Predictive Coding Explains Asymmetric Connectivity in the Brain: A Neural Network Study

Romesa Khan[1], Hongsheng Zhong[2], Shuvam Das[2], Jack Cai[3], Matthias Niemeier[1,4,5,*]

[1] Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada, M1C1A4
[2] Department of Computer Science, University of Toronto, Toronto, ON, Canada, M5S1A1
[3] Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada, M5S1A1
[4] Centre for Vision Research, York University, Toronto, ON, Canada, M4N3M6
[5] Vision: Science to Applications, York University, Toronto, ON, Canada, M4N3M6
[*] Correspondence concerning this article should be addressed to Matthias Niemeier at m.niemeier@utoronto.ca

Conflict of interest statement: The authors declare no competing financial interests.

Acknowledgement: This research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

Abstract

Seminal frameworks of predictive coding propose a hierarchy of generative modules, each attempting to infer the neural representation of the module one level below; the predictions are carried by top-down feedback projections, while the predictive error is propagated by reciprocal forward pathways. Such symmetric feedback connections support visual processing of noisy stimuli in computational models. However, neurophysiological studies have yielded evidence of asymmetric cortical feedback connections. We investigated the contribution of neural feedback during sensorimotor processes, in particular visual processing during grasp planning, by utilizing convolutional neural network models that had been augmented with predictive feedback and were trained to compute grasp positions for real-world objects. After establishing an ameliorative effect of symmetric feedback on grasp detection performance when evaluated on noisy stimuli, we characterized the performance effects of asymmetric feedback, similar to that observed in the cortex. Specifically, we tested model variants extended with *short*-, *medium*- and *long*-range feedback connections (i) originating at the same source layer or (ii) terminating at the same target layer. We found that the performance-enhancing effect of predictive coding under adverse conditions was optimal for *medium*-range asymmetric feedback. Moreover, this effect was most prominent when *medium*-range feedback originated at a level of representational abstraction that was proximal to the input layer, in contrast to more distal layers. To conclude, our simulations show that introducing biologically realistic asymmetric predictive feedback improves model robustness to noisy visual stimuli in a neural network model optimized for grasp detection.

Introduction

Descending neural pathways transmit neural signals from higher cortical areas back to earlier processing stages, enabling top-down modulation of neuronal activity in the lower-level circuits. Such feedback connections have been recognized as central for perceptual and cognitive processes, enabling the brain to implement learning mechanisms (Roelfsema & Holtmaat, 2018), attentional modulation (Baluch & Itti, 2011; Desimone & Duncan, 1995) and refinement of sensory representations based on context, expectations, and prior experience (Gilbert & Li, 2013; Kok, Jehee, et al., 2012; Mumford, 1992). Crucially, feedback is believed to be instrumental for conveying predictions.

That is, the brain has been widely conceptualized as a "prediction machine" that relies on internal generative models to actively construct explanations for the causes of noisy sensory inputs (Friston, 2005). More specifically, it is held that predictive feedback signals are conveyed via cortico-cortical top-down projections that are then compared to bottom-up signals such that the deviations from these signals, in the form of predictions errors, are carried forward. For example, an influential framework of such predictive coding has proposed a hierarchy of generative modules, each attempting to infer the neural representation of the module one level below in the hierarchy with topographically reciprocal feedforward connections (Lee & Mumford, 2003; Pennartz et al., 2019; Rao & Ballard, 1999a; Shipp, 2016). This idea maps well on previous anatomical data: Felleman and Van Essen (1991) proposed a hierarchical model of the visual cortex in which adjacent areas often maintain both feedforward and feedback pathways, implying a measure of reciprocity in the network, particularly evident between V1 and V2 (Markov et al., 2011; Rockland & Pandya, 1979). Indeed, additional tracing studies in macaques have revealed that reciprocal links can account for a substantial fraction of interareal

communication, especially between closely related visual areas (Markov et al., 2011). Such symmetric feedback connections have been found to be particularly important for object recognition under noisy or ambiguous viewing conditions. Hupé et al. (1998) have shown that area V2 modulates V1 responses, enhancing the detection of figures embedded in cluttered or noisy images. Furthermore, recurrent interactions between V1 and higher visual areas, including V2, are also crucial for figure–ground segregation (Lamme & Roelfsema, 2000). Similarly, in computer simulations symmetric feedback connections have been shown to support object discrimination (Wen et al., 2018), particularly when processing noisy stimuli (Choksi et al., 2021; Huang et al., 2020).

However, there is growing evidence of asymmetric, non-reciprocated cortical feedback connections from quantitative mapping studies (Markov et al., 2014; Salin & Bullier, 1995). Long-range descending pathways have been shown to cascade over multiple cortical areas (Kravitz et al., 2013; Rockland & Van Hoesen, 1994), and an advantage of long-latency over short-latency visuomotor feedback has been observed during cortical reward processing (Codol et al., 2023). Furthermore, experiments studying visuospatial attention in primates indicate a functional role of medium-range predictive feedback, extending from area V4 to area V1, in encoding accuracy of input stimuli (Debes & Dragoi, 2023), suggesting a potential benefit of medium-range feedback in signal processing under noise. This is similar to the previously suggested role of medium-range feedback in disambiguating signal from noise during global contour integration (Chen et al., 2014).

To investigate the contribution of asymmetrical neural feedback, here we simulated sensorimotor functions utilizing convolutional neural networks (CNNs) as a modelling framework. CNN architectures, inspired by the organization of the mammalian visual cortex

(Hubel & Wiesel, 1962), have been commonly used for object recognition tasks (Girshick et al., 2014) where the constituent convolutional layers extract simple to complex features in an image in a hierarchical fashion, with advanced layer representations becoming increasingly abstract. Hence, these networks have been posited as suitable models of vision in the primate brain (Guclu & van Gerven, 2015; Yamins et al., 2014).

In detail, in the present study we trained CNNs to process visual images of real-world objects to predict grasp points as required for picking them up with a hand. The images of these objects combined color information with a depth channel that served as a proxy of stereovision (Watt & Bradshaw, 2000). Crucially, rather than solely relying on the feedforward flow of information in the canonical architecture of CNNs, we augmented them with symmetric, or asymmetric generative feedback loops that carried advanced representations from later to earlier layers of the networks, mimicking various aspects of feedback in the biological brain. The aim of the study was to understand (a) if, similar to the suggested role of feedback in visual recognition, an ameliorative effect of feedback can be isolated during action-guided visual tasks when the incoming signal is corrupted with adversarial noise. (b) We explored the layer-dependence of such predictive feedback in our model to shed light on the relative functional contribution of cortico-cortical connectivity of asymmetric *short*-, *medium*- and *long*-range feedback denoting varying levels of representational abstraction. (c) We compared the benefits of this feedback at proximal and distal levels of abstraction along the network backbone. Our data are the first to show that predictive feedback aids sensorimotor computations and that *medium*-range top-down projections at proximal levels mark an optimal trade-off of neural informativeness and abstraction of the feedback source. This intimates a previously uncharted role of intermediate areas within visual cortex.

Methods

Neural network architecture

We first trained two versions of a feedforward model, a short and a long backbone, on the task of grasp point generation. As a second step we then augmented the networks with feedback connections (feedback model) that conveyed predictive signals.

Short feedforward model. In detail, for step 1 we obtained the architecture of the short backbone model by modifying a custom convolutional neural network (CNN) architecture (see Figure 1(a)) into which we inputted photographic images with 4 channels (aside from red, green, and blue there was a depth channel, RGB-D, as an approximation of stereovision, that is known to be relevant for grasp performance (Watt & Bradshaw, 2000b)). The RGB-D images were passed into an input layer with two arms, specialized for receiving RGB and depth channels, respectively. The backbone contained a 'feature compression' module consisting of 4 convolution blocks that were connected through maximum pooling (MaxPool) layers to attain dimensionality reduction and invariant feature extraction (Simonyan & Zisserman, 2014). That is, each convolution block consisted of two or three convolution (Conv) layers, each of which was sandwiched together with a batch normalization (BN) and a rectified linear unit (ReLU) activation layer. Next, we added a 'feature expansion' module, composed of Conv and ConvTranspose2d layers (Paszke et al., 2019) that upscaled the feature maps back to the same spatial dimensions as the input (i.e., the entire network had an autoencoder style architecture). The output layer was a regression head that generated a 6-channel array of pixel-wise maps, each corresponding to one of six grasp parameters required for grasping (grasp centre coordinates (x, y); orientation; grasp opening; gripper size; grasp quality score), similar to (Kumra et al., 2022). The sole purpose of the feature expansion module together with its map-style output was not to

attain biological realism but to facilitate interpolation of the infinite number of correct grasp solutions from the finite set of ground truth labels thereby facilitating training of the feedforward network (e.g., (Kumra et al., 2022)). The network architecture of the short backbone is detailed in Table 1.

Long feedforward model. The long backbone once again had a compression module followed by an expansion module similar to the short backbone, except, here the compression module comprised of a VGG16 model (Simonyan & Zisserman, 2014) with 5 convolution blocks where each convolution block contained two or three Conv layers with ReLU activation layers, followed by a MaxPool layer, and BN layers interspersed between each Conv and ReLU layer. Just like before the model architecture was modified with a two-arm input layer to receive 4-channel RGB-D input images. Also, the classification head of the canonical VGG16 architecture was replaced with the same regression head as the short backbone model to return the required parametric map output (see Table 1 for details).

Table 1. Neural network architecture of the short and long feedforward models. Layer-wise architecture of the feedforward models is described. All layers were implemented in PyTorch. The Conv2d (output channels, kernel size, stride) and ConvTranspose2d (output channels, kernel size, stride) refer to convolutional and deconvolutional layers, respectively. The MaxPool (kernel size, stride) layer implements feature pooling. BatchNorm2d: batch normalization; ReLU: rectifying linear unit; Tanh: hyperbolic tangent; Sigmoid: sigmoid function.

