bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Mice use robust and common strategies to discriminate natural scenes

Yiyi Yu*[,1], Riichiro Hira*[,1], Jeffrey N. Stirman[1], Waylin Yu[2], Ikuko T. Smith[2], Spencer L. Smith[1,3,4]

** Co-first authors*

[1] Neuroscience Center

[2] Department of Pharmacology

[3] Department of Cell Biology and Physiology

[4] Carolina Institute for Developmental Disabilities

University of North Carolina - Chapel Hill

Chapel Hill, NC 27517

Correspondence should be addressed to S.L.S.

slab@unc.edu

+1 919-966-2349

111 Mason Farm Road

Campus Box 7545

Chapel Hill, NC 27599-7545

USA

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Visual perception is essential for animal survival in natural environments. Mice use vision to navigate and avoid predators. However, the spatial resolution of mouse vision is poor compared to primates and carnivores, and mice lack a fovea. Thus, it is unclear how well mice can discriminate ethologically relevant scenes. Here, we examined natural scene discrimination in mice using an automated touch-screen system, which measured the discriminability for sets of target and distractor scenes. We found that a conventional metric for estimating image resemblance, structural similarity (SSIM), reasonably predicted the discrimination performance. Mouse-to-mouse consistency was high, and the performance of each mouse was better predicted by the population mean than SSIM. This high inter-mouse agreement indicates that mice use common and robust strategies to discriminate natural scenes. We tested several alternative image metrics to find an alternative to SSIM for predicting mouse discrimination performance. We found that a primary visual cortex (V1) model-inspired approach predicted mouse performance with fidelity comparable to the inter-mouse agreement. The model was based on convolving the images with Gabor filters, and its performance was orientation-specific. The orientation-specificity was stimulus-dependent. These results indicate that, compared to artificial parameters like SSIM, neurophysiological-based models of V1 processing can better predict visual discrimination behavior of naturalistic stimuli.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1. Introduction

Visual processing of natural scenes is essential for animal survival. Mammalian visual systems including primates and rodents evolved to efficiently process natural stimuli (Olshausen 1996, Bell and Sejnowski 1997, Itti, Koch et al. 1998, Vinje and Gallant 2000, Froudarakis, Berens et al. 2014). Mice use vision to hunt prey (Hoy, Yavorska et al. 2016) and navigate (Harvey, Collman et al. 2009).

A number of studies have characterized the behavioral performance of mice in discriminating visual stimuli including simple shapes (Bussey, Saksida et al. 2001), gratings, (Prusky, West et al. 2000), shapes (Glickfeld, Histed et al. 2013, Long, Jiang et al. 2015, Nithianantharajah, McKechanie et al. 2015), and random dot kinematograms (Stirman, Townsend et al. 2016). However, these results cannot be extrapolated to natural scene discrimination because visual coding is substantially different between natural images and artificial ones (Vinje and Gallant 2000, Froudarakis, Berens et al. 2014). Moreover, the spatial resolution of mouse vision is orders of magnitude lower than that of primates and carnivores. Even when natural scenes might be discriminated, individual mice might focus on different regions of the images to discriminate them, and this would lead to high mouse-to-mouse variability. Thus, the behavioral investigation using natural images provides essential information for understanding evolutionally optimized encoding strategy of visual systems.

The perception of visual information depends on a hierarchical processing by visual cortical areas (Felleman and Van Essen 1991, Andermann, Kerlin et al. 2011, Marshel, Garrett et al. 2011). Visual information is first processed in the primary visual cortex (V1). One remarkable feature of the primary visual cortex is the orientation specific processing

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Hubel and Wiesel 1962, Ohki, Chung et al. 2005). This would facilitate the sparse coding of natural images by V1 neurons (Olshausen 1996, Bell and Sejnowski 1997, Itti, Koch et al. 1998). Orientation specific features are further transformed and integrated in higher visual areas to extract higher order statistical structures of the image and detect objects (Young and Yamane 1992, Quiroga, Reddy et al. 2005, Serre, Wolf et al. 2007). Thus, the orientation feature selectivity is the foundation of visual perception in many contexts. However, it is unclear how orientation features in naturalistic images are perceived and whether they are relevant to the animal behavior.

Here, we developed a natural image discrimination task for freely moving mice using an automated touchscreen-based system that we recently established for global motion discrimination task (Stirman, Townsend et al. 2016). The mice successfully and quickly learned to discriminate images of natural scenes. We investigated which features of natural images including orientation features explained the discriminability by mice, and discussed how the mice discriminated the natural images.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2. Methods

2.1 Subjects

Ten adult C57BL/6 mice (2 males and 4 females in main experiments and 2 males and 2 females in additional experiments) were used in the experiments reported here. Animals were between 80 and 160 days old at the start of training, which lasted for approximately 1 month. Mice were housed in rooms on reversed light-dark cycle (dark during the day, room light on at night), and training and testing were performed during the dark cycle of the mice. All training and procedures were reviewed and approved by the Institutional Animal Care and Use Committee of the University of North Carolina, which ensures compliance with the standards of the American Association for Laboratory Animal Science (AALAS).

2.2 Apparatus

The operant chamber and controlling devices were the same as previously described (Stirman et al., 2016), and are based on work by Saksida and Bussey (Bussey, Saksida et al. 2001, Bussey, Padain et al. 2008, Nithianantharajah, McKechanie et al. 2015). Briefly, a touchscreen panel was on the long size of a trapezoidal chamber, opposite of liquid reward port (a strawberry-flavored yogurt-based drink; Kefir). Correct responses were indicated by a brief auditory tone, and the reward port was illuminated. During time outs, a brief burst of white auditory noise was played, and the chamber light was illuminated for the duration of the time-out.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.3 Stimuli

For the natural scene discrimination task, we selected pictures of natural scenes (430 by 430 pixels static images, JPEG format). The original set of images was taken from three naturalistic image databases: UPenn (http://tofu.psych.upenn.edu/~upennidb/) (Tkačik, Garrigan et al. 2011), McGill (http://tabby.vision.mcgill.ca/) (Olmos and Kingdom 2004), and MIT (http://cvcl.mit.edu/database.htm) (Oliva and Torralba 2001). Average luminance of images was almost the same. One image was set as the target image during NID (naturalistic image discrimination) training sessions. Ten images which have low structure similarity (SSIM) with the target image, and eleven images with varied SSIM plus one image which is the same as the target image were selected as distractor images on the NID training and NID testing sessions, respectively (**Fig.1; Supplementary table 1.)**. All images were masked by a circular aperture. During NID testing sessions, 5 interleaved training trials were provided after every 12 testing trials. One image with low SSIM of the eleven distractors was used as the distractor for the interleaved training.

2.4 Behavioral Training

Food restriction and training stages 1-3 (FR, free reward; MT, must touch; IM image discrimination) were conducted as previously described (Stirman, Townsend et al. 2016). Briefly, during DR (Free Reward) phase (training stage 1), mice learned to lick the reward spout to receive a reward. As a result, mice associated the tone with the delivery of a reward, and learned the location of the reward. During MT (Must Touch) phase (training stage 2), mice had to touch any location on the screen at the front of the box to receive a reward. The goal of this phase was to associate touching the screen with delivery of a reward. During IM (Image Discrimination) phase (training stage 3), a dot and fan image

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

pair stimulus (Bussey, Saksida et al. 2001) was presented and mice required to touch a specific target stimulus on the screen to earn a reward.

