Deep Predictive Learning as a Model of Human Learning

Randall C. O'Reilly[a,1], **Jacob L. Russin**[a], **and John Rohrlich**[a]

[a]Department of Psychology, Computer Science, and Center for Neuroscience, University of California Davis

How does the human brain learn new concepts from raw sensory experience, without explicit instruction? This longstanding mystery remains unsolved, despite recent demonstrations of the impressive learning power of deep convolutional neural networks (DCNN's), which notoriously require explicit training from massive human-labeled datasets. The plausibility of the error backpropagation powering these models has also long been questioned on biological grounds, although various related biologically plausible mechanisms have been proposed. Here, we show that a biologically based form of *predictive* error-driven learning, where error signals arise from differences between a prediction and what actually occurs, learns to systematically categorize 3D objects according to invariant shape properties from raw visual inputs alone. We found that these categories match human judgments on the same stimuli, and are consistent with neural representations in inferotemporal (IT) cortex in primates. Biologically, we propose that distinctive patterns of connectivity between the neocortex and thalamus drive alternating top-down prediction and bottom-up outcome representations over the pulvinar nucleus, at the alpha frequency (10 Hz), with the temporal difference driving error-driven learning throughout neocortex. We show that comparison predictive DCNN models lacking these biological features did not learn object categories that go beyond the visual input structure. Thus, we argue that incorporating these biological properties of the brain can potentially provide a better understanding of human learning at multiple levels relative to existing DCCN models.

Computational Modeling | Predictive Learning | Object Recognition | Pulvinar | Neocortex

The fundamental epistemological conundrum of how knowledge emerges from raw experience has plagued philosophers and scientists for centuries. Computational models with powerful learning mechanisms driven by raw images or other sensory inputs provide an attractive way to approach this problem, yet many of the current models based on deep convolutional neural networks (DCNN's) notoriously require explicit training from massive human-labeled datasets (1–3). Such models are cognitively implausible, as non-human primates and human infants learn to recognize and categorize objects without the benefit of such labeled data (4). Furthermore, the biological plausibility of the core learning mechanism, *error backpropagation* (5), has also long been questioned on biological grounds (6), although various related biologically plausible mechanisms have been proposed (7–9).

Here we propose a form of *predictive* error-driven learning (10, 11) that learns directly on raw sensory inputs without the need for explicit human-generated labels. This learning mechanism leverages distinctive patterns of connectivity between the neocortex and thalamus (12) (Figure 1) to achieve a biologically based form of predictive learning. In contrast to existing predictive learning frameworks (13–16), we suggest that error signals, as differences between a prediction and what actually occurs, remain as a *temporal difference* in activation states in the network, and are not explicitly represented through error-coding neurons. Specifically, the pulvinar nucleus of the thalamus receives both top-down predictions and bottom-up sensory outcome signals, alternating within an *alpha* frequency cycle (10 Hz, 100 msec), via two distinctive pathways. Thus, our framework has many testable differences from these existing theories, and we argue that existing data is more consistent with our framework.

Through large-scale simulations based on the known structure of the visual system, we found that this biologically based predictive learning mechanism developed high-level abstract representations that systematically categorize 3D objects according to invariant shape properties, based on raw visual inputs alone. We found that these categories match human judgments on the same stimuli, and are consistent with neural representations in inferotemporal (IT) cortex in primates (17). Furthermore, we show that comparison predictive DCNN models lacking these biological features (18) did not learn object categories that go beyond the visual input structure. Thus, we argue that incorporating

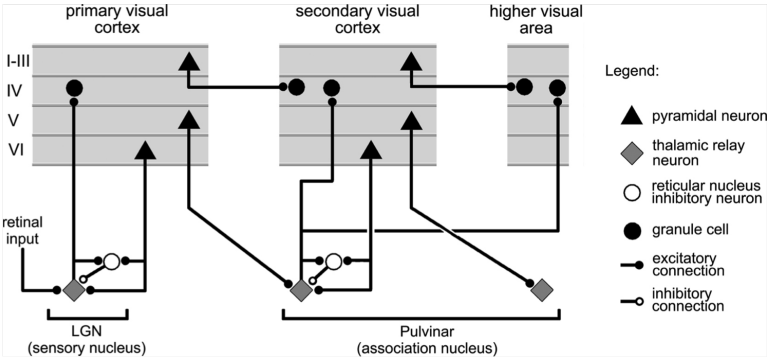

Fig. 1. Summary figure from Sherman & Guillery (2006) showing the strong feedforward driver projection emanating from layer 5IB cells in lower layers (e.g., V1), and the much more numerous feedback "modulatory" projection from layer 6CT cells. We interpret these same connections as providing a prediction (6CT) vs. outcome (5IB) activity pattern over the pulvinar.

these biological properties of the brain can potentially provide a better understanding of human learning at multiple levels relative to existing DCCN models.