Module		Layers	Short	Long
Input	RGB	`Conv2d (64, 3, 1)`	•	•
	Depth	`Conv2d (64, 3, 1)`	•	•
Feature compression		`BatchNorm2d | ReLU`	•	•
	Conv block 1	`Conv2d (64, 3, 1) | BatchNorm2d | ReLU`	•	•
		`MaxPool (2, 2)`	•	•
	Conv block 2	`Conv2d (128, 3, 1) | BatchNorm2d | ReLU`	•	•
		`Conv2d (128, 3, 1) | BatchNorm2d | ReLU`	•	•
		`MaxPool (2, 2)`	•	•
	Conv block 3	`Conv2d (256, 3, 1) | BatchNorm2d | ReLU`	•	•
		`Conv2d (256, 3, 1) | BatchNorm2d | ReLU`	•	•
		`Conv2d (256, 3, 1) | BatchNorm2d | ReLU`	•	•
		`MaxPool (2, 2)`	•	•
	Conv block 4	`Conv2d (256, 3, 1) | BatchNorm2d | ReLU`	•	
		`Conv2d (256, 3, 1) | BatchNorm2d | ReLU`	•	
		`Conv2d (512, 3, 1) | BatchNorm2d | ReLU`		•
		`Conv2d (512, 3, 1) | BatchNorm2d | ReLU`		•
		`Conv2d (512, 3, 1) | BatchNorm2d | ReLU`		•
		`MaxPool (2, 2)`		•
	Conv block 5	`Conv2d (512, 3, 1) | BatchNorm2d | ReLU`		•
		`Conv2d (512, 3, 1) | BatchNorm2d | ReLU`		•
		`Conv2d (512, 3, 1) | BatchNorm2d | ReLU`		•
Feature expansion		`ConvTranspose2d (256, 3, 2) | ReLU`		•
		`Conv2d (256, 5, 1) | ReLU`		•
		`ConvTranspose2d (128, 3, 2) | ReLU`	•	•
		`Conv2d (128, 5, 1) | ReLU`	•	•
		`ConvTranspose2d (64, 3, 2) | ReLU`	•	•
		`Conv2d (128, 5, 1) | ReLU`	•	•
		`ConvTranspose2d (64, 3, 2) | ReLU`	•	•
Output	Grasp	`ConvTranspose2d (5, 3, 1) | Tanh`	•	•
	Confidence	`ConvTranspose2d (1, 3, 1) | Sigmoid`	•	•

Predictive coding dynamics for feedback models

After training the feedforward backbone models, we augmented them with generative feedback connections adapting the PyTorch *Predify* library (Choksi et al., 2021) with some modifications that we made to the original code to introduce custom feedback connectivity. The library is designed to add predictive coding dynamics to existing deep neural networks. The predictive coding framework posits that the brain maintains an internal model of the world to actively predict sensory inputs (Rao & Ballard, 1999b). In this hierarchical model, higher areas generate predictions for lower areas, and discrepancies between predicted and actual inputs (prediction errors) are used to update and refine the higher-layer representations. This iterative process allows the network to minimize prediction errors and enhance perception of sensory information. For example, CNNs have been recently combined with feedback mechanisms for robust object perception (Choksi et al., 2021; Huang et al., 2020).

In the present study, we selected N encoding modules, e_n ($n \in \{1,2,\dots,N\}$ with $N \in \mathbb{Z}^+$) from the network backbone of the feedforward model and added N corresponding decoding modules, d_n. An encoding module, e_n, and the corresponding decoding module, d_n, collectively constituted a *PCoder*. Each e_n (except for e_1) consisted of the MaxPool layer of a given convolution block and the two or three Conv/BN/ReLU layers of the subsequent convolution block. The output of e_n was then passed through a feedback layer (deconvolution layer) d_n that predicted the output of the last Conv/BN/ReLU layer of the previous convolution block (target). To this end, each d_n consisted of a ConvTranspose2d layer that upscaled the input feature map to match the spatial dimensions of the target with feedback weights connecting module $n+1$ to module n being denoted by $W^b_{n+1,n}$. This way, when an input image initially activated all encoding modules through a feedforward pass, over subsequent successive recurrent iterations

(timesteps t), both the decoding and encoding module representations were updated using the following equations:

$$d_n(t) = W^b_{n+1,n} e_{n+1}(t) \tag{1}$$

$$e_n(t+1) = e_n(\beta_n e_{n-1}(t+1) + \lambda_n d_{n-1}(t) + \gamma_n e_n(t) - \alpha_n \nabla \epsilon_{n-1}(t) \tag{2}$$

Here, β_n, λ_n, γ_n (with $\beta_n + \lambda_n + \gamma_n = 1$) and α_n served as balancing coefficients for the feedforward, feedback, recurrence and error-correction terms, respectively. The recurrence term, γ_n, functioned as a memory buffer for retaining the encoding representation at the current timestep. The reconstruction error at module $n-1$, denoted $\epsilon_{n-1}(t)$, was defined as the mean squared error (MSE) between the feedforward representation $e_{n-1}(t)$ and the predicted reconstruction $d_{n-1}(t)$ at that timestep. The feedforward and feedback weights remained frozen across iterations.

Feedback models. Modifications were made to the original code from (Choksi et al., 2021) to introduce custom feedback connectivity to the forward models.

Experiment 1. To test whether predictive coding aids not only object classification (Choksi et al., 2021; Huang et al., 2020) but also visuomotor processes, the network backbone of the short backbone model was augmented with 4 consecutive PCoders, similar to the feedback connectivity pattern attempted in (Choksi et al., 2021). The feedback connectivity for Experiment 1 is shown in Figure 1(a).

Experiment 2. In order to identify the most effective source of feedback, feedback connections were implemented with a consistent feedback target layer and varying feedback source layers across model variants. Specifically, for each feedback type, 2 PCoders were designed. For Experiment 2, e_1 was kept fixed as the first convolution block, across all feedback model variants. Notably, the feedback target layer for e_1 was the input layer of the network; the

input representations were configured to be static and were not updated by iterative recurrent mechanisms. The associated feedback loop was merely an appendage, which had no functional effect but was required in the model architecture for technical reasons. Additionally, the feedback source layer of e_2 did not receive any recurrent feedback. Therefore, all model variants of feedback connectivity effectively contained a single functional loop of feedback, shown as solid red lines in Figure 2, for controlled comparisons between different feedback types.

Short backbone. Three variants of feedback connectivity were compared: *short*-range, *medium*-range and *long*-range feedback, as shown in Figure 2(a). e_1 comprised of the Conv/BN/ReLU layers of convolution block 1 for all model variants. e_2 comprised of the following: the MaxPool layer of block 1 and the Conv/BN/ReLU layers of block 2 (*short*-range); the MaxPool layers of blocks 1 and 2, and the Conv/BN/ReLU layers of blocks 2 and 3 (*medium*-range); the MaxPool layers of blocks 1, 2 and 3, and the Conv/BN/ReLU layers of blocks 2,3 and 4 (*long*-range).

Long backbone. Four variants of feedback connectivity were compared: *short*-range, *medium*-range, *long*-range and *longer*-range feedback, illustrated in Figure 2(a). Similar to the short backbone, e_1 comprised of the Conv/BN/ReLU layers of convolution block 1 for all model variants. e_2 comprised of the following: the MaxPool layer of block 1 and the Conv/BN/ReLU layers of block 2 (*short*-range); the MaxPool layers of blocks 1 and 2, and the Conv/BN/ReLU layers of block 2 and 3 (*medium*-range); the MaxPool layers of blocks 1 ,2 and 3, and the Conv/BN/ReLU layers of block 2, 3 and 4 (*long*-range); the MaxPool layers of blocks 1, 2, 3 and 4, and the Conv/BN/ReLU layers of block 2, 3, 4 and 5 (*longer*-range).

Experiment 3. To further isolate any effects of the feedback target layer, the feedback source layer was kept constant, while varying feedback target layers across model variants.

Again, 2 PCoders were introduced into the feedforward network. Crucially, the feedback connection emanating from e_2 was the only functional feedback loop. For all model variants, e_1 spanned a single Conv layer that immediately precedes e_2.

Short backbone. Three variants of feedback connectivity were compared: *short*-range, *medium*-range and *long*-range feedback. The feedback source layer of e_2 was kept fixed as the final ReLU layer of convolution block 4 for all feedback variants. e_2 comprised of the following: the MaxPool layer of block 3 and the Conv/BN/ReLU layers of block 4 (*short*-range); the MaxPool layers of blocks 2 and 3, and the Conv/BN/ReLU layers of blocks 3 and 4 (*medium*-range); the MaxPool layers of blocks 1,2 and 3, and the Conv/BN/ReLU layers of blocks 2, 3 and 4 (*long*-range). These model variants of feedback connectivity are shown in Figure 3(a).

Long backbone. Four variants of feedback connectivity were compared: *short*-range, *medium*-range, *long*-range and *longer*-range feedback. The feedback source layer of e_2 was maintained as the final ReLU layer of convolution block 5 for all feedback variants. e_2 comprised of the following: the MaxPool layer of block 4 and the Conv/BN/ReLU layers of block 5 (*short*-range); the MaxPool layers of blocks 3 and 4, and the Conv/BN/ReLU layers of blocks 4 and 5 (*medium*-range); the MaxPool layers of blocks 2, 3 and 4, and the Conv/BN/ReLU layers of blocks 3, 4 and 5 (*long*-range); the MaxPool layers of blocks 1, 2, 3 and 4, and the Conv/BN/ReLU layers of blocks 2, 3, 4 and 5 (*longer*-range). Model variants of feedback connectivity are shown in Figure 3(a).

Experiment 4. The effect the abstraction level of the feedback source on model performance was also investigated. *Medium*-range feedback (i.e., the most effective feedback distance in Experiments 2 and 3) was implemented at three different levels of abstraction along the long backbone: *proximal*, *intermediate* and *distal* feedback loops. The naming convention is

in relation to the distance from the input layer of the network. For each level of source abstraction, 2 PCoders were integrated into the network backbone. Like Experiment 2, feedback emerging from e_2 was the only operant feedback connection, whereas the feedback loop across e_1 traversed a single layer and had no predictive function.

For the three model variants, e_2 comprised of the MaxPool layers of blocks 1 and 2, and the Conv/BN/ReLU layers of blocks 2 and 3 (*proximal*); the MaxPool layers of blocks 2 and 3, and the Conv/BN/ReLU layers of blocks 3 and 4 (only the first two Conv/BN/ReLU layers of block 4 were included in order to implement *medium*-range feedback distance) (*intermediate*); the MaxPool layers of blocks 3 and 4, and the Conv/BN/ReLU layers of blocks 4 and 5 (only the first two Conv/BN/ReLU layers of block 5 were included in order to maintain *medium*-range feedback distance) (*distal*). The model variants of medium-range feedback that were implemented are detailed in Figure 4(a).

Model training and hyperparameters

Both, short and long backbone models were initialized with random weights (Xavier initialization) and trained on the task of grasp parameter estimation. Training involved error back-propagation and gradient descent to minimize the mean squared error (MSE) loss for the pixelwise regression output.

After hyperparameter optimization using a grid search strategy, both, the short and long forward models were trained using a Ranger optimizer, with a learning rate of 0.01, weight decay of 0, and a batch size of 75. The short and long backbone models were trained for 17 and 9 epochs, respectively.

The feedback models were created by adding recurrent feedback to the already trained backbone models. The activations of all encoding modules, e_n were initiated with a feedforward

pass. Then the forward weights were frozen, and the weights of the feedback deconvolution layers, d_n, were trained using an unsupervised reconstruction objective. This seeks to minimize the reconstruction loss (prediction error), modelled as MSE between the outputs of e_{n-1} and d_n.

Hyperparameters for predictive coding dynamics, including layer-dependent balancing coefficients for the feedforward (β_n), feedback (λ_n), recurrence (γ_n) and error-correction (α_n) terms, were optimized using a grid search strategy. The respective hyperparameters were conserved across all model variants within each experiment and are presented in Table 2.

Model training was performed on the NVIDIA T4 and the NVIDIA QUADRO RTX 8000 GPUs.

Table 2. Predictive coding hyperparameters. The balancing coefficients for the feedforward (β_n), feedback (λ_n), recurrence (γ_n) and error-correction (α_n) terms that were selected after hyperparameter optimization, are stated for the short and/or long feedforward model(s), for all experiments.