Training stage 4: Natural image discrimination (NID)

Before this phase, mice were trained to discriminate two static images (dots vs fan). On this phase, the mice had to touch the target natural image, avoiding one of 10 distractors. SSIM index of all pairs were less than 0.2 (low SSIM implies that the two images are not similar). This stage of training incorporated correction trials as described previously (Stirman, Townsend et al. 2016).

2.5 *Behavioral Testing*

This condition was similar to the testing condition of the kinematogram we previously described (Stirman, Townsend et al. 2016). Images for testing phase consisted of one target image and 12 distractor images whose SSIM indices were 0.074 to 0.54. Testing consisted of interleaved blocks of testing and training. Mice were not cued as to whether they were in a testing or training block. During testing blocks, all 12 distractors (12 trials per block) were presented in a random order, and all answers were rewarded. During training blocks (5 trials per block), stimuli with SSIM = 0.13 were presented and normal performance feedback (including timeout periods and correction trials) was provided. Performance during these interleaved training blocks served as an internal control to ensure the mice were still working to perceive the direction of the NID stimuli in the testing blocks. Testing data was analyzed only if the animal performed on average at criteria (≥ 85% correct) during the interleaved training blocks. This criterion excluded one mouse out of the six mice in the main experiments. Testing sessions were conducted for 120 -

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

150 minutes per day. In additional experiments, the target and rotated images were rotated only in the NID training and testing sessions. All 5 mice generated testing data in the additional experiments.

2.6 *SSIM (Structural similarity)*

SSIM indices for all pairs of presented natural stimuli were calculated as reported by Wang et al 2004 (Wang, Bovik et al. 2004). First, the pair of two images were smoothed with Gaussian filter (σ = 1.5 pixels). SSIM index was obtained for each window of a pair of images. First, the SSIM for each pixel was calculated for square windows, centered at the same pixel (w,h) of two images. The length of a side of square window was 8. The SSIM (*x,y*) was then obtained for two windows in a target image (*x*) and a distractor image (*y)* as follows,

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

,

where μ_x, μ_y are pixel intensity averages of window *x* and *y*, σ_x, σ_y, σ_{xy} are standard deviation of window *x* and *y*, and covariance of two windows, c_1 and c_2 are $(0.01 \times L)^2$ and $(0.03 \times L)^2$, respectively, where *L* is 255 (corresponding to the dynamic range of 8-bit monochrome images).

SSIM(*x,y*) was obtained for all (w,h) and was averaged over the circular area where the natural images are located. The average of SSIM(*x,y*) was called simply as SSIM index for each pair of target and distractor in this paper (**Fig.1**). One distractor was the same as a target image, so that its SSIM index is 1. The pairs of the other distractors and a target image had SSIM index less than 1.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

2.7 Psychometric curve

Psychometric curve was obtained with regression of the following function (Weibull function) to the data,

$$CorrectRate(SSIM) = 0.5 + \frac{0.5}{1 + exp(\frac{SSIM-\alpha}{\beta})}$$

where α and β are parameters determined by the regression to each data set. The threshold was set to 70 % accuracy value of SSIM estimated by this function.

2.8. Image processing

Pixel-wise correlation and root-mean-squared error (RMSE) between a target image and each distractor image were obtained as candidate parameters that may capture the similarity of two images. The registration of two images were done with a MSE based registration algorithm (Turboreg) (Thevenaz, Ruttimann et al. 1998).

2.9. Gabor filter

We used Gabor filtering to analyze the orientation specific features of the image.

$$g_{\lambda,\theta,\gamma,\sigma}(x,y) = \exp\left(-\frac{x' + \gamma y'}{2\sigma^2}\right) * \exp(i(2\pi\frac{x'}{\lambda}))$$

The Gabor filter bank was generated using the built-in Matlab function 'gabor'. The wavelength (λ) of the Gabor filters ranged from 5 – 28 pixels per cycle. The orientation (θ) of Gabor filters is either 0°, 30°, 60°, 90°, 120°, and 150°. We set the aspect ratio (γ) of Gabor filter to 0.5. The standard deviation of the Gaussian envelops (σ) is $\sigma = \frac{\lambda}{\pi} * \sqrt{\frac{\log(2)}{2}} * 3$ is determined by the wavelength. We computed the Gabor feature magnitude ($G_{\lambda,\theta}(x,y)$) by convolving Gabor filter at specific orientation and wavelength with each

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

image. Then we compute the orientation specific similarity (OSS) between two images by the following function.

$$OSS(\lambda,\theta) = \frac{1}{N} * \sum_{N\ pixels} \frac{2 * G_{\lambda,\theta}(x,y)_{distractor} * G_{\lambda,\theta}(x,y)_{target}}{(G_{\lambda,\theta}(x,y)_{distractor})^2 + (G_{\lambda,\theta}(x,y)_{target})^2}$$

.

2.10. Statistics.

Student's t-test and Spearman's correlation test were used for statistical comparisons. Pairwise comparisons were two-tailed unless otherwise noted. Error bars in graphs represent the mean ± s.e.m unless otherwise noted. Analysis of variance (ANOVA) was used for multiple comparisons, followed by t-test (presented p-values for t-test were Bonferroni corrected). No statistical tests were performed to predetermine sample size. No blinding or randomization were performed.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

3. Results

3.1. *Training*

All mice trained in the main experiment (6 of 6) successfully passed the pre-training phases (FR, MT, & IM), meeting criteria to advance to NID training in 14.5 ± 2.9 days (mean ± S.D.) (**Fig. 2a**). These mice also readily acquired the NID training task (6 of 6), ultimately discriminating correctly between a natural target image and 10 distractor images on 85% or more of trials (**Fig. 2b, c**). Two out of 6 mice were trained for an hour per day, and the other 4 mice were trained for two hours per day. Once mice performed the NID training task with 85% accuracy for 2 consecutive days, they moved to the NID testing phase. Mice required fewer training sessions to reach criterion for NID compared to the mice trained in the random dot kinematogram (RDK) task we previously reported (Stirman, Townsend et al. 2016) (3, 3, 3, 5, 8, 9 days for NID vs 5, 10, 11, 14, 15 and 18 days for RDK; $p = 0.0088$, *t*-test).

3.2. *Testing*

Five out of six mice retained the correct trial rate with > 85 % in the interleaved training trials during testing phase, but one mouse (Mouse 6) had a lower correct rate for interleaved testing trials. This mouse seemed to recognize that it does not have to touch the target image to get rewards during testing phase, and it always tended to select the left panel (**Fig. 3a,b**). Accordingly, we excluded this mouse, and analyzed data from five mice. We obtained the correct trial rate with all five animals for the testing sessions (range: 1920 - 2712 trials, over 4 – 8 sessions per mouse). Many numbers of repeated trials with the same sets of a target and distractor enabled us to precisely estimate the correct trial rate. Psychometric curves were obtained using correct trial rate and SSIM

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

indices for 12 distractor images (**Fig. 4a**). The performances of the mice were remarkably similar (the thresholds of the psychometric curves were 0.29, 0.30, 0.27, 0.34 and 0.32; **Fig. 4b**). In total, the SSIM index approximated the correct trial rate, and the threshold was ~0.30 (**Fig. 4c**).

Next, we explored the relation between response time (RT) and correct rate. The mean RT varied mouse to mouse, but each mouse tended to have long RTs for the distractor images with high SSIMs and high correct rates (**Fig. 5a**). To examine the relative RT values of different mice, we calculated the z-score of the RT for each mouse. The normalized RT was negatively correlated with the SSIM index, and positively correlated with the correct rate (**Fig. 5b,c**). Thus, the mice spent a longer time to discriminate easy trials than difficult ones. This unexpected result will be examined in the Discussion section.