Figure 1 shows the thalamocortical circuits characterized by Sherman & Guillery (12) and others, which have two distinct projections converging on the principal thalamic relay cells (TRCs) of the pulvinar (which is interconnected with all higher-level posterior cortical visual areas; (19)). The numerous, weaker projections originating in deep layer VI of the neocortex (the 6CT corticothalamic projecting cells) appear ideal for establishing a top-down prediction state in the pulvinar, based on extensive learning in this pathway and the deep cortical layers that drive it. In contrast, the very sparse (typically one-to-one; (20, 21)) and very strong *driver* inputs originate from lower-level layer V intrinsic bursting cells (5IB), and these can provide a *phasic*, strong bottom-up *ground truth* signal against which the top-down prediction is compared. The 5IB neurons burst at the alpha frequency (22–24), providing a natural timing to the overall predictive learning cycle, consistent with the large and growing literature on alpha properties and effects on perception (25–28).

Based on this and other biological evidence, we hypothesize that this distinctive thalamocortical circuit supports predictive error-driven learning in a way that shapes learning throughout the posterior neocortex (29) (Figure 2a). Specifically, sensory predictions in posterior neocortex are generated roughly every 100 msec at the alpha rhythm, and the pulvinar represents this top-down

Significance Statement

We present a significant advance in understanding how the human brain learns, based on the idea that canonical circuits between the neocortex and thalamus drive alternating phases of prediction and bottom-up outcomes, and the resulting prediction errors (as differences in activation states over time) can drive powerful learning. Critically, we show for the first time that learning based solely on predicting raw visual inputs can generate higher-level abstract categorical representations of 3D objects, which previously has required explicit human-labeled training. This captures the seemingly magic way in which human learning can create knowledge out of raw experience, without explicit teaching.

RCO developed the model, performed the non-PredNet simulations, and drafted the paper. JLR performed the PredNet simulations and analysis, and edited the paper. JR contributed to developing the model and edited the paper.

