = 10^{-3}$ ; 'Familiar 695  
696 Objects' in Lee et al. 2006,  $\beta = .02$ ,  $F(1, 19) = 4.56$ ,  $P = 10^{-3}$ , 'Different Scences' in Lee et al. 2005, 696  
697  $\beta = .01$ ,  $F(1, 19) = 6.09$ ,  $P = 10^{-5}$ ) but not others ('Faces' stimuli in Stark et al. 2000,  $\beta = .01$ , 697  
698  $F(1, 19) = 2.62$ ,  $P = .05$ , 'High Snow' stimuli in Stark et al. 2000,  $\beta = .00$ ,  $F(1, 19) = .06$ ,  $p > .05$ , 698  
699 'High Ambiguity Familiar' stimuli in Barense et al. 2007  $\beta = -.00$ ,  $F(1, 19) = -.05$ ,  $p > .05$ , 'High 699  
700 Ambiguity Novel' stimuli in Barense et al. 2007,  $\beta = -.01$ ,  $F(1, 19) = -3.31$ ,  $P = 4 \times 10^{-3}$ , 'Faces'

701 in Lee et al. 2006, experiment 1,  $\beta = -.01$ ,  $F(1, 19) = -4.38$ ,  $P = 3 \times 10^{-4}$ , ‘Faces’ in Lee et al. 701  
 702 2006, experiment 2,  $\beta = .00$ ,  $F(1, 19) = .07$ ,  $p > .05$  ‘Different Faces’ in Lee et al. 2005,  $\beta = -.00$ , 702  
 703  $F(1, 19) = -.38$ ,  $p > .05$ ). We inspect the behavior of PRC-lesioned participants across all experi- 703  
 704 ments, separated according to whether each experiment exhibited depth-dependent improvements. 704  
 705 PRC-lesioned participants performed significantly better ( $t(6) = 5.17$ ,  $P = .001$ ) on experiments 705  
 706 that exhibited depth improvements ( $\mu = .88$ ) than those that did not ( $\mu = .52$ ). This latter group 706  
 707 of experiments are those experiments with the most substantial differences between PRC-lesioned 707  
 708 and -intact behaviors. We then determine whether deeper *architectures* are able to better per- 708  
 709 form these experiments with the biggest difference between PRC-intact and -lesioned behavior. We 709  
 710 recruit a family of deep residual neural networks<sup>49</sup> (i.e. “resnets”) optimized to perform object 710  
 711 classification on a large-scale image classification task (ImageNet<sup>48</sup>). The model enables us to 711  
 712 preserves the same computational motif across models while increasing the number of layers from 712  
 713 18 to 152 in an effort to examine the effect of depth on model performance. We implement this 713  
 714 analysis using pretrained architectures from pytorch’s model zoo, and conduct the retrospective 714  
 715 analysis (Methods: Model Performance on Retrospective Dataset) using the penultimate layer as 715  
 716 the readout used to determine model performance. The MSPE between model performance and 716  
 717 PRC-intact behavior (Methods: VVS Reliance) decrease with model depth ( $t(4) = 2.56$ ,  $p > .05$ ), 717  
 718 nor does the slope of the line of best fit (a measure of how ‘on diagonal’ PRC-intact behavior is 718  
 719 from model performance) change with model depth ( $t(4) = 2.76$ ,  $p > .05$ ). More directly, the main 719  
 720 findings observed in the original model are replicated across these novel, deeper architectures, such 720  
 721 that the interaction between PRC-intact and -lesioned participants is observed in all models (18 721  
 722 layers:  $\beta = -.51$ ,  $F(3, 24) = -3.32$ ,  $P = .005$ ; 34 layers:  $\beta = -.45$ ,  $F(3, 24) = -3.07$ ,  $P = .005$ ; 50 722  
 723 layers:  $\beta = -.49$ ,  $F(3, 24) = -3.25$ ,  $P = .005$ ; 101 layers:  $\beta = -.56$ ,  $F(3, 24) = -3.66$ ,  $P = .005$ ; 723  
 724 152 layers:  $\beta = -.55$ ,  $F(3, 24) = -3.97$ ,  $P = .005$ ). Deeper models do not perform these behaviors 724  
 725 more like PRC-intact participants. 725

## 726 4.11 Content-Specific Optimization Procedure

727 We optimize a computational proxy for the VVS to perform putatively ‘complex’ tasks (e.g. face 727  
 728 discrimination) by changing the distribution of training data: instead of training to perform an 728  
 729 object classification task on a dataset with millions of common objects, as per prior models used 729  
 730 in this study, we use a large-scale face-classification dataset which approximates a face individua- 730  
 731 tion task<sup>50</sup>. With a pytorch implementation, we use a pretrained model to extract features from 731  
 732 experimental stimuli as in prior analyses. In the retrospective dataset, we extract face-trained 732  
 733 model responses and determine model performance as outlined in model performance on Retro- 733  
 734 spective Dataset (Fig. 7a-c). In the novel stimulus set, we first employ the same leave-one-out 734  
 735 cross-validation strategy to determine model performance, simply using the face-trained model in 735  
 736 place of the object-trained model. However, this results in model performance on faces that is not 736  
 737 statistically different from object-trained model performance ( $t(46) = 1.23$ ,  $p > .05$ )—that is, there 737  
 738 appears to be no improvement for faces and significantly worse performance for all other objects 738  
 739 ( $t(167) = 10.65$ ,  $p = 1.51^{-19}$ ; S4d). Additionally, there is a complete lack of correspondence be- 739  
 740 tween face-trained model performance and human performance ( $\beta = .25$ ,  $F(1, 30) = 1.50$ ,  $p > .05$ ; 740  
 741 S4f), IT-supported performance ( $\beta = .30$ ,  $F(1, 30) = .61$ ,  $p > .05$ ; S4h), and human reaction time 741  
 742 ( $\beta = -884.36$ ,  $F(1, 30) = -.71$ ,  $p > .05$ : S4i). We note that the dataset used to optimize the 742  
 743 face-trained model presents all stimuli at central field of view with cropped backgrounds, while 743  
 744 the novel stimulus set presents stimuli at random locations and sizes. To address this, we add 744  
 745 one additional image preprocessing step in order to make the testing data more closely resemble 745  
 746 the viewing conditions in the training dataset: ‘foveating’ the object within the image, prior to 746  
 747 presenting it to the model. Using meta-data available for this stimulus set, the center of the object 747  
 748 is identified and a bounding box is placed around the object, with minimal background included. 748  
 749 This serves to crop the image, creating a synthetic ‘foveating’ process. The centered, cropped 749  
 750 object is then rescaled to match the dimensions of the model inputs and passed to the model. In 750  
 751 the main results section, we report results from this ‘foveated’ face-trained model performance and 751  
 752 observe a significant increase in the performance of these models on face tasks. For consistency, we 752  
 753 perform this additional ‘foveating’ step for both the object-trained model reported in these data as 753  
 754 well (Fig. 7e, g, i). We can conclude that while this content-specific optimization procedure leads 754  
 755 to increased performance on ‘within distribution’ tasks, this procedure does not generalize across 755  
 756 these viewing conditions, further corroborating the restricted performance enhancements observed 756  
 757 with this approach. 757

## 758 References

- 759 [1] Howard Eichenbaum and Neal J Cohen. *From conditioning to conscious recollection: Memory* 759  
 760 *systems of the brain*. Oxford University Press on Demand, 2004. 760