PCoder	**Feedback hyperparameters**	**Experiments**					
		1	2		3		4
		Short	Short	Long	Short	Long	Long
1	β_n	0.2	0.3	0.35	0.35	0.2	0.35
	λ_n	0.05	0.3	0.25	0.25	0.4	0.25
	γ_n	0.75	0.4	0.4	0.4	0.4	0.4
	α_n	0.01	0.01	0.01	0.01	0.01	0.01
2	β_n	0.4	0.4	0.4	0.4	0.4	0.4
	λ_n	0.1	0	0	0	0	0
	γ_n	0.5	0.6	0.6	0.6	0.6	0.6
	α_n	0.01	0.1	0.1	0.1	0.01	0.1
3	β_n	0.4	-	-	-	-	-
	λ_n	0.1	-	-	-	-	-
	γ_n	0.5	-	-	-	-	-
	α_n	0.01	-	-	-	-	-
4	β_n	0.4	-	-	-	-	-
	λ_n	0	-	-	-	-	-
	γ_n	0.6	-	-	-	-	-
	α_n	0.01	-	-	-	-	-

Dataset

Both, the backbone and feedback models, were trained on the grasp detection task using the large-scale Jacquard dataset for robotic grasp detection (Depierre et al., 2018). The Jacquard dataset is composed of 11,619 distinct objects with a total of ~5,000,000 possible grasp annotations. The grasp annotations corresponding to each image constituted a list of grasp candidate labels. Each grasp candidate was described by parameters including center pixel coordinates (x, y), orientation (theta), and dimensions (pixel width, pixel height) of the grasp. For each object instance (a distinct camera viewpoint of an object), a rendered RGB image, a segmentation mask, two depth images and the grasps annotations are available. The dataset was split into three parts: 66% for model training, 17% for model validation, and 17% for model testing.

Preprocessing. The data preprocessing pipeline transformed lists of grasp candidates into pixel-wise grasp maps. A grasp map array, of spatial dimensions corresponding to the image size, was initialized with six channels to represent the five grasp parameters and a count of overlapping grasp bounding boxes for each pixel. The preprocessing operation converted grasp annotations to bounding boxes, identified the bounding box boundaries, and iterated over each pixel within these boundaries to determine the grasp candidate closest to the pixel based on Euclidean distance. Each pixel was then labeled with the attributes of the nearest grasp and the count of valid bounding boxes covering it. The final channel, representing the count of bounding boxes per pixel, was normalized by dividing by the maximum count across all pixels. This preprocessed data format was used for training the backbone and feedback neural network models on the task of predicting grasp locations in an input image.

Feature normalization was applied to all images (RGB and depth) and grasp labels, to promote training stability and model convergence. Image pixel intensities and label values were normalized to a range between 0 and 1.

Additive noise. During model inference, all test images were injected with varying levels of additive Gaussian noise of standard deviation, $\sigma \in \{0,\ 0.25, 0.5, 0.75, 1\}$.

Statistics

A repeated measures analysis of variance (ANOVA) (Girden, 1992) was conducted to assess the effects of feedback type and noise level on the grasp detection accuracy of the model, using the Jamovi statistical software (The jamovi project, 2024). This analysis accounted for within-object effects and included tests for sphericity. Mauchly's test of sphericity (Mauchly, 1940) was performed to check the assumption of sphericity for feedback type, noise level, and their interaction. Where the assumption was violated, Greenhouse-Geisser corrections (Greenhouse & Geisser, 1959) were applied to adjust the degrees of freedom. Post hoc comparisons were conducted using Tukey's HSD test (Tukey, 1949) to evaluate pairwise differences between the levels of feedback type.

A One-Way ANOVA (David C. Howell, 2002) was conducted to compare the difference in model performance between the first and final timestep during the test, across noise different noise levels (The jamovi project, 2024).

Results

We investigated the contribution of feedback to vision for action, such as grasp planning, using a neural network model that was optimized to detect grasp points on objects.

Feedback improves robustness of sensorimotor control to noise

We evaluated the effect of predictive feedback on the performance of the trained model, when presented with adversarial images that were injected with varying levels of Gaussian noise ($\sigma = \{0, 0.25, 0.5, 0.75, 1.0\}$) (Experiment 1 in Methods). In effect, the performance of a feedback model at the first timestep ($t = 1$) of the iterative predictive coding dynamics is identical to the performance of the corresponding feedforward backbone. This backbone performance was compared to the performance of the feedback model after several recurrent timesteps ($t = 15$) after which performance no longer changed. We observed an improvement in the grasp detection accuracy of the model after predictive coding iterations, across all levels of Gaussian noise (Figure 1B; n.b., there was a small decline in performance for zero noise). The average difference in model performance between the first ($t = 0$) and the final ($t = 15$) timestep was significant across noise levels, $F(4, 22267) = 220036.00, p < .001$ (Figure 1C). Figure 1E depicts progressive noise removal in the model output (represented by the grasp quality score map) across successive predictive coding timesteps, for a sample object from the test dataset (Figure 1D).

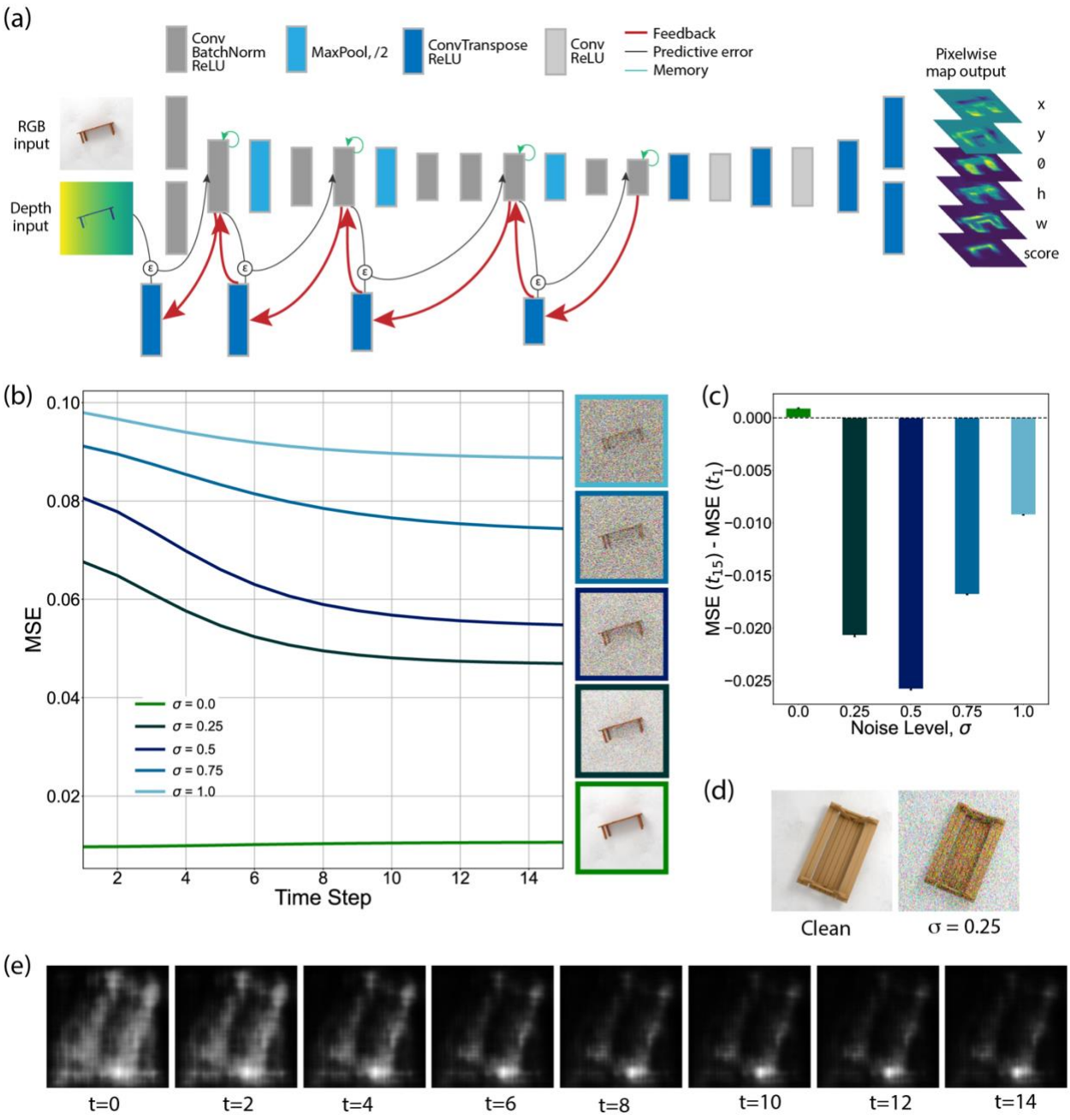


Figure 1. Predictive coding dynamics improve model performance for noisy stimuli. (a) Architecture of the grasping neural network, derived from the *short* feedforward backbone. Red arrows denote top-down feedback connections. Black arrows represent the error-correction process, whereby the predictive error (ε) drives the higher layer (feedback source) representation to better match the early layer representation (feedback target). Green arrows denote recurrent

memory processes, for conserving the layer representation over consecutive predictive cycles. Conv: convolutional layer; BatchNorm: batch normalization layer; ReLU: rectifying linear unit; MaxPool: pooling layer; ConvTranspose: deconvolutional layer. The network received a 3-channel RGB image and a 1-channel depth image of the input object. The network output consisted of a 6-channel array, each channel representing a pixelwise map with the same dimensions as the input images and corresponding to one of the 6 grasp parameters: grasp centre coordinates (x, y); orientation (θ); grasp opening (h); gripper size (w); grasp quality (confidence) score. **(b)** Model performance during the inference phase, across 15 predictive coding timesteps, measured as the average mean squared error at each timestep, when presented with the test dataset that had been injected with varying levels of additive Gaussian noise. The right panel shows a representative sample object, corrupted with the different noise levels. **(c)** The average difference in test performance of the model between the initial (t=1) and final (t=15) timestep, for each noise level. Black bars denote standard errors of mean. **(d)** Left panel: an RGB image of a sample object from the test dataset. Right panel: the RGB image with Gaussian noise ($\sigma = 0.25$). **(e)** The grasp quality score maps of the sample object, at even timesteps.

