Distance estimation from monocular cues in an ethological visuomotor task

Philip R. L. Parker[1], Elliott T. T. Abe[1], Natalie T. Beatie[1], Emmalyn S. P. Leonard[1], Dylan M. Martins[1], Shelby L. Sharp[1], David G. Wyrick[1], Luca Mazzucato[1,2], Cristopher M. Niell[1,3]

[1]Institute of Neuroscience, [2]Department of Mathematics, [3]Department of Biology, University of Oregon, Eugene, OR 97403

ABSTRACT

In natural contexts, sensory processing and motor output are closely coupled, which is reflected in the fact that many brain areas contain both sensory and movement signals. However, standard reductionist paradigms decouple sensory decisions from their natural motor consequences, and head-fixation prevents the natural sensory consequences of self-motion. In particular, movement through the environment provides a number of depth cues beyond stereo vision that are poorly understood. To study the integration of visual processing and motor output in a naturalistic task, we investigated distance estimation in freely moving mice. We found that mice use vision to accurately jump across a variable gap, thus directly coupling a visual computation to its corresponding ethological motor output. Monocular eyelid suture did not affect performance, thus mice can use cues that do not depend on binocular disparity and stereo vision. Under monocular conditions, mice performed more vertical head movements, consistent with the use of motion parallax cues, and optogenetic suppression of primary visual cortex impaired task performance. Together, these results show that mice can use monocular cues, relying on visual cortex, to accurately judge distance. Furthermore, this behavioral paradigm provides a foundation for studying how neural circuits convert sensory information into ethological motor output.

INTRODUCTION

Vision is an active process - we continuously move our eyes, head, and body to gain information about the world around us. One core function of active vision is to determine the distance between the observer and objects in its environment. This ability is so critical that many species have evolved to use multiple distinct cues to estimate depth, including retinal image size, motion and position parallax, and binocular disparity (Kral, 2003; Shinkman, 1962). In particular, depth perception through stereo vision has been heavily studied, but

other cues that provide important complements are less well understood. Furthermore, some of these monocular cues, such as motion parallax and loom, are closely integrated with movement. How does the brain make use of these diverse cues to guide different behaviors? For instance, is distance explicitly computed and represented in neural activity for some behaviors and implicitly encoded for others? Furthermore, how is this sensory representation converted into the appropriate motor output? Neurophysiological studies are often performed in head-fixed subjects, limiting the range of depth cues and behaviors that can be studied. Addressing these questions requires behaviors where experimental subjects amenable to neural circuit interrogation can engage in distance estimation behaviors unrestrained (Leopold and Park, 2020; Parker et al., 2020).

The mouse is an important model for vision, yet relatively few behavioral paradigms exist for studying natural, active vision in the mouse (Boone et al., 2021; Hoy et al., 2016; Yilmaz and Meister, 2013). Previous work in other rodent models, including rats and gerbils, showed that animals will accurately jump to distant platforms for a reward, and that changing experimental conditions can bias animals toward the use of certain depth cues, including monocular ones (Carey et al., 1990; Ellard et al., 1984; Goodale et al., 1990). Here we report that mice are capable of using vision to estimate the distance across a variable gap and execute an accurate ballistic jump. Using this behavior, we show that mice can use monocular vision to judge distance, and this depends on primary visual cortex (V1). Furthermore, this paradigm provides a foundation for studying various visual computations related to depth, and the corresponding motor output, in a species amenable to measurement and manipulation of neural activity in genetically-defined populations.

RESULTS

Mouse distance estimation (jumping) task

In order to establish a framework for studying distance estimation in the mouse, we adapted a gerbil/rat jumping task (Ellard et al., 1984; Legg and Lambert, 1990; Richardson, 1909), where animals were rewarded for successfully jumping across a variable gap (Figure 1A). Mice were free to roam around the arena, then initiated trials by mounting a take-off platform. An occluding barrier was introduced to block the mouse's view while the experimenter randomly placed one of three landing platforms at one of five distances from the take-off platform (Figure 1B). We used landing platforms of variable size to minimize the use of retinal image size cues, which may not require visual cortex for accurate distance estimation after learning (Carey et al.,

1990). The trial began as soon as the occluding barrier was removed, and after a decision period the mouse performed one of three outcomes (Figure 1C). On “success” trials, the mouse jumped and landed on the landing platform, and received a reward (see Supplemental Video 1). On “failure” trials, the mouse jumped and missed the landing platform, landing on the arena floor, and received no reward. On “abort” trials, mice dismounted the take-off platform onto the arena floor and received a mild air puff. Training, which usually took one to two weeks, was complete when mice successfully jumped to each of the three landing platforms at the maximum distance (24 cm). To quantify behavior, markerless pose estimation was performed on side- and top-view video with DeepLabCut (DLC; Mathis et al., 2018).

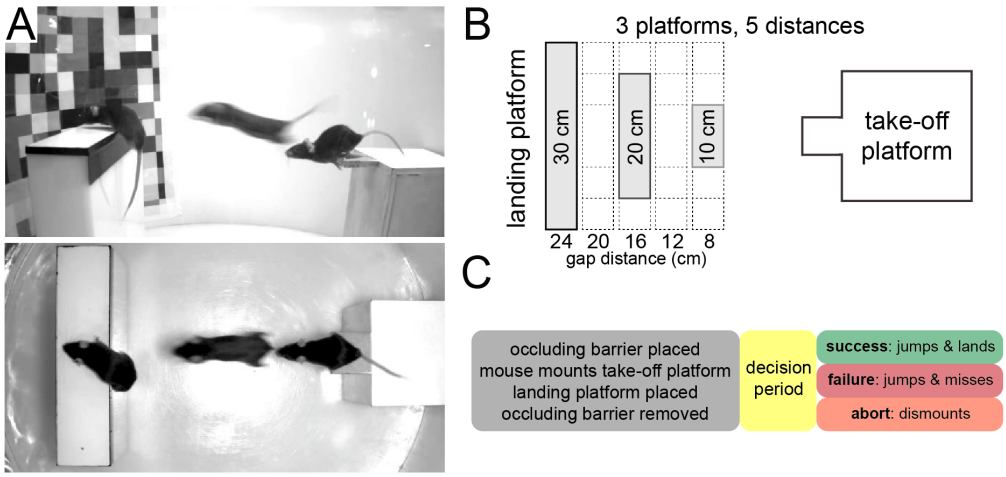


Figure 1: Mouse distance estimation (jumping) task. A) Example side and top-down video frames (three overlaid) from a single trial. B) A random combination of landing platform width (three sizes) and gap distance (five distances) is chosen for each trial. C) Trial logic.