- 761 [2] Daniel J Felleman and DC Essen Van. Distributed hierarchical processing in the primate 761  
762 cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.  
763 [3] Shimon Ullman et al. *High-level vision: Object recognition and visual cognition*, volume 2. 763  
764 MIT press Cambridge, MA, 1996.  
765 [4] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object 765  
766 recognition? *Neuron*, 73(3):415–434, 2012.  
767 [5] Larry R Squire and Stuart Zola-Morgan. The medial temporal lobe memory system. *Science*, 767  
768 253(5026):1380–1386, 1991.  
769 [6] Joseph R Manns and Howard Eichenbaum. Evolution of declarative memory. *Hippocampus*, 769  
770 16(9):795–808, 2006.  
771 [7] Wendy A Suzuki and David G Amaral. Functional neuroanatomy of the medial temporal lobe 771  
772 memory system. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, 772  
773 2004.  
774 [8] Elisabeth A Murray, Timothy J Bussey, and Lisa M Saksida. Visual perception and memory: 774  
775 a new view of medial temporal lobe function in primates and rodents. *Annu. Rev. Neurosci.*, 775  
776 30:99–122, 2007.  
777 [9] Wendy A Suzuki. Perception and the medial temporal lobe: evaluating the current evidence. 777  
778 *Neuron*, 61(5):657–666, 2009.  
779 [10] Wendy A Suzuki and Mark G Baxter. Memory, perception, and the medial temporal lobe: a 779  
780 synthesis of opinions. *Neuron*, 61(5):678–679, 2009.  
781 [11] Yasushi Miyashita. Perirhinal circuits for memory processing. *Nature Reviews Neuroscience*, 781  
782 20(10):577–592, 2019.  
783 [12] William Beecher Scoville and Brenda Milner. Loss of recent memory after bilateral hippocam- 783  
784 pal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11, 1957.  
785 [13] John P Aggleton and Malcolm W Brown. Interleaving brain systems for episodic and recog- 785  
786 nition memory. *Trends in cognitive sciences*, 10(10):455–463, 2006.  
787 [14] Thackery I Brown, Bernhard P Staresina, and Anthony D Wagner. Noninvasive functional 787  
788 and anatomical imaging of the human medial temporal lobe. *Cold Spring Harbor perspectives 788  
789 in biology*, 7(4):a021840, 2015.  
790 [15] Martine Meunier, Jocelyne Bachevalier, Mortimer Mishkin, and Elizabeth A Murray. Effects 790  
791 on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex 791  
792 in rhesus monkeys. *Journal of Neuroscience*, 13(12):5418–5432, 1993.  
793 [16] MJ Eacott, D Gaffan, and EA Murray. Preserved recognition memory for small sets, and 793  
794 impaired stimulus identification for large sets, following rhinal cortex ablations in monkeys. 794  
795 *European Journal of Neuroscience*, 6(9):1466–1478, 1994.  
796 [17] D Gaffan and EA Murray. Monkeys with rhinal cortex lesions succeed in object discrimination 796  
797 learning despite 24-hour intertrial intervals and fail at match to sample despite double sample 797  
798 presentations. *Behavioral Neuroscience*, 106:30–38, 1992.  
799 [18] Timothy J Bussey, Lisa M Saksida, and Elisabeth A Murray. Perirhinal cortex resolves feature 799  
800 ambiguity in complex visual discriminations. *European Journal of Neuroscience*, 15(2):365– 800  
801 374, 2002.  
802 [19] Mark J Buckley and David Gaffan. Perirhinal cortex ablation impairs visual object identifica- 802  
803 tion. *Journal of Neuroscience*, 18(6):2268–2275, 1998.  
804 [20] MJ Buckley and D Gaffan. Impairment of visual object-discrimination learning after perirhinal 804  
805 cortex ablation. *Behavioral neuroscience*, 111(3):467, 1997.  
806 [21] MJ Buckley and D Gaffan. Perirhinal cortex ablation impairs configural learning and paired- 806  
807 associate learning equally. *Neuropsychologia*, 36(6):535–546, 1998.  
808 [22] Elisabeth A Murray and Timothy J Bussey. Perceptual–mnemonic functions of the perirhinal 808  
809 cortex. *Trends in cognitive sciences*, 3(4):142–151, 1999.

- 810 [23] Timothy J Bussey and Lisa M Saksida. The organization of visual object representations: a 810  
811 connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience*, 811  
812 15(2):355–364, 2002. 812
- 813 [24] Elizabeth A Buffalo, Lisa Stefanacci, Larry R Squire, and Stuart M Zola. A reexamination of 813  
814 the concurrent discrimination learning task: the importance of anterior inferotemporal cortex, 814  
815 area te. *Behavioral neuroscience*, 112(1):3, 1998. 815
- 816 [25] Elizabeth A Buffalo, Seth J Ramus, Robert E Clark, Edmond Teng, Larry R Squire, and 816  
817 Stuart M Zola. Dissociation between the effects of damage to perirhinal cortex and area te. 817  
818 *Learning & Memory*, 6(6):572–599, 1999. 818
- 819 [26] Elizabeth A Buffalo, Paul J Reber, and Larry R Squire. The human perirhinal cortex and 819  
820 recognition memory. *Hippocampus*, 8(4):330–339, 1998. 820
- 821 [27] Mark J Buckley, Michael CA Booth, Edmund T Rolls, and David Gaffan. Selective perceptual 821  
822 impairments after perirhinal cortex ablation. *Journal of Neuroscience*, 21(24):9824–9836, 2001. 822
- 823 [28] Timothy J Bussey, Lisa M Saksida, and Elisabeth A Murray. Impairments in visual discrim- 823  
824 ination after perirhinal cortex lesions: testing ‘declarative’ vs. ‘perceptual-mnemonic’ views of 824  
825 perirhinal cortex function. *European Journal of Neuroscience*, 17(3):649–660, 2003. 825
- 826 [29] Andy CH Lee, Tim J Bussey, Elisabeth A Murray, Lisa M Saksida, Russell A Epstein, Narinder 826  
827 Kapur, John R Hodges, and Kim S Graham. Perceptual deficits in amnesia: challenging the 827  
828 medial temporal lobe ‘mnemonic’ view. *Neuropsychologia*, 43(1):1–11, 2005. 828
- 829 [30] Andy CH Lee, Mark J Buckley, David Gaffan, Tina Emery, John R Hodges, and Kim S 829  
830 Graham. Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond 830  
831 long-term declarative memory: a double dissociation in dementia. *Journal of Neuroscience*, 831  
832 26(19):5198–5203, 2006. 832
- 833 [31] Morgan D Barense, David Gaffan, and Kim S Graham. The human medial temporal lobe 833  
834 processes online representations of complex objects. *Neuropsychologia*, 45(13):2963–2974, 2007. 834
- 835 [32] Marika C Inhoff, Andrew C Heusser, Arielle Tambini, Chris B Martin, Edward B O’Neil, 835  
836 Stefan Köhler, Michael R Meager, Karen Blackmon, Blanca Vazquez, Orrin Devinsky, et al. 836  
837 Understanding perirhinal contributions to perception and memory: Evidence through the lens 837  
838 of selective perirhinal damage. *Neuropsychologia*, 124:9–18, 2019. 838
- 839 [33] Craig EL Stark and Larry R Squire. Intact visual perceptual discrimination in humans in the 839  
840 absence of perirhinal cortex. *Learning & Memory*, 7(5):273–278, 2000. 840
- 841 [34] Daniel A Levy, Yael Shrager, and Larry R Squire. Intact visual discrimination of complex 841  
842 and feature-ambiguous stimuli in the absence of perirhinal cortex. *Learning & memory*, 12(1): 842  
843 61–66, 2005. 843
- 844 [35] Larry R Squire, Yael Shrager, and Daniel A Levy. Lack of evidence for a role of medial 844  
845 temporal lobe structures in visual perception. *Learning & Memory*, 13(2):106–107, 2006. 845
- 846 [36] Ashley R Knutson, Ramona O Hopkins, and Larry R Squire. Visual discrimination perfor- 846  
847 mance, memory, and medial temporal lobe function. *Proceedings of the National Academy of 847  
848 Sciences*, 109(32):13106–13111, 2012. 848
- 849 [37] Santiago A Cadena, George H Denfield, Edgar Y Walker, Leon A Gatys, Andreas S Tolias, 849  
850 Matthias Bethge, and Alexander S Ecker. Deep convolutional models improve predictions of 850  
851 macaque v1 responses to natural images. *PLoS computational biology*, 15(4):e1006897, 2019. 851
- 852 [38] Pouya Bashivan, Kohitij Kar, and James J DiCarlo. Neural population control via deep image 852  
853 synthesis. *Science*, 364(6439):eaav9436, 2019. 853
- 854 [39] Daniel LK Yamins, Ha Hong, Charles F Cadieu, Ethan A Solomon, Darren Seibert, and 854  
855 James J DiCarlo. Performance-optimized hierarchical models predict neural responses in higher 855  
856 visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624, 2014. 856
- 857 [40] Najib J Majaj, Ha Hong, Ethan A Solomon, and James J DiCarlo. Simple learned weighted 857  
858 sums of inferior temporal neuronal firing rates accurately predict human core object recognition 858  
859 performance. *Journal of Neuroscience*, 35(39):13402–13418, 2015. 859
- 860 [41] Russell A Poldrack and Martha J Farah. Progress and challenges in probing the human brain. 860  
861 *Nature*, 526(7573):371–379, 2015. 861