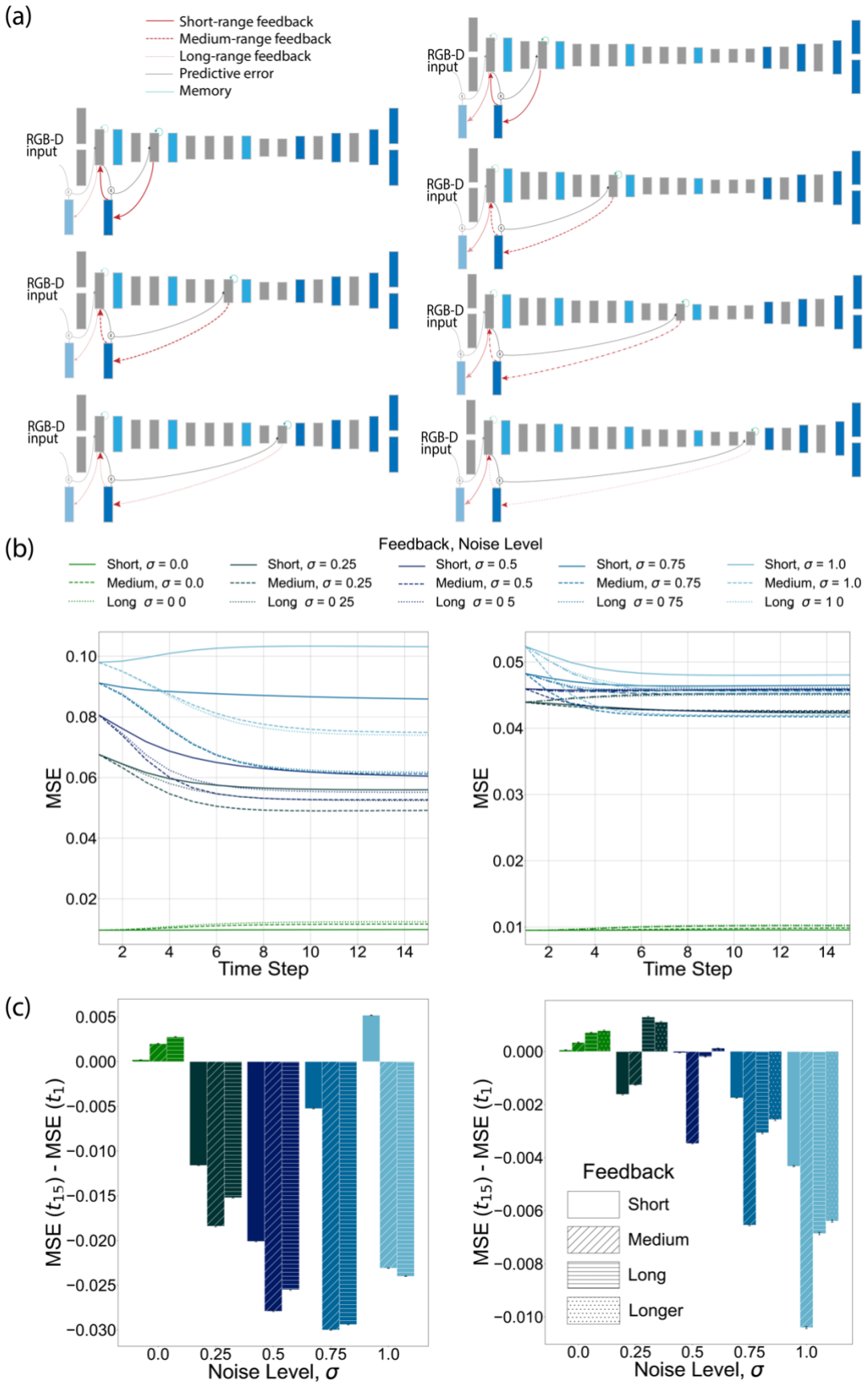
Performance-related effects of feedback are layer-dependent

Next, we tested whether the performance-enhancing effects of feedback were contingent on layer-specific feedback connectivity. To probe this question, we contrasted different patterns of feedback connectivity.

Constant feedback target. First, we set up model variants augmented with a constant feedback target (final ReLU layer of convolution block 1) that received *short*-, *medium*-, or *long*-range feedback, for the short backbone and an additional *longer*-range feedback variant for the long backbone, as detailed in the Methods (see Experiment 2). The average difference in model performance between the first (t=0) and the final (t=15) for all feedback model variants was compared.

For the short backbone, *medium*-range feedback confers the largest advantage on model performance on average under Gaussian noise ($\sigma = \{0, 0.25, 0.5, 0.75\}$), as seen in Figure 2(b,c). A repeated measures ANOVA after Greenhouse-Geisser correction yielded a significant main effect of noise-dependent benefit of feedback on model performance (see Table 3) reflecting that performance declined with increasing noise, as expected. More importantly, there was a significant main effect of feedback distance (see Table 4). Post hoc comparisons using Tukey's HSD test revealed significant differences in model performance between all pairs of feedback types with medium-range feedback being the most effective on average (see Table 6). Additionally, there was a significant interaction between feedback distance and noise level (see Table 5), indicating that most effective feedback distance depended on the noise-dependent benefit (*medium*-range feedback for intermediate levels of noise, *short*-range for zero noise, *long*-range for high levels of noise).

Similarly, in the case of the long backbone, *medium*-range feedback showed on average the greatest robustness to noise ($\sigma = \{0.5, 0.75, 1.0\}$), as shown in Figure 2(b,c). A Greenhouse-Geisser corrected repeated measures ANOVA showed significant main effects of noise level (see Table 3), and feedback distance (see Table 4), where Tukey's HSD test confirmed that *medium*-range feedback was most effective establishing significant differences between various pairs of feedback types (see Table 6). Once again, the interaction between feedback distance and noise-dependent feedback benefit was also significant (see Table 3), indicating that *medium*-range feedback was most effective for intermediate and high levels of noise whereas *short*-range feedback was more effective for zero and low levels of noise.


(a)
Short-range feedback
Medium-range feedback
Long-range feedback
Predictive error
Memory
RGB-D input
(b)
Feedback, Noise Level
Short, σ = 0.0
Medium, σ = 0.0
Long σ = 0 0
Short, σ = 0.25
Medium, σ = 0.25
Long σ = 0 25
Short, σ = 0.5
Medium, σ = 0.5
Long σ = 0 5
Short, σ = 0.75
Medium, σ = 0.75
Long σ = 0 75
Short, σ = 1.0
Medium, σ = 1.0
Long σ = 1 0
MSE
Time Step
(c)
MSE (t15) - MSE (t1)
Noise Level, σ
Feedback
Short
Medium
Long
Longer

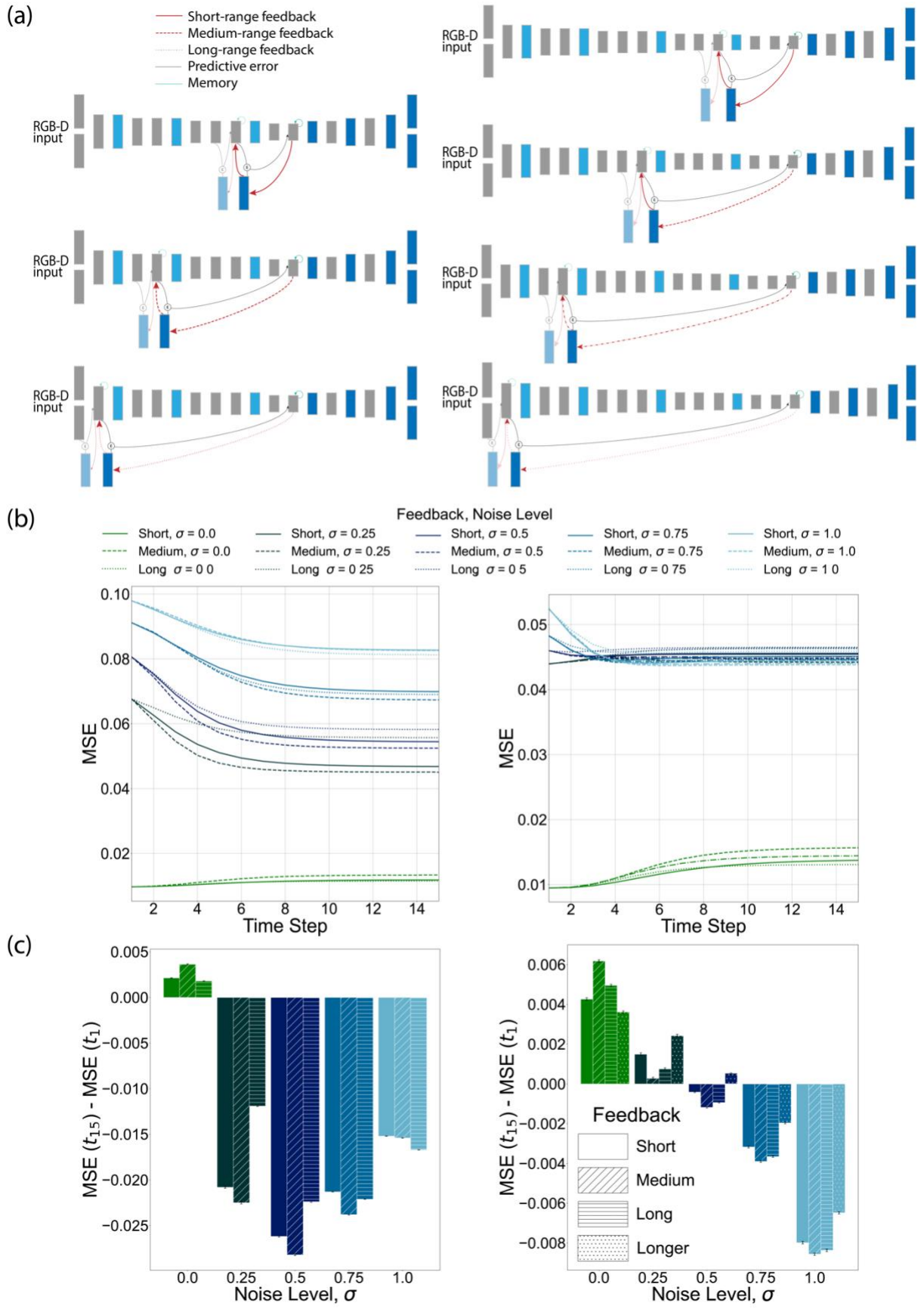
397

Figure 2. Layer-dependent effects of feedback with a uniform target. (a) Left panel: short feedforward model, augmented with *short*-range (first row; solid red arrow), *medium*-range (second row; dashed red arrow) and *long*-range (third row; dotted red arrow) feedback. Right panel: long feedforward model, augmented with *short*-range (first row; solid red arrow), *medium*-range (second row; dashed red arrow), *long*-range (third row; dash-dotted red arrow) and *longer*-range (fourth row; dotted red arrow) feedback. All effective feedback connections, pertaining to the second PCoder in each model variant, terminate at the same target layer: the final ReLU layer of convolution block 1. The first PCoder (translucent) does not represent a functional feedback connection. **(b)** Performance of feedback model variants across 15 predictive coding timesteps, on test dataset injected with varying levels of Gaussian noise (standard deviation, $\sigma = \{0, 0.25, 0.5, 0.75, 1.0\}$). Left panel: short feedforward backbone. Right panel: long feedforward backbone. **(c)** The average difference in test performance of the feedback model variants, between the initial (t=1) and final (t=15) timestep, for each noise level. Black bars denote standard errors of mean.