Mice accurately estimate distance under binocular and monocular conditions

Mice successfully jumped to all three platforms at all distances, however failures were more common at the farthest distance (24 cm; Figure 2A, B top). On success trials, the distance jumped matched very closely to the actual gap distance (Figure 2B, bottom), showing that mice accurately jumped rather than adapting an alternative strategy (e.g. picking one of two jump forces across the five distances). The gap was too large for mice to reach across with their whiskers, preventing the use of somatosensation to judge the distance. Furthermore, mice did not perform any jumps in the dark (n=4 mice, 4 sessions), suggesting they relied on vision.

A number of depth cues are available in natural contexts. To test the need for stereopsis, we performed monocular eyelid suture, after which mice performed equally well at the task (Figure 2B, gray lines), suggesting binocular vision is not required for accurate distance estimation under these conditions (two-way ANOVA: performance F=0.128, p= 0.722; distance jumped F=0.112, p= 0.739), and demonstrating that mice can use monocular cues to accurately judge distance. We also tested for a role of retinal image size by analyzing performance across the three different landing platforms. Mice performed similarly across all three sizes, suggesting they did not rely primarily on retinal image size, although binocular animals showed a significant interaction between distance jumped and platform width (two-way ANOVA: binocular performance F=1.558, p=0.215; monocular performance F=0.004, p=0.996; binocular distance jumped F=6.846, p=0.001; monocular distance jumped F=0.036, p=0.96).

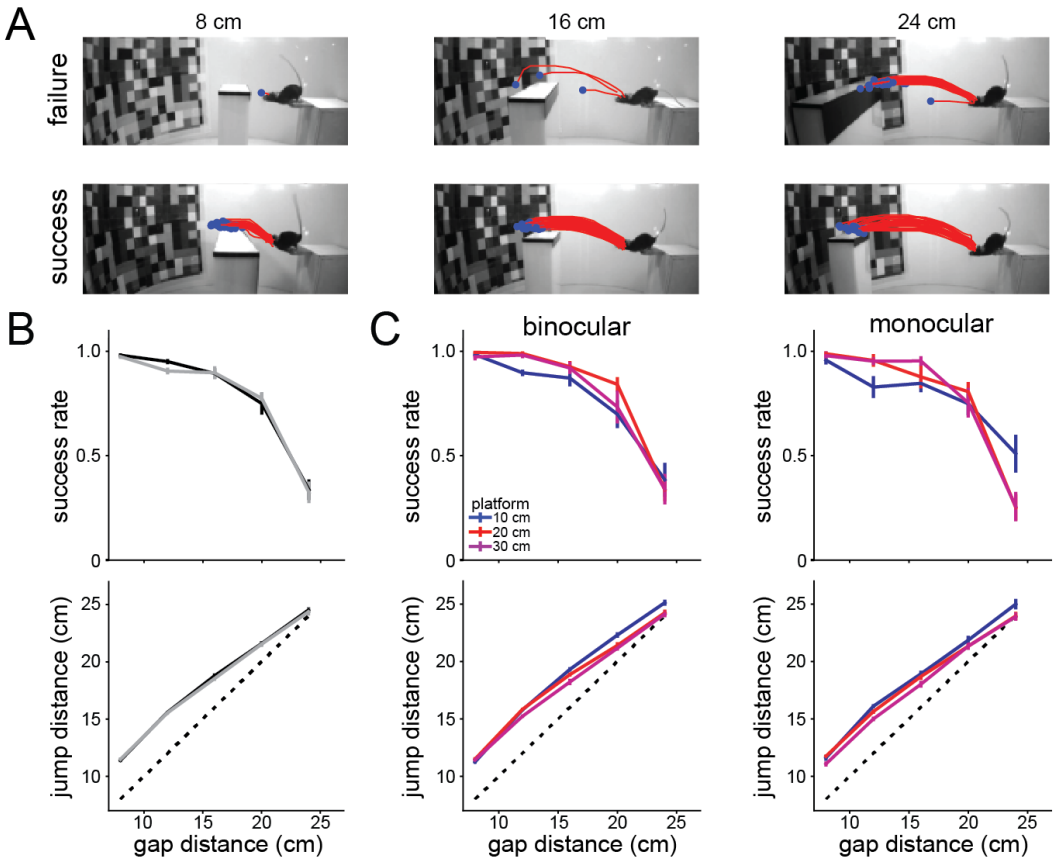


Figure 2: Mice accurately judge distance under binocular and monocular conditions. A) Example jump trajectories from a single mouse (red, blue dot is end point) at three distances for failure (top row) and success (bottom row) trials. B) Performance (top) and accuracy (bottom) in binocular (black, n=8 mice) and monocular (gray, n=8 mice) conditions averaged across landing platform widths. C) Performance and accuracy for bi/monocular conditions by landing platform width.

Mice perform more head movements under monocular conditions

To quantify the fine-scale structure of behavior leading up to the jump, we inputted the decision period x/y values of DLC markers of the nose, left eye, and left ear for all trials (both binocular and monocular) into an auto-regressive hidden markov model (ARHMM; Linderman et al., 2019). A five-state model readily summarized the data, where each state represented a common movement trajectory that together formed a largely feed-forward chain leading up to the jump (Figure 3A, arrow thickness indicates transition probability). The most common sequence of transitions was state 3 to 4 to 5, qualitatively described as a forward/upward movement, followed by a forward/downward movement, then a very small amplitude 'wiggle' immediately preceding the jump (Figure 3A, B). Two larger amplitude upward (state 1) and downward (state 2) head movements tended to precede this common sequence.

Next we sought to determine whether the behavior of the mice leading up to the jump was different between the binocular and monocular conditions. In both groups, states 1 and 2 were less frequent than states 3-5, and frequencies of states 2 and 3 were higher under monocular conditions (Figure 3C; binocular vs. monocular state frequency, s1 $p=0.237$, s2 $p=0.004$, s3 $p=0.005$, s4 $p=0.030$, s5 $p=0.058$; alpha=0.010 Bonferroni corrected paired t-test). However, the overall order of these states relative to the jump was similar between binocular and monocular conditions (Figure 3D). A simple decoder was able to distinguish between binocular and monocular conditions based on the transition count matrices (Figure 3E). This was largely due to an increase in state transitions between later states and earlier states; in other words, the monocular group chose to revisit earlier states before executing a jump (Figure 3F, G).