- 862 [42] Jackson C Liang, Jonathan Erez, Felicia Zhang, Rhodri Cusack, and Morgan D Barense. 862  
863 Experience transforms conjunctive object representations: Neural evidence for unitization after 863  
864 visual expertise. *Cerebral Cortex*, 30(5):2721–2739, 2020. 864
- 865 [43] James J DiCarlo and David D Cox. Untangling invariant object recognition. *Trends in cognitive 865  
866 sciences*, 11(8):333–341, 2007. 866
- 867 [44] Michael J Arcaro, Peter F Schade, Justin L Vincent, Carlos R Ponce, and Margaret S Living- 867  
868 stone. Seeing faces is necessary for face-domain formation. *Nature neuroscience*, 20(10):1404, 868  
869 2017. 869
- 870 [45] Krishna Srihasam, Justin L Vincent, and Margaret S Livingstone. Novel domain formation 870  
871 reveals proto-architecture in inferotemporal cortex. *Nature neuroscience*, 17(12):1776–1783, 871  
872 2014. 872
- 873 [46] Krishna Srihasam, Joseph B Mandeville, Istvan A Morocz, Kevin J Sullivan, and Margaret S 873  
874 Livingstone. Behavioral and anatomical consequences of early versus late symbol training in 874  
875 macaques. *Neuron*, 73(3):608–619, 2012. 875
- 876 [47] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale 876  
877 image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 877
- 878 [48] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large- 878  
879 scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern 879  
880 recognition*, pages 248–255. Ieee, 2009. 880
- 881 [49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image 881  
882 recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 882  
883 pages 770–778, 2016. 883
- 884 [50] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. *British 884  
885 Machine Vision Association*, 2015. 885

## 886 5 Acknowledgements 886

887 This work is supported by a National Science Foundation Graduate Research Fellowship under 887  
888 Grant No. DGE-1656518, Stanford’s Center for Mind Brain Behavior and Technology, and the 888  
889 Marcus and Amelia Wallenberg Foundation (MAW2015.0043). We thank the all of the original 889  
890 authors from those studies in the retrospective dataset—providing stimuli when possible, and their 890  
891 assistance even when the stimuli were not accessible. Specifically, we would like to thank Morgan 891  
892 Barense, Elizabeth Buffalo, Tim Bussey, Lila Davachi, Andy Lee, Elizabeth Murray, Craig Stark, 892  
893 and Larry Squire, as well as Mona Hopkins and Jennifer Frascino for their diligent efforts securing 893  
894 multiple stimulus sets. We thank Mark Eldridge, Nathan Kong, Heather Kosakowski, Russel 894  
895 Poldrack, Emily Mackevicius, and Natalia Veléz for their comments and feedback on previous 895  
896 versions on this manuscript. 896

## 897 6 Author contributions statement 897

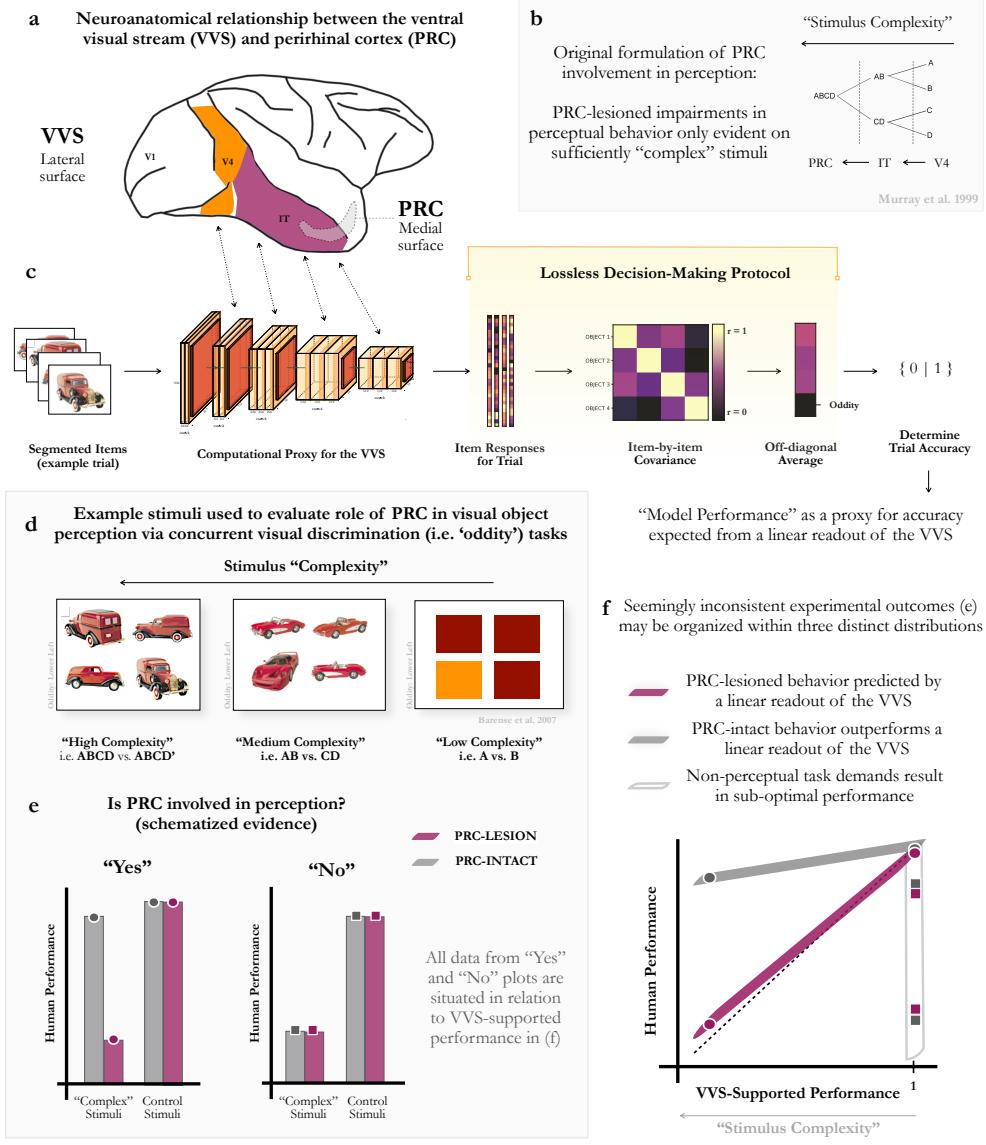
898 T.B. and A.D.W. conducted the literature review, and reached out to original authors. T.B. con- 898  
899 ceived of the modeling approach and performed all modeling work. D.L.K.Y. provided technical 899  
900 advice on neural fitting. T.B. designed, implemented, and analyzed novel experiments. T.B. de- 900  
901 signed, implemented, and analyzed the in silico experiments. T.B., D.L.K.Y., and A.W.D. discussed 901  
902 results, then wrote and revised the manuscript. 902