The SSIM index approximated the correct rate, but correct rate of some of the distractor images were better or worse than predicted by the psychometric curve based on SSIM (**Fig. 6a**). Notably, this deviation was not due to outliers of a few mice, but all mice similarly behaved. In fact, correct rate of each mouse was highly predictable by the mean correct rate of the other four animals (**Fig. 6b**). Root mean squared error (RMSE) for linear regression of correct rate for each mouse by the mean correct rate of the other mice were significantly smaller than the RMSE values between psychometric curve and real data (**Fig. 6c**; $p = 0.0037$, paired *t*-test). These results indicate that the mice visual system has rules that is not fully captured by SSIM values.

The mouse visual system does not have sufficient acuity with which to distinguish each pixel of the touch screen in this apparatus. Thus, mice might have to discriminate two images with lower resolution representations than those we presented

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

on the video screen. To investigate the visual acuity dependency, we filtered the image with Gaussian filter with standard deviation 2 to 64 pixels (**Fig. 7a, b**), and obtained SSIM after Gaussian filtering (fltSSIM). The fltSSIM reached minimal when the standard deviation was 4.6 pixels, but the improvement was small.

SSIM is commonly used for estimating image similarity, but is vulnerable to image translation. To robustly estimate the similarity between two images, we translated images so that the mean squared difference between two images was the lowest. The SSIM values were then obtained between target image and translated distractor images (SSIM after registration, regSSIM). However, the predictability of regSSIM was comparable to that of original SSIM, and not as good as the correct rate of the other mice (**Fig. 7c**).

We used other metrics to measure the similarity of two images (pixel-wise cross-correlation, and RMSE). These parameters were also not as good predictor of correct rate as the correct rate of the other mice (**Fig. 7d,e**). Finally, we suspected that mice tended to avoid images similar to the distractor of the interleaved training phase ('anti-target'), because mice were only punished when they mistakenly selected interleaved 'anti-target' images during the NID testing. In this case, larger the SSIM index between distractor and the anti-target should lead to higher the correct trial rate. However, SSIM index between 11 the other distractors and anti-target was not positively correlated with correct rate (**Fig. 7f**), indicating that the mice did not have tendency to avoid the image which is similar to the distractor of training phase during testing sessions. These results indicate that the mouse visual system has unpredictable rules, by which the mouse judges the similarity between two images.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

So far, we have explored the parameters that may capture the global similarity between two images. SSIM and the other parameters predicted mice behavior, but it is not an intuitive model for the mouse visual pathway. Accordingly, we tried to predict the behavior performance with a model mimicking neuronal activity of mouse V1. Previous studies showed that receptive field of the simple neurons in mouse V1 can be modeled as Gabor filters (Bonin, 2011). Each Gabor filter extract a small patch of image with specific orientation and wavelength (**Fig. 8a**). We filtered the target and distractor images with a Gabor filter with various orientation and wavelength, and calculated the similarity of the filtered target image and each distractor image. We defined this similarity as the orientation specific similarity (OSS) (**Fig. 8b**), which is a function of the orientation and wavelength. The prediction of mouse behavior OSS was fit by psychometric curve similarly as the SSIM model (**Fig. 8c**). We obtained RSME value for each set of wavelength and orientation parameters (**Fig. 8e,f**). The RMSE of OSS model averaged over all orientations reached minimal value when the wavelength of Gabor filters is 7.07 – 10 pixels (**Fig. 8d**), or the sigma of the Gaussian envelop if 2.6 – 3.7 (methods), which is comparable to the optimal Gaussian filter size for the SSIM model (standard deviation, 4.6 pixels; **Fig. 7b**). Average of the OSS over all orientations with optimal wavelength predicted animal behavior was similar to the SSIM model after optimally filtered with Gaussian function (RMSE of OSS, 0.082 ± 0.003 (λ = 7.07); RMSE of SSIM, 0.074 ± 0.003; t-test, p=0.08). These results indicated that Gabor filtering averaging over all orientations had similar predictability of mouse behavior.

If the orientation specific information was used in the image discrimination task, we wondered whether either of the individual orientation would predict the performance better than other orientations. Interestingly, the prediction RMSE of OSS at different

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

orientation was significantly varied (**Fig. 8e**, ANOVA $p = 9.0 \times 10^{-11}$). The horizontal OSS was the worst parameter for predicting the correct trial rate regardless of the spatial scale of the Gabor fitters, while the near-vertical OSS predicted the animal behavior more accurately than the SSIM model and the averaged OSS over all orientations. The orientation bias was consistent over a wide range of Gabor filter wavelength (**Fig. 8f**). These results suggest that mice use orientation specific features in the naturalistic image discrimination task.

The orientation specific prediction accuracy could result from intrinsic biases of orientation selectivity in mouse visual system, or acquired through the learning of specific images. To investigate these possibilities, we trained a new cohort of mice using the same set of images but rotated by 90° (n = 4 mice; **Fig. 9a**). If the orientation specific biases were intrinsic, the results of orientation biases with rotated images will be preserved; while if the orientation biases were acquired, the biases will be rotated by 90°. All mice trained in the additional experiments (four of four) successfully passed the pre-training phases as the mice in the main experiments did. In NID testing phase, we found that the orientation bias was shifted by ~90° (**Fig. 9b**). These results suggest that mouse visual system can learn to extract specific orientation information based on the presented natural image. Although, we did not find out exactly what features of images the mice attended to, this feature may be one of the additional factors that the correct rates of the different mice were similar than predicted by similarity parameters such as SSIM.

4. Discussion

In this paper, we investigated the ability of mice to discriminate natural scenes, using an automated touchscreen-based system. We found that the mice learned to

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

discriminate natural scenes quickly, and that the psychometric function with SSIM predicted the correct rate. However, the behavioral performance of each mouse was highly consistent, and deviated in significant and reproducible ways from the prediction based on SSIM. Thus, we searched for a parameter or model that more accurately predicts the performance of the mice on this task. We found that a V1-inspired model provides a prediction whose accuracy can approach that of the inter-animal agreement. Thus, V1 processing may partly explain the way in which the mouse vision discriminates natural images.

Natural scenes were compared using the SSIM index, which is commonly used to estimate the similarity of two images. SSIM indices correlated with the correct rate of the mice in our task. However, the psychometric curves plotted against SSIM provided only marginal fits to the data. In particular, there were several images that had significantly higher or lower correct rates across all mice than predicted by SSIM (**Fig. 4a**). In addition, the correct rates of mice were highly correlated, which indicated that the deviations from the SSIM-based psychometric curves were not due to mouse-to-mouse variance, but rather that the mice perceived the similarity of images in ways not captured by SSIM.

Reducing the spatial resolution (i.e., blurring) of the images to better match mouse vision did not improve the fits of SSIM-based psychometric functions, nor did several other manipulations that we explored. Instead, we found that a Gabor filter-based approach provided the best fit. Notably, the improvement was specific to the orientation angle of the Gabor patch used. Horizontal orientation filtering did not provide a good fit, but a peri-vertical one did. In addition, the orientation bias was shifted accordingly when

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mice were required to discriminate rotated images. These results indicate that mice could extract orientation specific information depending on the task context.

