

Sensory and Motor Systems

# Optimized but Not Maximized Cue Integration for 3D Visual Perception

Ting-Yu Chang,<sup>1</sup> Lowell Thompson,<sup>1</sup> Raymond Doudlah,<sup>1</sup> Byoungsoon Kim,<sup>1</sup> Adhira Sunkara,<sup>2</sup> and Ari Rosenberg<sup>1</sup>

<https://doi.org/10.1523/ENEURO.0411-19.2019>

<sup>1</sup>Department of Neuroscience, School of Medicine and Public Health, University of Wisconsin–Madison, Madison, WI 53705 and <sup>2</sup>Department of Surgery, School of Medicine and Public Health, University of Wisconsin–Madison, Madison, WI 53705

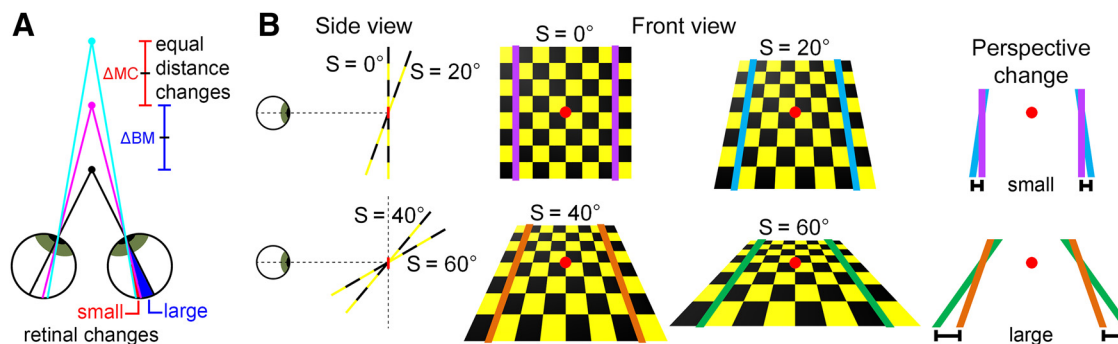
## Abstract

Reconstructing three-dimensional (3D) scenes from two-dimensional (2D) retinal images is an ill-posed problem. Despite this, 3D perception of the world based on 2D retinal images is seemingly accurate and precise. The integration of distinct visual cues is essential for robust 3D perception in humans, but it is unclear whether this is true for non-human primates (NHPs). Here, we assessed 3D perception in macaque monkeys using a planar surface orientation discrimination task. Perception was accurate across a wide range of spatial poses (orientations and distances), but precision was highly dependent on the plane's pose. The monkeys achieved robust 3D perception by dynamically reweighting the integration of stereoscopic and perspective cues according to their pose-dependent reliabilities. Errors in performance could be explained by a prior resembling the 3D orientation statistics of natural scenes. We used neural network simulations based on 3D orientation-selective neurons recorded from the same monkeys to assess how neural computation might constrain perception. The perceptual data were consistent with a model in which the responses of two independent neuronal populations representing stereoscopic cues and perspective cues (with perspective signals from the two eyes combined using nonlinear canonical computations) were optimally integrated through linear summation. Perception of combined-cue stimuli was optimal given this architecture. However, an alternative architecture in which stereoscopic cues, left eye perspective cues, and right eye perspective cues were represented by three independent populations yielded two times greater precision than the monkeys. This result suggests that, due to canonical computations, cue integration for 3D perception is optimized but not maximized.

**Key words:** 3D visual perception; canonical computations; divisive normalization; optimal cue integration; perspective; stereoscopic

## Significance Statement

Our eyes sense two-dimensional (2D) projections of the world, like a movie on a screen, but we perceive the world as three-dimensional (3D). Here, we show that non-human primates (NHPs), like humans, achieve more precise 3D vision by perceptually integrating distinct 3D cues. We also present evidence that perception is influenced by 3D natural scene statistics, and that priors over 3D orientation are subjectively encoded. Using simulations, we examine how neural computation can constrain 3D perception and estimate that perception is half as precise as theoretically possible. Our findings suggest that the concurrence of multiple canonical computations simultaneously optimizes and curbs 3D visual perception, and highlight that what constitutes optimal task performance depends on the underlying neural architecture.