## 903 7 Competing Interests 903

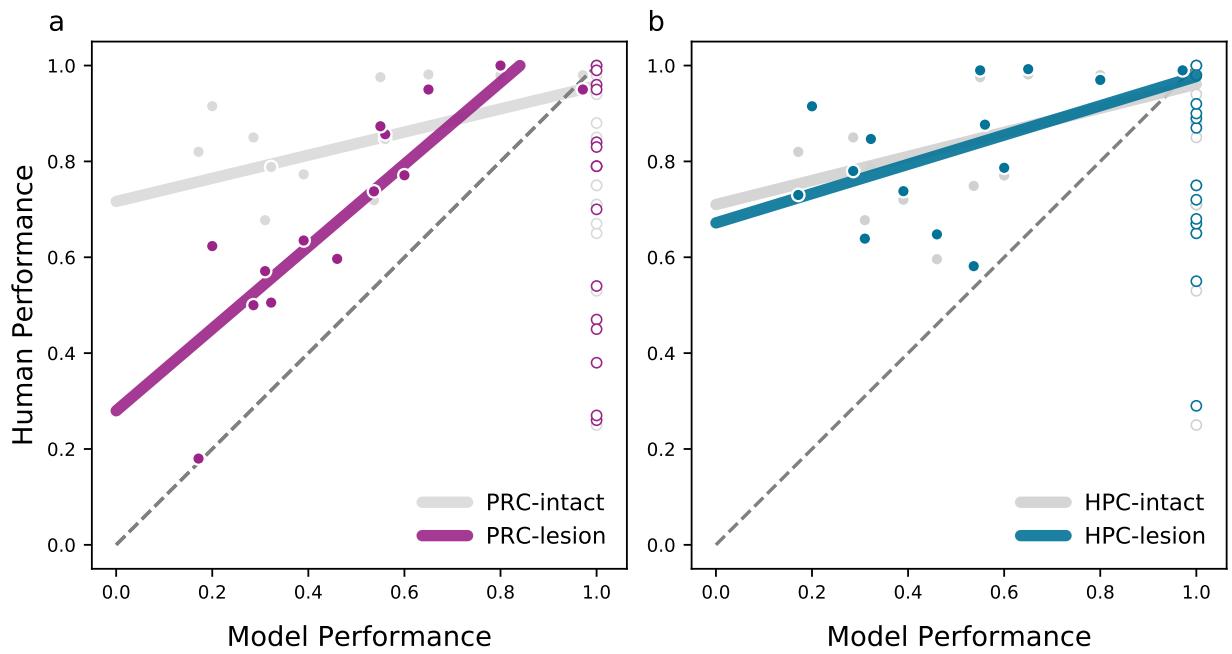
904 The authors declare no competing interests. 904

## 905 8 Code & Data Availability 905

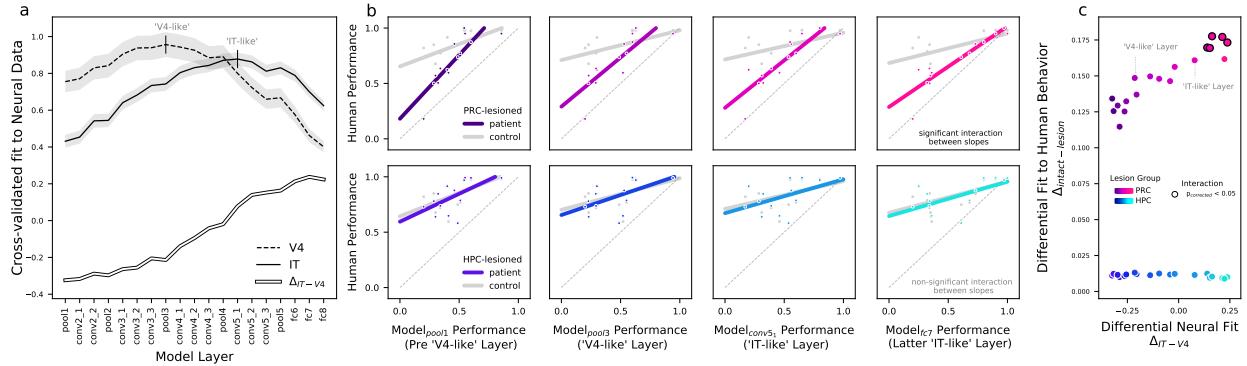
906 Code for all analyses can be found at [https://github.com/tzler/mlt\\_perception](https://github.com/tzler/mlt_perception). Stimulus sets can 906  
907 be obtained by contacting the corresponding author. 907



**Figure 1: Resolving seemingly inconsistent experimental findings with a computational proxy for the ventral visual stream. (a)** Perirhinal cortex (PRC) is a neuroanatomical structure within the medial temporal lobe (MTL) situated at the apex of the ventral visual system (VVS), downstream of ‘high-level’ visual structures such as inferior temporal (IT) cortex. **(b)** A perceptual-mnemonic hypothesis posits that PRC enables perceptual behaviors not supported by canonical sensory cortices, in addition to its mnemonic functions. Critically, PRC-related perceptual impairments are only expected on so-called “complex” perceptual stimuli. **(c)** Our trial-level protocol formalizes perceptual demands on PRC in concurrent visual discrimination (i.e. ‘oddity’) tasks. We segment each stimulus screen containing  $N$  objects into  $N$  independent images, pass them to a computational proxy for the VVS, and extract  $N$  feature vectors from an ‘IT-like’ layer. After generating a item-by-item covariance matrix for each trial, the item with the least off-diagonal covariance is marked as the ‘oddity.’ Critically, this is a lossless decision-making protocol which is agnostic to extra-perceptual task demands (i.e. memory, attention, motivation). **(d)** Example stimuli used to evaluate the perceptual-mnemonic hypothesis that span the range of stimulus ‘complexity.’ **(e)** Evaluating PRC involvement in perception has historically been formatted in categorical terms, and been forced to rely on with *descriptive* accounts of stimulus properties (e.g. stimulus “complexity”). This has generated seemingly inconsistent experimental evidence both for (left) and against (right) PRC involvement in perception. **(f)** Here we propose to resolve these apparent inconsistencies using this null model of PRC involvement in oddity tasks by identifying three distinct distributions in the literature: PRC-lesioned behavior that is predicted by a linear readout of the VVS, PRC-intact behavior that outperforms a linear readout of the VVS, and stimuli for which non-perceptual task demands result in sub-optimal performance. We consider experiments described as ‘complex’ but which the model performs at ceiling (i.e.  $x=1$ ) to be misclassified.