In mammals, the orientation information itself could be modulated depending on the current demands. For instance, human visual cortex showed increased activity during being forced to attend to specific orientation, which is similar to the preferred orientation of the cortical areas (Warren, Yacoub et al. 2014). In monkey, neurons in visual cortical areas with orientation specificity increased its stimulus evoked firing rate when the monkey attends to the receptive field of the neurons (McAdams and Maunsell 1999). The direction and orientation specificity of mouse V1 neurons increased during learning visual discrimination tasks using grating stimuli when the preferred direction of the neuron was task relevant (Goltstein, Coffey et al. 2013, Poort, Khan et al. 2015). Thus, the orientation specific enhancement of the visual processing would be common features of human, monkey and mouse visual systems, and visual processing to specific orientation might explain our results. The changes of the neural activity to specific orientation, if it exists, might not arise by nature but by learning. This might be also consistent with the fact that while the orientation selective neurons observed in mouse V1 has horizontal bias during developing, it has an almost equal distribution in adults (Hagihara, Murakami et al. 2015).

On the RT analysis, we found that the mice spent longer time to discriminate dissimilar distractors (Fig. 5). This may suggest that mice made a decision based on the judgement of one figure, rather than extensively comparing two figures, which were displayed side by side. Here is the reasoning. Mice probably select the image they first see, when it is similar enough to the target image, while avoided the image they first see, when it is not similar at all to the target. The RT would be longer for the latter case, as mice had to move to the other side of the screen to finish the task.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Visual processing tasks can be categorized into scene/object recognition and motion/action, which are represented along different “streams” or subnetworks of cortical visual areas (Goodale and Milner 1992, Wang, Gao et al. 2011). In that context, this NID task complements the random dot kinematogram task we recently developed (Stirman, Townsend et al. 2016). We can use these experimental paradigms to investigate stream-specific higher visual processing of mice (Andermann, Kerlin et al. 2011, Marshel, Garrett et al. 2011, Stirman, Smith et al. 2016).

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

This work was supported by grants to S.L.S. from the Human Frontier Science Program, NIH (R01NS091335 and R01EY024294), the Simons Foundation (SCGB 325407SS), and the Whitehall Foundation. Y.Y. was supported by a Helen Lyng White fellowship, R.H. was supported by fellowships from Naitoh and the Japan Society for the Promotion of Science, and J.N.S. was support by a Career Award from Burroughs-Wellcome.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and legends

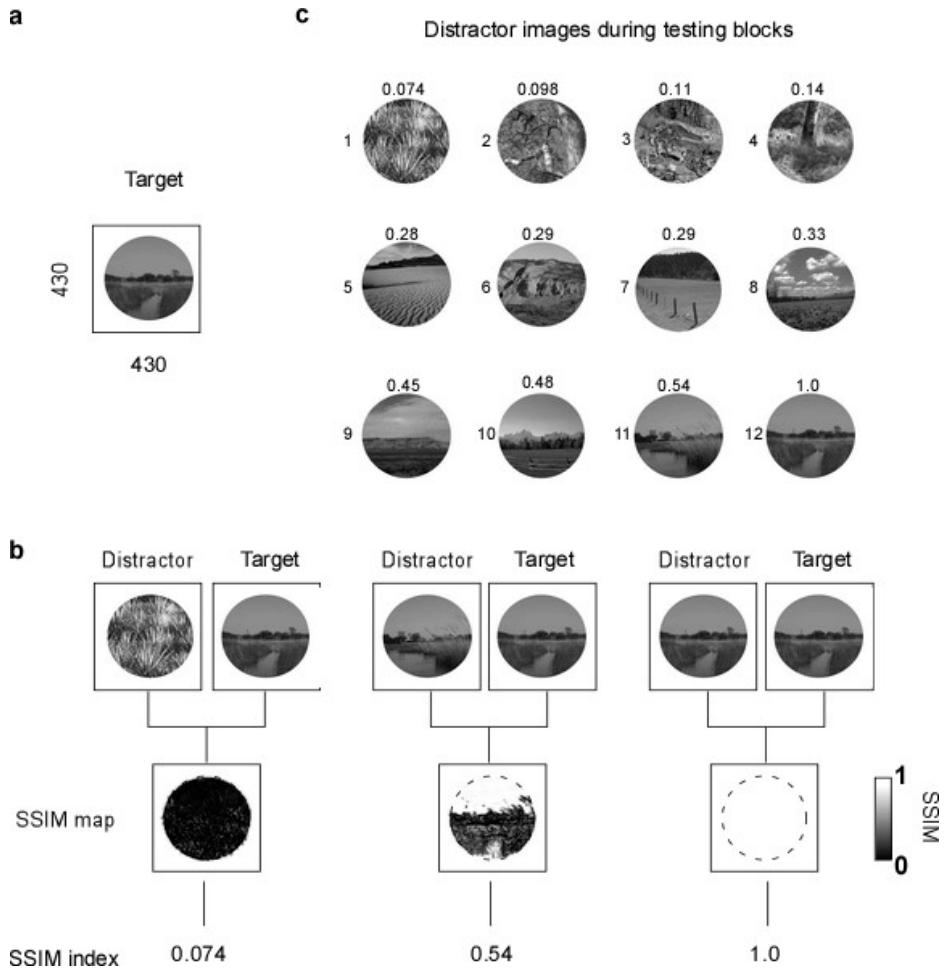

Fig. 1. SSIM index for quantifying expected discrimination difficulty

a. Target image for testing sessions. **b**. Example pairs of a distractor image and a target image. SSIM map is obtained for each pair. SSIM index is obtained by averaging the SSIM map inside the dotted area. When the distractor is the same as target, the SSIM index is 1 by definition (right most). Dotted line indicates the edge of the photo. **c**. 12 distractor images and corresponding SSIM indices (top). Each number on the left is ID of each distractor. A distractor (ID, 12) is the same as a target image.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

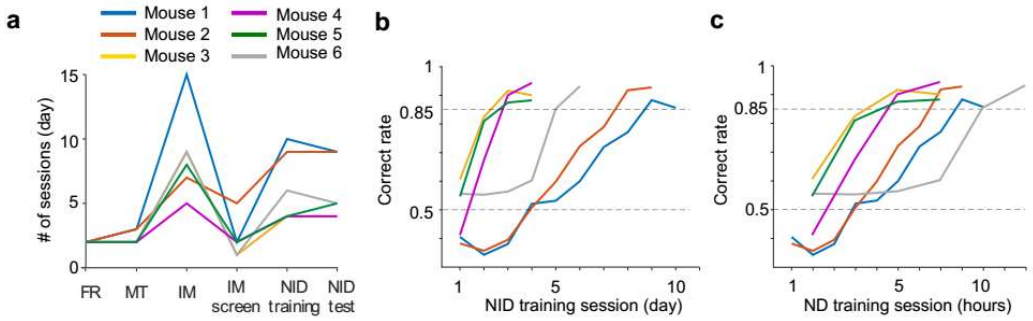

Fig. 2. Learning rates for each mouse

a. Number of sessions that required mice to reach criteria for each phase. **b**. Correct rate of each mouse as a function of days of training during NID training phase. (c) Correct rate of each mouse as a function of incremental hours of training during NID training phase.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

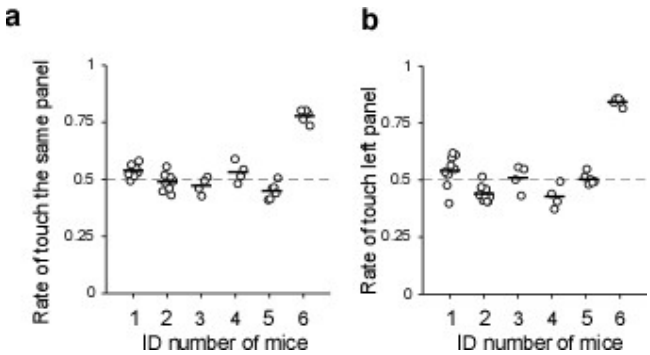

Fig. 3. Touch rate analysis disqualified one mouse

a. Rate of the trials the left panel was touched. Response bias is clearly observed in Mouse 6. **b**. Rate of the trials the same panel as the previous trial was touched. History dependent choice was not observed mice except for Mouse 6.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

