Deep hierarchical sensory processing accounts for effects of arousal state on perceptual decision-making

Lynn K. A. Sörensen[1,2], Sander M. Bohté[3], Heleen A. Slagter[4,5*], H. Steven Scholte[1,2*]

[1] Department of Psychology, University of Amsterdam, The Netherlands
[2] Amsterdam Brain & Cognition (ABC), University of Amsterdam, The Netherlands
[3] Machine Learning Group, Centrum Wiskunde & Informatica, The Netherlands
[4] Department of Experimental and Applied Psychology, Vrije Universiteit Amsterdam, The Netherlands
[5] Institute of Brain and Behaviour Amsterdam, Vrije Universiteit Amsterdam, The Netherlands
[*] Shared senior authorship

Corresponding authors: lynn.soerensen@gmail.com; h.s.scholte@uva.nl

Abstract

An organism's level of arousal strongly affects task performance. Yet, what level of arousal is optimal for performance depends on task difficulty. For easy tasks, performance is best at higher arousal levels, whereas arousal levels show an inverted-U-shaped relationship with performance for difficult tasks, with best performance at medium arousal levels. This interaction between arousal and task difficulty is known as the Yerkes-Dodson effect (1908) and is thought to reflect sensory decision-making in the locus coeruleus and associated widespread release of noradrenaline. Yet, this account does not explain why perceptual performance decays with high levels of arousal in difficult, but not in simple tasks. Recent studies suggest that arousal may also affect performance by modulating sensory processes. Here, we augment a deep convolutional neural network (DCNN) with a global gain mechanism to mimic the effects of arousal on sensory processing. This allowed us to reproduce the Yerkes-Dodson effect in the model's performance. Investigating our network furthermore revealed that for easy tasks, early network features contained most task-relevant information during high global gain states, resulting in model performance on easy tasks being best at high global gain states. In contrast, later layers featured most information at medium global gain states and were essential for performance on challenging tasks. Our results therefore establish a novel account of the Yerkes-Dodson effect, where the interaction between arousal state and task difficulty directly results from an interaction between arousal states and hierarchical sensory processing.

Significance statement

Over a hundred years ago, it was first observed that the effect of arousal on performance depends on task difficulty: the Yerkes-Dodson effect. Difficult tasks are best solved at intermediate arousal levels, whereas easy tasks benefit from a high arousal state. Current theories on how arousal affects neural processing cannot explain this effect of task difficulty. Here, we implement a key effect of arousal on cortical processing, a change in gain, in a computational model of visual processing capable of object recognition. Across a series of experiments, we find that our model can reproduce the Yerkes-Dodson effect behaviorally and that this effect can be explained by where in the processing hierarchy different arousal states optimize sensory information encoding.

Introduction

Cognitive performance is not stable but varies over time and across situations. Arousal state, the overall level of activation of the nervous system, is thought to be a key determinant of variability in cognitive performance (1). As an example, consider a student taking the same difficult test in two situations: in the first situation, it is a practice test. The student is calm. In the second situation, it is the final exam. The student is very nervous. The difference between these situations is that in the latter, the student might be overly aroused, leading him to perform more poorly on the test. Indeed, a large body of work indicates that arousal level is an important determinant of task performance and perceptual decision-making across species (2, 3, 4, 5, 6, 7, 8, 9). Yet, arousal state does not affect performance on any task in the same way (10, for a review, see 11). Specifically, while performance on easy tasks improves with arousal level, performance on difficult tasks exhibits an inverted U-shape relationship with arousal level, with performance peaking at intermediate arousal levels. Thus, optimal performance occurs at different arousal states as a function of task difficulty, a phenomenon known as the Yerkes-Dodson effect (see Figure 2A for a schematic). At the neural level, the inverted U-shape relationship has been associated with the functioning of the locus coeruleus (LC), the central release site of noradrenaline (Figure 1A). The phasic noradrenergic response (upon target detection) in particular also follows an inverted U-shape in its responsivity across LC baseline firing levels. Based on this observation, Aston-Jones and Cohen (2005) have suggested that this change in responsivity brings about the inverted U-shape in performance. Yet, this influential view does not address why arousal levels differentially impact performance on relatively easy versus difficult tasks. Indeed, the heterogeneity in findings of recent studies suggests a more complex interplay between arousal state, perceptual performance and other factors such as task difficulty: While some studies report enhanced perceptual performance with increasing baseline arousal levels (5, 8), others find performance increases with decreasing baseline arousal levels (12) or observe a curvilinear relationship between performance and arousal (2, 3). Possibly, differences in arousal states explored and in task difficulty across studies can account for this diversity in findings. Yet, systematically characterizing the relationship between arousal level and task difficulty is hard to achieve experimentally, in part because it is difficult to cover the full range of possible arousal states and task difficulties within one experimental setting. While this might explain the diversity in findings, it also highlights the need for a suitable modelling framework that can encompass factors such as task difficulty.

Deep convolutional neural networks (DCNN) allow us to address this gap in knowledge. In particular, these networks optimized for object recognition not only parallel human performance on some object recognition tasks (17), but they also feature processing characteristics that bear a remarkable resemblance to the visual ventral stream in primates (18, 19, 20, 21, 22, 23). Both of these characteristics make them an attractive modelling framework for testing computational hypotheses about the link between neural processing and behavior (24, 25). Notably, recent studies show that heightened arousal increases the signal-to-noise ratio (SNR) of sensory neurons (2, 4, 26, 27), both by decreasing variability in spontaneous activity and by increasing neuronal responsivity or gain (2, 3, 26, 27, 28), with consequences for perceptual performance (2, 3). Furthermore, stimulus-unrelated neural activity in early visual cortices is closely related to arousal (29). This raises the intriguing possibility that arousal may affect task performance by modulating the processing of sensory features that perceptual decisions are grounded in. Here, we use DCNNs to test this notion and systematically investigate how a wide range of arousal states modulates sensory processing and perceptual decision-making for tasks of varying difficulty.

In line with the early findings of Yerkes and Dodson (10) and others, we recover the same interaction between arousal state and task difficulty with our DCNN model. That is, also in our model, easier tasks were best solved at high global gain states, whereas difficult tasks were best solved at medium global gain states. This relationship was specific to perceptual difficulty and absent for other types of difficulty (e.g., in response complexity). Moreover, we found that how
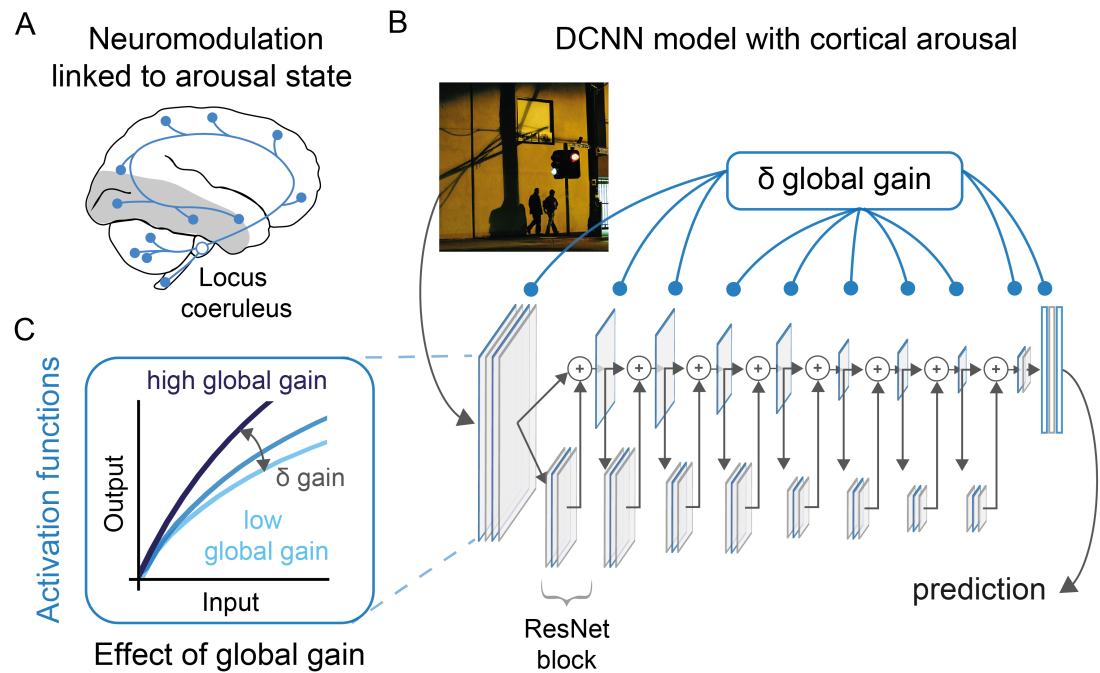
Figure 1. Global gain as a model for cortical arousal. (A) Schematic illustration of the projections of the locus coeruleus following (13, 14). The locus coeruleus is located in the pons of the brain stem and features efferent projections throughout the brain. Note that subcortical projections have been omitted here for clarity. The grey shaded areas depict the early visual cortex and the visual ventral stream that we sought to model in this work. (B) Illustration of the DCNN architecture and the locations of the activation functions where global gain, our implementation of cortical arousal, was manipulated (blue framed rectangles). A DCNN takes images as an input and produces a prediction as an output. Importantly, the global gain of the model can be altered with a single parameter that is applied to all activation functions. This takes inspiration from the widespread noradrenergic projections reported for the locus coeruleus (e.g.,(13)) and their effects on sensory processing (e.g.,(15)). (C) Changes in global gain resulted in a multiplicative scaling of the activation function. Higher gain levels (dark blue) resulted in higher values compared to the baseline (blue), which corresponds to a gain of 1, and reduced gain levels (light blue) lead to lower activation values. The example picture in B is licensed under CC BY-SA 2.0 and was adapted from Flickr (16).

global gain states affected performance on a given task could be explained by their effects on information encoding across the processing hierarchy. High global gain states enhanced encoding of information in early network features that easy tasks capitalized more on, while intermediate global gain states enhanced encoding of information that was more important for performance on difficult tasks higher in the processing hierarchy. These findings critically inform current debate as to how arousal impacts perceptual decision-making.

Results

We augmented a DCNN with a global gain mechanism to investigate how arousal state changes in sensory areas may affect performance and relate to the Yerkes-Dodson effect more specifically. To this end, we used a ResNet18-architecture (30) with a biologically inspired activation function (31, 32, see Figure 1C for an illustration). The global gain mechanism targeted all activation

functions in the network simultaneously. For every activation function, a change in global gain resulted in a change of response gain (δ global gain, see Figure 1). All networks were trained in a neutral gain state of 1, which corresponds to a standard DCNN without global gain modulation. Subsequently, these trained weights were evaluated across a range of global gain states (without any further training). Using this model allowed us to control global gain state so as to isolate the mechanisms underlying the interaction between arousal state and task difficulty as captured by the Yerkes-Dodson effect.