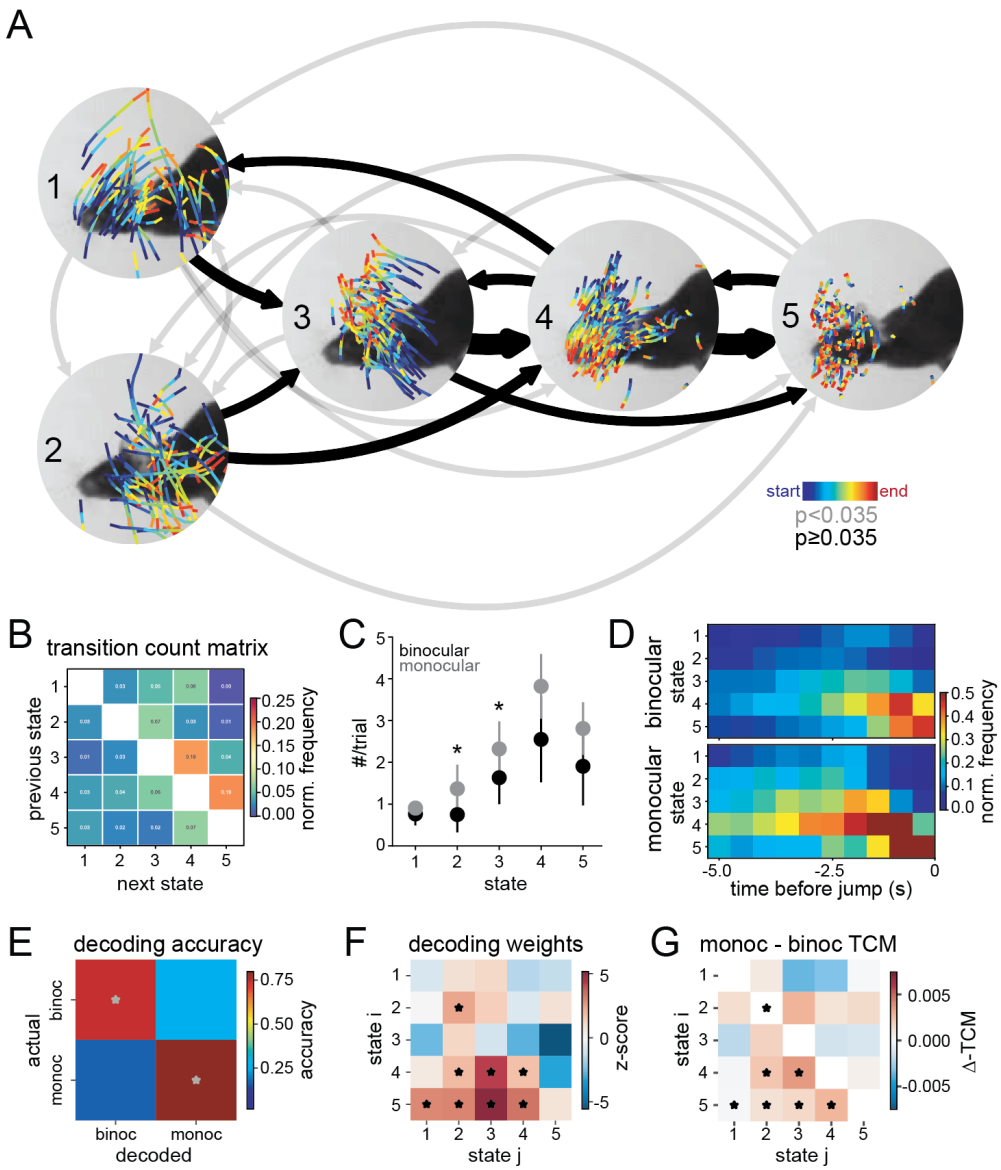


Figure 3: Mice perform more head movements during the decision period under monocular conditions. A) Example traces of eye position from five movement states labeled with auto-regressive hidden Markov modeling of DeepLabCut-tracked points during the decision period (progressing blue to red in time) in average temporal order. Arrow line widths are proportional to transition probabilities between states (gray < 0.035 ≤ black). B) Transition count matrix for binocular condition, showing the frequency of transitioning from one state (y-axis) to another state (x-axis) as a fraction of all unique state transitions; these values were used to generate the arrows in panel A. C) Frequency of each state for binocular (black) and monocular (gray) conditions. Asterisk indicates p<0.01, paired t-test. D) Heatmaps of start time histograms for each state normalized to the total number of trials for binocular (top) and monocular (bottom) conditions. E) Two-fold decoding analysis on transition count matrices for binocular vs. monocular conditions (performed within-animal, averaged across animals). F) Z-scored weights used to decode binocular vs. monocular condition. G) Difference between monocular and binocular transition count matrices; red transitions are more frequent in monocular, blue in binocular. n=8 mice for all plots.

Eye movements compensate for head movements to stabilize gaze

Previous work shows that the majority of eye movements in rodents, including mice, are compensatory for head movements, and that saccades occur primarily as a consequence of large amplitude head movements (Meyer et al., 2020; Michaiel et al., 2020; Wallace et al., 2013). Some species decrease horizontal vergence angle to increase binocular overlap during behaviors such as prey capture (Bianco et al., 2011). To determine how mice target their gaze during distance estimation, we performed bilateral eye tracking using miniature head-mounted cameras (Michaiel et al., 2020), then used DLC to track the pupil in order to quantify horizontal and vertical eye movements (Figure 4A; see Supplementary Video 2). Importantly, mice continued to accurately perform the task despite the head-mounted hardware and tether (~3 g weight; Figure 4B, C). Head pitch (vertical head angle) was anti-correlated with both eye vergence (horizontal angle of the two eyes) and eye phi (vertical eye movements) both during the early portion of the decision period when mice were approaching the jump (start of trial to 2 sec before jump; pitch vs. vergence R^2=0.51, phi R^2=0.28) and in the late portion of the decision period immediately prior to the jump (2 sec prior to jump; pitch vs. vergence R^2=0.70, phi R^2=0.54; Figure 4D, E). Thus, upward head movements caused the eyes to move down and toward the nose, while downward head movements caused upward and outward eye movements, consistent with vestibulo-ocular reflex-mediated gaze maintenance throughout the decision period. Additionally, while there was a slight change in vergence between the early and late periods of the trial (vergence early 1.77 ± 0.21 deg, late -1.66 ± -0.29 deg; p=1.57e-4), this was explained by a similar difference in head pitch between these two periods (pitch early -45.45 ± 1.24 deg, late -38.47 ± 1.43 deg; p=4.97e-5), demonstrating that mice do not move their eyes to increase binocular overlap preceding the jump.