Constant feedback source. Secondly, we compared model performance under adversarial noise, when the feedback connections originated from the same source layer across all model variants (for details of the model architecture, see Experiment 3 in Methods). Model variants were evaluated after timestep $t = 15$. Once again for the short backbone we found *medium*-range feedback was most ameliorative, when presented with stimuli injected especially with intermediate levels of Gaussian noise ($\sigma = \{0.25, 0.5, 0.75\}$), as shown in Figure 3(b,c). A repeated measures ANOVA with Greenhouse-Geisser correction demonstrated a significant main effect of noise-dependent benefit of feedback (see Table 3). More importantly, we observed a significant main effect of feedback distance on model performance (see Table 4), with *medium*-range feedback yielding the greatest benefit on average. In detail, Tukey's HSD test revealed significant differences between various pairs of feedback types (see Table 6). Finally, there was a significant interaction between feedback distance and noise-dependent benefit (see Table 5), , indicating that *medium*-range feedback was optimal for intermediate levels of noise whereas *long*-range feedback was more ameliorative for high and zero levels of noise.

Similar effects of feedback connectivity on model performance were observed in the case of the long backbone, Figure 3(b,c). The network augmented with *medium*-range feedback was most resilient to all levels of additive noise tested, except for zero noise ($\sigma = \{0.25, 0.5, 0.75, 1.0\}$). A repeated measures ANOVA with Greenhouse-Geisser correction showed significant main effects of noise-dependent feedback benefit (see Table 3), and feedback distance (see Table 4). Post hoc comparisons using Tukey's HSD test revealed significant differences between the various pairs of feedback types (see Table 6). These results are consistent with *medium*-range feedback being most beneficial across noise levels on average. Furthermore, a significant interaction between feedback distance and noise-dependent benefit of feedback (see

Table 6) reflected that *medium*-range feedback was optimal for most noise levels except zero noise.


(a)
Short-range feedback
Medium-range feedback
Long-range feedback
Predictive error
Memory
RGB-D input
(b)
Feedback, Noise Level
Short, σ = 0.0
Medium, σ = 0.0
Long σ = 0 0
Short, σ = 0.25
Medium, σ = 0.25
Long σ = 0 25
Short, σ = 0.5
Medium, σ = 0.5
Long σ = 0 5
Short, σ = 0.75
Medium, σ = 0.75
Long σ = 0 75
Short, σ = 1.0
Medium, σ = 1.0
Long σ = 1 0
MSE
Time Step
(c)
MSE (t15) - MSE (t1)
Noise Level, σ
Feedback
Short
Medium
Long
Longer

440

Figure 3. Layer-dependent effects of feedback with a uniform source. (a) Left panel: short feedforward model, augmented with *short*-range (first row; solid red arrow), *medium*-range (second row; dashed red arrow) and *long*-range (third row; dotted red arrow) feedback. Right panel: long feedforward model, augmented with *short*-range (first row; solid red arrow), *medium*-range (second row; dashed red arrow), *long*-range (third row; dash-dotted red arrow) and *longer*-range (fourth row; dotted red arrow) feedback. All effective feedback connections, pertaining to the second PCoder in each model variant, originate at the same source layer; short feedforward backbone: final ReLU layer of convolution block 4; long feedforward backbone: final ReLU layer of convolution block 5. The first PCoder (translucent) does not represent a functional feedback connection. **(b)** Performance of feedback model variants across 15 predictive coding timesteps, on test dataset injected with varying levels of Gaussian noise (standard deviation, $\sigma = \{0, 0.25, 0.5, 0.75, 1.0\}$). Left panel: short feedforward backbone. Right panel: long feedforward backbone. **(c)** The average difference in test performance of the feedback model variants, between the initial (t=1) and final (t=15) timestep, for each noise level. Black bars denote standard errors of mean.

Proximal neural feedback loops improve robustness to noise

Having established the precedence of medium-range neural feedback loops in two distinct neural network architectures, we investigated if the performance-enhancing effects of such feedback were sensitive to the level of abstraction of the feedback source, i.e., feedback originating from early vs. late areas in the dorsal visual stream. Using the long backbone architecture, we set up three different model variants, augmented with *medium*-range feedback connections, originating at *proximal*, *intermediate* or *distal* layers relative to the input (see Experiment 4 in Methods), which were then presented with input images corrupted with Gaussian noise. Feedback most *proximal* to the input layer, had the greatest performance-enhancing effect, as seen in Figure 4(b). A repeated measures ANOVA with Greenhouse-Geisser correction examined the effects of feedback source abstraction level, as well as noise level on the change in performance of the model between timestep $t = 0$ and timestep $t = 15$. The analysis identified significant main effects of noise-dependent benefit of feedback (see Table 3), as well as feedback abstraction level (Table 4). The latter effect reflected proximal feedback to be optimal as shown with post hoc comparisons using Tukey's HSD test indicating significant differences between various pairs of feedback abstraction levels (see Table 6). Moreover, the interaction between abstraction and noise level was significant (see Table 5) due to the influence of abstraction level being less prominent for zero noise than for other noise levels.

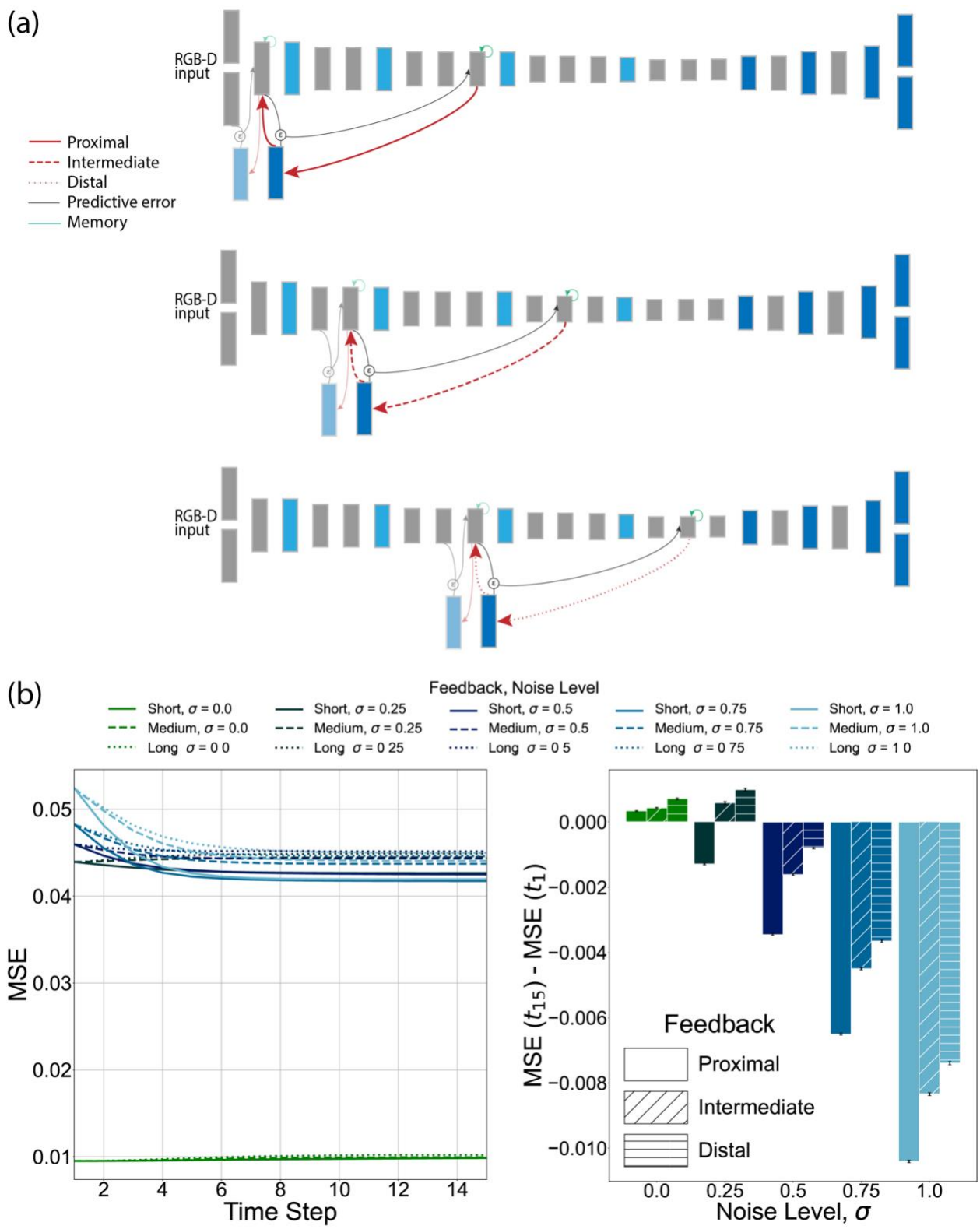


Figure 4. Effect of feedback source abstraction level. (a) Long feedforward model, augmented with *medium*-range feedback connections (pertaining to the second PCoder in each model

variant), originating at layers *proximal* (first row; solid red arrow), *intermediate* (second row; dashed red arrow), or *distal* (third row; dash-dotted red arrow) to the input layer. The first PCoder (translucent) does not represent a functional feedback connection. **(b)** Left panel: performance of feedback model variants across 15 predictive coding timesteps, on test dataset injected with varying levels of Gaussian noise (standard deviation, $\sigma = \{0, 0.25, 0.5, 0.75, 1.0\}$). Right panel: the average difference in test performance of the feedback model variants, between the initial (t=1) and final (t=15) timestep, for each noise level. Black bars denote standard errors of mean.

Table 3. Within Subjects Effects - Noise-dependent benefit

Experiment			df	F	p	η^2
2	Short backbone	Noise-dependent benefit	1.68	109649	< .001	0.454
		Residual	15772.36			
	Long backbone	Noise-dependent benefit	1.64	43410	< .001	0.426
		Residual	15398.52			
3	Short backbone	Noise-dependent benefit	1.56			
		Residual	14671.10	59300	< .001	0.652
	Long backbone	Noise-dependent benefit	2.71	24519	< .001	0.376
		Residual	25388.71			
4	Long backbone	Noise-dependent benefit	2.13	48395	< .001	0.520
		Residual	19937.27			

Table 4. Within Subjects Effects - Feedback distance (abstraction level)

Experiment			df	F	p	η^2
2	Short backbone	Feedback distance	1.18	87764	< .001	0.194
		Residual	11072.25			
	Long backbone	Feedback distance	1.64	10542	< .001	0.080
		Residual	15429.00			
3	Short backbone	Feedback distance	1.10	5365	< .001	0.010
		Residual	10297.82			
	Long backbone	Feedback distance	2.61	2908	< .001	0.004
		Residual	24479.40			

4	Long backbone	Feedback abstraction level	1.35	13523	< .001	0.039
		Residual	12629.91			

Table 5. Within Subjects Effects - Feedback distance (abstraction level) x Noise-dependent benefit.