A DCNN with global gain replicates the Yerkes-Dodson effect in performance

A hallmark of the effect of arousal state on performance is the interaction between task difficulty and arousal state, the Yerkes-Dodson effect (see Figure 2A for an illustration). In our first set of analyses, we show that our DCNN with a global gain mechanism can replicate this effect. That is, we show that easy and difficult tasks require different global gain values for optimal performance across a range of performance measures, with easy tasks yielding the best performance at high global gain states, and difficult tasks performing best at intermediate global gain states.

To study the Yerkes-Dodson effect, one needs to effectively manipulate task difficulty. Task difficulty is commonly manipulated by changing stimulus strength or complexity in perceptual decision-making tasks, for instance, by altering the signal-to-noise ratio of a pure tone embedded in a sequence of noise (3) or by phase-scrambling black and white images (8). In a similar vein, we adopt a task, in which the model has to decide whether an image is real or an average image, created by averaging over a variable number of images, taken from a pool of images consisting half of real and half of average images. Task difficulty was manipulated by varying the number of images that were averaged for making the average images. This makes that an average image consisting of few images (e.g., 1.25 images) was relatively similar to and therefore hard to distinguish from a real image and vice versa, an average image consisting of many images was relatively dissimilar and hence easier to distinguish (see Figure 2B for example stimuli). We chose to use averaging as our difficulty manipulation reasoning that it would result in difficult images closer to the training images of the DCNN, thereby increasing the task difficulty. Comparing performance at a neutral gain state (i.e., global gain is 1) in Figure 2C shows that our manipulation of perceptual difficulty was effective in producing performance differences. For every level of perceptual difficulty, we trained a separate output layer, while keeping the rest of the model weights unchanged. After training, we tested the model on a new set of images across a wide range of global gain states. This allowed us to get high-resolution estimates of the performance-gain profile for every level of perceptual difficulty.

Based on the Yerkes-Dodson effect, we expected to see changing performance-gain profiles across task difficulties and in particular, a shift in the arousal state that is linked to the best performance (see Figure 2A for our hypothesis). Strikingly, we indeed reproduce the Yerkes-Dodson effect with our global gain manipulation, as reflected in a right-ward shift in the performance-gain profile with a decrease in task difficulty (Figure 2C) that translates to a U-shaped relationship between gain and performance for difficult tasks, but a linear relationship for easier tasks when a more limited range of arousal is taken into account (see below, Figure 2D). Identifying the global gain level associated with peak performance for every task difficulty (Figure 2E), further reveals that there is a quasi-linear negative relationship between task difficulty and the global gain state associated with the best performance, with easy tasks being performed best at high gain states and difficult tasks being performed best at medium gain states. To further dissociate which aspects in the model's performance were changed by the global gain mechanism, we also analyzed performance with regard to changes in sensitivity and bias, two measures from signal detection theory (SDT). Whereas sensitivity refers to the ability to distinguish a signal from noise, bias describes the propensity to answer irrespective of the
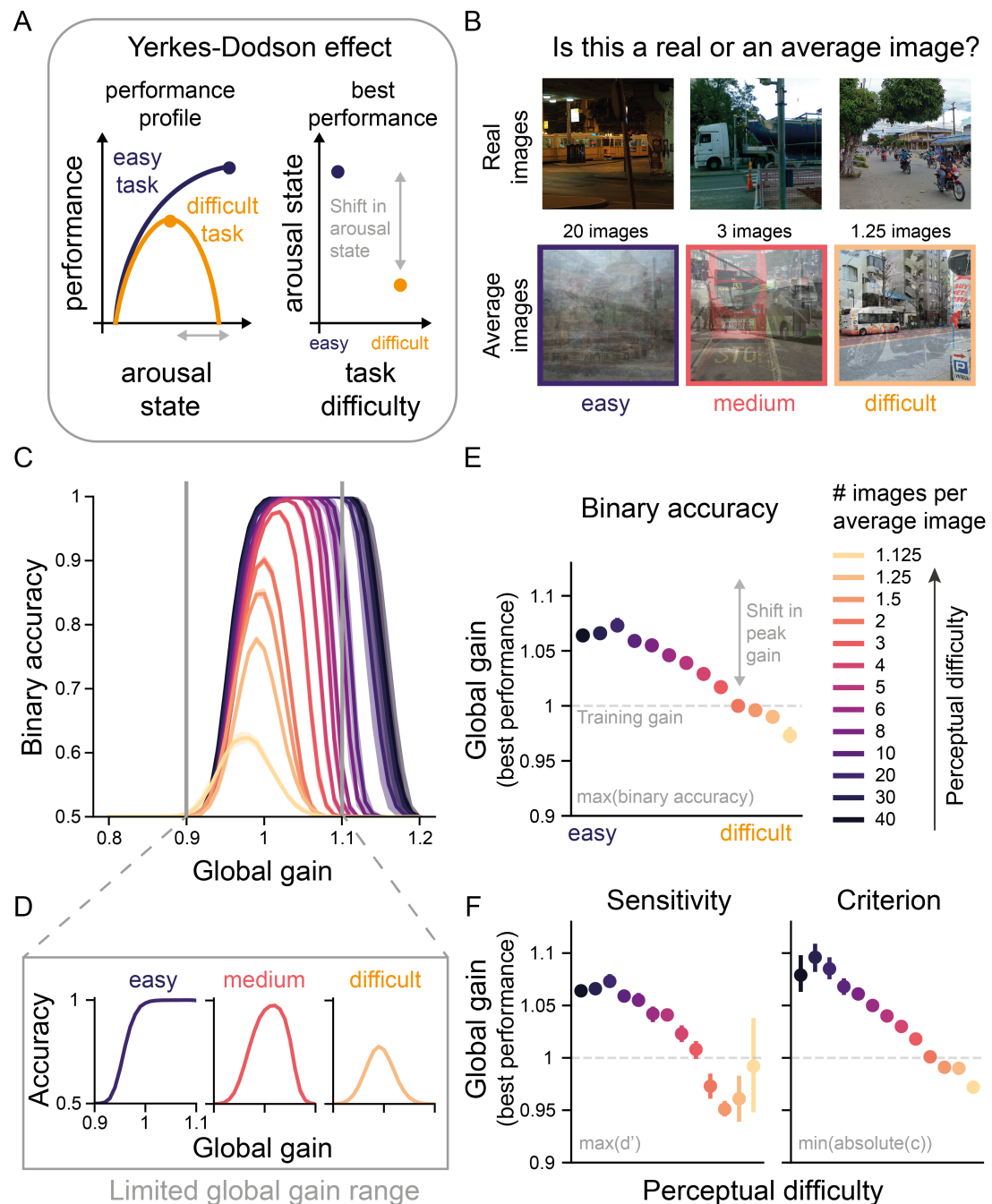
Figure 2. Incorporating global gain in a DCNN reproduces the Yerkes-Dodson effect. (A) Left: Schematic of the Yerkes-Dodson effect, showing performance-gain profiles for difficult and easy tasks. While difficult tasks (orange) typically follow an inverted U-shape relationship between arousal state and performance, easier tasks have been associated with an increasing relationship between arousal state and performance. The dots on the performance-gain profiles denote the peak performance. Right: The Yerkes-Dodson effect can be also expressed as a difference in arousal state linked to the peak performance across task difficulty conditions. From this perspective, easy tasks are linked to high arousal states and difficult tasks to medium arousal states for the best performance. (B) Illustration of the network's task. The network was tasked to distinguish between real and average images. To manipulate perceptual difficulty, these average images consisted of an increasing number of images per average image. During training, we fine-tuned an output node for every perceptual difficulty level without applying

any global gain changes. This allowed us to dissociate between the network's trained ability and the effects of global gain. For every perceptual difficulty level, we fine-tuned ten network instances. During testing, we presented the fine-tuned networks repeatedly with the same dataset (new set of test images), while adjusting the global gain parameter. This resulted in a binary accuracy result for every global gain value for each of the model instances. Here, we show example images from the real image condition and from the average image condition for three levels of difficulty: difficult (average of 1.25 images), medium (average of 3 images) and easy (average of 20 images). (C) Binary accuracy for all levels of perceptual difficulty as a function of global gain state. The shaded areas represent the 95% confidence interval (CI) across the ten network instances. (D) As in (C) but for the three levels of difficulty (easy, medium, difficult) and a limited global gain range. It can clearly be seen that as expected, the easy perceptual difficulty condition shows a monotonically increasing relationship between gain and performance, while the difficult perceptual difficulty condition shows an inverted U-shape relationship between gain and performance. Our gain manipulation thus reproduced the Yerkes-Dodson effect in DCNN performance. (E) Global gain level linked to peak performance per difficulty condition in C. As predicted by the Yerkes-Dodson effect, peak performance was associated with reductions in global gain level with increasing task difficulty. The error bars correspond to 95% CI across instances. If multiple gain states were linked to the best performance, the median was used to summarize them. The dashed line corresponds to the neutral gain state from training, during which no global gain changes are applied. (F) As in (E) but now global gain states linked to the performance for sensitivity and bias. Hit and false alarm rates, sensitivity and bias values for all perceptual difficulties and global gain states can be found in Figure S1. The example pictures in B are licensed under CC BY-NC-SA 2.0 and were adapted from Flickr (33, 34, 35).

signal. In turn, accurate performance is defined by high values in sensitivity and a bias close or equal to zero. Following this definition, we identified the global gain state linked to the best performance for both of these measures. As with accuracy, we observe for both sensitivity and bias that the best performance for easy tasks is observed at high gain states, whereas for difficult tasks it is obtained at medium gain states close to the neutral gain (Figure 2F). These findings show that global gain did not merely result in a change in the networks responsivity or bias, but rather also changed its sensitivity, or its ability to distinguish between real and average images across different task difficulties.

It is hard to assess a wide range of arousal states experimentally, not only because of experimental constraints, but also because organisms are very unlikely to visit these extreme arousal states due to homeostatic constraints (hypo- and hyperarousal). Thus, the experimental data so far available likely stem from a much narrower range of arousal states than we explored here. Indeed, it has been argued that some studies might have observed a linear increase in performance as a function of arousal state, because they only sampled from the left-side of the inverted U-shape (e.g. 36, 37). Mimicking a limited global gain range with our results (Figure 2D) reveals a qualitative correspondence to the experimental data observed for arousal state manipulations across a set of task difficulties (10) and shows how plausibly an increasing relationship between arousal state and performance would emerge for easy tasks, if only low and medium arousal states are experimentally sampled. For our data, inspecting a wider range of arousal states clarifies that, in fact, all performance gain profiles could be readily described as an inverted U (see Figure 2C). This finding can explain how a diverse set of curvilinear relationship can be identified experimentally if the highest arousal states are not accessible.