R. C. O'Reilly is Chief Scientist at eCortex, Inc., which may derive indirect benefit from the work presented here.

[1]To whom correspondence should be addressed. E-mail: oreilly@ucdavis.edu

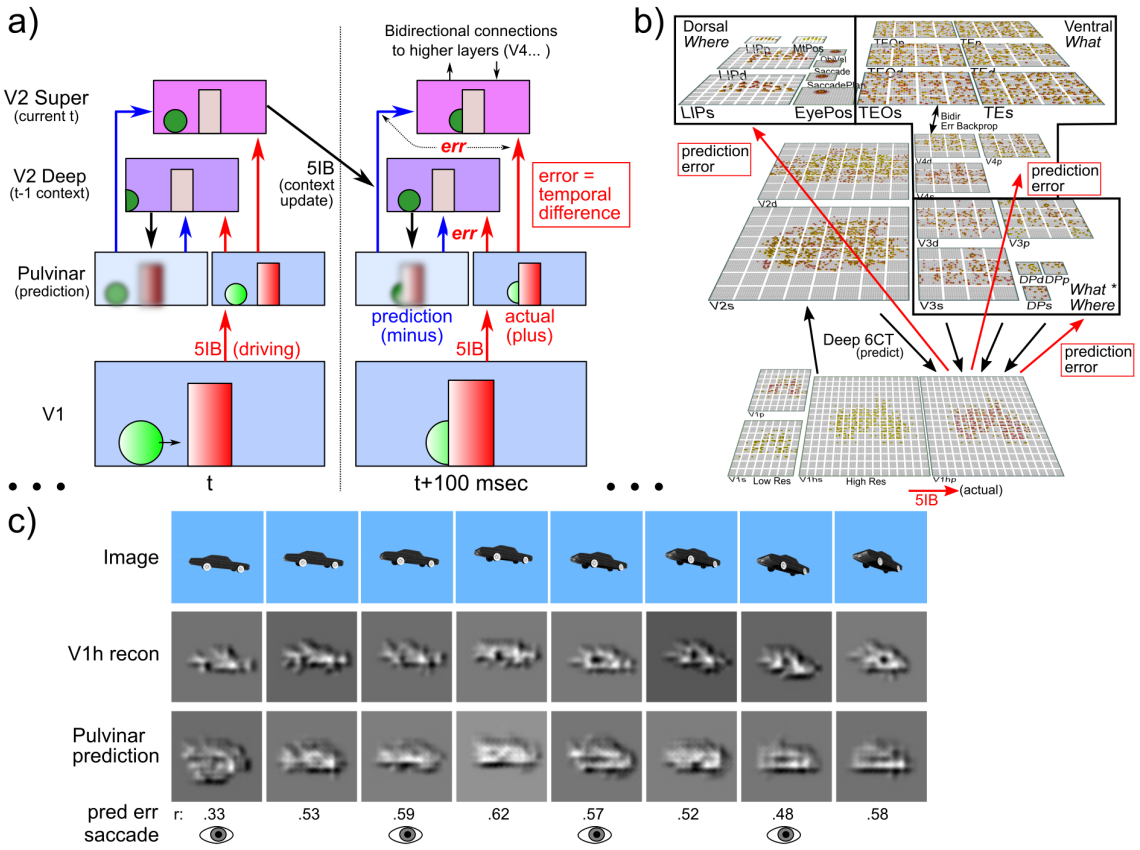

Fig. 2. **a)** Temporal evolution of information flow in the DeepLeabra algorithm predicting visual sequences, over two alpha cycles of 100 msec each. In each alpha cycle, the V2 Deep layer (lamina 5, 6) uses the prior 100 msec of context to generate a prediction (*minus* phase) on the pulvinar thalamic relay cells (TRC). The bottom-up outcome is driven by V1 5IB strong driver inputs (*plus* phase); error-driven learning occurs as a function of the *temporal difference* between these phases, in both superficial (lamina 2, 3) and deep layers, sent via broad pulvinar projections. 5IB bursting in V2 drives update of temporal context in V2 Deep layers, and also the plus phase in higher area TRC, to drive higher-level predictive learning. See supporting information (SI) for more details. **b)** The *What-Where-Integration, WWI* model. The dorsal *Where* pathway learns first, using easily-abstracted *spatial blobs*, to predict object location based on prior motion, visual motion, and saccade efferent copy signals. This drives strong top-down inputs to lower areas with accurate spatial predictions, leaving the *residual* error concentrated on *What* and *What * Where* integration. The V3 and DP (dorsal prelunate) constitute the *What * Where* integration pathway, binding features and locations. V4, TEO, and TE are the *What* pathway, learning abstracted object category representations, which also drive strong top-down inputs to lower areas. *s* suffix = superficial, *d* = deep, *p* = pulvinar. **c)** Example sequence of 8 alpha cycles that the model learned to predict, with the reconstruction of each image based on the V1 gabor filters (*V1 recon*), and model-generated prediction (correlation r prediction error shown). The low resolution and reconstruction distortion impair visual assessment, but r values are well above the r's for each V1 state compared to the previous time step (mean = .38, min of .16 on frame 4 – see SI for more analysis). Eye icons indicate when a saccade occurred.

prediction for roughly 75 msec of the alpha cycle as it develops, after which point the layer 5IB intrinsic-bursting neurons send strong, bottom-up driving input to the pulvinar, representing the actual sensory stimulus. Critically, the prediction error is implicit in the temporal difference between these two periods of activity within the alpha cycle over the pulvinar, which is consistent with the biologically plausible form of error-driven cortical learning used in our models (7). The pulvinar sends broad projections back up to all of the areas that drive top-down predictions into it (19, 30), thus broadcasting this error signal to drive local synaptic plasticity in the neocortex. This mathematically approximates gradient descent to minimize overall prediction errors (7). This computational framework makes sense of otherwise puzzling anatomical and physiological properties of the cortical and thalamic networks (12), and is consistent with a wide range of detailed neural and behavioral data (29).

A critical question for predictive learning is whether it can develop high-level, abstract ways of representing the raw sensory inputs, while learning from nothing but predicting these low-level visual inputs. For instance, can predictive learning really eliminate the need for human-labeled image datasets where abstract category information is explicitly used to train object recognition models via error-backpropagation? Existing predictive-learning models based on error backpropagation (18) have not demonstrated the development of abstract, categorical representations. Previous work has shown that predictive learning can be a useful method for pretraining networks that are subsequently trained using human-generated labels, but here we focus on the formation of systematic categories *de-novo*.

To determine if our biologically based predictive learning model (Figure 2b) can naturally form such categorical encodings in the complete absence of external category labels, we showed the model brief movies of 156 3D object exemplars drawn from 20 different basic-level categories (e.g., car, stapler, table lamp, traffic cone, etc.) selected from the CU3D-100 dataset (31). The objects moved and rotated in 3D space over 8 movie frames, where each frame was sampled at the alpha frequency (Figure 2c). There were also saccadic eye movements every other frame, introducing an additional predictive-learning challenge. An efferent copy signal enabled full prediction of the effects of the eye movement, and allows the model to capture *predictive remapping* (a widely-studied signature of predictive learning in the brain) (32, 33), and introduces additional predictive-learning challenge. The only learning signal available to the model was a prediction error generated by the temporal difference between what it predicted to see in the next frame and what was actually seen.