**Figure 2: After excluding PRC-irrelevant stimuli, a computational proxy of the VVS predicts PRC-lesioned performance directly from experimental stimuli, while each are outperformed by PRC-intact participants.** We collect previously published ‘oddity’ tasks administered to PRC-lesioned and -intact human participants. We then build a linear decoder off ‘IT-like’ layers from a computational proxy for the VVS in order to determine the average performance across all trials in each experiment. This single value, model performance, corresponds to the experimental accuracy expected from a linear readout of IT cortex under a lossless decision-making protocol. Stimuli where model performance is at ceiling ( $x=1$ , open dots) are not relevant for evaluating the role of PRC in perception: As VVS responses should support perfect discrimination between these stimuli, any below ceiling performance in the human is attributed to extra-perceptual task demands (i.e. memory). **(a)** This computational proxy for IT cortex predicts the behavior of PRC-lesioned participants, while PRC-intact participants outperform both model and PRC-lesioned participants. **(b)** HPC-lesioned and intact participants all outperform this computational model on relevant stimuli; both for participants with an entirely intact medial temporal lobe, which includes PRC, as well as participants with selective damage to the hippocampus that spare PRC. Together, these results suggest that PRC-lesioned behavior reflects a linear readout of the VVS, neurotypical behaviors on these tasks outperform the VVS, and this behavior is dependent on PRC.



**Figure 3: VVS reliance in a PRC-lesioned state: While ‘IT-like’ model layers predict perirhinal-lesioned behaviors, the available stimuli do not clearly separate IT- from V4-supported performance.** There has long been concern that concurrent damage in PRC-adjacent cortical structures (such as IT) leads to perceptual deficits, not damage to PRC per se. These concerns are allayed by the observation that IT-like layers fail to perform ‘complex’ oddity tasks. Nonetheless, a question remains: where in the VVS is PRC-lesioned behavior reliant on? To address this question, we leverage the model’s differential correspondence with V4 and IT electrophysiological responses across layers. **(a)** For each layer, we estimate the noise-corrected, cross-validated fit to electrophysiological responses in macaque IT and V4. We then compute each layer’s differential fit to IT ( $\Delta_{IT-V4}$ : hollow). **(b)** Using the retrospective stimulus set, we determine each layer’s differential fit to lesioned behavior, both for PRC- and HPC-lesioned participants (top and bottom panels, respectively), using the mean squared prediction error (MSPE) between human and model behavior. We then compute the difference between lesioned and intact model fits at each layer ( $\Delta_{lesion} = MSPE_{intact} - MSPE_{lesion}$ ), for both PRC- and HPC-lesioned groups (e.g.  $\Delta_{prc} = MSPE_{prc.intact} - MSPE_{prc.lesion}$ ). Additionally, we determine whether the interaction between lesioned and intact subject behavior is significant, repeating previous analyses (from Fig. 2a) across all layers. **(c)** Model layers that better fit IT cortex ( $\Delta_{IT-V4}$ ) are better predictors of differential fit with PRC-reliant behavior ( $\Delta_{prc}$ , top). Additionally, the interaction between PRC-intact and -lesioned performance is only significant in ‘IT-like’ layers, after correcting for multiple comparisons (black outlined circles). There is no correspondence between ( $\Delta_{IT-V4}$ ) and HPC-lesioned behavior ( $\Delta_{hpc}$ , bottom). However, when directly comparing the model fit to PRC-lesioned participants in ‘IT-like’ and ‘V4-like’ model layers, there is not a significant difference, as can be seen in the relative similarity in the model fit to PRC-lesioned behaviors across all layers in (b). While these data suggest that PRC-lesioned behavior is reliant on high-level visual cortex, the available stimuli in the retrospective dataset do not enable focal anatomical claims.

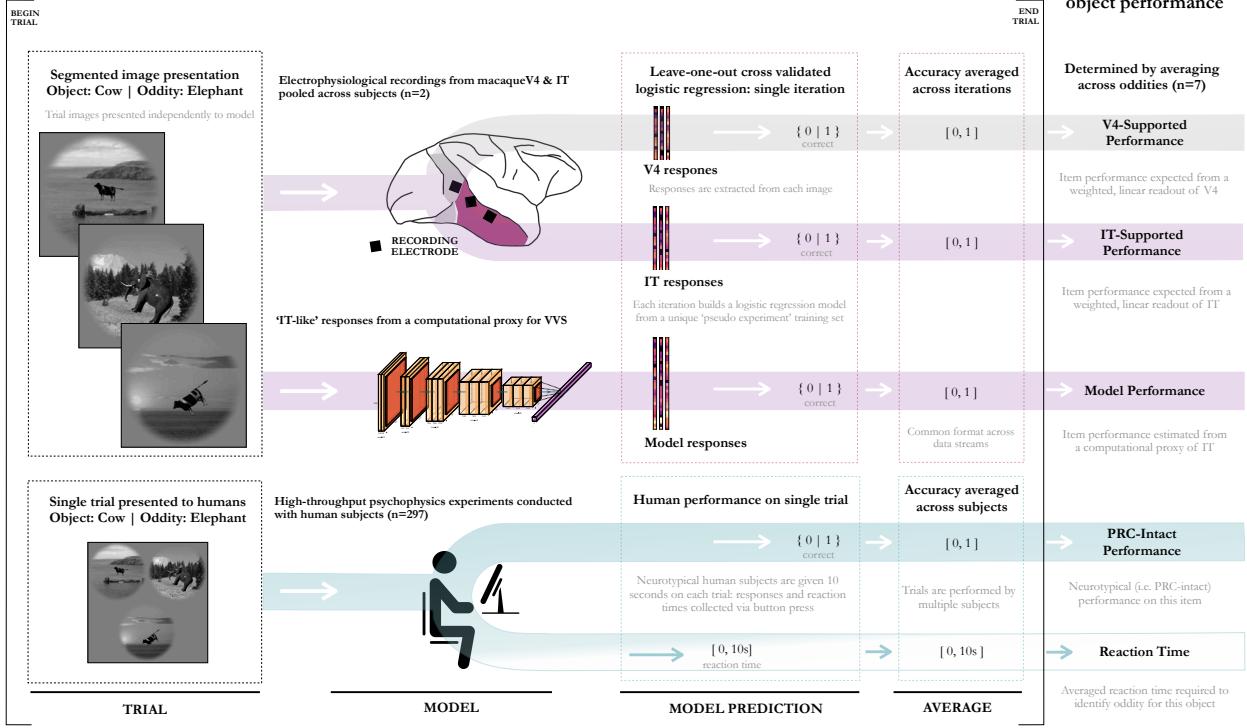
**a Example stimuli from novel concurrent visual discrimination dataset: one trial from each of four categories**

'Animal' trial : Object: Cow | Oddity: Elephant    'Chair' trial : Object: Soft-back | Oddity: Hard-back    'Plane' trial : Object: Metallic | Oddity: Duster    'Face' trial : Object: Face01 | Oddity: Face03