To summarize these findings, we have replicated the behavioral signatures associated with the Yerkes-Dodson effect, across a set of three performance measures. Further, our data offer an explanation for the diversity in experimental findings showing how these could result from sampling a limited range of arousal states covered here by our global gain manipulation.
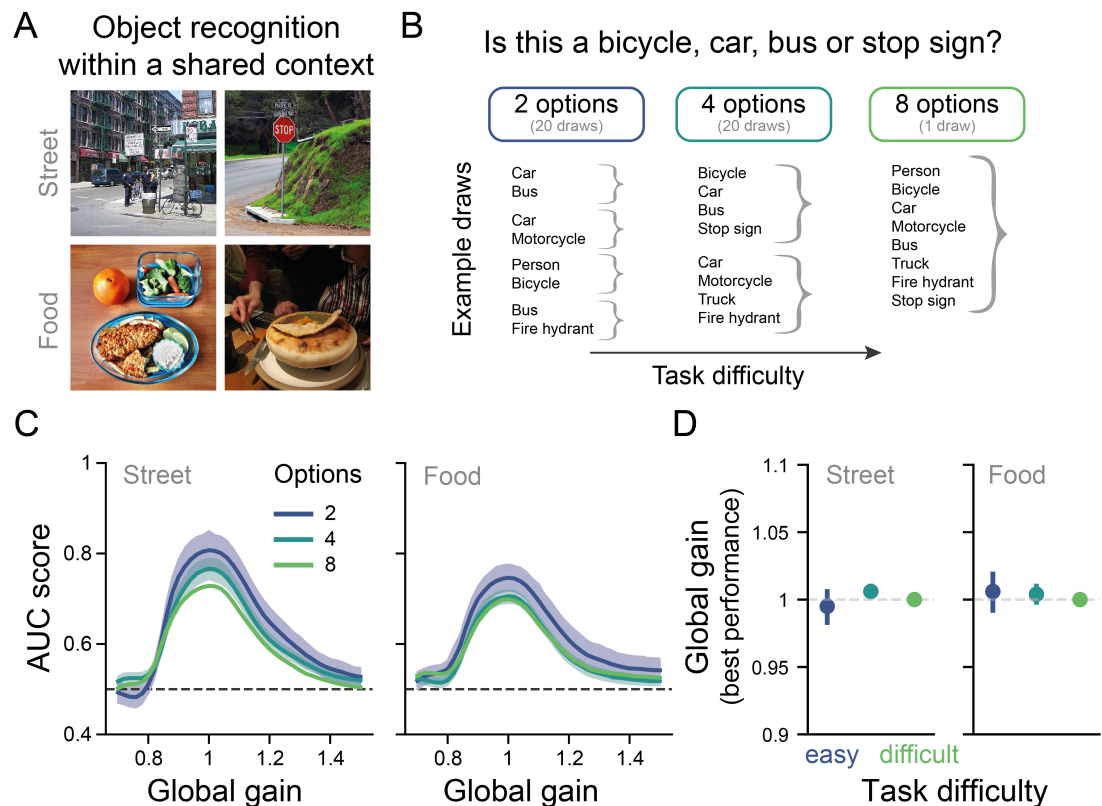
Figure 3. Assessing the Yerkes-Dodson effect for another kind of task difficulty: response complexity. (A) Example images from the two visual search datasets curated from the COCO database (38). (B) Illustration of how task difficulty was increased by increasing the number of answer options for the street dataset. Easy tasks consisted of only two categories to be distinguished, while difficult tasks entailed choosing from 8 answer options. Different easy tasks were obtained by subsampling from the 8 possible target categories. Task difficulty increased with the number of categories that had to be distinguished. (C) Performance-gain profiles across the three task difficulty conditions for both datasets. The global gain level associated with peak performance was unaffected by task difficulty (number of answer options). Performance was assessed with the area under the curve metric of the receiver operating characteristic curve. The lines represent the mean performance across 20 trained networks for the conditions with 2 and 4 answer options. The shaded areas depict the 95% CI across model instances. The dashed dark grey line depicts chance performance. (D) Global gain state associated with the best model performance was similar for every task difficulty level for both datasets. The dashed light grey line depicts the neutral model, at which no global gain is applied. The example pictures in A are licensed under CC BY 2.0 and were adapted from Flickr (39, 40, 41, 42).

Yerkes-Dodson effect is absent for other factors of task difficulty

Our results so far show that the Yerkes-Dodson effect can be replicated in our model with a manipulation of perceptual difficulty. Yet, in other studies, the Yerkes-Dodson effect is linked to task difficulty more broadly (43). In turn, an obvious question to ask is whether our observations also hold for other manipulations of task difficulty. With task difficulty manipulations, we here simply refer to changes to the task that result in performance changes at a neutral global gain state (i.e., 1, without applying any global gain changes). In our next analysis, we show that another task difficulty manipulation, response complexity, does not reproduce a shift towards

higher gain states with decreases in task difficulty, suggesting the Yerkes-Dodson effect to be specific for perceptual difficulty in our model.

To introduce a new task difficulty manipulation, we altered the number of answer options available to the model (Figure 3B), while keeping all other factors constant such as perceptual difficulty. We applied this manipulation in the context of two different object recognition datasets. Both datasets were curated to be challenging visual search tasks by reducing the amount of informative background differences across categories (for further information, see 32). As can be seen in Figure 3C, this manipulation resulted in differences for the baseline networks (with global gain set to 1), as indexed by AUC, thus showing that the manipulation was effective. This was true for both datasets that we applied this manipulation to (food and street scene images; Figure 3A). If this manipulation also produces an interaction between task difficulty and global gain state, one expects the global gain states associated with the best model performance to differ between the conditions with different numbers of answer options. Figure 3D shows that this is not the case, instead for both data sets, peak performance occurs at the same, medium global gain state independent of the number of answer options. Thus, manipulating the number of answer options for the network merely scaled the gain-performance profile up- or downwards without qualitatively shifting the optimal gain state.

These results also corroborate our findings on perceptual difficulty. Since both datasets were curated to be challenging visual search tasks, they required the analysis of complex visual features. Based on this, one would expect an inverted U-shaped gain profile (as was observed in Figure 2 for perceptually difficult tasks), which we indeed observed across all six conditions (both datasets, all answer options).

In sum, these findings complement our understanding of the Yerkes-Dodson effect in our model in two ways. First, we observed that it is specific to perceptual difficulty. Second, we have shown that visual search in natural scenes is associated to an inverted U-shape across global gain states.

Global gain differentially affects task information across the model hierarchy

The above analyses described how the model's performance, that is, its output, was shaped by changes in global gain states across two tasks. In our next set of analyses, we leveraged the full observability of our model to address how sensory processing may give rise to the Yerkes-Dodson effect. To this end, we first characterized the different stages of processing with regard to this interaction between global gain state and perceptual difficulty. In particular, we tracked how task-relevant information throughout the network hierarchy was affected by these factors, using a decoding analysis. In brief, this analysis showed that with an increase in model depth there was a shift from higher to medium global gain states for representing the most task-relevant information.

To quantify the presence of task-relevant information across model depth (ResNet block; see Figure 1C), we used a linear decoder. Specifically, we adopted logistic regression and predicted whether a model input had been a real or an average image based on the block's activations on a held-out test set (see methods). Our results show that this can be done almost perfectly for the easier tasks across all model blocks and global gain states (see Figure 4A for examples). Furthermore, it became clear that task difficulty as well as global gain state modulated decoding accuracy across all model blocks (see Figure S2A). In contrast, mean activations did not contain such modulations of task difficulty and were only driven by changes in global gain resulting in an overall increase in activation (see Figure S2B).

To understand how the decoding-gain profiles may be linked to the shift in optimal performance across task difficulty more directly, we identified the peak global gain state linked to the best decoding performance (see Figure 4B). This analysis indicated that, here as well, easy task information tended to be best decodable at high gain states, whereas difficult task
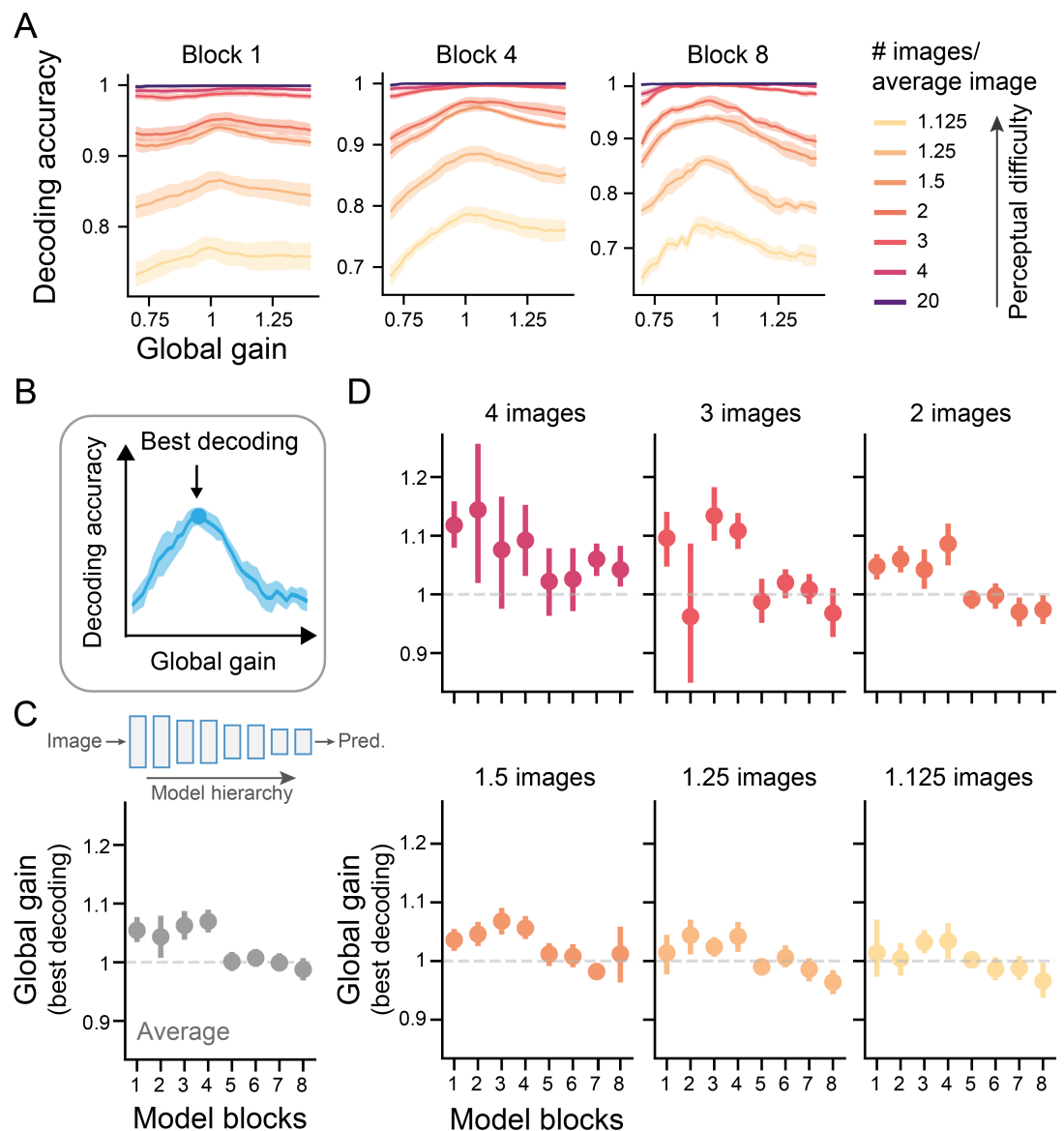
Figure 4. Global gain differentially affects the representational quality along the model hierarchy. (A) Mean linear decoding accuracy recorded after the last activation function in a given ResNet block across five repetitions and shown as a function of perceptual difficulty and global gain (for an overview of all blocks, see Figure S2A). The shaded areas depict the 95% CI across repetitions. (B) Analogous to the behavioral analysis, we here identified the global gain state linked to the best decoding accuracy. Since the easiest task (20 images/average image) was solved almost at ceiling across all evaluated blocks and irrespective of global gain state, the peak estimation was unreliable, and we thus excluded it from this analysis (see Figure S2A). (C) Model blocks feature most decodable task-relevant information at different global gain states across the hierarchy. The decoding peak refers to the median global gain values linked to the best decoding accuracy across repetitions. To obtain an estimate across task conditions, we averaged across perceptual difficulties. As in figure 2D-F and 3C, the dotted line refers to the neutral gain states and the error bars describe the 95% CI. The top part of the figure is a schematic of the DCNN architecture shown in Figure 1 in more detail. (D) The global gain level associated with peak decoding for each model block and perceptual difficulty level separately. This panel shows that easy task information tended to be best decodable at high gain states in early network blocks, whereas difficult task information was best decoded at medium gain states and later model blocks.