We performed a representational similarity analysis (RSA) on the learned activity patterns at each layer in the model, and found that the highest IT layer (TE) produced a systematic organization of the 156 3D objects into 5 categories (Figure 3a), which visually correspond to the overall shape of the objects (pyramid-shaped, vertically-elongated, round, boxy / square, and horizontally-elongated). This organization of the objects matches that produced by humans making shape similarity judgments on the same set of objects, using the V1 reconstruction as shown in Figure 2c to capture the model's coarse-grained perception (Figure 3b; see supporting information for methods and further analysis). Critically, Figure 3c shows that the overall similarity structure present in IT layers (TEO, TE) of the biological model is significantly different from the similarity structure at the level of the V1 primary visual input. Thus the model, despite being trained only to generate accurate visual input-level predictions, has learned to represent these objects in an abstract way that goes beyond the raw input-level information. Furthermore, this abstract category organization reflects the overall visual shapes of the objects as judged by human participants, suggesting that the model is extracting geometrical shape information that is invariant to the differences in motion, rotation, and scaling that are present in the V1 visual inputs. We further verified that at the highest IT levels in the

Fig. 3. **a)** Category similarity structure that developed in the highest layer, TE, of the biologically based predictive learning model, showing *1-correlation* similarity of the TE representation for each 3D object against every other 3D object (156 total objects). Blue cells have high similarity, and model has learned block-diagonal clusters or categories of high-similarity groupings, contrasted against dissimilar off-diagonal other categories. Clustering maximized average *within - between* correlation distance (see SI). All items from the same basic-level object categories (N=20) are reliably subsumed within learned categories. **b)** Human similarity ratings for the same 3D objects, presented with the V1 reconstruction (see Fig 1c) to capture coarse perception in model, aggregated by 20 basic-level categories. Each cell is 1 - proportion of time given object pair was rated more similar than another pair (see SI). The human matrix shares the same centroid categorical structure as the model (confirmed by permutation testing and agglomorative cluster analysis, see SI). **c)** Emergence of abstract category structure over the hierarchy of layers. Red line = correlation similarity between the TE similarity matrix (shown in panel a) and all layers; black line shows correlation similarity between V1 against all layers (1 = identical; 0 = orthogonal). Both show that IT layers (TEO, TE) progressively differentiate from raw input similarity structure present in V1, and, critically, that the model has learned structure beyond that present in the input.

model, a consistent, spatially-invariant representation is present across different views of the same object (e.g., the average correlation across frames within an object was .901). This is also evident in Figure 3a by virtue of the close similarity across multiple objects within the same category.

Further evidence for the progressive nature of representation development in our model is shown in Figure 4, which compares the similarity structures in layers V4 and IT in macaque monkeys (17) with those in corresponding layers in our model. In both the monkeys and our model, the higher IT layer builds upon and clarifies the noisier structure that is emerging in the earlier V4 layer. Considerable other work has also compared DCNN representations with these same data from monkeys (17), but it is essential to appreciate that those DCNN models were explicitly trained on the category labels, making it somewhat less than surprising that such categorical representations developed. By contrast, we reiterate that our model has discovered its categorical representations entirely on its own, with no explicit categorical inputs or training of any kind.

Figure 5 shows the results from a purely backpropagation-based (Bp) version of the same model architecture, and a standard PredNet model (18) with extensive hyperparameter optimization (see SI). In the Bp model, the highest layers in the network form a simple binary category structure overall, and the detailed item-level similarity structure does not diverge significantly from that present at the lowest V1 inputs, indicating that it has not formed novel systematic structured representations, in contrast to those formed in the biologically based model. Similar results were found in the PredNet model, where the highest layer representations remained very close to the V1 input structure. Thus, it is clear that the additional biologically derived properties are playing a critical role in the development of abstract categorical representations that go beyond the raw visual inputs. These properties include: excitatory bidirectional connections, inhibitory competition, and an additional Hebbian form of learning that serves as a regularizer (similar to weight decay) on top of predictive error-driven learning (34, 35).

Each of these properties could promote the formation of categorical representations. Bidirectional connections enable top-down signals to consistently shape lower-level representations, creating significant attractor dynamics that cause the entire network to settle into discrete categorical attractor states. By contrast, backpropagation networks typically lack these kinds of attractor dynamics, and this could contribute significantly to their relative lack of categorical learning.