	Experiment		df	F	p	η^2
2	Short backbone	Feedback distance x Noise-dependent benefit	2.79	60426	< .001	0.176
		Residual	26216.36			
	Long backbone	Feedback distance x Noise-dependent benefit	4.94	5272	< .001	0.057
		Residual	46387.80			
3	Short backbone	Feedback distance x Noise-dependent benefit	1.94	13862	< .001	0.030
		Residual	18161.08			
	Long backbone	Feedback distance x Noise-dependent benefit	7.35	1884	< .001	0.009
		Residual	68946.23			
4	Long backbone	Feedback abstraction level x Noise-dependent benefit	5.61	1290	< .001	0.008
		Residual	52603.38			

Table 6. Post Hoc Comparisons - Feedback distance (abstraction level)

Experiment		Comparison		Mean Difference	SE	df	t	p_{tukey}
		Feedback distance (Exp 2-3); Feedback source abstraction (Exp 4)	Feedback distance (Exp 2-3); Feedback source abstraction (Exp 4)					
2	Short backbone	1	2	0.01316	4.05e-5	9381	324.5	< .001
			3	0.01195	4.21e-5	9381	284.1	< .001
		2	3	-0.00121	1.43e-5	9381	-85.0	< .001
	Long backbone	1	2	0.00274	1.97e-5	9381	138.88	< .001
			3	8.92e-5	2.56e-5	9381	3.48	0.003
			4	-1.42e−4	2.42e-5	9381	-5.87	< .001
		2	3	-0.00265	1.57e-5	9381	-168.80	< .001
			4	-0.00288	1.56e-5	9381	-185.02	< .001
		3	4	-2.31e−4	7.16e-6	9381	-32.33	< .001
3	Short backbone	1	2	9.75e-4	1.77e-5	9381	55.3	< .001
			3	-0.00200	2.55e-5	9381	-78.4	< .001
		2	3	-0.00297	4.01e-5	9381	-74.1	< .001
	Long backbone	1	2	2.80e-4	1.29e-5	9381	21.81	< .001
			3	2.91e-4	1.38e-5	9381	21.03	< .001
			4	-7.83e−4	1.62e-5	9381	-48.41	< .001
		2	3	1.06e-5	1.06e-5	9381	1.01	0.745
			4	-0.00106	1.32e-5	9381	-80.30	< .001
		3	4	-0.00107	1.21e-5	9381	-88.67	< .001
4	Long backbone	1	2	-0.00157	1.54e-5	9381	-101.8	< .001
			3	-0.00223	1.68e-5	9381	-132.5	< .001
		2	3	-6.59e−4	7.83e-6	9381	-84.2	< .001

Discussion

In the present study, we investigated the contribution of neural feedback during sensorimotor function, in particular visual processing during grasp planning, by utilizing convolutional neural network models that had been augmented with predictive feedback and were trained to compute grasp positions for real-world objects. After establishing an ameliorative effect of symmetric feedback on grasp detection from noisy images, we characterized the performance effects of asymmetric feedback, similar to that observed in the cortex. We found that the performance-enhancing effect of predictive coding under adverse conditions was optimal, across a range of noise levels, for *medium*-range asymmetric feedback. Moreover, this effect was most prominent when *medium*-range feedback originated at a level of representational abstraction that was proximal to the input layer, in contrast to more distal layers.

Prior computational work has demonstrated that symmetric feedback, originating from the immediate next higher area in the cortical hierarchy, helps suppress irrelevant noise to enhance performance in an object classification task (Choksi et al., 2021; Huang et al., 2020). Importantly, our work extends the role of predictive feedback in noise-cleanup to the more general function of sensorimotor processing. This involves progressively denoising the feature representations during the test phase, essentially by projecting the representations onto the learned manifolds in the representational space of the trained model. Consistent with this idea, iterative predictive coding dynamics worsen the model's performance when presented with clean images, possibly because the model, after finding a suitable solution after the first forward pass, converged onto a less optimal average solution within the latent representational space.

However, accumulating neurophysiological evidence has suggested that cortical feedback pathways are not strictly reciprocal (Salin & Bullier, 1995) and may traverse multiple areas at

different latencies (Bullier, 2001; Markov et al., 2014). By systematically comparing different feedback path lengths and levels of representational abstraction, we illustrated that too short a feedback loop may lack the broader contextual information needed to mitigate noise effectively, while longer loops may carry overly abstract predictions. We propose that *medium*-range feedback marks a unique sweet spot, putatively incorporating an optimal neural transmission delay and an informative level of abstraction of the feedback source. Moreover, our results suggest that *medium*-range feedback benefits noise clean-up the most when it originates from layers that are proximal to the input, suggesting that relatively early representations are most effective in removing noise.

Of note is the observed U-shaped effect of noise-dependent performance benefit of predictive feedback for the short backbone experiments (Fig 2C, left; Fig3C, left) vs. a linear decrease observed for the long backbone models (Fig 2C, right; Fig3C, right; Fig 4B, left), across the five discrete levels of additive Gaussian noise $\sigma = \{0, 0.25, 0.5, 0.75, 1.0\}$ that were tested. We speculate that the linear trend seen for the long backbone is the arm of a shifted, wider U-shape reflecting robustness to noise perturbation and recovery for the higher-dimensional (in terms of parametric space) long backbone models. The hypothesized U-shaped may emerge if the long backbone model variants are evaluated on higher levels of additive Gaussian noise e.g. $\sigma = \{1.25, 1.5, 1.75, \dots\}$.

Together our findings suggest a previously uncharted role of intermediate cortical areas along the dorsal stream, e.g., roughly similar to area V6A in the primate brain, as a source of predictive feedback in the brain. V6A is primarily involved in encoding action-guided vision during grasping behaviors (Galletti & Fattori, 2018). It plays a key role in guiding prehension movements (Galletti et al., 2003). Importantly, V6A integrates both visual inputs and

somatosensory feedback from the upper limbs (Gamberini et al., 2011). Although it is unclear to which extent the present neural networks exhibit V6a-like properties, our results nevertheless indicate that V6a might be at a strategically optimal position to exert dynamic top-down influence onto early visual areas during visuomotor processing.

Arguably, the observed beneficial effect of *medium*-range feedback might reflect a limitation of the functional approximator corresponding to the feedback layer (defined as the decoding module of a PCoder). If that is true, reducing the representational capacity (or richness of the feature information e.g. the number of feature channels) of the feedback layer may result in a greater beneficial effect of short-range feedback. On the other hand, if we increase the representational capacity of the feedback layer we might see an optimal effect of long-range feedback. Hence, the Goldilocks effect of *medium* feedback might depend on the representational capacity, and hence reconstruction capacity, of the feedback layer. Future research could examine this hypothesis by deepening or diminishing the feedback layer and observing the ensuing effects on model performance.

More research is also required to further investigate our observation that feedback arising from proximal layers of the network is more beneficial than feedback coming from distal layers. Interestingly, the opposite pattern of effectiveness was recently reported for simulated feedback emulating attentional effects in a VGG16 network (Lindsay & Miller, 2018). These authors found that attentional effects were most noticeable when injected into more distal layers. This difference cannot be due to structural differences in the feedforward network that was identical to the long backbone in the present study. It is possible that difference arose from differences in tasks. However, Lindsay and Miller (2018) observed distal benefits across various tasks. Instead, the proximal vs. distal differences might, and probably do, come from the disparity in feedback.

Whereas in the present study we implemented predictive feedback signals, Lindsay and Miller (2018) simulated attentional feedback according to the feature similarity gain model (Martinez-Trujillo & Treue, 2004; Treue & Trujillo, 1999) where neuronal activity changes as a function of similarity of a neuron's tuning to an attended feature. Whether this particular kind of attentional modulation or more generally the function of attention as a form of task-relevant resource allocation in contrast to task-independent predictive coding (Kok, Rahnev, et al., 2012) can explain the differences between proximal vs. distal effectiveness should be explored in the future.

In conclusion, our simulations show that introducing biologically realistic asymmetric predictive feedback improves model robustness to noisy visual stimuli in a neural network model optimized for sensorimotor transformations. Our findings highlight the possible functional significance of mid-tier dorsal stream areas (e.g., V6A) that lie between low-level (V1) and high-level (parietal) representations, suggesting these areas might serve as ideal "predictors" for early visuomotor processing.

References

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences*, *34*(4), 210–224. https://doi.org/10.1016/j.tins.2011.02.003

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2–3), 96–107. https://doi.org/10.1016/S0165-0173(01)00085-6

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., & Li, W. (2014). Incremental Integration of Global Contours through Interplay between Visual Cortical Areas. *Neuron*, *82*(3), 682–694. https://doi.org/10.1016/j.neuron.2014.03.023

Choksi, B., Mozafari, M., Biggs O'May, C., Ador, B., Alamia, A., & VanRullen, R. (2021). Predify: Augmenting deep neural networks with brain-inspired predictive coding dynamics. *NeurIPS*.

Codol, O., Kashefi, M., Forgaard, C. J., Galea, J. M., Pruszynski, J. A., & Gribble, P. L. (2023). Sensorimotor feedback loops are selectively sensitive to reward. *ELife*, *12*. https://doi.org/10.7554/eLife.81325

David C. Howell. (2002). *Statistical Methods for Psychology* (5th ed.). Duxbury Press.

Debes, S. R., & Dragoi, V. (2023). Suppressing feedback signals to visual cortex abolishes attentional modulation. *Science*, *379*(6631), 468–473. https://doi.org/10.1126/science.ade1855

Depierre, A., Dellandréa, E., & Chen, L. (2018). Jacquard: A Large Scale Dataset for Robotic Grasp Detection. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3511–3516. https://api.semanticscholar.org/CorpusID:215827111

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Galletti, C., & Fattori, P. (2018). The dorsal visual stream revisited: Stable circuits or dynamic pathways? *Cortex*, *98*, 203–217. https://doi.org/10.1016/j.cortex.2017.01.009

Gamberini, M., Galletti, C., Bosco, A., Breveglieri, R., & Fattori, P. (2011). Is the Medial Posterior Parietal Area V6A a Single Functional Area? *The Journal of Neuroscience*, *31*(13), 5145–5157. https://doi.org/10.1523/JNEUROSCI.5489-10.2011

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Girden, E. R. (1992). ANOVA: Repeated measures. In *ANOVA: Repeated measures.* Sage Publications, Inc.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. https://doi.org/10.1007/BF02289823

Guclu, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Huang, Y., Gornet, J., Dai, S., Yu, Z., Nguyen, T., Tsao, D. Y., & Anandkumar, A. (2020). Neural Networks with Recurrent Generative Feedback. *NeurIPS*.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787. https://doi.org/10.1038/29537

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cerebral Cortex*, *22*(9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kumra, S., Joshi, S., & Sahin, F. (2022a). GR-ConvNet v2: A Real-Time Multi-Grasp Detection Network for Robotic Grasping. *Sensors*, *22*(16), 6208. https://doi.org/10.3390/s22166208

Kumra, S., Joshi, S., & Sahin, F. (2022b). GR-ConvNet v2: A Real-Time Multi-Grasp Detection Network for Robotic Grasping. *Sensors*, *22*(16), 6208. https://doi.org/10.3390/s22166208

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434. https://doi.org/10.1364/JOSAA.20.001434

Lindsay, G. W., & Miller, K. D. (2018). How biological attention mechanisms improve task performance in a large-scale visual system model. *ELife*, *7*. https://doi.org/10.7554/eLife.38105