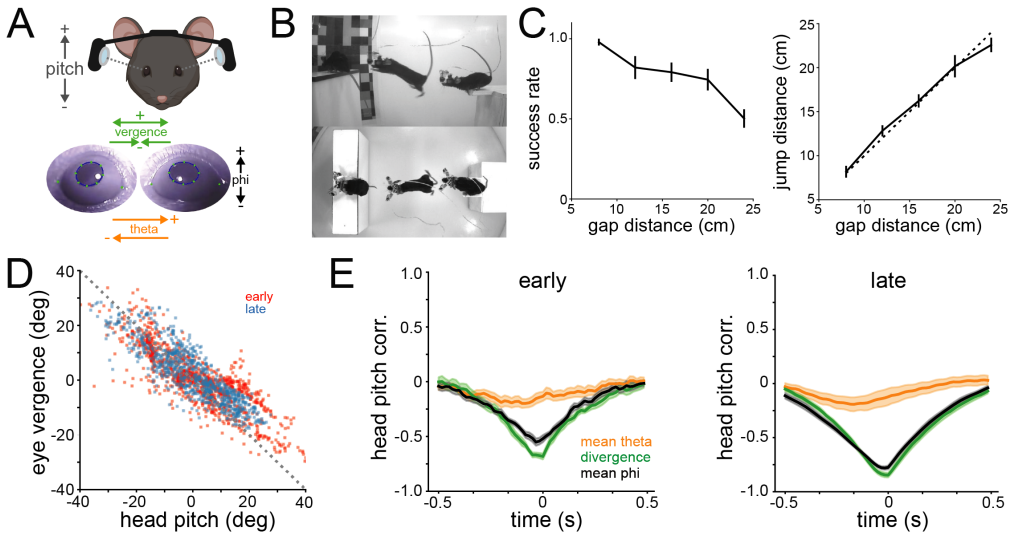


Figure 4: Eye movements compensate for head movements to stabilize gaze. A) Schematic of experimental setup for measuring head and eye movements; bilateral eye tracking with miniature head-mounted cameras (top) and ellipse fitting of DLC-tracked pupil points (bottom). B) Side and top-view images of a mouse performing the task with the eye tracking system (3 frames overlaid). C) Performance (left) and distance jumped (right) for eye-tracking experiments. D) Horizontal angle between the two eyes (eye theta divergence) as a function of head pitch during the decision period. 'Early' is from the start of the trial to 2 sec before the jump, and 'late' is the 2 sec preceding the jump. E) Mean eye theta, eye theta vergence, and eye phi cross correlations with head pitch angle for early (left) and late (right) portions of the decision period; n=8 mice for all plots.

V1 optogenetic suppression disrupts distance estimation

Finally, we tested whether distance estimation under monocular conditions requires visual cortex. Bilateral optic fibers were implanted at the surface of the cortex above monocular V1 in either control mice (wild type) or mice expressing channelrhodopsin-2 (ChR2) in parvalbumin-expressing inhibitory interneurons (PV-Cre:Ai32, referred to as PV-ChR2 here; Hippenmeyer et al., 2005; Madisen et al., 2012), all of which underwent monocular eyelid suture in order to isolate the use of monocular cues (Figure 5A). On a third of trials, light was delivered through the implanted optic fibers during the decision period (470 nm, 5 mW/mm^2, 40 Hz, 50% duty cycle). Control animals showed no change in performance with the laser on ($F = 0.971$, $p=0.332$; ANOVA), whereas PV-ChR2 animals (see Supplementary Videos 3 and 4) showed a significant reduction in performance across distances (Figure 5B; $F=38.938$, $p=2.94e-8$, ANOVA). This decrease in performance was largely due to an increase in aborts (control $F=1.000$, $p=0.325$; PV-ChR2 $F=34.187$, $p=1.46e-7$; ANOVA) with little change in failure rate (control $F=0.799$, $p=0.379$; PV-ChR2 $F=0.812$, $p=0.371$, ANOVA). Interestingly, PV-ChR2

performance was somewhat intact for the medium-sized landing platform (Figure 5C) and worse for the wider and narrower platforms, suggesting that animals may have resorted to using retinal image size as a cue, using the average platform width as an estimate. ARHMM analysis suggested that mice in both groups did not change their behavior when the laser was on, and a simple decoder could not accurately differentiate between laser off and laser on trials based on the transition count matrix of either group (Supplemental Figure S1). Together, these results suggest that V1 is required for accurate distance estimation under conditions of monocular vision.

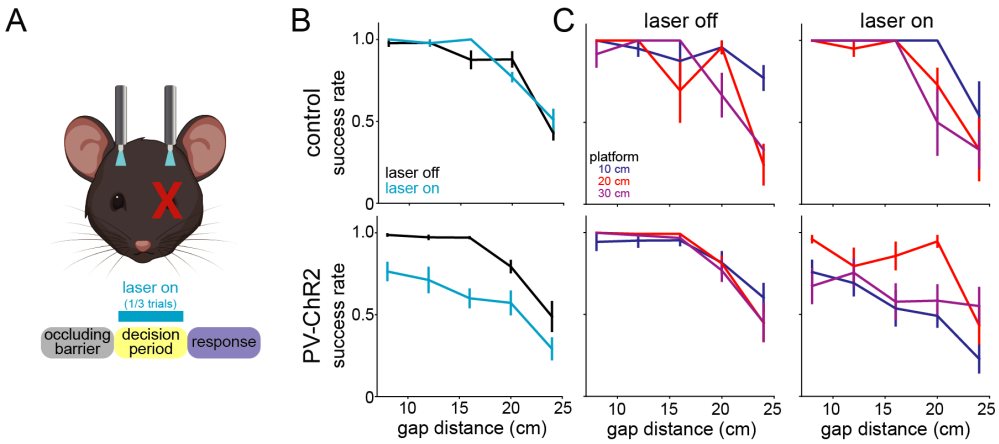


Figure 5: V1 optogenetic suppression disrupts distance estimation. A) Schematic of experimental setup for optogenetic experiments; bilateral illumination of monocular V1 in monocular animals during the decision period. B) Performance curves for laser off (black) and laser on (cyan) conditions in control (WT, top, n=4 mice) and experimental (PV-Cre:Ai32, bottom, n=8) mice expressing ChR2 in PV+ inhibitory interneurons. C) Performance curves by landing platform width for control (top) and PV-ChR2 (bottom) groups with laser off (left) and laser on (right).