information was best decoded at medium gain states for most blocks (see Figure 4D), thereby capturing a key property of the performance data (Figure 2E). Furthermore, comparing across blocks shows that the global gain state associated with peak decoding was higher for early blocks than later blocks, which exhibited peak decoding at more medium global gain states (Figure 4C). In addition to these general trends, inspecting the global gain states linked to peak decoding performance across task difficulties also indicated an interaction with model depth: Whereas easy tasks showed a negative linear link between model block and peak gain state, this relationship flattened out with increasing perceptual difficulty (see Figure 4D). These findings suggests that later network blocks contained most task-relevant information at medium global gain states, while early network blocks best represented task information at high gain states, in particular for easier tasks.

Taken together, these findings indicate that different stages of hierarchical processing were linked to different optimal global gain states for encoding task-relevant information. While global gain states facilitated sensory encoding in early network blocks at high gain states, late network blocks contained most task-relevant sensory information at medium gain states.

Task difficulty differentially engages model features across the model hierarchy

In the last analysis, we have shown that also decodable information is associated with different optimal global gain states across perceptual difficulties. Crucially, this link varied along the model hierarchy. This observation begs the question whether this linearly decodable information is actually used during network performance and how this factor might interact with perceptual difficulty more directly. To address this, we devised a manipulation that allowed us to selectively reduce the contribution of each model block. In particular, we took advantage of the architecture of our model that has two main branches, one for maintaining and one for processing information (see Figure 5A). By interfering with the information in the processing branch, we were able to inspect the performance-gain profiles with and without the contribution of a given block. In brief, we find that indeed early network blocks are the main driver behind performance being optimal during high global gain states for easy tasks. Furthermore, this analysis makes it evident that the difficult tasks particularly rely on late network features for performance, which biases them towards medium global gain states in contrast to easy tasks.

To selectively control the contribution of a model block to the overall performance, we developed the spatial scrambling method (see Figure 5A). This approach selectively perturbs the information for a given block. As a result, this model block's features cannot effectively contribute to performance anymore and thus the remaining unaffected blocks will drive the model's performance (see methods for an in-depth description of spatial scrambling). We leveraged this principle to dissect the network's performance and this helped us understand how the features along the model hierarchy differentially contributed to performance of easy and difficult tasks. Specifically, we either increasingly added spatial scrambling from the top to the bottom of the network, thereby making the network more and more reliant on the early blocks or we applied it from the bottom to the top of the network, so that late network features drove performance (see Figure 5A).

Our results so far suggest that high global gain states optimize sensory processing in early blocks. Therefore, we expected that increasingly disrupting late network function (*Late to early*) and giving more weight to earlier blocks, would lead to a right-ward shift in the performance gain profiles, that is, higher gain states being linked to the best performance. Figure 5B shows that this expectation was met for both perceptually difficult and easy tasks. In particular, we observe that greater reliance on processing in early network leads to peak performance at higher global gain states (see Figure 5C, *Late to early*). Notably, this results in very high global gain states compared to the baseline network (grey lines and dots in Figure 5B and C). Across both conditions, we see that peak performance shifted to the highest global gain states
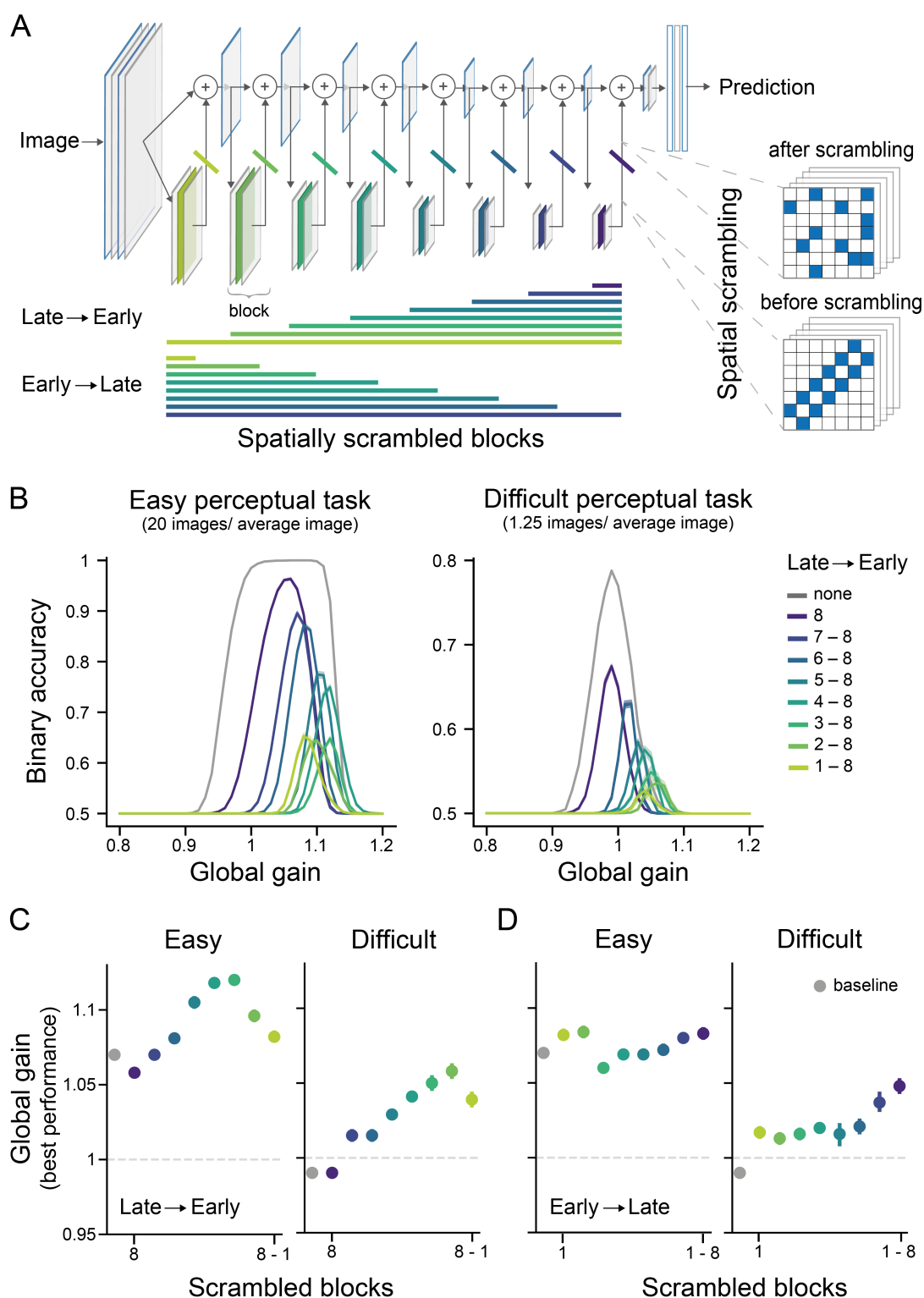
Figure 5. Task difficulty differentially engages model features across the hierarchy, leading to differences in optimal global gain state in performance. (A) Spatial scrambling was used to control the contribution of the different model blocks to performance. Spatial scrambling spatially randomizes the activations within a predefined proportion of feature maps in a given model block. This manipulation was either applied starting from late blocks and then progressing to early model blocks or it was first applied to early blocks and then it progressed to later network blocks. Crucially, after a block is scrambled its informative contribution to the model's performance is strongly reduced (see Methods for an explanation). (B) Starting at later blocks and successively scrambling earlier blocks reveals a pronounced right-ward shift in the performance-gain profiles for both task difficulties. The grey line depicts the performance-gain profile for the baseline network (without any spatial scrambling). Shaded areas show the 95% confidence intervals across repetitions. (C) Determining the global gain states linked to the best performance across different scrambling states of the model shows that in particular the scrambling of later blocks leads to a shift towards higher global gain states for the best performance. (D) In contrast, first scrambling early blocks had little effect on the global gain state linked to peak performance. This only changed for the difficult task once late blocks were scrambled, which resulted in peak performance being again associated to higher global gain states (as in C). This indicates that while these later network blocks are deployed during recognition, the model operates best overall at medium gain states, since later network blocks operate best at medium global gain states.

if almost all late network features were scrambled. This finding suggests that higher global gain states facilitate network performance by enhancing feature content in early network blocks. As perceptually easy tasks relied more on sensory processing in early network features than perceptually difficult tasks, scrambling late network features had less of an effect on performance on easy tasks.