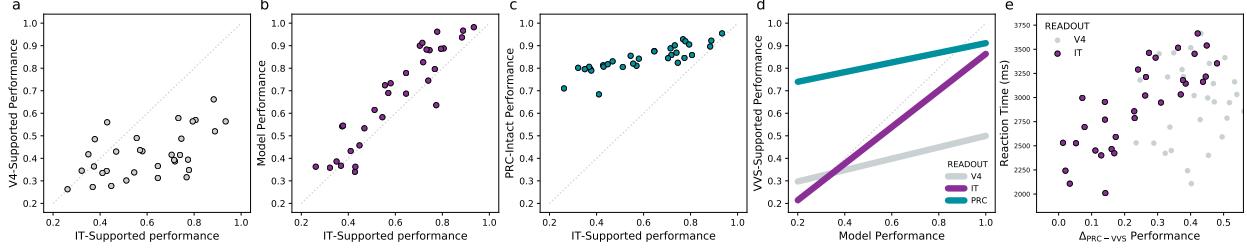


**b**

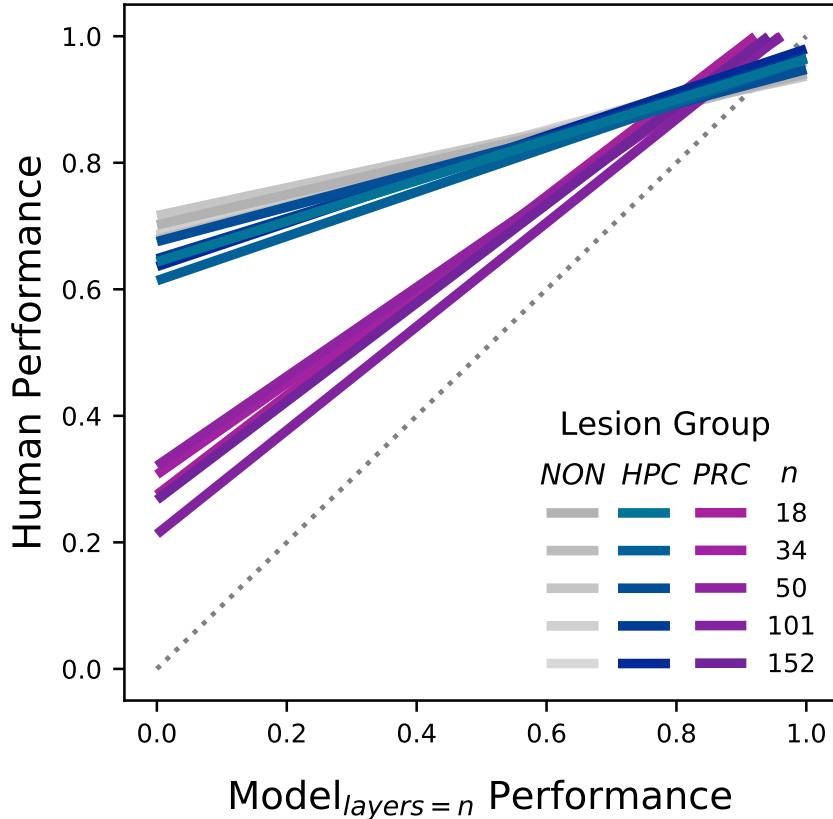
Parallel data streams for estimating Human, Model, V4-, and IT-Supported Performance on a given trial



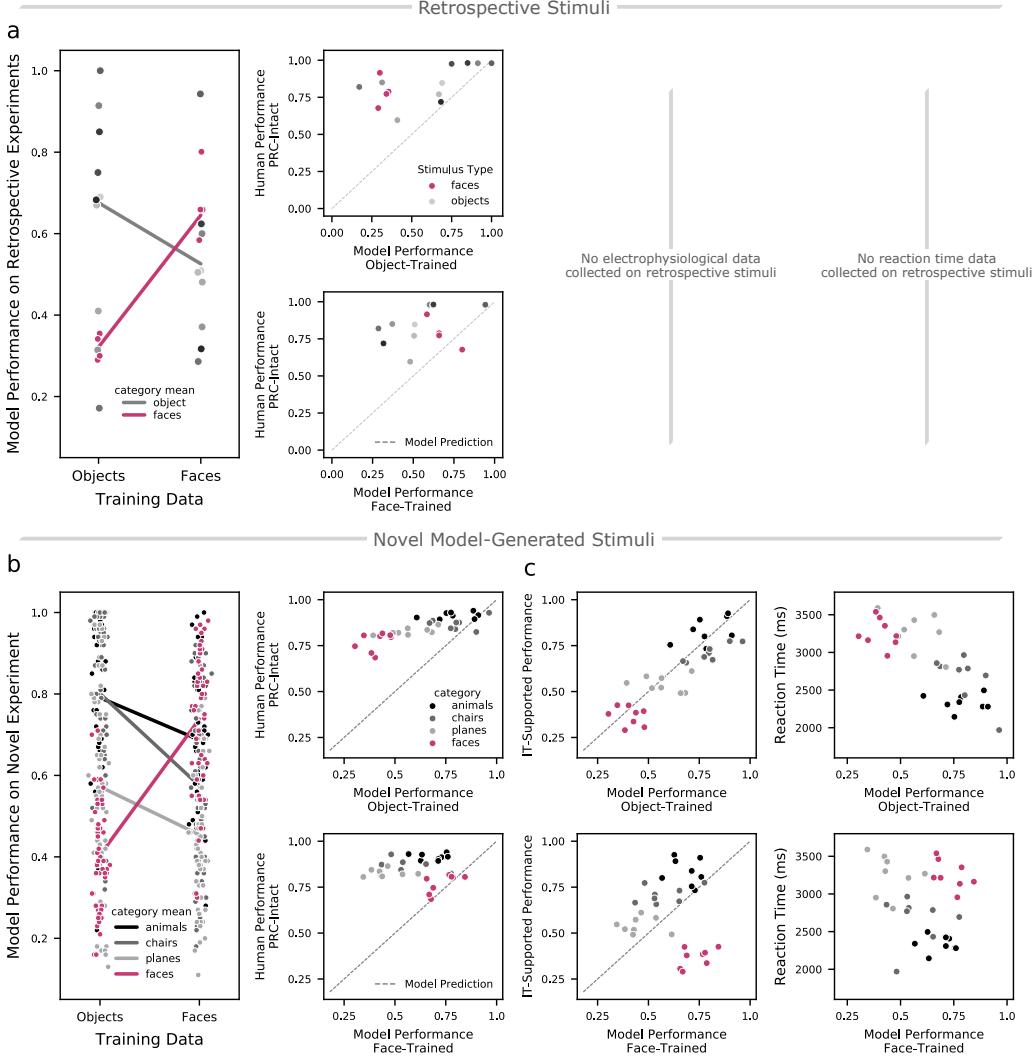
**Figure 4: Parallel data processing streams enable comparison of VVS-supported performance, model performance, and PRC-Intact Performance on novel experimental stimuli** (a) Example stimuli from four categories used in the novel, model-driven concurrent visual discrimination experiment: together, this stimulus set contains 32 unique objects used to generate 224 unique within-category object combinations. (b) For each trial, given the same object and oddity images (left), there are parallel data processing streams to estimate human performance and Reaction time (RT), model performance, as well as V4- and IT-Supported Performance. Human data (bottom) are collected via high-throughput psychophysics experiments online: for a given trial, accuracy and RT data are collected, which are averaged across participants. To estimate model performance on these same stimuli (middle), objects are segmented and presented to the model, responses are extracted from an IT-like layer, a prediction is made using a modified leave-one-out cross-validated approach, and the average accuracy across iterations is taken as this trial's estimate of model performance. To estimate V4- and IT-Supported Behavior (top), we use the same protocol developed for the model, but predictions are made over electrophysiological recordings collected from the macaque<sup>40</sup> instead of model responses. (c) To estimate performance on each unique object in this stimulus set (n=32), we take the average value collected across that object with all seven of its oddities. This yields human performance, model performance, as well as V4- and IT-Supported Performance on the same experimental stimuli. Colors matched to Fig. 5.