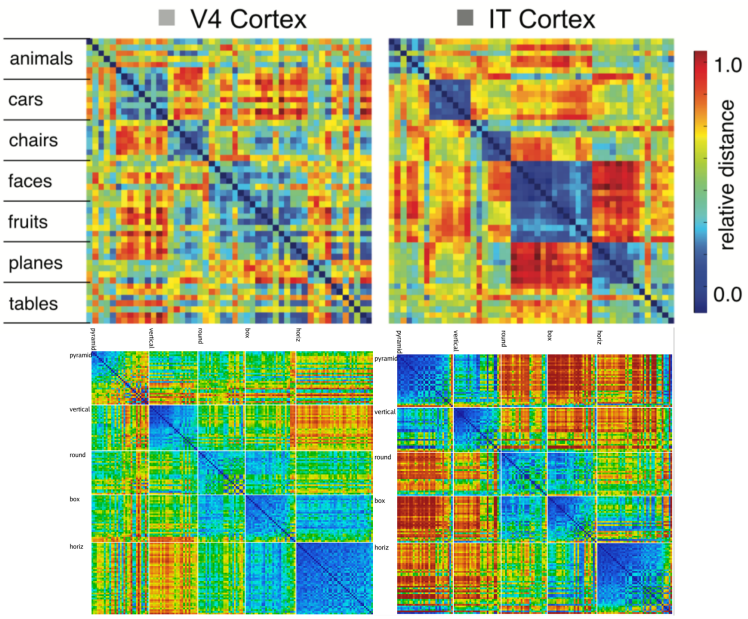

Fig. 4. Comparison of progression from V4 to IT in macaque monkey visual cortex (top row, from Cadieu et al., 2014) versus same progression in model (replotted using comparable color scale). Although the underlying categories are different, and the monkeys have a much richer multi-modal experience of the world to reinforce categories such as foods and faces, the model nevertheless shows a similar qualitative progression of stronger categorical structure in IT, where the block-diagonal highly similar representations are more consistent across categories, and the off-diagonal differences are stronger and more consistent as well (i.e., categories are also more clearly differentiated). Note that the critical difference in our model versus those compared in Cadieu et al. 2014 and related papers is that they explicitly trained their models on category labels, whereas our model is *entirely self-organizing* and has no external categorical training signal.