**Figure 1.** 3D cue reliabilities depend on object pose. **A**, Stereoscopic cue reliability decreases with distance. Equivalent changes in object distance produce smaller retinal image changes at greater distances. This is illustrated with an observer fixating the black dot. The distance between black and magenta dots ( $\Delta BM$ ) is equal to the distance between magenta and cyan dots ( $\Delta MC$ ), but the retinal change is larger for  $\Delta BM$  than  $\Delta MC$ . **B**, The reliability of perspective cues increases with orientation-in-depth (slant). Equivalent slant changes produce larger changes in the rate at which parallel lines converge in the 2D projection at larger base slants. This is illustrated with a checkerboard rotated about the horizontal axis passing through the red dot. Colored lines are parallel in the world. A 20° slant ( $S$ ) rotation produces a smaller perspective change between 0° and 20° (top row) than between 40° and 60° (bottom row).

## Introduction

Three-dimensional (3D) visual perception is a significant achievement of the primate brain (Barton, 2004). Because eyes sense two-dimensional (2D) projections of the world, 3D structure must be estimated. Creating 3D percepts from 2D images is a nonlinear optimization problem plagued by ambiguities and noise (Hartley and Zisserman, 2003). Studies with humans have shown that integrating distinct visual cues resolves ambiguities and improves 3D estimates (Knill and Saunders, 2003; Hillis et al., 2004; Welchman et al., 2005; Preston et al., 2009; Murphy et al., 2013; Rideaux and Welchman, 2018). In particular, stereoscopic and perspective cues have prominent roles in human 3D vision (Stevens, 1981; Howard and Rogers, 1995; Knill, 1998a). The reliability of these cues depends on an object's spatial pose (position and orientation). Stereoscopic cue reliability decreases with distance from the observer (Fig. 1A). Perspective cue reliability increases with orientation-in-depth (slant; Fig. 1B). Humans perceptually integrate these cues near-optimally, with weights that match the reliabilities. Although non-human primates (NHPs) are an essential model system for study-

ing the neural basis of 3D vision, surprisingly little is known about how they perceptually integrate these cues.

Here, we quantified 3D perception in NHPs using an eight-alternative forced choice (8AFC) surface orientation discrimination task. The contributions of stereoscopic and perspective cues to perception were assessed by presenting cue-isolated and combined-cue stimuli. With stereoscopic cues, performance decreased with distance from the fixation plane, consistent with the geometry of stereovision and the physiology of stereopsis (Howard and Rogers, 1995; Cumming and DeAngelis, 2001; Prince et al., 2002). With both cues, performance increased with slant. The perception of combined-cue stimuli was consistent with an optimal integration strategy (Murray and Morgenstern, 2010; Landy et al., 2011). Moreover, errors in perception could be explained by a prior resembling the 3D orientation statistics of natural scenes (Adams et al., 2016; Burge et al., 2016).

It is also unclear how the architecture of the visual system and circuit-level computations constrain 3D perception. The neuronal implementation of optimal cue integration is theoretically linear (Ma et al., 2006), but nonlinear computations such as quadratics and divisive normalization are widely implicated in neural processing (Busse et al., 2009; Beck et al., 2011; Carandini and Heeger, 2012; Ni et al., 2012; Louie et al., 2013; Qamar et al., 2013; Rosenberg et al., 2015). These nonlinearities can introduce dependencies between the representations of individual cues, and may therefore limit the precision of perception. Specifically, even if the individual cues were optimally integrated, the perception of combined-cue stimuli would be less precise if their representations were dependent than if they were independent (Oruç et al., 2003). We therefore performed simulations to assess the impact of the neural architecture and computations on perception.

We found that combined-cue perception was consistent with a neural architecture in which stereoscopic cues and perspective cues were represented by independent neuronal populations. To account for perception, a key

Received October 2, 2019; accepted December 8, 2019; First published December 13, 2019.

The authors declare no competing financial interests.

Author contributions: T.-Y.C., B.K., A.S., and A.R. designed research; T.-Y.C., L.T., R.D., B.K., and A.R. performed research; T.-Y.C., L.T., R.D., B.K., and A.R. analyzed data; T.-Y.C., L.T., R.D., A.S., and A.R. wrote the paper.

This work was supported by the Alfred P. Sloan Foundation Grant FG-2016-6468, the Whitehall Foundation Research Grant 2016-08-18, a Shaw Scientist grant from the Greater Milwaukee Foundation, and the National Institutes of Health Grant EY029438. Further support was provided by the National Institutes of Health Grant P51OD011106 to the Wisconsin National Primate Research Center, University of Wisconsin–Madison.