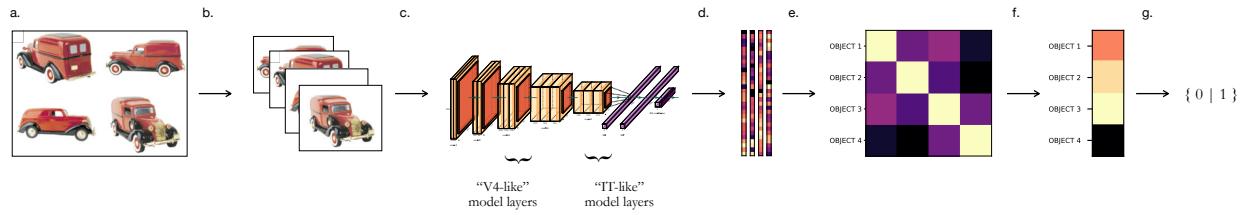
**Figure 5: A model-driven stimulus set separates perirhinal-dependent behaviors from multiple stages of processing throughout the ventral visual system.** Here we evaluate the relationship between model, electrophysiological, and human performance on a novel stimulus set, generated within this modeling approach. **(a)** A weighted, linear readout of IT outperforms V4, clearly separating early from late stage processing within the VVS. **(b)** Model performance on these stimuli corresponds to IT-Supported Performance, validating the use of this model as a computational proxy for IT in oddity tasks. **(c)** Neurotypical (i.e. PRC-intact) human participants outperform V4- and IT-supported behavior, replicating findings from the retrospective analysis with a stimulus set that more densely and continuously samples the space of VVS-supported behavior. Additionally, these predictions are at the item level (averaged across oddities,  $N = 7$ ), not experimental averages. **(d)** The model provides a basis space to situate human behavior in relationship with VVS-supported performance, enabling more focal neuroanatomical claims about VVS-reliance in this and future experiments. **(e)** The difference between PRC-intact and IT-supported performance on each item scales with reaction time. These data suggest that in order to outperform a linear readout of IT cortex, PRC-intact participants require more time.



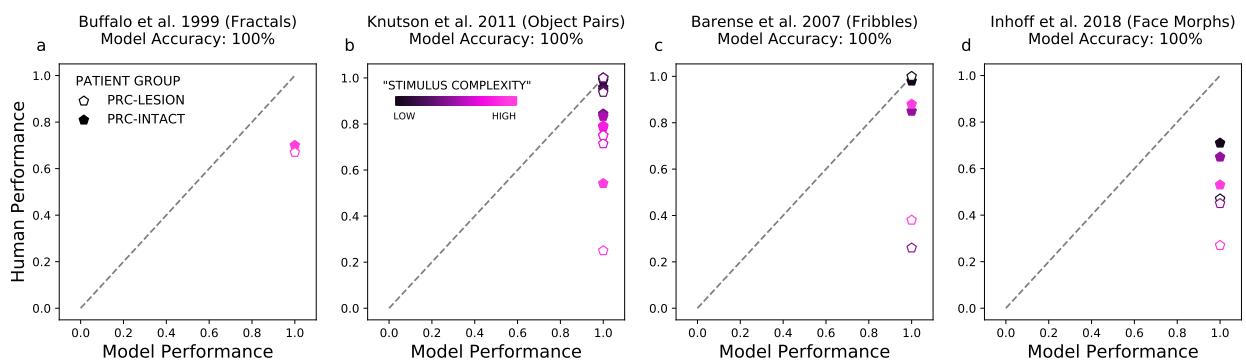
**Figure 6: Increasing architecture depth does not achieve PRC-intact performance.** Here we repeat previous analyses but systematically vary model ‘depth’ from 18–152 layers. For each of these architectures, we determine model performance for each experiment within the retrospective dataset. First we determine the model-selected oddity in each trial by identifying the item with the lowest off-diagonal correlation to the other items—as described in the retrospective analysis—using a penultimate, ‘IT-like’ model layer; we then average the model’s accuracy across all trials within an experiment. We compare model performance to PRC-intact (greys), HPC-lesioned (blues) and PRC-lesioned (purples) behavior for each model. Solid lines correspond to the best fit across all experiments. The interaction between PRC-intact and -lesioned subject behavior is persistent across all models. Increasing the number of VVS-like computations over a given stimulus—that is, by adding more layers—does not appear to approximate PRC-supported behaviors.



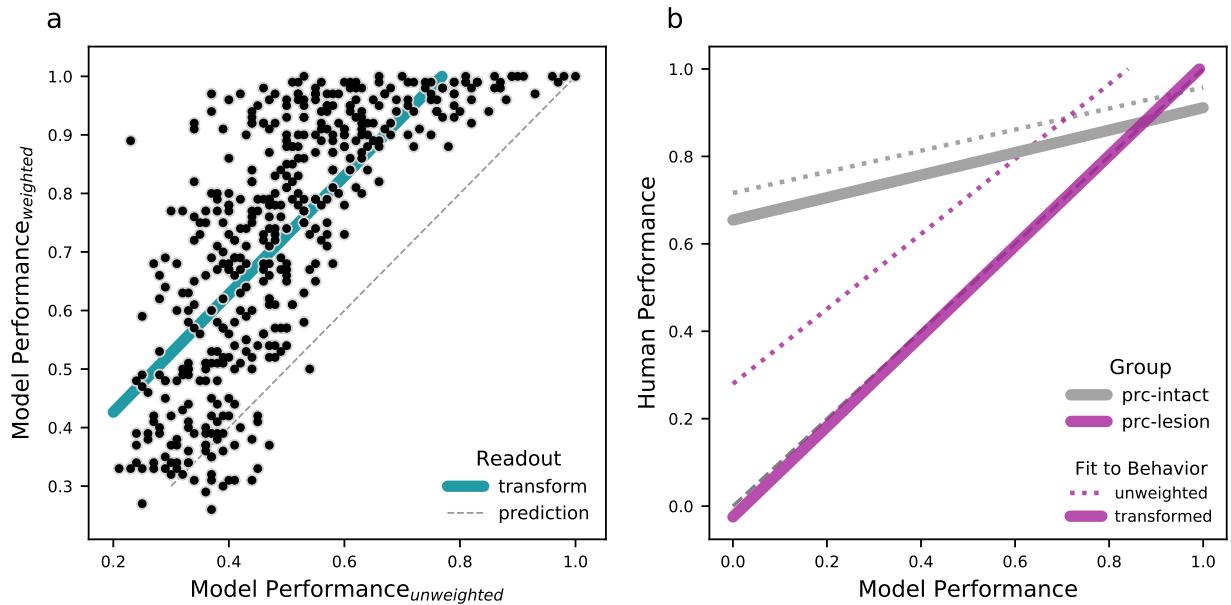
**Figure 7: A computational proxy for the VVS achieves PRC-intact performance on ‘complex’ stimuli, with training, but fails to generalize.** Faces are an example of putatively ‘complex’ experimental stimuli: VVS-like models, PRC-lesioned participants, and IT-supported performance all fail to approximate PRC-intact behaviors on this stimulus class. We optimize a computational proxy for the VVS to perform these ‘complex’ tasks by changing the distribution of its training data (i.e. using a dataset with millions of faces) and compare its behavior with a model optimized for a more domain general task of object categorization. **(a)** This content-specific optimization leads to increased model performance on ‘within distribution’ stimuli in the retrospective dataset, while not generalizing to out-of-distribution stimuli; models optimized for face discrimination perform face-oddity tasks better than models optimized to perform object classification (red, left), while models optimized for objects classification better perform object-oddity tasks (grey, left). Comparing to human performance on faces in the retrospective dataset, object-trained models are significantly outperformed by PRC-intact participants (top right), while face-trained models exhibit performance that is not significantly different from PRC-intact behavior (bottom right). This pattern of results suggests that performance gains scale with the relative similarity of testing and training data, not stimulus properties, per se. **(b)** We replicate findings from the retrospective analysis using the novel, model-driven experimental stimuli: Models optimized for ‘complex’ visual content significantly outperform other models on ‘within distribution’ stimuli in the novel experiment (red, left), while exhibiting degraded performance on out-of-distribution stimuli (grey, left). Comparing to human performance on faces in the novel experiment, while object-trained models are significantly outperformed by PRC-intact participants (top right), face-trained models exhibit performance that is not significantly different from PRC-intact behavior (bottom right). **(c)** Model’s optimized for object classification recapitulate the performance supported by IT (top left) and reaction time of PRC-intact human subjects (top right). In contrast, this content-specific optimization breaks the correspondence between the model and IT-supported behavior (bottom left) and reaction time (bottom right); while this optimization procedure leads to performance comparable to PRC-intact behavior on the trained stimulus type, these models should be considered to offer a solution to these tasks unlike PRC-dependent computations. Together, these results demonstrate that a content-specific optimization procedure enables VVS-like architectures to discriminate between ‘complex’ stimuli, reflecting numerous findings from perceptual learning in the biological system. These computational results suggests that PRC-dependence on ‘complex’ stimuli is not about stimulus properties, per se, but the interaction between stimulus properties and stimulus-relevant experience.