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huissoud, C., Clavagnier, S., Giroud, P., Sappey-Marinier, D., Barone, P., Dehay, C., Toroczkai, Z., … Kennedy, H. (2014). A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex. *Cerebral Cortex*, *24*(1), 17–36. https://doi.org/10.1093/cercor/bhs270

Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M. A., Giroud, P., Ercsey-Ravasz, M., Pilaz, L. J., Huissoud, C., Barone, P., Dehay, C., Toroczkai, Z., Van Essen, D. C., Kennedy, H., & Knoblauch, K. (2011). Weight Consistency Specifies Regularities of Macaque Cortical Networks. *Cerebral Cortex*, *21*(6), 1254–1272. https://doi.org/10.1093/cercor/bhq201

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-Based Attention Increases the Selectivity of Population Responses in Primate Visual Cortex. *Current Biology*, *14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Mauchly, J. W. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209. https://doi.org/10.1214/aoms/1177731915

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

Pennartz, C. M. A., Dora, S., Muckli, L., & Lorteije, J. A. M. (2019). Towards a Unified View on Pathways and Functions of Neural Recurrent Processing. *Trends in Neurosciences*, *42*(9), 589–603. https://doi.org/10.1016/j.tins.2019.07.005

Rao, R. P. N., & Ballard, D. H. (1999a). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rao, R. P. N., & Ballard, D. H. (1999b). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, *179*(1), 3–20. https://doi.org/10.1016/0006-8993(79)90485-2

Rockland, K. S., & Van Hoesen, G. W. (1994). Direct Temporal-Occipital Feedback Connections to Striate Cortex (V1) in the Macaque Monkey. *Cerebral Cortex*, *4*(3), 300–313. https://doi.org/10.1093/cercor/4.3.300

Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiological Reviews*, *75*(1), 107–154. https://doi.org/10.1152/physrev.1995.75.1.107

Shipp, S. (2016). Neural Elements for Predictive Coding. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01792

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, *abs/1409.1556*. https://api.semanticscholar.org/CorpusID:14124313

The jamovi project. (2024). *jamovi* (2.5). Retrieved from https://www.jamovi.org.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, *5*(2), 99. https://doi.org/10.2307/3001913

Watt, S. J., & Bradshaw, M. F. (2000a). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia*, *38*(11), 1473–1481. https://doi.org/10.1016/S0028-3932(00)00065-8

Watt, S. J., & Bradshaw, M. F. (2000b). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia*, *38*(11), 1473–1481. https://doi.org/10.1016/S0028-3932(00)00065-8

Wen, H., Han, K., Shi, J., Zhang, Y., Culurciello, E., & Liu, Z. (2018). *Deep Predictive Coding Network for Object Recognition*.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Model training was performed on the NVIDIA T4 and the NVIDIA QUADRO RTX 8000 GPUs.

PCoder	**Feedback hyperparameter**	**Experiment**			
		1	**2**	**3**	**4**
1	β_n				
	λ_n				
	γ_n				
	α_n				
2					
	β_n				
	λ_n				
	γ_n				
	α_n				
3	β_n				
	λ_n				
	γ_n				
	α_n				
4	β_n		-	-	-
	λ_n		-	-	-
	γ_n		-	-	-
	α_n		-	-	-

Table 2

Dataset

Both, the backbone and feedback models, were trained on the grasp detection task using the large-scale Jacquard dataset for robotic grasp detection (Depierre et al., 2018)(Depierre et al., 2018)(Depierre et al., 2018)(Depierre et al., 2018). The Jacquard dataset was composed of 11,619 distinct objects with a total of ~5,000,000 possible grasp annotations. The grasp annotations corresponding to each image constituted a list of grasp candidate labels. Each grasp candidate was described by parameters including center pixel coordinates (x, y), orientation (theta), and dimensions (pixel width, pixel height) of the grasp. For each object instance (a distinct camera viewpoint of an object), a rendered RGB image, a segmentation mask, two depth images and the grasps annotations were available. The dataset was split into three parts: 66% for model training, 17% for model validation, and 17% for model testing.

Preprocessing. The data preprocessing pipeline transformed lists of grasp candidates into pixel-wise grasp maps. A grasp map array, of spatial dimensions corresponding to the image size, was initialized with six channels to represent the five grasp parameters and the count of overlapping grasp bounding boxes for each pixel. The preprocessing operation converted grasp annotations to bounding boxes, identified the bounding box boundaries, and iterated over each pixel within these boundaries to determine the grasp candidate closest to the pixel based on Euclidean distance. Each pixel was then labeled with the attributes of the nearest grasp and the count of valid bounding boxes covering it. The final channel, representing the count of bounding boxes per pixel, was normalized by dividing by the maximum count across all pixels. This preprocessed data format was used for training the backbone and feedback neural network models on the task of predicting grasp locations in an input image.

Feature normalization was applied to all images (RGB and depth) and grasp labels, to promote training stability and model convergence. Image pixel intensities and label values were normalized to a range between 0 and 1.

Additive noise. During model inference, all test images were injected with varying levels of additive Gaussian noise of standard deviation, $\sigma \in \{0,\ 0.25, 0.5, 0.75, 1\}$.

Statistics

A repeated measures ANOVA (Girden, 1992)(Girden, 1992)(Girden, 1992)(Girden, 1992) was conducted to assess the effects of feedback type and noise level on the grasp detection accuracy of the model, using the Jamovi statistical software (The jamovi project, 2024)(The jamovi project, 2024)(The jamovi project, 2024)(The jamovi project, 2024). This analysis accounted for within-object effects and included tests for sphericity. Mauchly's test of sphericity (Mauchly, 1940)(Mauchly, 1940)(Mauchly, 1940)(Mauchly, 1940) was performed to check the assumption of sphericity for feedback type, noise level, and their interaction. Where the assumption was violated, Greenhouse-Geisser corrections (Greenhouse & Geisser, 1959)(Greenhouse & Geisser, 1959)(Greenhouse & Geisser, 1959)(Greenhouse & Geisser, 1959) were applied to adjust the degrees of freedom. Post hoc comparisons were conducted using Tukey's HSD test (Tukey, 1949)(Tukey, 1949)(Tukey, 1949)(Tukey, 1949) to evaluate pairwise differences between the levels of feedback type.

Code Availability

All code used for this study is available at https://github.com/khanrom/Feedback_Grasping.

Results

We investigated the contribution of feedback during action-guided vision, such as grasp planning, using a neural network model that is optimized on a grasp-detection task.

Feedback improves robustness to noise

We evaluated the effect of predictive feedback on the performance of the trained model, when presented with adversarial images that were injected with varying levels of Gaussian noise (Experiment 1 in Methods). In effect, the performance of a feedback model at timestep ($t = 0$) of the iterative predictive coding dynamics is identical to the performance of the corresponding feedforward backbone. This backbone performance was compared to the performance of the feedback model after several recurrent timesteps ($t = 11$). We observed an improvement in the grasp detection accuracy of the model after predictive coding iterations, across all levels of Gaussian noise. A repeated measures ANOVA revealed a significant main effect of noise level, $F(2.14, 20081.92) = 220036, p < .001$, after applying the Greenhouse-Geisser correction. Post hoc comparisons using Tukey's HSD test revealed significant differences between various pairs of noise levels. Performance is negligibly worsened by predictive coding in the absence of noise.

Performance-related effects of feedback are layer-dependent

The contribution of feedback in the ventral stream has been shown to be region specific (Debes & Dragoi, 2023b)(Debes & Dragoi, 2023b)(Debes & Dragoi, 2023b)(Debes & Dragoi, 2023b). We investigated if the observed performance-enhancing effects of feedback in the grasping neural network model of the dorsal stream, when presented with noisy stimuli, were contingent on layer specific feedback connectivity. To probe this question, we contrasted different patterns of feedback connectivity.

Constant feedback target. First, we set up model variants augmented with *short*-, *medium*-, and *long*-range feedback, for the short backbone and an additional *longer*-range feedback variant for the long backbone, as detailed in the Methods (see Experiment 2). The

feedback target layer was kept constant across all model variants. Performance of all feedback model variants was compared after predictive coding timestep $t = 15$.

For the short backbone, *medium*-range feedback confers the largest advantage on model performance under Gaussian noise ($\sigma = \{0.5, 0.75, 1.0\}$). A repeated measures ANOVA, with Greenhouse-Geisser corrections applied due to violations of sphericity, indicated a significant main effect of feedback type, $F(1.28,12040.05) = 82765, p < .001$, and a significant main effect of noise level, $F(1.90,17821.85) = 196066, p < .001$, on model performance. Additionally, there was a significant interaction between feedback type and noise level, $F(3.31,31091.79) = 40589, p < .001$. Post hoc comparisons using Tukey's HSD test revealed significant differences in model performance between all pairs of feedback types. Specifically, the mean difference between *short*- and *medium*-range feedback was 0.01013 ($SE = 3.03e-5, t(9381) = 333.9, p < .001$), between *short*- and *long*-range feedback was 0.00913 ($SE = 3.37e-5, t(9381) = 271.1, p < .001$), and between *medium*- and *long*-range feedback was -0.000997 ($SE = 1.43e-5, t(9381) = -69.8, p < .001$).

Similarly, in the case of the long backbone, *medium*-range feedback showed the greatest robustness to noise ($\sigma = \{0.5, 0.75, 1.0\}$). A Greenhouse-Geisser corrected repeated measures ANOVA showed significant main effects of feedback type, $F(1.50,14087.81) = 10471, p < .001$, and noise level, $F(1.27,11891.35) = 51565, p < .001$. The interaction between feedback type and noise level was also significant, $F(4.23,39668.91) = 5130, p < .001$. Post hoc comparisons using Tukey's HSD test established significant differences between various pairs of feedback types. In particular, the mean difference between *short*- and *medium*-range feedback was 0.00273 ($SE = 2.02e-5,\ t(9381) = 135.45,\ p < .001$), between *short*- and *long*-range feedback was $7.68e-5$ ($SE = 2.64e-5, t(9381) = 2.91, p = 0.019$), and between

short- and *longer*-range feedback was $-1.40e-4$ ($SE = 2.50e-5, t(9381) = -5.62, p < .001$). The comparisons between *medium*- and *long*-range feedback, and *medium*- and *longer*-range feedback, also showed significant differences, with mean differences of -0.00266 ($SE = 1.45e-5, t(9381) = -182.83, p < .001$) and -0.00287 ($SE = 1.45e-5, t(9381) = -197.76, p < .001$), respectively. Finally, the difference between *long*- and *longer*-range feedback was $-2.17e-4$ ($SE = 5.92e-6, t(9381) = -36.66, p < .001$).