In contrast to easy tasks, the network's performance on difficult tasks was best at medium global gain states. As we found optimal decoding of task-relevant information in later blocks at medium global gain states for difficult tasks (Figure 4), we expected that first scrambling early blocks would not affect the relationship between gain and performance on difficult tasks until scrambling also included later blocks. To test this, we next increasingly scrambled the network blocks increasingly from early to late blocks (see Figure 5C, *Early to late*). Indeed, the global gain state associated with the best performance remained relatively stable and visibly lower compared to the early-to-late condition for both task difficulty levels (see Figure S3 for performance-gain profiles). Only once also the very last network features were affected, there was a notable increase towards higher gain states for peak performance on the difficult task. This pattern of results suggests that sensory processing in later network blocks is more critical for solving perceptually difficult tasks and that medium global gain states facilitate later processing and hence performance on difficult tasks. In contrast, easy tasks rely more on sensory processing in early blocks, which operate best at higher global gain states.

Taken together, these findings indicate that the effects of arousal states on sensory processing can explain at least in part the effects of arousal state on perceptual performance as a function of task difficulty, commonly referred to as the Yerkes-Dodson effect. More concretely, global gain level may differentially affect computational features across model depth, which, depending on the level of global gain and task difficulty, benefits performance on some tasks, but impairs performance on other tasks.

Discussion

Since the pioneering work by Yerkes and Dodson (10), a large body of research has demonstrated the prominent role of arousal, not just in sleep and wakefulness, but in task-related neural

processing and performance (e.g. 4, 11, 13, 29, 44). Yet, precisely how arousal states affect task performance, and specifically, why high arousal states impair performance on difficult, but not easy tasks, as reflected in the Yerkes-Dodson effect, is still unclear. We here capitalized on recent insights from neuroscience to model the Yerkes-Dodson relationship, exploiting the unique ability of DCNN modelling to systematically simulate the full range of possible arousal states while performing a task. Traditionally viewed to reflect the outcome of perceptual decision-making in the locus coeruleus (13), we here show that we cannot only reproduce the Yerkes-Dodson effect in our model by augmenting it with a global gain mechanism, but also that sensory effects of arousal states are sufficient to account for the Yerkes-Dodson effect. Specifically, investigating the inner workings of our model, we show that different global gain states optimize sensory encoding of task-relevant information at different stages of hierarchical processing thereby in particular affecting performance on tasks that more critically rely those stages of processing. That is, high global gain states facilitated early model processing, thereby specifically enhancing performance on perceptually easy tasks, while intermediate global gain states facilitated late model processing, specifically enhancing performance on perceptually difficult tasks. Altogether, we provide a new perspective on the long-standing question as to how arousal can facilitate and impair task performance by highlighting the complex interaction between arousal and hierarchical sensory processing.

DCNNs as promising avenue for modelling the effects of cortical arousal

To our knowledge, this is the first large-scale simulation of the effect of arousal state on visual processing. While other computational accounts have described the effects of gain in small-scale models (45, 46), these could not address factors such as perceptual task difficulty. Here we critically extend this work by directly linking global gain to effects on perceptual performance using natural images and a task that varied parametrically in perceptual difficulty. As any model, our model omits some crucial details such as the temporal signature of these effects typically used to delineate phasic from tonic arousal effects in pupillometry (for a discussion, see 47) or oscillatory signatures linked to synchronized and desynchronized states (48). Yet, our findings provide a starting point for understanding the complex interaction between an unspecific neuromodulator such as noradrenaline and sensory processing. As apparent from our results, working with such a model both affords an in-depth investigation of model processing (Figure 4) and more importantly, allows for directed causal manipulation (Figure 5). Our study highlights how this approach can advance our understanding, both by accounting for perceptual complexity and by leveraging computational mechanisms observed in neural data. Furthermore, our model also introduces state changes into the DCNN modelling framework, thereby contributing to the effort to develop computational models approximating the computations in the ventral visual stream and beyond (49). While current efforts are mainly directed at describing trial-averaged performance (50) and neural data (22), this approach may overlook both the computational mechanisms and effects of interindividual state changes (1, 13, 48, 51). For instance, our results reveal how changes in global gain effectively reconfigure computational features and thereby dramatically change model performance. This reweighting of computational features as a function of global gain state may be a strategy that could also subserve visual processing in the brain, benefitting some visual tasks and while hampering others, as we have shown here.

Link between the model's global gain states and biological arousal states

We could investigate the effect of arousal state across a wide range of arousal states allowing us to carefully describe the performance-gain profiles associated with different task difficulties. While not all gain values examined here are biologically plausible, it was our goal to also describe the tails of the distributions that are usually not accessible experimentally. A biological system bound to homeostasis likely never visits such extreme arousal states. The decay in performance on easy tasks observed for the highest gain states is therefore likely not commonly observed in

biological systems and in practice may more resemble a continuous increase as has been also suggested by others (36, 37). Moreover, importantly, using our parametric perceptual task, we did find that the performance-gain profiles observed as a function of perceptual difficulty in fact lie on a continuum (see Figure 2). Thus, one may expect to find a shift in peak performance while varying arousal and perceptual difficulty. Importantly, our data also suggests that most variation across levels of task difficulty is expected to be observed at high arousal states. Future behavioral studies adopting a continuous perceptual difficulty manipulation across a wide range of arousal states will be key in testing our predictions.

DCNNs with global gain reproduce key behavioral signatures linked to arousal states

In addition to the Yerkes-Dodson effect, our model also replicates a number of other recent findings from studies that assessed arousal state by measuring pupil size before or during stimulus processing (15, 27). In line with our model, across many studies, the highest perceptual sensitivity has repeatedly been linked to medium arousal states for both visual and auditory tasks in mice and humans (2, 3, 37). All of these tasks were challenging and linked to the characteristic inverted U-shape across sensitivities. Crucially, this connects our model to a larger framework in which arousal regulates the efficiency of information processing (4, 5, 13, 52).

DCNNs as a model for visual processing and perceptual difficulty

We here found that the Yerkes-Dodson effect in our model was specific to a manipulation of perceptual difficulty and was insensitive to another manipulation of task difficulty (i.e., number of answer options). At first sight, this finding may seem in contrast to a wide range of studies describing links between the Yerkes-Dodson effect and a range of other cognitive factors (e.g. 43). Yet, we do not wish to suggest that our model can serve as a model for the brain as a whole, since DCNNs largely capture processing in the visual ventral stream (18, 19, 20, 21, 22, 23). Thus, our study should be interpreted as a minimal model for reproducing the Yerkes-Dodson effect, and we cannot exclude that other task factors, which mainly engage other cortical areas, would not produce or contribute to a similar pattern of results. Rather, our findings are a proof of principle of the complex interactions arising from the combination of global neuro-modulatory signals such as noradrenaline and hierarchical sensory processing.

Indeed, another indication for the fact that our model may have specifically captured more initial stages of perceptual decision-making is suggested by our observation of a continuous decrease in criterion with increasing global gain for all task difficulty levels, related to continuous increases in both hit and false alarm rates with increasing gain (Figure S1). This finding contrasts with a study by McGinley et al. (3) in mice that found a decrease in hit rate during high arousal states and correspondingly, a U-shaped link between the mice's criterion and arousal state. Our model does not reproduce this decrease in hit rate during high arousal states, suggesting that it may not well capture all stages of perceptual decision-making. Being a model for visual processing, it is not designed to mimic processes linked to other aspects of decision-making, such as flexibly criterion setting or behavioral strategy (e.g. 51, 53). Notably, a dissociation between perceptual and other aspects of decision-making is supported by a number of recent studies, which attribute a reduction in choice bias to phasic arousal and computations in the brain stem arousal systems and prefrontal cortices (9, 54, 55). Nonetheless, our model provides a promising starting point for also modelling these more complex aspects of decision-making.

Concluding remarks

To conclude, our results suggest that DCNNs with a global gain mechanism can serve as a computational model of the sensory effects of cortical arousal. We here showcased how such a model can capture the behavioral signatures linked to arousal states across species as well as

provide mechanistic insights, thereby providing a new account of the Yerkes-Dodson effect. Our results also illustrate the value of neuroscience-inspired computer vision algorithms in the study of brain-behavior relationships by revealing the complex interplay between global gain effects and hierarchical sensory processing.

Methods & Materials

Overview

We modelled cortical arousal by augmenting a DCNN with a global gain mechanism (see *Model*) and tested it on two tasks (see *Tasks*). In addition, we also investigated effects on processing by both assessing the model's decodable task information and its activations (see *Block information & activations*) and disrupting its processing (see *Model disruption – Spatial Scrambling*).

Model

As a base architecture for our DCNN, we adopted a first generation ResNet18 (30) that was trained with a sigmoidal-like activation function (31) on the ImageNet database (56) and achieved a top-1 accuracy of 64.04% on the validation set. Commonly, DCNNs use a rectified linear unit (ReLU) as activation function and due to the use of this alternative activation function, performance was slightly lower than models trained with ReLU (typically around ca. 69%). We chose here to use a sigmoidal-like transfer function instead because it better captures the saturating property of biological neurons at extreme values, a characteristic we deemed to be central for studying the effects of gain across a wide range of gain levels, including very high gain levels (57, see Figure 1A).

Global gain mechanism

Arousal has been shown to increase firing rates in early visual cortex in response to sensory stimulation, that is, to increase sensory gain (e.g. 26). These increases in gain are thought to be a result of a global release of noradrenaline, making them spatially unspecific (4, 13, 14, 44, 48).

To simulate such global gain increases in our DCNN, we scaled all activation functions with one global gain parameter γ:

$$f(S) = \gamma \max\left(0, \frac{h}{\exp\left(\frac{c_1 S + c_2}{c_3 S + c_4}\right) - 1} - c_0 + \frac{h}{2}\right) \ ,$$

where $f(S)$ describes the outgoing activation and S is the incoming activation. The second part of the equation describes the sigmoidal-like properties of the activation function where the constants h, c_0, c_1, c_2, c_3 and c_4 were derived from a spiking neuron model, capturing the mapping between the incoming current and the average postsynaptic potential over infinite time steps (see 31, for more details). In brief, we changed the global gain of our model with a single parameter γ, which scaled the output of the activation functions, thereby increasing or reducing response gain everywhere in the model.