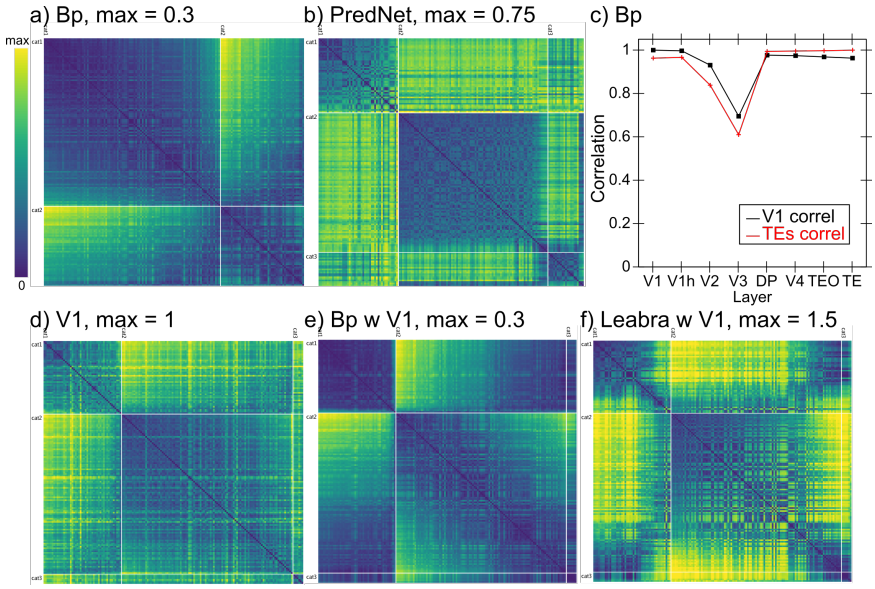

Fig. 5. **a)** Best-fitting category similarity for TE layer of the backpropagation (Bp) model with the same What / Where structure as the biological model. Only two broad categories are evident, and the lower *max* distance (0.3 vs. 1.5 in biological model) means that the patterns are highly similar overall. **b)** Best-fitting similarity structure for the PredNet model, in the highest of its layers (layer 6), which is more differentiated than Bp (max = 0.75) but also less cleanly similar within categories (i.e., less solidly blue along the block diagonal), and overall follows a broad category structure similar to V1. **c)** Comparison of similarity structures across layers in the Bp model (compare to Figure 2c): unlike in the biological model, the V1 structure is largely preserved across layers, and is little different from the structure that best fits the TE layer shown in panel **a**, indicating that the model has not developed abstractions beyond the structure present in the visual input. Layer V3 is most directly influenced by spatial prediction errors, so it differs from both in strongly encoding position information. **d)** The best fitting V1 structure, which has 2 broad categories and banana is in a third category by itself. The lack of dark blue on the block diagonal indicates that these categories are relatively weak, and every item is fairly dissimilar from every other. **e)** The same similarities shown in panel **a** for Bp TE also fit reasonably well sorted according to the V1 structure (and they have a similar average within - between contrast differences, of 0.0838 and 0.0513 – see SI for details). **f)** The similarity structure from the biological model resorted in the V1 structure does *not* fit well: the blue is not aligned along the block diagonal, and the yellow is not strictly off-diagonal. This is consistent with the large difference in average contrast distance: 0.5071 for the best categories vs. 0.3070 for the V1 categories.

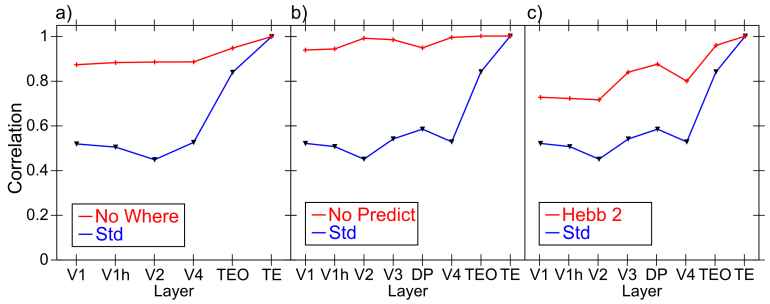

Fig. 6. Effects of various manipulations on the extent to which TE representations differentiate from V1. *Std* is the same result shown in Figure 2c from the intact model for ease of comparison. All of the following manipulations significantly impair the development of abstract TE categorical representations (i.e., TE is more similar V1 and the other layers). **a)** Dorsal *Where* pathway lesions, including lateral inferior parietal sulcus (LIP), V3, and dorsal prelunate (DP). This pathway is essential for regressing out location-based prediction errors, so that the residual errors concentrate feature-encoding errors that train the *What* pathway. **b)** Allowing the deep layers full access to current-time information, thus effectively eliminating the prediction demand and turning the network into an auto-encoder, which significantly impairs representation development, and supports the importance of the challenge of predictive learning for developing deeper, more abstract representations. **c)** Reducing the strength of Hebbian learning by 20% (from 2.5 to 2), demonstrating the essential role played by this form of learning on shaping categorical representations. Eliminating Hebbian learning entirely (not shown) prevented the model from learning anything at all, as it also plays a critical regularization and shaping role on learning.

Hebbian learning drives the formation of representations that encode the principal components of activity correlations over time, which can help more categorical representations coalesce (and results below already indicate its importance). Inhibition, especially in combination with Hebbian learning, drives representations to specialize on more specific subsets of the space. Ongoing work is attempting to determine which of these is essential in this case (perhaps all of them) by systematically introducing some of these properties into the backpropagation model, though this is difficult because full bidirectional recurrent activity propagation, which is essential for conveying error signals top-down in the biological network, is incompatible with the standard efficient form of error backpropagation, and requires much more computationally intensive and unstable forms of fully recurrent backpropagation (36, 37). Furthermore, Hebbian learning requires inhibitory competition which is difficult to incorporate within the backpropagation framework.

Figure 6 shows just a few of the large number of parameter manipulations that have been conducted to develop and test the final architecture. For example, we hypothesized that separating the overall prediction problem between a spatial *Where* vs. non-spatial *What* pathway (38, 39), would strongly benefit the formation of more abstract, categorical object representations in the *What* pathway. Specifically, the *Where* pathway can learn relatively quickly to predict the overall spatial trajectory of the object (and anticipate the effects of saccades), and thus effectively regress out that component of the overall prediction error, leaving the residual error concentrated in object feature information, which can train the ventral *What* pathway to develop abstract visual categories. Figure 6a shows that, indeed, when the *Where* pathway is lesioned, the formation of abstract categorical representations in the intact *What* pathway is significantly impaired. Figure 6b shows that full predictive learning, as compared to just encoding and decoding the current state (which is much easier computationally, and leads to much better overall accuracy), is also critical for the formation of abstract categorical representations — prediction is a "desirable difficulty" (40). Finally, Figure 6c shows the impact of reducing Hebbian learning, which impairs category learning as expected.