Correspondence should be addressed to Ari Rosenberg at ari.rosenberg@wisc.edu.

<https://doi.org/10.1523/ENEURO.0411-19.2019>

Copyright © 2020 Chang et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

element was to combine perspective cues detected by the two eyes using quadratic nonlinearities and divisive normalization before their integration with stereoscopic cues. The integration of stereoscopic cues and perspective cues was statistically optimal since the neuronal representations were linearly summed (Ma et al., 2006). However, cue integration was not maximized. An alternative architecture in which stereoscopic cues, left eye perspective cues, and right eye perspective cues were each represented by independent populations yielded two times greater precision. Our findings suggest that cue integration is a conserved computation by which primates achieve robust 3D vision, generate testable hypotheses about neural architectures responsible for 3D perception, and indicate that the concurrence of multiple canonical neural computations (linear summation, quadratics, and divisive normalization) may simultaneously optimize and curb perception.

## Materials and Methods

### Subjects and preparation

All surgeries and experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Wisconsin–Madison, and in accordance with the National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Two male rhesus macaques (*Macaca mulatta*) participated (Monkey L: five years of age, ~7.8 kg in weight; Monkey F: four years, ~5.5 kg). A Delrin ring for stabilizing the head during training and experimental sessions was attached to the skull under general anesthesia. After recovery, the monkeys were trained to sit in a custom primate chair with head restraint, and to fixate targets for 2 s at simulated depths ranging between –20 and 40 cm from the screen (located at 57 cm) within 2° version and 1° vergence windows for a liquid reward. Eye positions were monitored optically at 1000 Hz (EyeLink 1000 plus, SR Research).

### Experimental control and stimulus presentation

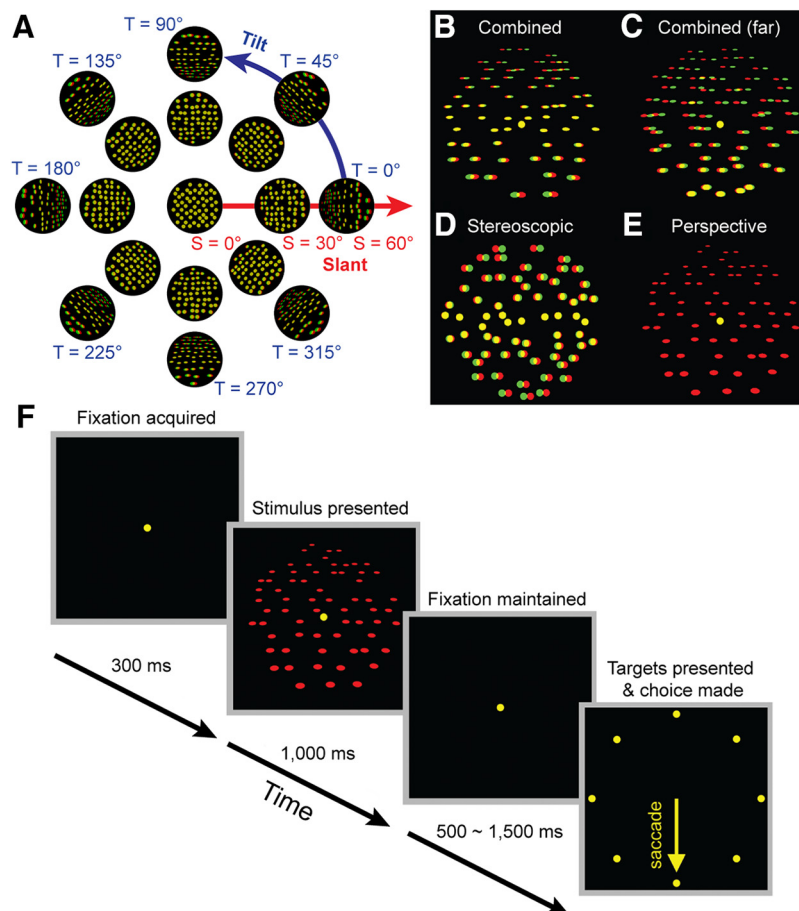
Experimental control was performed using an open-source, network-based parallel processing framework (Kim et al., 2019). Stimuli were created in MATLAB using Psychtoolbox 3 (Kleiner et al., 2007), and rendered with anti-aliasing using an NVIDIA GeForce GTX 970 graphics card on a Linux workstation (Ubuntu 16.04 LTS, Intel Xeon Processor, 24 GB RAM). A DLP LED projector (VPixx Technologies, Inc.) rear projected the stimuli at 1280 × 720-pixel resolution with a 240-Hz refresh rate onto a polarization preserving screen (Stewart Film Screen, Inc.). The projected area subtended 70° × 43° of visual angle. Stereoscopic presentation was achieved by sequencing the presentation of “half-images” to each eye (120 Hz/eye) using a circular polarizer synchronized to the projector. Polarized glasses were worn.