Tasks

Parametric perceptual difficulty task

The Yerkes-Dodson effect - the relationship between arousal level and performance as a function of task difficulty - is commonly studied by manipulating sensory difficulty (4, 10, 58). To recreate an analogous situation for our model, we developed a task in which we could parametrically manipulate the visual complexity required to solve a binary discrimination task. This task

required the model to distinguish a real image from an average image, an image that was created by averaging over different numbers of scene images. That is, an average image could either be based on only 1.25 images (average of a normal image and another image with an alpha level of 25%), which rendered it very similar to a real image and thus perceptually difficult to distinguish from a real natural scene image, or it could be based on averaging over up to 40 images, making the average image very dissimilar from a natural scene and thus the discrimination task easier (see Figure 2B for examples). The average images were constructed from a separate set of images than those used as real images, thereby avoiding overlap in image features. All images were taken from a dataset consisting of street scenes, described here (32).

For every level of perceptual difficulty, a separate sigmoidal output node of the DCNN was fine-tuned, while the rest of the weights were kept unchanged and the global gain parameter was set to 1. The respective models were fine-tuned for 50 epochs on 6040 training images and validated using 2592 images, each containing 50% natural scenes for 10 network instances. The data shown in Figure 2 for a global gain of 1 show the performance of these models on the held-out test set (2159 images). To construct the performance-gain profiles depicted in Figure 2, the same test set was evaluated using the same fine-tuned DCNNs.

For evaluating the signal detection properties, we calculated d' and the *criterion* as follows:

$$d' = z(\frac{hits}{hits + misses}) - z(\frac{FA}{FA + CR})$$

$$criterion = -0.5(z(\frac{hits}{hits + misses}) + z(\frac{FA}{FA + CR})),$$

where z corresponds to a z-transform and FA to false alarms and CR to correct rejections.

Visual search with a varying number of options

To assess how the results from the parametric perceptual difficulty task generalized to other scenarios, we also tested the model on a visual search task across two scene contexts in which we varied the number of answer options. The benefit of this was two-fold. First, we could test how our the DCNN responds to another factor of task difficulty, with an eight-way classification being more difficult than a two-way classification. Secondly, we could assess whether we would also observe an inverted U-shaped gain-performance profile for visual search, a perceptually difficult task. We tested the DCNN across two search contexts (street scenes and food scenes). These datasets were curated to be challenging for visual search, while reducing other informative features such as the background (for a detailed description, see 32).

We fine-tuned an output layer with sigmoid nodes separately for each number of answer options (2, 4, 8 categories), while keeping the remaining model weights unchanged. The models were trained on multi-label images for 75 epochs and tested on single-label images. To reduce the number of answer options, but prevent differences between the different answer option conditions in categories included, we randomly sampled with replacement from the 8 categories over 20 iterations thus providing us with the depicted standard deviation in Figure 3 for the conditions 2 and 4.

Block information

We examined the effects of global gain on model block information to determine how our manipulation of arousal, at the level of sensory processing, may have affected performance as a function of perceptual difficulty. To assess the amount of task-related information that is linearly decodable, we used a logistic regression classifier. In particular, we used this decoding approach on the activations from the last activation layer in each ResNet block, thus just before the residual and skip connections split. We ran 5 iterations with randomly drawn training and test datasets of 1000 images each for every combination of level of perceptual difficulty and

gain value. For the logistic regression, we used the standard parameters as implemented in the scikit-learn package (59).

Block activation

Next to examining effects of global gain on block information, we also examined gain effects on task-related activation. To obtain a measure of activation along the hierarchy of the model, we recorded the mean activation across all dimensions for the test set used in Figure 2 for varying degrees of global gain and perceptual difficulty for all trained model instances in the same layers as included in the decoding analyses.

Model disruption – Spatial scrambling

To manipulate the model's processing, we developed a method to selectively reduce the contribution of a given block to the model's performance, named spatial scrambling (Figure 5A). In particular, we leveraged the organizational principle of ResNet blocks, consisting of a residual and a skip branch (see Figure 5A), where the residual branch contains the majority of the additional task-relevant features extracted for a given block (in the convolutional layers). In a functioning ResNet, the features from the residual branch are added back to the skip branch, which did not undergo this extra computational step. The information transmitted via the skip connection is thus identical to the model activations before they entered the residual branch (see Figure 5A). Spatial scrambling interferes with these residual branch features to effectively reduce their contribution to the model's performance. Specifically, we spatially scrambled the feature maps and did so with an increasing proportion of all feature maps. Thereby spatial scrambling maintained the distributional statistics within a feature map (in contrast to other methods such as Dropout (60) or lesioning), but interfered with the information contained therein.

By increasing the proportion of randomly chosen feature maps affected by spatial scrambling while probing the network for its performance, it is possible to estimate how robust the network is to these manipulations while at a neutral gain of 1. This is illustrated in Figure S3, which shows how performance drops as a function of features maps being targeted by spatial scrambling (importance curves). The gradual decay of these curves informs about the relevance of these model features for performance. For instance, a model's block importance curve that quickly falls off as a function of an increased proportion indicates that this feature was essential to the model's overall performance. For the main experiments in which we also varied global gain, we determined the proportion of randomly chosen feature maps at which a model block dropped to 20% of its possible contribution to network performance (dots in Figure S3). This ensured that applying spatial scrambling to every targeted ResNet-block had approximately the same relative impact on model performance.

In a next step, we used this approach to dissect which part of the model drove the performance-gain profiles to be optimal at medium or heightened global gain states. For this analysis, we now only applied the same spatial scrambling rate identified based on the importance curves for every block, but changed the global gain of the model. In particular, we were interested to dissociate the influence of early versus later network blocks on the performance-gain profiles. To this end, we either gradually disrupted later model blocks first and successively added earlier model blocks or the other way around, thereby either leaving mostly earlier or later model blocks intact. Doing this for different task difficulties then allowed us to inspect the performance-gain profiles as a function of disrupted early or later network block function. All spatial scrambling experiments were repeated ten times on the same set of training weights.

Software

All code for reproducing these results will be made available upon publication. Furthermore, the results presented here were obtained while relying on the following Python packages: NumPy (61),

keras (62), TensorFlow (63), Pandas (64), Scikit-Learn (59) and SciPy (65). Data visualization was done using matplotlib (66) and, in particular, seaborn (67).

Acknowledgements & funding sources

References

1. Leonhard Waschke, Niels A Kloosterman, Jonas Obleser, and Douglas D Garrett. Behavior needs neural variability. *Neuron*, February 2021.

2. Garrett T Neske, Dennis Nestvogel, Paul J Steffan, and David A McCormick. Distinct waking states for strong evoked responses in primary visual cortex and optimal visual detection performance. *J. Neurosci.*, 39(50):10044–10059, December 2019.

3. Matthew J McGinley, Stephen V David, and David A McCormick. Cortical membrane potential signature of optimal states for sensory signal detection. *Neuron*, 87(1):179–192, July 2015.

4. Matthew J McGinley, Martin Vinck, Jacob Reimer, Renata Batista-Brito, Edward Zagha, Cathryn R Cadwell, Andreas S Tolias, Jessica A Cardin, and David A McCormick. Waking state: Rapid variations modulate neural and behavioral responses. *Neuron*, 87(6): 1143–1161, September 2015.

5. Hagar Gelbard-Sagiv, Efrat Magidov, Haggai Sharon, Talma Hendler, and Yuval Nir. Noradrenaline modulates visual perception and late visually evoked activity. *Curr. Biol.*, 28(14):2239–2249.e6, July 2018.

6. T Pfeffer, A Ponce-Alvarez, T Meindertsma, C Gahnström, R L van den Brink, G Nolte, K Tsetsos, A K Engel, G Deco, and T H Donner. Dissociation between catecholamines and acetylcholine in the human cerebral cortex.

7. T Pfeffer, Arthur-Ervin Avramiea, Guido Nolte, Andreas K Engel, Klaus Linkenkaer-Hansen, and Tobias H Donner. Catecholamines alter the intrinsic variability of cortical population activity and perception. *PLoS Biol.*, 16(2):e2003453, February 2018.

8. Ella Podvalny, Matthew W Flounders, Leana E King, Tom Holroyd, and Biyu J He. A dual role of prestimulus spontaneous neural activity in visual object recognition. *Nat. Commun.*, 10(1):3910, September 2019.

9. Jan Willem de Gee, Konstantinos Tsetsos, Lars Schwabe, Anne E Urai, David McCormick, Matthew J McGinley, and Tobias H Donner. Pupil-linked phasic arousal predicts a reduction of choice bias across species and decision domains. *Elife*, 9, June 2020.

10. Robert M Yerkes and John D Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Punishment: Issues and experiments*, pages 27–41, 1908.

11. Karl Halvor Teigen. Yerkes-Dodson: A law for all seasons. *Theory Psychol.*, 4(4):525–547, November 1994.

12. Jochem van Kempen, Gerard M Loughnane, Daniel P Newman, Simon P Kelly, Alexander Thiele, Redmond G O'Connell, and Mark A Bellgrove. Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. *Elife*, 8, March 2019.

13. Gary Aston-Jones and Jonathan D Cohen. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.*, 28:403–450, 2005.

14. Susan J Sara. The locus coeruleus and noradrenergic modulation of cognition. *Nat. Rev. Neurosci.*, 10(3):211–223, March 2009.

15. Jacob Reimer, Matthew J McGinley, Yang Liu, Charles Rodenkirch, Qi Wang, David A McCormick, and Andreas S Tolias. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.*, 7:13289, November 2016.

16. http://farm2.staticflickr.com/1196/1089845176_c9b801237d_z.jpg. `http://farm2.staticflickr.com/1196/1089845176_c9b801237d_z.jpg`, . Accessed: 2021-5-17.

17. Rufin VanRullen. Perception science in the age of deep neural networks. *Front. Psychol.*, 8:142, February 2017.

18. Michael Eickenberg, Alexandre Gramfort, Gaël Varoquaux, and Bertrand Thirion. Seeing it all: Convolutional network layers map the function of the human visual system. *Neuroimage*, 152:184–194, May 2017.

19. Umut Güçlü and Marcel A J van Gerven. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.*, 35(27): 10005–10014, July 2015.

20. Seyed Mahdi Khaligh-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Comput. Biol.*, 10 (11):e1003915, November 2014.

21. Jonas Kubilius, Martin Schrimpf, Aran Nayebi, Daniel Bear, Daniel L K Yamins, and James J DiCarlo. CORnet: Modeling the neural mechanisms of core object recognition. September 2018.

22. Martin Schrimpf, Jonas Kubilius, Michael J Lee, N Apurva Ratan Murty, Robert Ajemian, and James J DiCarlo. Integrative benchmarking to advance neurally mechanistic models of human intelligence. *Neuron*, September 2020.