In conclusion, we have demonstrated that learning based strictly on predicting what will be seen next is, in conjunction with a number of critical biologically motivated network properties and mechanisms, capable of generating abstract, invariant categorical representations of the overall shapes of objects. The nature of these shape representations closely matches human shape similarity judgments on the same objects. Thus, predictive learning has the potential to go beyond the surface structure of its inputs, and develop systematic, abstract encodings of the "deeper" structure of the

environment. Relative to existing machine-learning-based approaches in "deep learning", which have generally focused on raw categorization accuracy measures using explicit category labels or other human-labeled inputs, the results here suggest that focusing more on the nature of what is learned in the model might provide a valuable alternative approach. Considerable evidence in cognitive neuroscience suggests that the primary function of the many nested ("deep") layers of neural processing in the neocortex is to *simplify* and aggressively *discard* information (41), to produce precisely the kinds of extremely valuable abstractions such as object categories, and, ultimately, symbol-like representations that support high-level cogntive processeses such as reasoning and problem-solving (42, 43). Thus, particularly in the domain of predictive or generative learning, the metric of interest should not be the accuracy of prediction itself (which is indeed notably worse in our biologically based model compared to the DCNN-based PredNet and backpropagation models), but rather whether this learning process results in the formation of simpler, abstract representations of the world that can in turn support higher levels of cognitive function.

Considerable further work remains to be done to more precisely characterize the essential properties of our biologically motivated model necessary to produce this abstract form of learning, and to further explore the full scope of predictive learning across different domains. We strongly suspect that extensive cross-modal predictive learning in real-world environments, including between sensory and motor systems, is a significant factor in infant development and could greatly multiply the opportunities for the formation of higher-order abstract representations that more compactly and systematically capture the structure of the world (44). Future versions of these models could thus potentially provide novel insights into the fundamental question of how deep an understanding a pre-verbal human, or a non-verbal primate, can develop (11, 45), based on predictive learning mechanisms. This would then represent the foundation upon which language and cultural learning builds, to shape the full extent of human intelligence.

ACKNOWLEDGMENTS. We thank Dean Wyatte, Tom Hazy, Seth Herd, Kai Krueger, Tim Curran, David Sheinberg, Lew Harvey, Jessica Mollick, Will Chapman, Helene Devillez, and the rest of the CCN Lab for many helpful comments and suggestions. Supported by: ONR grants ONR N00014-19-1-2684 / N00014-18-1-2116, N00014-14-1-0670 / N00014-16-1-2128, N00014-18-C-2067, N00014-13-1-0067, D00014-12-C-0638. This work utilized the Janus supercomputer, which is supported by the National Science Foundation (award number CNS-0821794) and the University of Colorado Boulder. The Janus supercomputer is a joint effort of the University of Colorado Boulder, the University of Colorado Denver and the National Center for Atmospheric Research. All data and materials will be available at https://github.com/ccnlab/deep-obj-cat upon publication.

1. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet Classification with Deep Convolutional Neural Networks in *Advances in Neural Information Processing Systems 25*, eds. Pereira F, Burges CJC, Bottou L, Weinberger KQ. (Curran Associates, Inc.), pp. 1097–1105.
2. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444.
3. Schmidhuber J (2015) Deep learning in neural networks: An overview. *Neural Networks* 61:85–117.
4. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ (2017/ed) Building machines that learn and think like people. *Behavioral and Brain Sciences* 40.
5. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323(9):533–536.
6. Crick F (1989) The recent excitement about neural networks. *Nature* 337:129–132.
7. O'Reilly RC (1996) Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Computation* 8(5):895–938.
8. Xie X, Seung HS (2003) Equivalence of backpropagation and Contrastive Hebbian Learning in a layered network. *Neural Computation* 15(2):441–454.
9. Bengio Y, Mesnard T, Fischer A, Zhang S, Wu Y (2017) STDP-compatible approximation of backpropagation in an energy-based model. *Neural Computation* 29(3):555–577.
10. Elman JL (1990) Finding Structure In Time. *Cognitive Science* 14(2):179–211.
11. Elman J, et al. (1996) *Rethinking Innateness: A Connectionist Perspective on Development.* (MIT Press, Cambridge, MA).
12. Sherman S, Guillery R (2006) *Exploring the Thalamus and Its Role in Cortical Function.* (MIT Press, Cambridge, MA).
13. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics* 66(3):241–251.
14. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* 2(1):79–87.
15. Kawato M, Hayakawa H, Inui T (1993) A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network: Computation in Neural Systems* 4(4):415–422.
16. Friston K (2005) A theory of cortical responses. *Philosophical Transactions of the Royal Society B* 360(1456):815–836.
17. Cadieu CF, et al. (2014) Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLoS Computational Biology* 10(12):e1003963.
18. Lotter W, Kreiman G, Cox D (2016) Deep predictive coding networks for video prediction and unsupervised learning. *arXiv:1605.08104 [cs, q-bio].*
19. Shipp S (2003) The functional logic of cortico-pulvinar connections. *Philosophical Transactions of the Royal Society of London B* 358(1438):1605–1624.
20. Rockland KS (1998) Convergence and branching patterns of round, type 2 corticopulvinar axons. *The Journal of Comparative Neurology* 390(4):515–536.
21. Rockland KS (1996) Two types of corticopulvinar terminations: Round (type 2) and elongate (type 1). *The Journal of comparative neurology* 368:57–87.