DISCUSSION

We have established a visual distance estimation task in mice that engages an ethological, freely moving behavior. Previous research using similar versions of this task suggests that gerbils and rats utilize multiple cues to determine the distance to objects in the environment, including retinal image size, binocular vision, and motion parallax (Carey et al., 1990; Ellard et al., 1984; Goodale et al., 1990; Legg and Lambert, 1990; Michaiel et al., 2020). Importantly, this task is distinct from ‘gap crossing’ tasks (Hutson and Masterton, 1986) where animals can use the whisker somatosensory system to determine the distance across a short gap. Furthermore, in contrast to other recently developed tasks that are designed to probe binocular depth perception and stereopsis (Boone et al., 2021), in this task mice are able to use monocular cues for depth,

including those that are generated by self-movement. It can therefore be flexibly used to investigate a variety of distance estimation tactics by manipulating experimental conditions.

Cues for distance estimation

Binocular vision (and therefore stereopsis) was not required for accurate distance estimation in this task, consistent with previous studies in gerbils (Ellard et al., 1984). This provides the first demonstration that mice are able to use depth cues that are available besides stereopsis. This does not rule out the use of binocular disparity under normal conditions, and it will be interesting to determine whether mice use binocular disparity cues to judge distance in this task. In contrast, vertical head movements sufficient to generate motion parallax cues are increased in frequency under monocular conditions, suggesting mice may use motion parallax under monocular vision in this task. Previous work found increasing frequency of vertical head movements as a function of gap distance in gerbils and rats (Ellard et al., 1984; Legg and Lambert, 1990). We did not see such a relationship, which could be due to species-specific differences, or differences in task design. For example, previous studies used longer landing platforms, resulting in few failure trials and more variance in landing accuracy. Ultimately, closed-loop control of the landing platform based on head movements would provide a causal test to determine whether mice use motion parallax cues, as was performed in locusts (Wallace, 1959). Importantly, one could study multiple depth cues using this task depending on experimental conditions, such as the number and sizes of platforms and the visual context. For example, experiments with a modified version of this task suggest that parietal cortex is necessary for context-dependent use of retinal image size cues (Ellard and Sharma, 1996).

Eye movements during distance estimation

Using miniature cameras to track the eyes, we found that eye movements compensate for head movements to stabilize gaze leading up to the jump. This would allow the mouse to both maintain gaze toward the platform and reduce motion blur throughout large amplitude head movements (Land, 1999). This is consistent with previous work showing that mouse eye movements stabilize gaze during both operant behavior (Meyer et al., 2020) and a natural behavior, prey capture (Michaiel et al., 2020). It will be interesting to determine whether jumping mice control their gaze to localize the platform on a specific subregion of the retina; i.e., whether there

exists a retinal specialization for determining distance. In the case of prey capture, the image of the cricket is stabilized in the retinal region with the highest concentration of alpha-ganglion cells (Holmgren et al., 2021). We also found that mice did not move their eyes to increase binocular overlap during the period immediately preceding the jump, similar to a previous finding demonstrating that rats do not align the gaze of the two eyes before crossing a short gap (Wallace et al., 2013). Smooth eye movements provide extra-retinal signals for computing depth from motion parallax in primates (Kim et al., 2017; Nadler et al., 2009, 2008), and future studies may address whether the compensatory movements we observed play a similar role in mice. Finally, these experiments show that mice are capable of performing this task with a tether and significant hardware weight on the head, which is a critical requirement for introducing additional techniques such as electrophysiology into this paradigm.

Neural circuits underlying distance estimation

We provide the first evidence for V1 specifically being important for distance estimation in mice under monocular conditions. V1 suppression also did not significantly alter vertical head movements prior to the jump. Mice were still successful on a fraction of trials even during inactivation, which could be due to a number of factors. First, animals could be jumping to the maximum possible distance and landing on the platform by chance at shorter distances. Second, some vision could remain intact during light delivery, as it is likely that suppression of V1 under these conditions was incomplete.

These results are consistent with previous work showing broad lesions of occipital cortex disrupt performance without affecting head movements, whereas lesions to superior colliculus and preoptic area had no effect on either (Ellard et al., 1986). This task could therefore be a useful tool for studying the specific computations performed in V1 that mediate accurate distance estimation, and both the visual and non-visual input signals required to perform these computations. Additionally, the neural circuits that convert visual information into a jump command are also not well understood. Most work has examined jumping in nocifensive and defensive contexts rather than navigation (Barik et al., 2018; Wang et al., 2015), although a recent behavioral study demonstrated that squirrels learn to integrate multiple factors, including gap distance and branch flexibility, in executing a jump (Hunt et al., 2021).

Utility of studying natural distance estimation behavior

Natural behavior is often a continuous control process, which is fundamentally closed-loop, unlike stimulus-response paradigms that dominate behavior literature (Cisek, 1999). This task accordingly permits investigation of both how movement through the environment generates sensory cues useful for judging distance, and how the visual information is directly converted into a motor output. Furthermore, perception of spatial layout is an embodied process, and thus body- and action-scaling cues that are not available under conditions of restraint could provide distance information under the freely moving conditions of this task (Fajen, 2021). Critically, natural behaviors may be the most appropriate tool for studying the neural basis of sensory processing, since theoretical considerations suggest that neural circuits may perform sub-optimal inference under non-natural conditions (Beck et al., 2012). Finally, beyond studies of visual distance estimation, this task could provide a framework for integrated studies of motivation, motor control, and decision making within an ethological context.

MATERIALS AND METHODS

Animals: Male and female mice between postnatal day 40 (P40) and P365 were bred in-house in a C57/BL6J background. For optogenetic experiments, transgenic mice were used to target the expression of channelrhodopsin-2 to parvalbumin-expressing neurons (PV-Cre [Jax #008069] crossed to Ai32 [Jax #012569]; Hippenmeyer et al., 2005; Madisen et al., 2012). Mice were housed in a reverse 12h light-dark cycle room. Mice were placed under a water restriction schedule at the start of training, only receiving fluids during training/task periods. All procedures were performed in accordance with the University of Oregon Institute for Animal Care and Use Committee (IACUC) and Animal Care Services standard operating procedures.