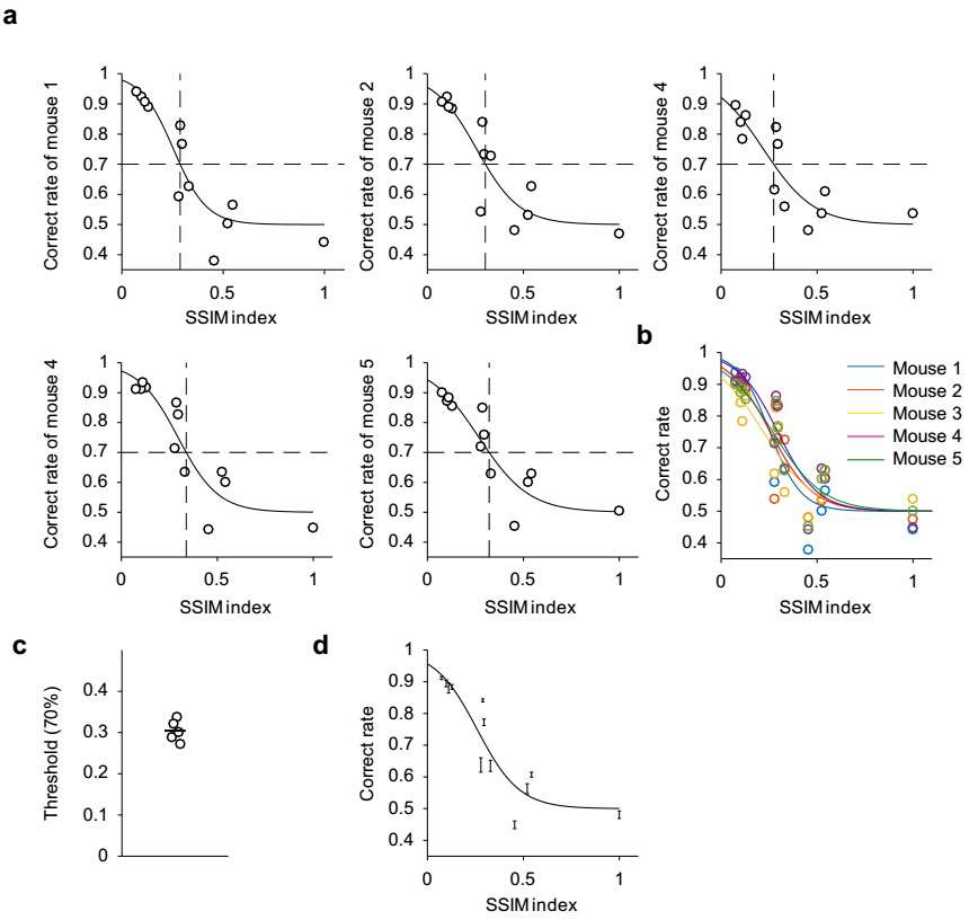

Fig. 4. Psychometric curves

a. Psychometric curves were obtained for five mice (Mouse 1-5, left to right and bottom left to center bottom) by data obtained during testing blocks. During the testing blocks, mice were presented stimulus pairs of SSIM indices between 0.074 and 1 in random order, allowing us to sample the full range of SSIM indices for each animal. Bottom right, the five psychometric curves were overlayed. **b**. SSIM threshold (percent coherence at 70% accuracy, horizontal dotted line in **a**) for the five mice. **c**. Each data point shows average ± s.e.m for each distractor. Psychometric curves for the average data were also overplayed. Threshold was 0.30. Note that for some distractors the correct rate was not approximated by the psychometric curve.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

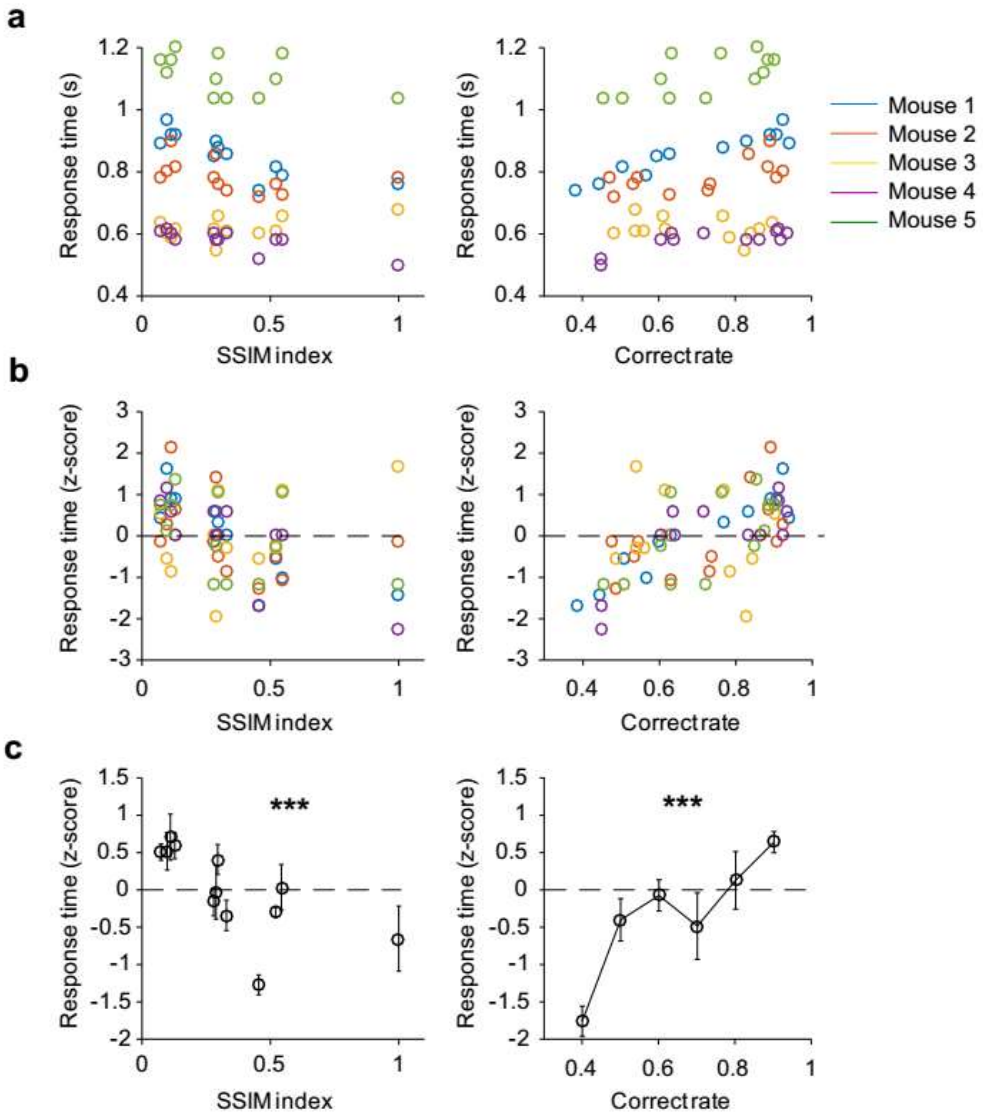

Fig. 5. Response time (RT) vs SSIM

a. RT were plotted as a function of SSIM indices (left) and performance (correct trial rate, right) for each mouse. **b**. Z-scored RT were plotted as a function of SSIM indices (left) and performance (correct trial rate, right) for each mouse. **c**. The same as (**b**). The same as in **c,** but shows mean and s.e.m. Left c= -0.43; right, c= 0.61; ***p<0.001 (Spearman's rank correlation test, two- tail).