### Visual stimuli

Planar surfaces were defined using random dot patterns ( $N = 250$  dots). The dots were bright (35.1 cd/m<sup>2</sup>) on a gray (11.06 cd/m<sup>2</sup>) background (PR-524 LiteMate,

Photo Research), measured through the glasses. On the screen, the stimulus envelope was circular and subtended 20°. We defined planar surface orientation using two angular variables: tilt and slant (Stevens, 1983b; Rosenberg et al., 2013). Tilt is a rotation about an axis parallel to the line of sight, and therefore specifies the direction that the plane is oriented in depth. Tilt was defined such that: right-near = 0°, top-near = 90°, left-near = 180°, and bottom-near = 270°. Slant is a rotation about an axis perpendicular to the line of sight, and therefore specifies the amount of depth variation. Together, these variables define a polar coordinate system (Fig. 2A). Planes were presented at all combinations of eight tilts (0° to 315° in 45° steps), four slants (15° to 60° in 15° steps), and either six (Monkey L: 37, 57, 77, 87, 107, and 137 cm) or eight (Monkey F: 37, 57, 77, 87, 97, 107, 117, and 137 cm) distances. Fixation was always at screen distance (57 cm), regardless of the stimulus distance. The stimulus distances corresponded to horizontal disparity pedestals of: –1.63° (37 cm), 0° (57 cm), 0.78° (77 cm), 1.04° (87 cm), 1.24° (97 cm), 1.41° (107 cm), 1.55° (117 cm), and 1.76° (137 cm). At 37 cm, all dots were in front of the plane of fixation. At 57 cm, dots were distributed in front of and behind the plane of fixation. At 77 cm and beyond, all dots were behind the plane of fixation. The protocol was thus similar to previous studies which varied the stimulus distance while holding the fixation distance constant (Nguyenkim and DeAngelis, 2003; Hegdé and Van Essen, 2005; Ban and Welchman, 2015; Alizadeh et al., 2018; Elmore et al., 2019; Henderson et al., 2019). Others have yoked the stimulus and fixation distances (Banks et al., 2001; Hillis et al., 2004). By presenting stimuli with the dots entirely in front of, distributed about, or entirely behind the plane of fixation, we prevented the monkeys from relying on local absolute disparity cues to perform the discrimination task (described below), ensuring that they judged the plane's 3D orientation (Elmore et al., 2019).

Three cue conditions were included. For combined-cue stimuli, both stereoscopic and perspective cues signaled the plane's orientation (Fig. 2B,C). The dots were uniformly distributed across the plane in the world. Left and right eye half-images were rendered using projective geometry, resulting in retinal density gradients, foreshortening, and scaling. The baseline dot size was 0.35°. For the largest slant tested (60°), the major and minor axes of dots on the axis of rotation were 0.35° × 0.18° (5.6 × 2.8 pixels). At the nearest point, they were 0.45° × 0.29° (7.3 × 4.8 pixels). At the furthest point, they were 0.24° × 0.08° (3.9 × 1.4 pixels). As in previous work (Hillis et al., 2004), the reliability of the perspective cues was held constant with distance by scaling the dots such that their screen size only depended on slant. Stereoscopic cue stimuli were created by defining a uniform distribution of dots on the screen and using ray tracing to assign each dot to a location on the plane in the world (Fig. 2D). All dots had a circular shape and subtended 0.35°, irrespective of the plane's pose. The perspective cue stimuli were the same as the combined-cue stimuli but only one eye saw the plane (pseudo-randomly selected each trial) while both eyes saw the fixation target (Fig. 2E).