Behavioral Apparatus and Jumping Task: The jumping arena was circular and roughly 30 cm high and 60 cm across. Mice self-initiated trials by mounting a take-off platform (15 cm height, 10 cm width, 10 cm depth, with 4 x 5 cm overhang in front). While blocking the mouse's view of the arena with a barrier, the experimenter then placed one of three platforms (10, 20, or 30 cm width, 19 cm height) at a random distance (8, 12, 16, 20, or 24 cm) from the edge of the take-off platform. Platforms were custom built from ¼" white acrylic, and tops were coated in white rubberized coating (Plasti-Dip) to prevent animals from slipping. A black strip was placed across the top leading edge of the landing platform, matched proportionally in height to platform width to

maintain height/width ratio). A static white noise background composed of grayscale squares (~1 deg each of visual angle from take-off platform) was mounted at the back of the arena. Six LED puck lights were evenly spaced around the top of the arena for even illumination. Cameras (FLIR BlackFly S USB3) were mounted above and to the side of the arena, and the entire behavioral session was recorded (720 x 540 pixels, 60 fps) with camera timestamps using a custom Bonsai workflow (Lopes et al., 2015). A custom Python script was used to generate randomized platform/distance combinations for the experimenter and to log trial outcomes and approximate jump times. The moment the barrier was lifted and the mouse was able to see the landing platform constituted the trial start, and the time elapsed until the mouse jumped was the "decision period." There were three possible trial outcomes: 1) the mouse jumped and successfully reached the landing platform and received a reward (success), 2) the mouse jumped and missed the landing platform and received no reward (failure), or 3) the mouse dismounted the take-off platform and received a light airpuff and a time-out (abort). A typical experiment lasted 30 minutes, during which mice performed ~30-60 trials.

Behavioral Training: Mice were habituated to the arena for 3 days with their cage mates, during which time they were individually handled and introduced to a clicker that indicated water reward (~25-50 ul), where each click is immediately followed by a reward. Mice were then individually clicker trained to mount a short take-off platform (10 cm height; click and reward upon mounting the platform), receiving water (administered by hand using a 1 ml syringe) and a small piece of tortilla chip (Juanita's). After 3-5 successful mounts, a landing platform (19 cm height) was placed against the take-off platform, and mice were clicker-rewarded for climbing up onto the landing platform. After three successful trials, the landing platform was moved slightly farther back, increasing the gap distance until jumping is required to reach the landing platform. At this point the clicker was typically no longer required. Once the mouse could jump to the maximum distance, the taller take-off platform used in the task was introduced, and landing platforms were again introduced at short distances and slowly moved farther away. Training was complete when mice could jump to all three landing platforms at the farthest distance, and typically took 1-2 weeks with all mice successfully learning the task.

Surgical Procedures: For all procedures, anesthesia was induced at 3% isoflurane and maintained at 1.5-2% in O_2 at a 1 l/min flow rate. Ophthalmic ointment was applied to both eyes, and body temperature was maintained using a closed-loop heating pad at 37°C. In order to minimize stress when plugging in optical tethers or

miniature cameras, a small steel headplate was mounted on the skull using dental acrylic (Unifast LC) to allow for brief head-fixation before the experiment.

Monocular suture: The area immediately surrounding the eye to be sutured was wiped with 70% ethanol before ophthalmic ointment was applied. Two to three mattress sutures were placed using 6-0 silk suture, opposing the full extent of the lid. The forepaw and hindpaw nails ipsilateral to the sutured eye were trimmed to help minimize post-procedural self-inflicted trauma.

Optic fiber implant: A minimal portion of scalp was resected bilaterally over visual cortex, and a small trepanation was made over each primary visual cortex (+1.0, +/- 2.5 mm from Lambda). Bilateral optic fibers (ceramic ferrules, thorlab fiber 0.5 mm length from end of ferrule) were stereotactically lowered into the burr hole and secured in place with dental acrylic. Vetbond was then applied to secure the skin in place around the implant. Fiber transmission rates were measured prior to implant and accounted for during experiments.

Miniature head-mounted cameras: To obtain high-resolution video of the eyes during behavior, a miniature camera (iSecurity), magnifying lens (12 mm focus, 6 mm diameter), and an infrared LED were mounted on a custom-designed 3D-printed plastic camera arm (Michaiel et al, 2020). Two miniature connectors (Mill Max 853-93-100-10-001000, cut to 2x4 pin) were glued to the headplate, and an equivalent connector on the camera arm was plugged in prior to the experiment. Camera power and data were passed through thin tethering wire (Cooner #CZ1174CLEAR) and acquired in Bonsai with system timestamps. The total hardware weight was approximately 3 g. Eye videos were deinterlaced to achieve 60 fps (matching top/side cameras) prior to analysis.

Data Analysis: Full task videos were first split into individual trials using custom Python software; the trial start, jump, and landing frame numbers were determined and individual trial videos were saved. These trial videos were then labeled using markerless pose estimation with DeepLabCut (DLC). A set of sample frames were manually labeled and used to train two networks (top/side cameras and eye cameras) that were then used to track features in all video data. DLC points from the decision period, defined as the last 5 s before the jump, were then passed through a median filter (n=3) and a convolutional smoothing filter (box, n=5). Values of the points tracking the nose, eye, and ear were then used as inputs to train an auto-regressive hidden Markov model (ARHMM) after centering across experiments by subtracting off values of the point that tracked the edge of the take-off platform. Model training was performed using the SSM package in python (Linderman et al.,

2019). Model selection was based on the elbow in two-fold cross validation log-likelihood curves across model iterations while balancing model interpretability with model fit (final model: K=6, lag=1, kappa=1e04, data temporally downsampled 2X, and one state discarded due to extremely low prevalence). ARHMM states were determined based on a posterior probability threshold of 0.8. Timepoints below the threshold were excluded from analysis. For lexical transition matrices, trials were first separated into binocular and monocular conditions. During the decision period of each trial, the transitions between unique ARHMM states were counted. The number of state transitions are then normalized by the total number of unique transitions per condition to calculate the relative frequency of transitions. For all summary analyses, data were first averaged within-animal (across days) and then across animals within a group (e.g. monocular, binocular). Statistical significance was determined using analysis of variance and the student's t-test with Bonferroni corrections for multiple comparisons. Unless otherwise noted, data are presented as mean ± standard error of the mean.