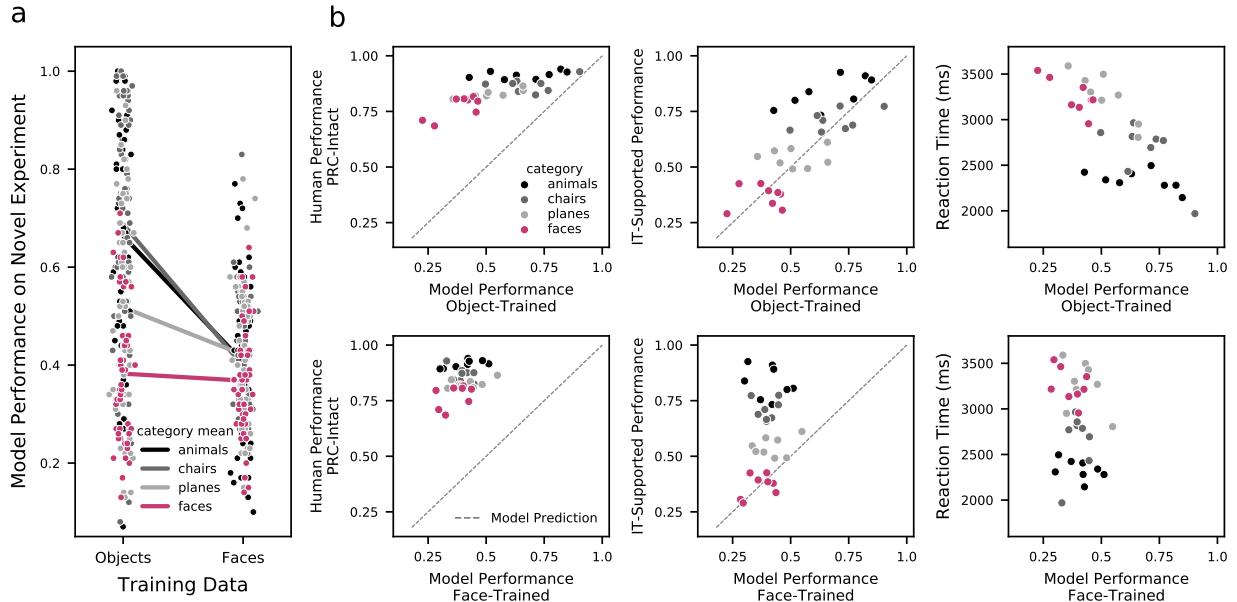
**Supplementary Figure S1: Experimental protocol for retrospective analyses.** (a) Each trial consists of a stimulus screen containing N objects. (b) These N objects are segmented into N object-centered images. (c) We pass these N object-centered images to the model, independently. (d) Using an “IT like” layer of the model, we extract model responses to the N objects, which are flattened into length F vectors, resulting in an  $F \times N$  response matrix for each trial. (e) To identify the item-by-item similarity between objects, we use the Pearson’s correlation between items in this  $F \times N$  response matrix, generating an  $N \times N$  correlation matrix. (f) We average over each item’s off-diagonal correlations, generating a single vector that corresponds to each item’s correlation with all other items. (g) We select the item with the lowest value as the model-selected oddity (e.g. bottom, the item least like the others). This model-selected oddity is labeled as either correct or incorrect, depending on its correspondence with ground truth.



**Supplementary Figure S2: Stimulus sets appear to have been misclassified in the retrospective dataset, on both sides of the perceptual-mnemonic debate.** While the original authors described these experiments as ‘complex,’ we find that they are perfectly computable by a computational proxy for the VVS (i.e. accuracy = 100%). Below ceiling human performance on these experiments can be attributed to extra-perceptual task demands (e.g. memory), and so these experiments are not able to adjudicate PRC-involvement in perception. We separate these misclassified experiments into two categories. (a-b) There are eight experiments across two studies that were argued as evidence against perirhinal involvement in perception because performance did not significantly differ between PRC-intact and -lesioned participants. For these experiments, modeling results suggest that canonical VVS regions should be sufficient to meet the perceptual demands in these tasks, and thus the observed matched performance is expected. (c-d) There were six experiments that were argued to reveal evidence in support of perirhinal involvement in perception. While the authors argued that the observed deficits in PRC-lesioned participants are due to the perceptual demands imposed by these stimulus sets, the model revealed that they are perfectly computable by a computational proxy for the VVS, and so these deficits can be attributed to extra-perceptual task demands. Data in a-d are presented in Fig. 2 at  $x=1$ .



**Figure S3: Learning a weighted, linear readout of model features on trial-by-trial concurrent visual discrimination tasks improves the correspondence between model performance and PRC-lesioned behavior.** (a) A novel concurrent visual discrimination stimulus set densely and continuously spans the space of model performance defined using an unweighted linear (i.e. distance-based) readout of model responses (x axis), as per the original retrospective dataset analysis, and a weighted, linear readout of model responses learned through a leave-one-out cross-validation strategy (y axis). As expected, the learned, weighted readout outperforms the distance metric. We learn the transformation that projects the unweighted performance into the performance expected for the same stimuli using a learned, weighted readout (green). (b) Using the transform learned in (a), we project model performance supported by a uniform readout of model responses (i.e. the original retrospective analysis) into the performance that would be expected were it possible to learn a weighted readout on these stimuli. This improves the correspondence between model performance and human performance, motivating the need to use stimuli that enable a learned, weighted readouts of model performance.



**Supplementary Figure S4: Without ‘foveating’ stimuli before being presented to the model, content-specific optimization does not improve performance on ‘complex’ experimental stimuli.** We optimize a computational proxy for the VVS to perform a ‘complex’ visual discrimination task—face identification—through perceptual training. In this approach, the images are presented to the model at central field of view, and encompass much of the available image. This is unlike the (previous) model trained through object classification, which receives images with objects whose locations and viewpoints are highly variable across images. The novel stimulus set, however, contains faces and other objects that are located at random locations across the stimulus screen—unlike the distribution of training data in the face-trained model. **(a)** Model performance<sub>faces</sub> is not better on face discrimination in the novel dataset than model performance<sub>objects</sub>—and it is significantly worse on all other object types. **(b)** While model performance<sub>objects</sub> is outperformed by PRC-intact participants across many items (top left), it nonetheless provides a good fit to IT-supported behavior (top center), and predicts human subject reaction time on these tasks (top right). Conversely, model performance<sub>faces</sub> is outperformed by PRC-intact participants across all items (bottom left), outperformed by IT-supported behavior across many items (bottom center), and demonstrates no correspondence with human reaction time on these tasks (bottom right). This content-specific optimization procedure fails to generalize to images with higher variance in object location, regardless of their stimulus type. These results further corroborating the restricted (i.e. ‘near transfer’) performance enhancements observed with this approach. All data reported in the main results, and in Fig. 7b-c are determined by ‘foveating’ the images in the novel dataset before presenting them to the model, rendering them more similar to the images used during model optimization.