**Figure 2.** Stimuli and discrimination task. **A**, Tilt ( $T$ ) and slant ( $S$ ) are polar coordinates describing planar surface orientation. Tilt specifies the direction that the plane is oriented in depth. Slant specifies how much it is oriented in depth. **B–E**, Example planes ( $T = 270^\circ$ ,  $S = 60^\circ$ ). For clarity, the dot size is exaggerated and the dot number is reduced from the actual experiments. **B**, Combined-cue stimulus at 57 cm (fixation distance). **C**, Combined-cue stimulus at 77 cm (all dots behind the plane of fixation). **D**, Stereoscopic cue stimulus at 57 cm. **E**, Perspective cue stimulus at 57 cm (left eye presentation). **F**, Eight alternative tilt discrimination task. Fixation was held on a target presented at 57 cm (screen distance) for 300 ms. A plane then appeared for 1000 ms. Fixation was then held for 500–1500 ms before the fixation target disappeared and eight choice targets appeared. The plane's tilt was reported through a saccade to a choice target. For example, the bottom target for a bottom-near plane ( $T = 270^\circ$ ). Planes are illustrated here using red-green anaglyphs.

### Tilt discrimination task

The monkeys were trained to report planar tilt in an 8AFC task. They first learned to perform a right-near versus left-near task with all slants, distances, and cue conditions interleaved. The correct choice target initially had a higher contrast than the distractor, and the contrast difference was reduced with training. Once an 80% correct rate with equal target contrasts was reached for planes at 57 cm, the four cardinal tilts were introduced with a target contrast difference. Once a 50% correct rate was reached with equal contrasts for planes at 57 cm, we started alternating training days between four cardinal and four oblique tilts. Once a 50% correct rate was reached with both sets at 57 cm, all eight tilts were introduced. Data collection began after performance in the 8AFC task stabilized. The correct rate depended on the surface pose and cue condition. Across all stimuli, each monkey had a mean correct rate of 55% and a SD of 27% (Monkey L:  $N = 576$ ; Monkey F:  $N = 768$ ). Both monkeys showed chance-level performance in a relatively

small number of stimulus conditions (Monkey L: 45/576, 7.8%; Monkey F: 9/768, 1.2%), but this had no obvious impact on their motivation to perform the task.

In the task (Fig. 2F), a monkey first acquired fixation of a target presented at 57 cm in the center of the screen. The target was a red dot ( $9.6 \text{ cd/m}^2$  through the glasses) subtending  $0.3^\circ$ . After fixating for 300 ms, a plane was presented at the center of the screen for 1000 ms while fixation was maintained. Fixation was then held for an additional 500–1500 ms (pseudo-random duration) before the fixation target disappeared and 8 choice targets (subtending  $0.7^\circ$ ,  $35.1 \text{ cd/m}^2$  through the glasses) appeared at  $11^\circ$  eccentricity and polar angles of  $0$ – $315^\circ$  in  $45^\circ$  steps. The side of the plane nearest the monkey was reported with a saccade to the choice target at the corresponding polar angle for a liquid reward. A trial was aborted if fixation was broken before the choice targets appeared or if a choice was not made within 500 ms, and the data were discarded. In NHP 3D vision studies, vergence is often enforced offline only (e.g., Sanada and



DeAngelis, 2014). Similarly, we enforced version and vergence with 2° windows during the task, and offline used a 0.5° window to eliminate trials with vergence errors  $\geq 0.25^\circ$ .

Stimuli were presented in a pseudo-random order in a block design. A block included one repetition of each combination of tilt, slant, distance, and cue condition (Monkey L:  $N = 576$ ; Monkey F:  $N = 768$ ). Aborted trials were reinserted at pseudo-random locations within the block, which did not finish until a response was provided for each stimulus. The percentage of trials that were aborted did not depend on the tilt, slant, distance, or cue condition for either monkey ( $\chi^2$ , all  $p > 0.91$ ). Monkey L completed 16,883 trials. Monkey F completed 52,586 trials.

### Stereoscopic cue controls

We tested whether the stereoscopic cue stimuli contained perspective cues that could be used to perform the task. Stereoscopic cue stimuli were presented binocularly as in the main experiment (both eyes saw the plane) as well as monocularly to eliminate stereoscopic cues (one eye saw the plane, both saw the fixation target) at 57 cm. To maximize potential perspective cues, the planes were presented at the largest tested slant (60°). Parameters were otherwise the same as in the main experiment. All planes were presented interleaved. Monkey L completed 1192 trials. Monkey F completed 1780 trials.

We also tested whether perception of the stereoscopic cue stimuli was affected by a potential conflict between the stereoscopically defined slant of the planes and the isotropic shape of the dots. To maximize the potential conflict, the planes were presented at 60° of slant. Eleven dot numbers ranging from 5 to 250 (steps of 25 starting at 25) were used. The planes were presented at 57 cm for Monkey L, and at 57 and 97 cm for Monkey F. Parameters were otherwise the same as in the main experiment. All planes were presented interleaved. Monkey L completed 3008 trials. Monkey F completed 5521 trials at 57 cm, and 4916 trials at 97 cm.