Decoding Analysis: We decoded the experimental condition (binocular vs. monocular, laser on vs. off) per animal from single-trial maximum a priori (MAP) motif sequences inferred using the ARHMM. Specifically, we trained binary decoders with a linear decision boundary (Linear Discriminant Analysis) to decode the above categorical variables from the single-trial empirical state transition probability matrices derived from the MAP sequence of each trial, thus providing not only state usage information, but transitions between states as information the classifier could use. For each animal, correct trials were pooled across distances to provide enough trials per class for decoding. Data were split into training and test datasets in a stratified 10-fold cross-validation manner, ensuring equal proportions of trials of different types (distance, platform width, visual condition, laser) in both datasets. To calculate the statistical significance of decoding accuracies, we performed an iterative shuffle procedure on each fold of the cross-validation, shuffling training labels and testing on unshuffled test labels 100 times to create a shuffle distribution for each fold of the cross-validation. From these distributions we calculated the z-score of decoding accuracy for each class in each cross-validation fold. These z-scores were then averaged across the folds of cross-validation and used to calculate the overall p-value of the decoding accuracy obtained on the original data. The decoding weights of the binary classifiers were examined as well to identify the significant transitions that contributed to decoding between visual conditions. The same shuffle procedure was used to assess significant elements of the classifier.

ACKNOWLEDGEMENTS

We thank members of the Niell lab for helpful discussions, and Dr. Michael Goard, Dr. David Leopold, Dr. Matt Smear, and Dr. Michael Stryker for feedback on the manuscript. This work benefited from access to the University of Oregon high performance computer, Talapas. This work was supported by grants from the National Institutes of Health: 5F32EY027696 (Parker), 1R21EY032708 (Niell, Parker), 1R34NS111669 (Niell). Some figure panels were made using Biorender.com.

REFERENCES

Barik A, Thompson JH, Seltzer M, Ghitani N, Chesler AT. 2018. A Brainstem-Spinal Circuit Controlling Nocifensive Behavior. *Neuron* **100**:1491–1503.e3.

Beck JM, Ma WJ, Pitkow X, Latham PE, Pouget A. 2012. Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron* **74**:30–39.

Bianco IH, Kampff AR, Engert F. 2011. Prey capture behavior evoked by simple visual stimuli in larval zebrafish. *Front Syst Neurosci* **5**:101.

Boone HC, Samonds JM, Crouse EC, Barr C, Priebe NJ, McGee AW. 2021. Natural binocular depth discrimination behavior in mice explained by visual cortical activity. *Curr Biol* **31**:2191–2198.e3.

Carey DP, Goodale MA, Sprowl EG. 1990. Blindsight in rodents: the use of a “high-level” distance cue in gerbils with lesions of primary visual cortex. *Behav Brain Res* **38**:283–289.

Cisek P. 1999. Beyond The Computer Metaphor: Behavior As Interaction. *Journal of Consciousness Studies* **6**:125-142.

Ellard CG, Goodale MA, Scorfield DM, Lawrence C. 1986. Visual cortical lesions abolish the use of motion parallax in the Mongolian gerbil. *Exp Brain Res* **64**:599–602.

Ellard CG, Goodale MA, Timney B. 1984. Distance estimation in the Mongolian gerbil: the role of dynamic depth cues. *Behav Brain Res* **14**:29–39.

Ellard CG, Sharma L. 1996. The effects of cortical lesions on recognition of object context in a visuomotor task in the Mongolian gerbil. *Behav Brain Res* **82**:13–22.

Fajen BR. 2021. Visual Control of Locomotion. doi:10.1017/9781108870474

Goodale MA, Ellard CG, Booth L. 1990. The role of image size and retinal motion in the computation of absolute distance by the Mongolian gerbil (Meriones unguiculatus). *Vision Res* **30**:399–413.

Hippenmeyer S, Vrieseling E, Sigrist M, Portmann T, Laengle C, Ladle DR, Arber S. 2005. A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol* **3**:e159.

Holmgren CD, Stahr P, Wallace DJ, Voit K-M, Matheson EJ, Sawinski J, Bassetto G, Kerr JND. 2021. Freely-moving mice visually pursue prey using a retinal area with least optic flow. *bioRxiv.* doi:10.1101/2021.06.15.448520

Hoy JL, Yavorska I, Wehr M, Niell CM. 2016. Vision Drives Accurate Approach Behavior during Prey Capture in Laboratory Mice. *Curr Biol* **26**:3046–3052.

Hunt NH, Jinn J, Jacobs LF, Full RJ. 2021. Acrobatic squirrels learn to leap and land on tree branches without falling. *Science* **373**:697–700.

Hutson KA, Masterton RB. 1986. The sensory contribution of a single vibrissa’s cortical barrel. *J Neurophysiol* **56**:1196–1223.

Kim HR, Angelaki DE, DeAngelis GC. 2017. Gain Modulation as a Mechanism for Coding Depth from Motion Parallax in Macaque Area MT. *J Neurosci* **37**:8180–8197.

Kral K. 2003. Behavioural-analytical studies of the role of head movements in depth perception in insects, birds and mammals. *Behav Processes* **64**:1–12.

Land MF. 1999. Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*. doi:10.1007/s003590050393

Legg CR, Lambert S. 1990. Distance estimation in the hooded rat: experimental evidence for the role of motion cues. *Behav Brain Res* **41**:11–20.

Leopold DA, Park SH. 2020. Studying the visual brain in its natural rhythm. *Neuroimage* **216**:116790.

Linderman S, Nichols A, Blei D, Zimmer M, Paninski L. 2019. Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in C. elegans. *bioRxiv*. doi:10.1101/621540

Lopes G, Bonacchi N, Frazão J, Neto JP, Atallah BV, Soares S, Moreira L, Matias S, Itskov PM, Correia PA, Medina RE, Calcaterra L, Dreosti E, Paton JJ, Kampff AR. 2015. Bonsai: an event-based framework for processing and controlling data streams. *Front Neuroinform* **9**:7.

Madisen L, Mao T, Koch H, Zhuo J-M, Berenyi A, Fujisawa S, Hsu Y-WA, Garcia AJ 3rd, Gu X, Zanella S, Kidney J, Gu H, Mao Y, Hooks BM, Boyden ES, Buzsáki G, Ramirez JM, Jones AR, Svoboda K, Han X, Turner EE, Zeng H. 2012. A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nat Neurosci* **15**:793–802.

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* **21**:1281–1289.