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

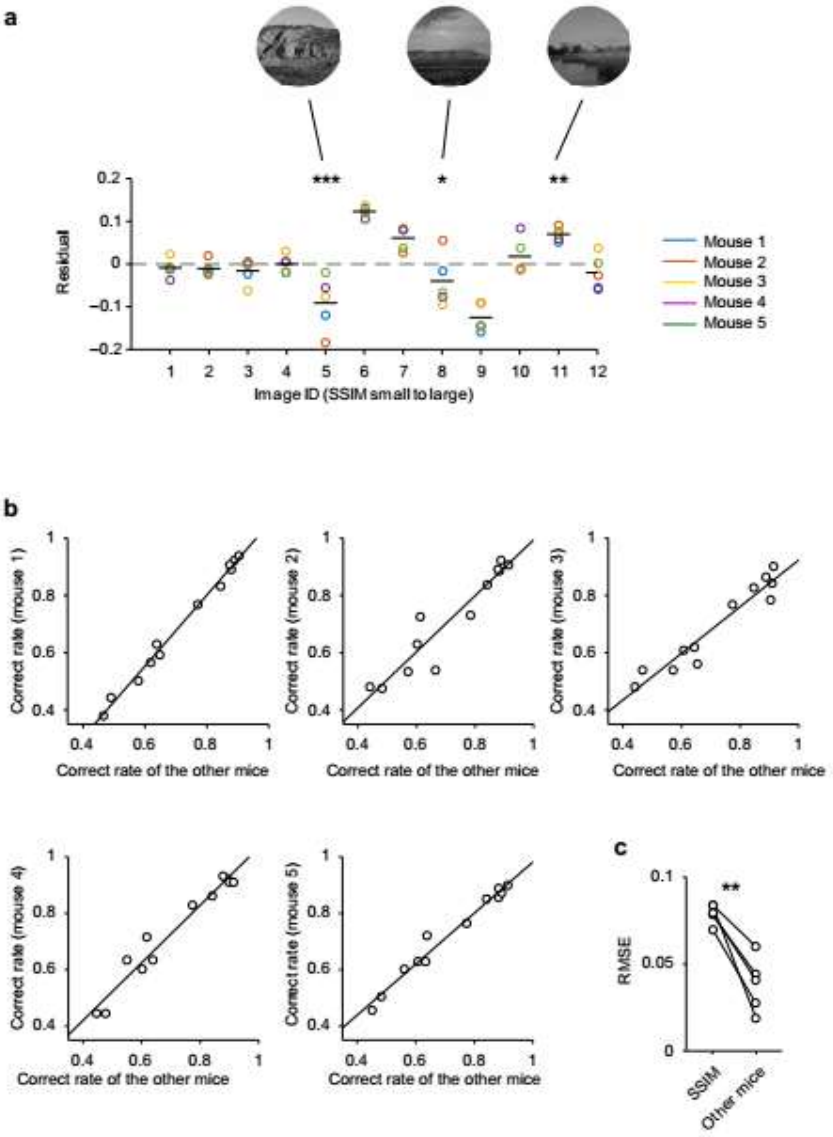

Fig. 6. High inter-mouse agreement

a. Difference between actual correct rate of each mouse and predicted correct rate by the SSIM-based psychometric curve for each distractor image. Note the behavior performances for three distractor images were not well predicted by the SSIM-based psychometric curve (ANOVA, $F(48, 11) = 17.8$, $p = 2.3 \times 10^{-13}$;post-hoc Tukey test, *p<0.05, **p<0.01, ***p<0.001). **b**. Correct rate of each mouse was plotted as a function

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of mean correct rate of the other mice. Note that correct rate of each mouse can be precisely predicted by that of other animals. **c**. RMSE obtained from the correct rate of the other mice was smaller than that from psychometric curves based on the SSIM index (**p<0.01; pared t-test).

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

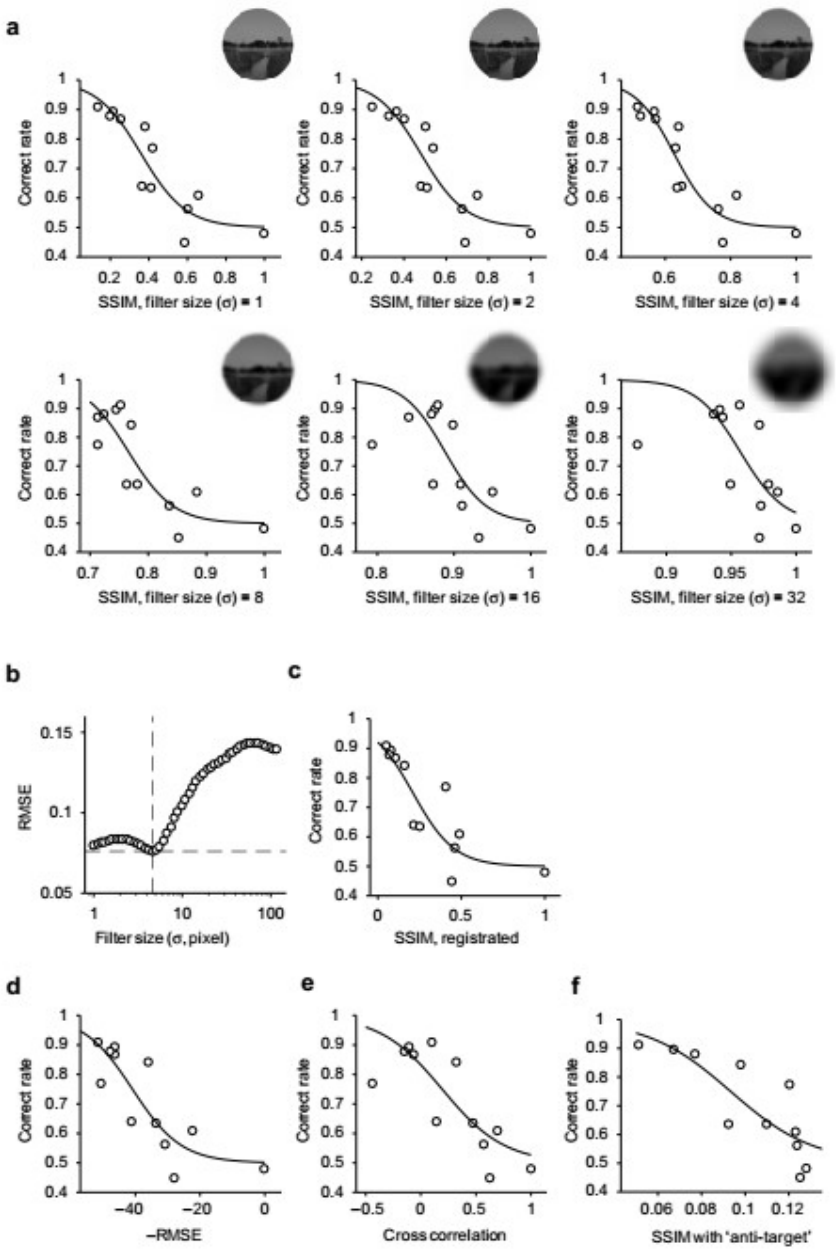

Fig. 7. Blurring the images and other attempts to reduce the fit error (RMSE)

a-f. Mean correct rate of the 12 images as a function of variety of parameters. The parameters are as follows. **a**; SSIM between distractors and a target image after filtering the images with Gaussian filter with specified standard deviation (sigma). The inset shows the filtered target image. **b**; SSIM between distractors and a target image after