23. Daniel L K Yamins, Ha Hong, Charles F Cadieu, Ethan A Solomon, Darren Seibert, and James J DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.*, 111(23):8619–8624, June 2014.

24. Deep neural networks in computational neuroscience. In Tim C Kietzmann, Patrick Mc-Clure, and Nikolaus Kriegeskorte, editors, *Oxford Research Encyclopedia of Neuroscience*. Oxford University Press, January 2019.

25. H Steven Scholte. Fantastic DNimals and where to find them. *Neuroimage*, 180(Pt A): 112–113, October 2018.

26. Martin Vinck, Renata Batista-Brito, Ulf Knoblich, and Jessica A Cardin. Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding. *Neuron*, 86(3):740–754, May 2015.

27. Jacob Reimer, Emmanouil Froudarakis, Cathryn R Cadwell, Dimitri Yatsenko, George H Denfield, and Andreas S Tolias. Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron*, 84(2):355–362, October 2014.

28. Patrick J Mineault, Elaine Tring, Joshua T Trachtenberg, and Dario L Ringach. Enhanced spatial resolution during locomotion and heightened attention in mouse primary visual cortex. *J. Neurosci.*, 36(24):6382–6392, June 2016.

29. Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, 364(6437):255, April 2019.

30. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. December 2015.

31. Davide Zambrano, Roeland Nusselder, H Steven Scholte, and Sander M Bohté. Sparse computation in adaptive spiking neural networks. *Front. Neurosci.*, 12:987, 2018.

32. Lynn K A Sörensen, Davide Zambrano, Heleen A Slagter, Sander M Bohté, and H Steven Scholte. Leveraging spiking deep neural networks to understand neural mechanisms underlying selective attention. December 2020.

33. http://farm1.staticflickr.com/89/225176986_5300ac2dd7_z.jpg. `http://farm1.staticflickr.com/89/225176986_5300ac2dd7_z.jpg`, . Accessed: 2021-5-17.

34. http://farm3.staticflickr.com/2481/3650599735_2d4cde89f8_z.jpg. `http://farm3.staticflickr.com/2481/3650599735_2d4cde89f8_z.jpg`, . Accessed: 2021-5-17.

35. http://farm8.staticflickr.com/7059/6773541522_c6f3d55e6f_z.jpg. `http://farm8.staticflickr.com/7059/6773541522_c6f3d55e6f_z.jpg`, . Accessed: 2021-5-17.

36. Josef Faller, Jennifer Cummings, Sameer Saproo, and Paul Sajda. Regulation of arousal via online neurofeedback improves human performance in a demanding sensory-motor task. *Proc. Natl. Acad. Sci. U. S. A.*, 116(13):6482–6490, March 2019.

37. Leonhard Waschke, Sarah Tune, and Jonas Obleser. Local cortical desynchronization and pupil-linked arousal differentially shape brain states for optimal sensory performance. *Elife*, 8, December 2019.

38. Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C Lawrence Zitnick, and Piotr Dollár. Microsoft COCO: Common objects in context. May 2014.

39. http://farm1.staticflickr.com/51/148868777_6c1ba30f06_z.jpg. `http://farm1.staticflickr.com/51/148868777_6c1ba30f06_z.jpg`, . Accessed: 2021-5-17.

40. http://farm1.staticflickr.com/159/330605195_1f467a4cdd_z.jpg. `http://farm1.staticflickr.com/159/330605195_1f467a4cdd_z.jpg`, . Accessed: 2021-5-17.

41. http://farm9.staticflickr.com/8334/8404801550_f68bf901ee_z.jpg. `http://farm9.staticflickr.com/8334/8404801550_f68bf901ee_z.jpg`, . Accessed: 2021-5-17.

42. http://farm6.staticflickr.com/5293/5538417341_f71df00d6b_z.jpg. `http://farm6.staticflickr.com/5293/5538417341_f71df00d6b_z.jpg`, . Accessed: 2021-5-17.

43. David M Diamond, Adam M Campbell, Collin R Park, Joshua Halonen, and Phillip R Zoladz. The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plast.*, 2007:60803, 2007.

44. James F A Poulet and Sylvain Crochet. The cortical states of wakefulness. *Front. Syst. Neurosci.*, 12:64, 2018.

45. M Usher, J D Cohen, D Servan-Schreiber, J Rajkowski, and G Aston-Jones. The role of locus coeruleus in the regulation of cognitive performance. *Science*, 283(5401):549–554, January 1999.

46. D Servan-Schreiber, H Printz, and J D Cohen. A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*, 249(4971):892–895, August 1990.

47. Siddhartha Joshi and Joshua I Gold. Pupil size as a window on neural substrates of cognition. *Trends Cogn. Sci.*, 24(6):466–480, June 2020.

48. David A McCormick, Dennis B Nestvogel, and Biyu J He. Neuromodulation of brain state and behavior. *Annu. Rev. Neurosci.*, 43:391–415, July 2020.

49. Daniel L K Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.*, 19(3):356–365, March 2016.

50. Rishi Rajalingham, Elias B Issa, Pouya Bashivan, Kohitij Kar, Kailyn Schmidt, and James J DiCarlo. Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *J. Neurosci.*, July 2018.

51. Angela J Yu and Peter Dayan. Uncertainty, neuromodulation, and attention. *Neuron*, 46 (4):681–692, May 2005.

52. Mara Mather, David Clewett, Michiko Sakaki, and Carolyn W Harley. Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behav. Brain Sci.*, 39:e200, January 2016.

53. Peter R Murphy, Evert Boonstra, and Sander Nieuwenhuis. Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nat. Commun.*, 7:13526, November 2016.

54. Jan Willem de Gee, Tomas Knapen, and Tobias H Donner. Decision-related pupil dilation reflects upcoming choice and individual bias. *Proc. Natl. Acad. Sci. U. S. A.*, 111(5): E618–25, February 2014.

55. Jan Willem de Gee, Olympia Colizoli, Niels A Kloosterman, Tomas Knapen, Sander Nieuwenhuis, and Tobias H Donner. Dynamic modulation of decision biases by brainstem arousal systems. *Elife*, 6, April 2017.

56. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.*, 115(3):211–252, December 2015.

57. Ken-Ichi Naka and William A H Rushton. S-potentials from luminosity units in the retina of fish (cyprinidae). *J. Physiol.*, 185(3):587–599, 1966.

58. P L Broadhurst. Emotionality and the Yerkes-Dodson law. *J. Exp. Psychol.*, 54(5): 345–352, November 1957.

59. Fabian Pedregosa. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12: 2825–2830, 2011.

60. Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. July 2012.

61. Charles R Harris, K Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández Del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E Oliphant. Array programming with NumPy. *Nature*, 585 (7825):357–362, September 2020.

62. F Chollet. keras. 2015.

63. Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-Scale machine learning on heterogeneous distributed systems. March 2016.

64. Wes McKinney and Others. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. conference.scipy.org, 2010.

65. Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J van der Walt, Matthew Brett, Joshua Wilson, K Jarrod Millman, Nikolay Mayorov, Andrew R J Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E A Quintero, Charles R Harris, Anne M Archibald, Antônio H Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A Nicholson, David R Hagen, Dmitrii V Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G Young, Gavin A Price, Gert-Ludwig Ingold, Gregory E Allen, Gregory R Lee, Hervé Audren, Irvin Probst, Jörg P Dietrich, Jacob Silterra, James T Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A Brodtkorb, Perry Lee, Robert T McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J Pingel, Thomas P Robitaille, Thomas Spura, Thouis R Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O Halchenko, Yoshiki Vázquez-Baeza, and SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods*, 17(3):261–272, March 2020.

66. J D Hunter. Matplotlib: A 2D graphics environment. *Computing in Science Engineering*, 9(3):90–95, May 2007.

67. Michael Waskom, Olga Botvinnik, Maoz Gelbart, Joel Ostblom, Paul Hobson, Saulius Lukauskas, David C Gemperline, Tom Augspurger, Yaroslav Halchenko, Jordi Warmenhoven, John B Cole, Julian de Ruiter, Jake Vanderplas, Stephan Hoyer, Cameron Pye, Alistair Miles, Corban Swain, Kyle Meyer, Marcel Martin, Pete Bachant, Eric Quintero, Gero Kunter, Santi Villalba, Brian, Clark Fitzgerald, C G Evans, Mike Lee Williams, Drew O'Kane, Tal Yarkoni, and Thomas Brunner. mwaskom/seaborn: v0.11.0 (sepetmber 2020), September 2020.
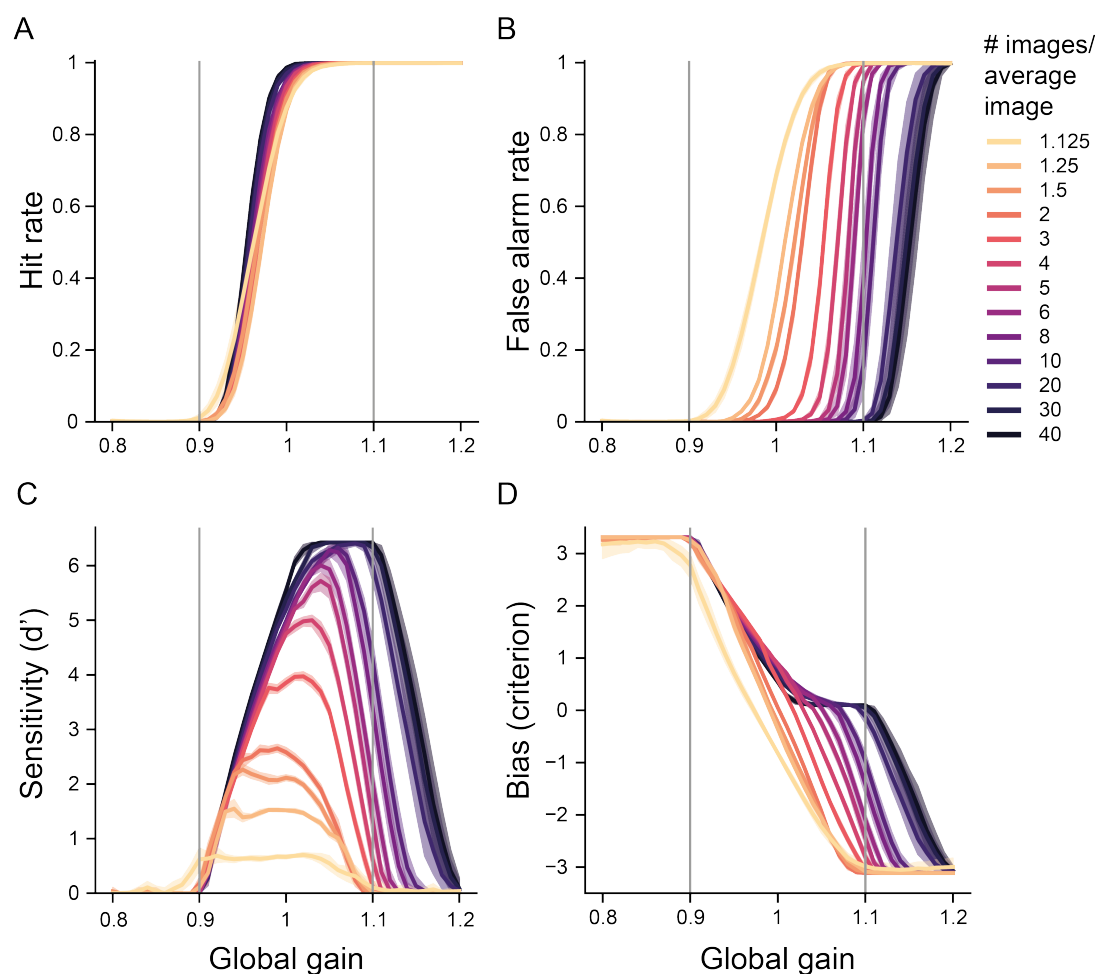
Supplementary figures