Meyer AF, O'Keefe J, Poort J. 2020. Two Distinct Types of Eye-Head Coupling in Freely Moving Mice. *Curr Biol* **30**:2116–2130.e6.

Michaiel AM, Abe ET, Niell CM. 2020. Dynamics of gaze control during prey capture in freely moving mice. *Elife* **9**. doi:10.7554/eLife.57458

Nadler JW, Angelaki DE, DeAngelis GC. 2008. A neural representation of depth from motion parallax in macaque visual cortex. *Nature* **452**:642–645.

Nadler JW, Nawrot M, Angelaki DE, DeAngelis GC. 2009. MT neurons combine visual motion with a smooth eye movement signal to code depth-sign from motion parallax. *Neuron* **63**:523–532.

Parker PRL, Brown MA, Smear MC, Niell CM. 2020. Movement-Related Signals in Sensory Areas: Roles in Natural Behavior. *Trends Neurosci* **43**:581–595.

Richardson F. 1909. A study of the sensory control in the rat. *The Psychological Review: Monograph Supplements*. doi:10.1037/h0093009

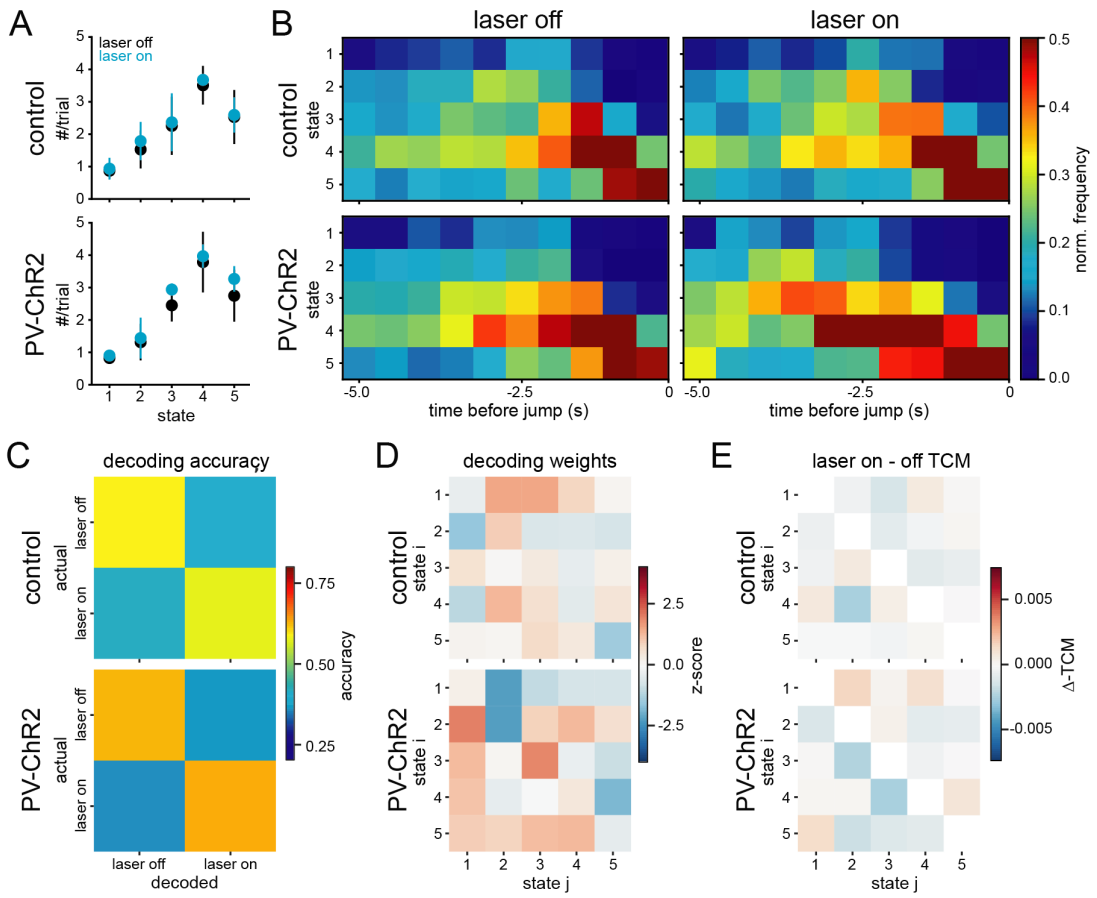
Shinkman PG. 1962. Visual depth discrimination in animals. *Psychological Bulletin*. doi:10.1037/h0040912

Wallace DJ, Greenberg DS, Sawinski J, Rulla S, Notaro G, Kerr JND. 2013. Rats maintain an overhead binocular field at the expense of constant fusion. *Nature* **498**:65–69.

Wallace GK. 1959. Visual Scanning in the Desert Locust Schistocerca Gregaria Forskål. *Journal of Experimental Biology*. doi:10.1242/jeb.36.3.512

Wang L, Chen IZ, Lin D. 2015. Collateral pathways from the ventromedial hypothalamus mediate defensive behaviors. *Neuron* **85**:1344–1358.

Yilmaz M, Meister M. 2013. Rapid innate defensive responses of mice to looming visual stimuli. *Curr Biol* **23**:2011–2015.



Supplemental Figure S1: V1 optogenetic suppression does not affect head movements. (Relates to Figure 5). A) Frequency of each ARHMM state for laser off (black) and laser on (cyan) conditions for control (top) and PV-ChR2 (bottom) animals. B) Heatmaps of start time histograms for each state normalized to the total number of trials for laser off (left) and laser on (right) conditions for control (top) and PV-ChR2 (bottom) animals. C) Two-fold decoding analysis on transition count matrices for laser on vs. laser off conditions (performed within-animal, averaged across animals) for control (top) and PV-ChR2 (bottom) animals. D) Z-scored weights used to decode laser off vs. laser on conditions for control (top) and PV-ChR2 (bottom) animals. E) Difference between laser off and laser on transition count matrices for control (top) and PV-ChR2 (bottom) animals; red transitions are more frequent in laser on, blue in laser off. n=4 mice for each group (control/PV-ChR2) for all plots.

Supplementary Video 1: Mouse performing the task under binocular conditions with DeepLabCut labels overlaid.

Supplementary Video 2: Mouse performing the task with miniature head-mounted cameras tracking both eyes.

Supplementary Video 3: PV-ChR2 mouse jumping during a laser-off trial.

Supplementary Video 4: PV-ChR2 mouse jumping during a laser-on trial.