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

registration of images. **c**; SSIM between distractors and a target image after registration of images. **d**; pixel wise RMSE (root mean squared error) between distractors and a target image. e; pixel wise cross correlation between distractors and a target image. **f**; SSIM between distractors and a distractor for training phase. RMSE for c-f was 0.089, 0.095, 0.11 and 0.10.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

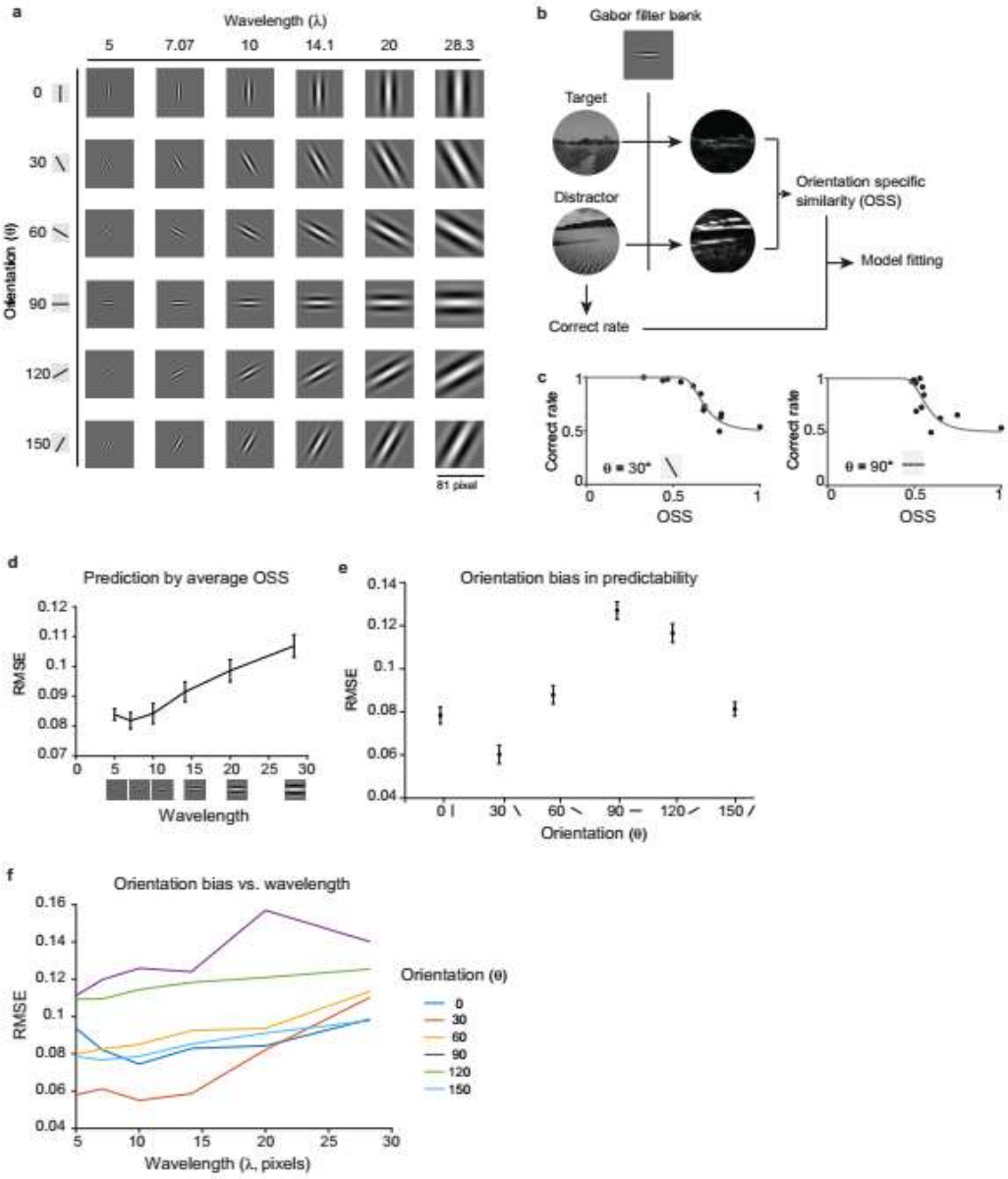

Fig. 8. A simple Gabor filter model predicts mouse performance on this task

a. Gabor filters with various orientations and wavelengths are shown. **b**. Illustration of the quantification of orientation specific similarity (OSS) between the target image and a distractor image. The Gabor filter presented here has orientation $\theta = 90°$, and wavelength $\lambda = 10$. **c**. The average correct rate of 5 animals was fit by a model of OSS with $\theta = 30°$,

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

λ = 10 (left, RMSE = 0.055); and a model of OSS with θ = 90°, λ = 10 (right; RMSE = 0.13). **d.** Mean and s.e.m. of RMSE values based on average OSS over all orientations were plotted against the wavelengths. Image patches are example Gabor filter patches of particular wavelength in 81x81 pixel squares. **e.** RMSE values based on OSS (wavelength = 10) was plotted against the orientations. Error bars are the s.e.m. (n = 5 mice). Orientation significantly affected the predictability of behavior ($p = 9*10^{-11}$, ANOVA). **f.** The prediction error (mean RMSE, n =5 mice) as in **c** were obtained with various wavelengths and orientations. Colors indicate different orientations

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

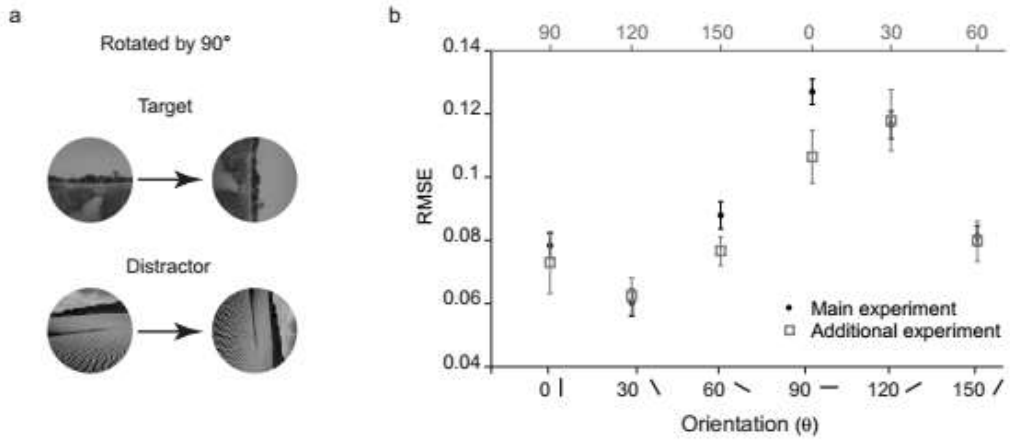

Fig. 9. Experiments with rotated images show that the orientation bias is not innate

a. Two example of original images and rotated images. The mice were trained with rotated images in NID training and testing sessions. **b.** Prediction error (RMSE) by OSS with each orientation (wavelength = 10). The black line shows the result from the mice trained with original images. The gray line is data from the experiments with rotated images, plotted against the top axis, which is shifted by 90 degrees relative to the bottom axis. This shows that the orientation specific prediction accuracy also shifted by 90 degrees.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Distractors (Testing)

ID	SSIM	Distractor
1	0.074	
2	0.098	
3	0.11	
4	0.14	
5	0.28	
6	0.29	
7	0.29	
8	0.33	
9	0.45	
10	0.48	
11	0.54	
12	1.0	

Distractors (Training)

ID	SSIM	Distractor
1	0.082	
2	0.085	
3	0.097	
4	0.10	
5	0.11	
6	0.12	
7	0.13	
8	0.14	
9	0.14	
10	0.16	

Supplementary table 1. Distractor images

All distractor images for testing phase (left) and for training phase (right). The SSIM index for each image is also listed.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Andermann, M. L., A. M. Kerlin, D. K. Roumis, L. L. Glickfeld and R. C. Reid (2011). "Functional specialization of mouse higher visual cortical areas." Neuron **72**(6): 1025-1039.