Constant feedback source. Secondly, we compared model performance under adversarial noise, when the feedback connections originated from the same source layer across all model variants (for details of the model architecture, see Experiment 3 in Methods). Model variants were evaluated after timestep $t = 15$. As observed for the experiments with a constant feedback source, medium-range feedback in a short backbone was most ameliorative after predictive coding iterations, when presented with stimuli injected with varying levels of Gaussian noise ($\sigma = \{0.25, 0.5, 0.75, 1.0\}$). A repeated measures ANOVA demonstrated a significant main effect of feedback type, $F(1.14, 10711.56) = 5175, p < .001$, after applying the Greenhouse-Geisser correction. Similarly, there was a significant main effect of noise level, $F(1.81, 16985.89) = 159095, p < .001$. There was also a significant interaction between feedback type and noise level, $F(2.47, 23139.00) = 11834, p < .001$. Tukey's HSD test revealed significant differences between various pairs of feedback types. Specifically, the mean difference between *short*- and *medium*-range feedback was 0.000955 ($SE = 1.85e-5, t(9381) = 51.5, p < .001$), between *short*- and *long*-range feedback was -0.00203 ($SE = 2.65e-5, t(9381) = -76.6, p < .001$), and between *medium*- and *long*-range feedback was -0.00299 ($SE = 4.06e-5, t(9381) = -73.5, p < .001$).

Similar effects of feedback connectivity on model performance were observed in the case of the long backbone. The network augmented with medium-range feedback was most resilient to all levels of additive noise tested ($\sigma = \{0.25, 0.5, 0.75, 1.0\}$). A repeated measures ANOVA showed significant main effects of feedback type, $F(1.60, 14971.52) = 5694, p < .001$, and noise level, $F(1.09, 10205.98) = 53645, p < .001$, after applying a Greenhouse-Geisser correction. A significant corrected interaction was also found between feedback type and noise level, $F(4.84, 45361.34) = 2379, p < .001$. Post hoc comparisons using Tukey's HSD test revealed significant differences between the various pairs of feedback types. The mean difference between *short*- and *medium*-range feedback was 0.000213 ($SE = 5.11e-6, t(9381) = 41.6, p < .001$), between *short*- and *long*-range feedback was 0.000131 ($SE = 6.55e-6, t(9381) = 19.9, p < .001$), and between *short*- and *longer*-range feedback was -0.000651 ($SE = 1.09e-5, t(9381) = -59.9, p < .001$). Additionally, the difference between *medium*- and *long*-range feedback was -0.0000823 ($SE = 4.15e-6, t(9381) = -19.8, p < .001$), between *medium*- and *longer*-range feedback was -0.000864 ($SE = 8.51e-6, t(9381) = -101.5, p < .001$), and between *long*- and *longer*-range feedback was -0.000782 ($SE = 6.93e-6, t(9381) = -112.9, p < .001$).

Proximal neural feedback loops improve robustness to noise

Having established the precedence of medium-range neural feedback loops in two distinct neural network architectures, we investigated if the performance-enhancing effects of such feedback were sensitive to the level of abstraction of the feedback source, i.e. feedback originating from early vs. late areas in the dorsal visual stream. We set up three different model variants, augmented with *medium*-range feedback connections, originating at *proximal*, *intermediate* or *distal* layers of the long backbone architecture (see Experiment 4 in Methods), which were

presented with input images corrupted with Gaussian noise. *Medium*-range feedback most *proximal* to the input layer, had the greatest performance-enhancing effect. A repeated measures ANOVA examined the effects of feedback source abstraction level, as well as noise level on performance of the long backbone model at timestep $t = 15$. The analysis identified significant main effects of feedback abstraction level, $F(1.19,11138.21) = 16545, p < .001$, as well as noise level, $F(1.18,11057.48) = 49839,\ p < .001$, following the application of the Greenhouse-Geisser correction. Moreover, the interaction between abstraction and noise level was significant $F(3.37,31611.95) = 1994, p < .001$. Post hoc comparisons using Tukey's HSD test indicated significant differences between various pairs of feedback abstraction levels. Specifically, the mean difference between *proximal* and *intermediate* was -0.00159 ($SE\ = 1.43e-5, t(9381) = -111, p < .001$), between *proximal* and *distal* was -0.00226 ($SE\ = 1.59e-5, t(9381) = -142, p < .001$), and between *intermediate* and *distal* was -0.00067 ($SE\ =\ 5.56e-6, t(9381) = -120, p < .001$).

References

Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends in Neurosciences*, *34*(4), 210–224. https://doi.org/10.1016/j.tins.2011.02.003

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2–3), 96–107. https://doi.org/10.1016/S0165-0173(01)00085-6

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., & Li, W. (2014). Incremental Integration of Global Contours through Interplay between Visual Cortical Areas. *Neuron*, *82*(3), 682–694. https://doi.org/10.1016/j.neuron.2014.03.023

Choksi, B., Mozafari, M., Biggs O'May, C., Ador, B., Alamia, A., & VanRullen, R. (2021). Predify: Augmenting deep neural networks with brain-inspired predictive coding dynamics. *NeurIPS*.

Choksi, B., Mozafari, M., O'May, C. B., Ador, B., Alamia, A., & VanRullen, R. (2021). Predify: Augmenting deep neural networks with brain-inspired predictive coding dynamics. *Neural Information Processing Systems*. https://api.semanticscholar.org/CorpusID:235358464

Codol, O., Kashefi, M., Forgaard, C. J., Galea, J. M., Pruszynski, J. A., & Gribble, P. L. (2023). Sensorimotor feedback loops are selectively sensitive to reward. *ELife*, *12*. https://doi.org/10.7554/eLife.81325

David C. Howell. (2002). *Statistical Methods for Psychology* (5th ed.). Duxbury Press.

Debes, S. R., & Dragoi, V. (2023a). Suppressing feedback signals to visual cortex abolishes attentional modulation. *Science*, *379*(6631), 468–473. https://doi.org/10.1126/science.ade1855

Debes, S. R., & Dragoi, V. (2023b). Suppressing feedback signals to visual cortex abolishes attentional modulation. *Science*, *379*(6631), 468–473. https://doi.org/10.1126/science.ade1855

Depierre, A., Dellandréa, E., & Chen, L. (2018). Jacquard: A Large Scale Dataset for Robotic Grasp Detection. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3511–3516. https://api.semanticscholar.org/CorpusID:215827111

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Galletti, C., & Fattori, P. (2018). The dorsal visual stream revisited: Stable circuits or dynamic pathways? *Cortex*, *98*, 203–217. https://doi.org/10.1016/j.cortex.2017.01.009

Gamberini, M., Galletti, C., Bosco, A., Breveglieri, R., & Fattori, P. (2011). Is the Medial Posterior Parietal Area V6A a Single Functional Area? *The Journal of Neuroscience*, *31*(13), 5145–5157. https://doi.org/10.1523/JNEUROSCI.5489-10.2011

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Girden, E. R. (1992). ANOVA: Repeated measures. In *ANOVA: Repeated measures.* Sage Publications, Inc.

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 580–587. https://doi.org/10.1109/CVPR.2014.81

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95–112. https://doi.org/10.1007/BF02289823

Guclu, U., & van Gerven, M. A. J. (2015). Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *Journal of Neuroscience*, *35*(27), 10005–10014. https://doi.org/10.1523/JNEUROSCI.5023-14.2015

Huang, Y., Gornet, J., Dai, S., Yu, Z., Nguyen, T. M., Tsao, D. Y., & Anandkumar, A. (2020a). Neural Networks with Recurrent Generative Feedback. *ArXiv*, *abs/2007.09200*. https://api.semanticscholar.org/CorpusID:220646910

Huang, Y., Gornet, J., Dai, S., Yu, Z., Nguyen, T., Tsao, D. Y., & Anandkumar, A. (2020b). Neural Networks with Recurrent Generative Feedback. *NeurIPS*.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787. https://doi.org/10.1038/29537

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C., & de Lange, F. P. (2012). Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cerebral Cortex*, *22*(9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. https://doi.org/10.1016/j.tics.2012.10.011

Kumra, S., Joshi, S., & Sahin, F. (2022a). GR-ConvNet v2: A Real-Time Multi-Grasp Detection Network for Robotic Grasping. *Sensors*, *22*(16), 6208. https://doi.org/10.3390/s22166208

Kumra, S., Joshi, S., & Sahin, F. (2022b). GR-ConvNet v2: A Real-Time Multi-Grasp Detection Network for Robotic Grasping. *Sensors*, *22*(16), 6208. https://doi.org/10.3390/s22166208

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434. https://doi.org/10.1364/JOSAA.20.001434

Lindsay, G. W., & Miller, K. D. (2018). How biological attention mechanisms improve task performance in a large-scale visual system model. *ELife*, *7*. https://doi.org/10.7554/eLife.38105

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huissoud, C., Clavagnier, S., Giroud, P., Sappey-Marinier, D., Barone, P., Dehay, C., Toroczkai, Z., … Kennedy, H. (2014). A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex. *Cerebral Cortex*, *24*(1), 17–36. https://doi.org/10.1093/cercor/bhs270

Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M. A., Giroud, P., Ercsey-Ravasz, M., Pilaz, L. J., Huissoud, C., Barone, P., Dehay, C., Toroczkai, Z., Van Essen, D. C., Kennedy, H., & Knoblauch, K. (2011). Weight Consistency Specifies Regularities of Macaque Cortical Networks. *Cerebral Cortex*, *21*(6), 1254–1272. https://doi.org/10.1093/cercor/bhq201

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-Based Attention Increases the Selectivity of Population Responses in Primate Visual Cortex. *Current Biology*, *14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Mauchly, J. W. (1940). Significance Test for Sphericity of a Normal n-Variate Distribution. *The Annals of Mathematical Statistics*, *11*(2), 204–209. https://doi.org/10.1214/aoms/1177731915

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf

Pennartz, C. M. A., Dora, S., Muckli, L., & Lorteije, J. A. M. (2019). Towards a Unified View on Pathways and Functions of Neural Recurrent Processing. *Trends in Neurosciences*, *42*(9), 589–603. https://doi.org/10.1016/j.tins.2019.07.005

Rao, R. P. N., & Ballard, D. H. (1999a). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rao, R. P. N., & Ballard, D. H. (1999b). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, *179*(1), 3–20. https://doi.org/10.1016/0006-8993(79)90485-2

Rockland, K. S., & Van Hoesen, G. W. (1994). Direct Temporal-Occipital Feedback Connections to Striate Cortex (V1) in the Macaque Monkey. *Cerebral Cortex*, *4*(3), 300–313. https://doi.org/10.1093/cercor/4.3.300

Salin, P. A., & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiological Reviews*, *75*(1), 107–154. https://doi.org/10.1152/physrev.1995.75.1.107

Shipp, S. (2016). Neural Elements for Predictive Coding. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01792

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, *abs/1409.1556*. https://api.semanticscholar.org/CorpusID:14124313

The jamovi project. (2024). *jamovi* (2.5). Retrieved from https://www.jamovi.org.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics*, *5*(2), 99. https://doi.org/10.2307/3001913

Watt, S. J., & Bradshaw, M. F. (2000a). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia*, *38*(11), 1473–1481. https://doi.org/10.1016/S0028-3932(00)00065-8

Watt, S. J., & Bradshaw, M. F. (2000b). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia*, *38*(11), 1473–1481. https://doi.org/10.1016/S0028-3932(00)00065-8

Wen, H., Han, K., Shi, J., Zhang, Y., Culurciello, E., & Liu, Z. (2018). *Deep Predictive Coding Network for Object Recognition*.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111