Figure S1. Signal detection properties as a function of perceptual difficulty and global gain. Shaded areas depict the 95% confidence interval across ten model instances. The grey vertical lines are in reference to Figure 2D and serve to illustrate the results for a more narrowly sampled global gain range.
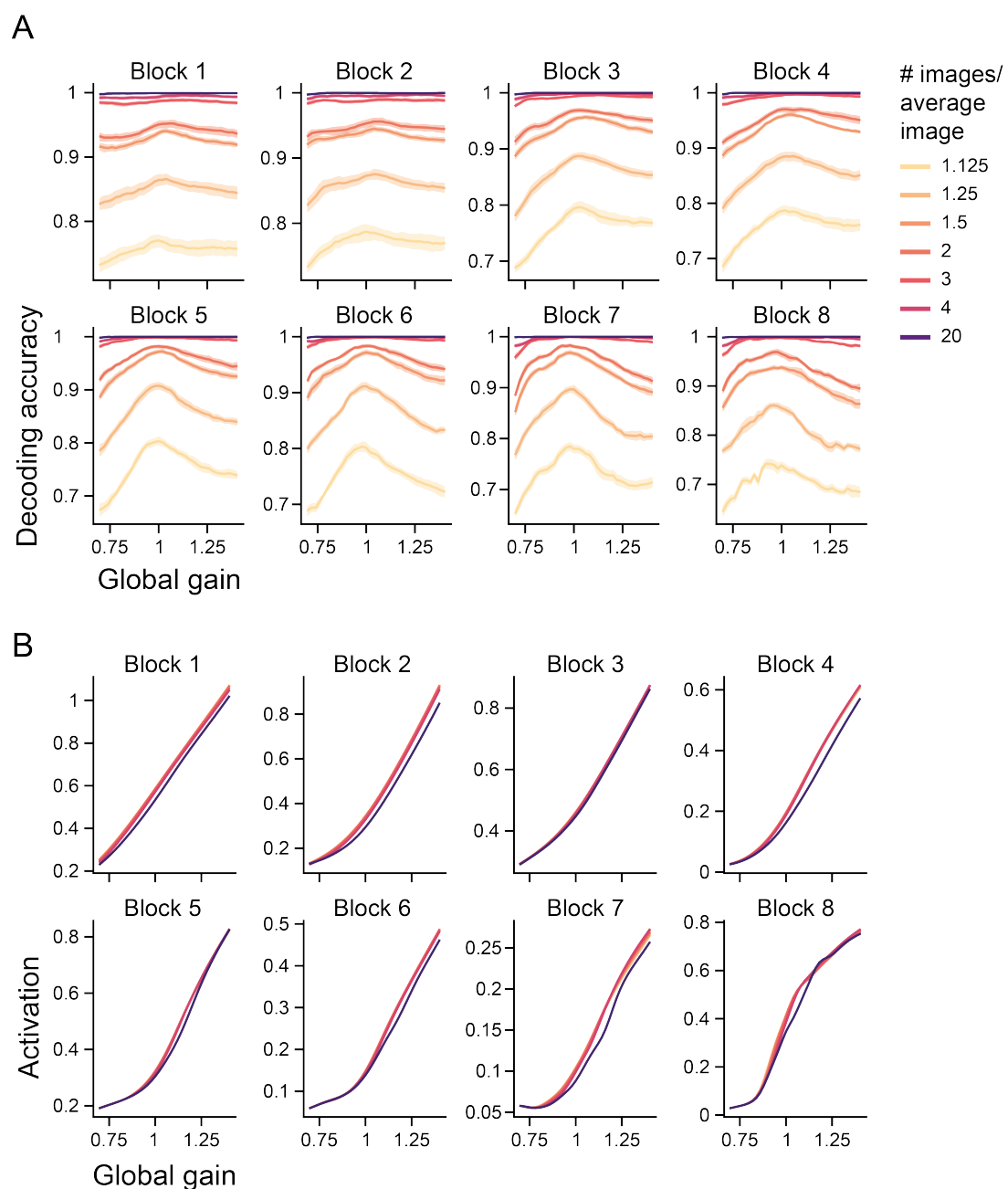
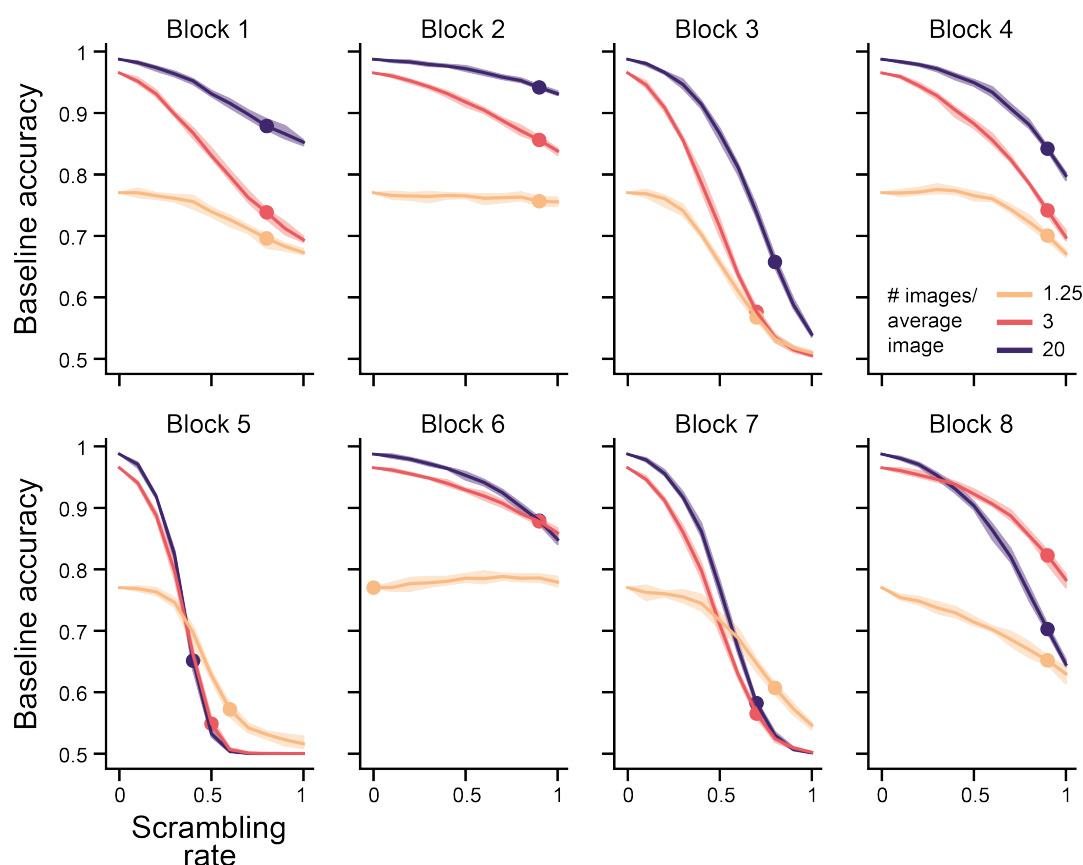**Figure S2. Global gain changes affect task-related information and activation.**

Figure S3. Importance curves obtained from spatial scrambling at neutral gain states. The dots represent the scrambling rate at which 20% of the baseline performance is maintained.
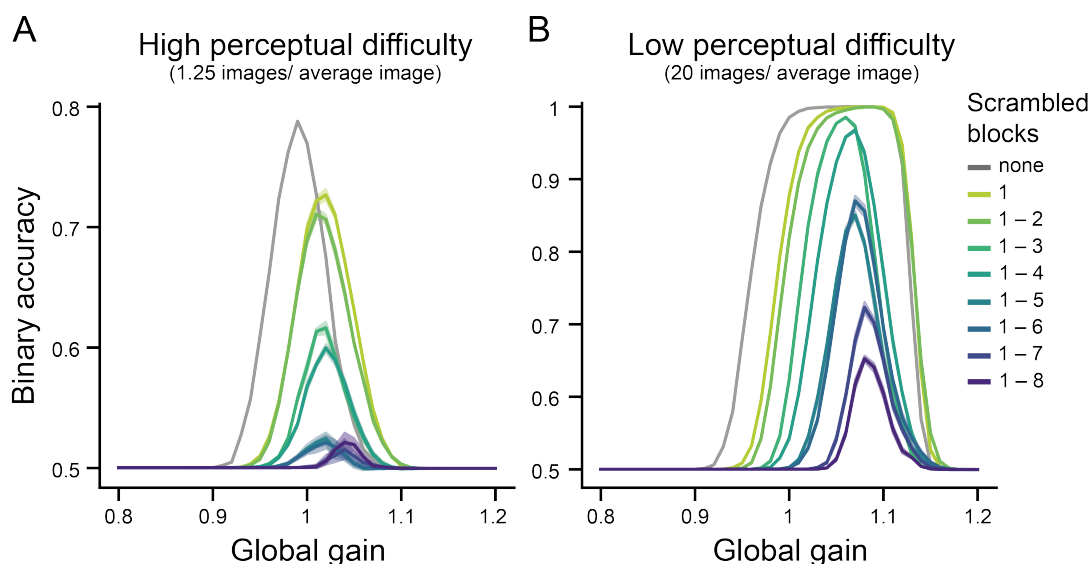

Figure S4. Performance-gain profiles for scrambling early to late network blocks. A shows a perceptually difficult task and B a perceptually easy task. All figure conventions are the same as in Figure 5.