Bell, A. J. and T. J. Sejnowski (1997). "The “independent components” of natural scenes are edge filters." Vision Research **37**(23): 3327-3338.

Bussey, T. J., T. L. Padain, E. A. Skillings, B. D. Winters, A. J. Morton and L. M. Saksida (2008). "The touchscreen cognitive testing method for rodents: how to get the best out of your rat." Learning & Memory **15**(7): 516-523.

Bussey, T. J., L. M. Saksida and L. A. Rothblat (2001). "Discrimination of computer-graphic stimuli by mice: a method for the behavioral characterization of transgenic and gene-knockout models." Behavioral Neuroscience **115**(4): 957.

Felleman, D. J. and D. C. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." Cerebral Cortex **1**(1): 1-47.

Froudarakis, E., P. Berens, A. S. Ecker, R. J. Cotton, F. H. Sinz, D. Yatsenko, P. Saggau, M. Bethge and A. S. Tolias (2014). "Population code in mouse V1 facilitates readout of natural scenes through increased sparseness." Nature Neuroscience **17**(6): 851-857.

Glickfeld, L. L., M. H. Histed and J. H. Maunsell (2013). "Mouse primary visual cortex is used to detect both orientation and contrast changes." The Journal of Neuroscience **33**(50): 19416-19422.

Goltstein, P. M., E. B. Coffey, P. R. Roelfsema and C. M. Pennartz (2013). "In vivo two-photon Ca2+ imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex." The Journal of Neuroscience **33**(28): 11540-11555.

Goodale, M. A. and A. D. Milner (1992). "Separate visual pathways for perception and action." Trends in Neurosciences **15**(1): 20-25.

Hagihara, K. M., T. Murakami, T. Yoshida, Y. Tagawa and K. Ohki (2015). "Neuronal activity is not required for the initial formation and maturation of visual selectivity." Nature Neuroscience **18**(12): 1780-1788.

Harvey, C. D., F. Collman, D. A. Dombeck and D. W. Tank (2009). "Intracellular dynamics of hippocampal place cells during virtual navigation." Nature **461**(7266): 941-946.

Hoy, J. L., I. Yavorska, M. Wehr and C. M. Niell (2016). "Vision drives accurate approach behavior during prey capture in laboratory mice." Current Biology **26**(22): 3046-3052.

Hubel, D. H. and T. N. Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." The Journal of Physiology **160**(1): 106-154.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Itti, L., C. Koch and E. Niebur (1998). "A model of saliency-based visual attention for rapid scene analysis." IEEE Transactions on pattern analysis and machine intelligence **20**(11): 1254-1259.

Long, M., W. Jiang, D. Liu and H. Yao (2015). "Contrast-dependent orientation discrimination in the mouse." Scientific Reports **5**.

Marshel, J. H., M. E. Garrett, I. Nauhaus and E. M. Callaway (2011). "Functional specialization of seven mouse visual cortical areas." Neuron **72**(6): 1040-1054.

McAdams, C. J. and J. H. Maunsell (1999). "Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4." The Journal of Neuroscience **19**(1): 431-441.

Nithianantharajah, J., A. McKechanie, T. Stewart, M. Johnstone, D. Blackwood, D. St Clair, S. Grant, T. Bussey and L. Saksida (2015). "Bridging the translational divide: identical cognitive touchscreen testing in mice and humans carrying mutations in a disease-relevant homologous gene." Scientific Reports **5**.

Ohki, K., S. Chung, Y. H. Ch'ng, P. Kara and R. C. Reid (2005). "Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex." Nature **433**(7026): 597-603.

Oliva, A. and A. Torralba (2001). "Modeling the shape of the scene: A holistic representation of the spatial envelope." International Journal of Computer Vision **42**(3): 145-175.

Olmos, A. and F. A. A. Kingdom (2004). "A biologically inspired algorithm for the recovery of shading and reflectance images." Perception **33**(12): 1463-1473.

Olshausen, B. A. (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." Nature **381**(6583): 607-609.

Poort, J., A. G. Khan, M. Pachitariu, A. Nemri, I. Orsolic, J. Krupic, M. Bauza, M. Sahani, G. B. Keller and T. D. Mrsic-Flogel (2015). "Learning enhances sensory and multiple non-sensory representations in primary visual cortex." Neuron **86**(6): 1478-1490.

Prusky, G. T., P. W. West and R. M. Douglas (2000). "Behavioral assessment of visual acuity in mice and rats." Vision Research **40**(16): 2201-2209.

Quiroga, R. Q., L. Reddy, G. Kreiman, C. Koch and I. Fried (2005). "Invariant visual representation by single neurons in the human brain." Nature **435**(7045): 1102-1107.

Serre, T., L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio (2007). "Robust object recognition with cortex-like mechanisms." IEEE transactions on pattern analysis and machine intelligence **29**(3): 411-426.

Stirman, J. N., I. T. Smith, M. W. Kudenov and S. L. Smith (2016). "Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain." Nature Biotechnology.

Stirman, J. N., L. B. Townsend and S. L. Smith (2016). "A touchscreen based global motion perception task for mice." Vision Research **127**: 74-83.

Thevenaz, P., U. E. Ruttimann and M. Unser (1998). "A pyramid approach to subpixel registration based on intensity." IEEE transactions on image processing **7**(1): 27-41.

bioRxiv preprint doi: https://doi.org/10.1101/156653; this version posted June 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tkačik, G., P. Garrigan, C. Ratliff, G. Milčinski, J. M. Klein, L. H. Seyfarth, P. Sterling, D. H. Brainard and V. Balasubramanian (2011). "Natural images from the birthplace of the human eye." PLoS One **6**(6): e20409.

Vinje, W. E. and J. L. Gallant (2000). "Sparse coding and decorrelation in primary visual cortex during natural vision." Science **287**(5456): 1273-1276.

Wang, Q., E. Gao and A. Burkhalter (2011). "Gateways of ventral and dorsal streams in mouse visual cortex." The Journal of Neuroscience **31**(5): 1905-1918.

Wang, Z., A. C. Bovik, H. R. Sheikh and E. P. Simoncelli (2004). "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image processing **13**(4): 600-612.

Warren, S. G., E. Yacoub and G. M. Ghose (2014). "Featural and temporal attention selectively enhance task-appropriate representations in human primary visual cortex." Nature Communications **5**.

Young, M. P. and S. Yamane (1992). "Sparse population coding of faces in the inferotemporal cortex." Science **256**(5061): 1327-1331.