22. Lorincz ML, Kekesi KA, Juhasz G, Crunelli V, Hughes SW (2009) Temporal framing of thalamic relay-mode firing by phasic inhibition during the alpha rhythm. *Neuron* 63(5):683–696.
23. Franceschetti S, et al. (1995) Ionic mechanisms underlying burst firing in pyramidal neurons: Intracellular study in rat sensorimotor cortex. *Brain Research* 696(1–2):127–139.
24. Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337(6095):753–756.
25. Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America* 108(27):11262–11267.
26. VanRullen R, Koch C (2003) Is perception discrete or continuous? *Trends in Cognitive Sciences* 7(5):207–213.
27. Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences* 16(4):200–206.
28. Fiebelkorn IC, Kastner S (2019) A rhythmic theory of attention. *Trends in Cognitive Sciences* 23(2):87–101.
29. O'Reilly RC, Wyatte D, Rohrlich J (2014) Learning Through Time in the Thalamocortical Loops. *arXiv:1407.3432 [q-bio]*.
30. Mumford D (1991) On the computational architecture of the neocortex. *Biological Cybernetics* 65(2):135–145.
31. O'Reilly RC, Wyatte D, Herd S, Mingus B, Jilk DJ (2013) Recurrent Processing during Object Recognition. *Frontiers in Psychology* 4(124).
32. Duhamel JR, Colby CL, Goldberg ME (1992) The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255(5040):90–92.
33. Cavanagh P, Hunt AR, Afraz A, Rolfs M (2010) Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences* 14(4):147–153.
34. O'Reilly RC (1998) Six Principles for Biologically-Based Computational Models of Cortical Cognition. *Trends in Cognitive Sciences* 2(11):455–462.
35. O'Reilly RC, Munakata Y (2000) *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain.* (MIT Press, Cambridge, MA).
36. Williams RJ, Zipser D (1992) Gradient-based learning algorithms for recurrent networks and their computational complexity in *Backpropagation: Theory, Architectures and Applications*, eds. Chauvin Y, Rumelhart DE. (Erlbaum, Hillsdale, NJ).
37. Pineda FJ (1987) Generalization of Backpropagation to Recurrent Neural Networks. *Physical Review Letters* 18:2229–2232.
38. Ungerleider LG, Mishkin M (1982) Two Cortical Visual Systems in *The Analysis of Visual Behavior*, eds. Ingle DJ, Goodale MA, Mansfield RJW. (MIT Press, Cambridge, MA), pp. 549–586.
39. Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. *Trends in Neurosciences* 15(1):20–25.
40. Bjork RA (1994) Memory and metamemory considerations in the training of human beings in *Metacognition: Knowing about Knowing.* (The MIT Press, Cambridge, MA, US), pp. 185–205.
41. Simons DJ, Rensink RA (2005) Change blindness: Past, present, and future. *Trends in cognitive sciences* 9(1):16–20.
42. Rougier NP, Noelle D, Braver TS, Cohen JD, O'Reilly RC (2005) Prefrontal Cortex and the Flexibility of Cognitive Control: Rules Without Symbols. *Proceedings of the National Academy of Sciences* 102(20):7338–7343.
43. O'Reilly RC, et al. (2014) How Limited Systematicity Emerges: A Computational Cognitive Neuroscience Approach in *The Architecture of Cognition: Rethinking Fodor and Pylyshyn*^{1}*s Systematicity Challenge*, eds. Calvo IP, Symons J. (MIT Press, Cambridge, MA).
44. Yu C, Smith LB (2012) Embodied attention and word learning by toddlers. *Cognition* 125(2):244–262.
45. Spelke E, Breinlinger K, Macomber J, Jacobson K (1992) Origins of Knowledge. *Psychological Review* 99(4):605–632.