### Analyses

#### Quantifying tilt perception

To quantify performance, we computed probability density functions describing the errors in reported tilts. First, we took the difference between the reported and presented tilt for each trial:  $\Delta\text{Tilt} = \text{reported tilt} - \text{presented tilt}$ . Second, we built error distributions by calculating the probability of each  $\Delta\text{Tilt}$ . The error distributions were calculated using either all tilts or just one tilt, depending on the analysis. Third, a von Mises probability density function was fit to each error distribution:

$$VM(\Delta\text{Tilt}) = e^{(\kappa \cdot \cos(\Delta\text{Tilt} - \mu))} / (2\pi \cdot I_0(\kappa)). \quad (1)$$

The mean ( $\mu$ ) and concentration ( $\kappa$ ) parameters describe the accuracy and precision of perception, respectively (Murray and Morgenstern, 2010; Seilheimer et al., 2014; Dakin and Rosenberg, 2018). The closer  $\mu$  is to 0, the more accurate (less biased) the judgments. The larger the  $\kappa$ , the more concentrated the distribution, indicating

more precise judgments. A modified Bessel function of order 0,  $I_0(\kappa)$ , normalizes the function to have unit area. The tilt sampling interval limits the maximum  $\kappa$  that can be estimated. Simulations showed that we could not distinguish between  $\kappa$ 's  $> 18$  ( $\geq 90\%$  of the density function within the sampling interval), so we set 18 as the upper bound in the estimation routine.

#### Optimal cue integration

We used optimal cue integration theory for circular variables to predict the combined-cue performance (Murray and Morgenstern, 2010). Predictions were derived from the stereoscopic and perspective cue biases ( $\mu_S$  and  $\mu_P$ , respectively) and precisions ( $\kappa_S$  and  $\kappa_P$ , respectively). The optimal combined-cue parameters (bias:  $\hat{\mu}_C$ ; precision:  $\hat{\kappa}_C$ ) are:

$$\hat{\mu}_C = \tan^{-1} \left( \frac{\kappa_S \cdot \sin(\mu_S) + \kappa_P \cdot \sin(\mu_P)}{\kappa_S \cdot \cos(\mu_S) + \kappa_P \cdot \cos(\mu_P)} \right) \quad (2)$$

and

$$\hat{\kappa}_C = \sqrt{\kappa_S^2 + \kappa_P^2 + 2 \cdot \kappa_S \cdot \kappa_P \cdot \cos(\mu_S - \mu_P)}. \quad (3)$$

#### Estimating a prior over tilt

For each stimulus condition, we used the product of a sensory likelihood function and a prior over tilt to model the perceptual data (Landy et al., 2011):

$$p(T | \hat{T}) \propto L(\hat{T} | T) \cdot p(T). \quad (4)$$

The posterior distribution,  $p(T | \hat{T})$ , indicates the probability that tilt  $T$  was presented given that tilt  $\hat{T}$  was sensed. The likelihood of sensing  $\hat{T}$  given  $T$ ,  $L(\hat{T} | T)$ , was assumed to be an unbiased von Mises function. The prior over tilt,  $p(T)$ , was constrained based on the natural scene statistics of surface tilt which includes more cardinal than oblique tilts, and equal left-near and right-near tilts (Adams et al., 2016; Burge et al., 2016). The prior was thus modeled as the mean of four von Mises densities centered on the cardinal tilts (Alberts et al., 2016). The concentration parameters of the densities centered at 0° and 180° were equal to reflect the symmetry of world tilts. Thus, the prior had three free parameters:  $\kappa_{0^\circ}$  &  $180^\circ$ ,  $\kappa_{90^\circ}$ , and  $\kappa_{270^\circ}$ .

A prior has the largest effect at low precisions. Since we found little to no bias in tilt perception for most conditions, we estimated the prior using a subset of the data, restricted to the first quartile of  $\kappa$ 's (Monkey L:  $\kappa \leq 1.59$ ,  $N = 144$ ; Monkey F:  $\kappa \leq 1.32$ ,  $N = 192$ ). For this subset of conditions, we fit each likelihood concentration parameter and a common prior by minimizing the root mean square error between the proportion of responses made by the monkeys and simulated with the model by binning 50,000 samples from the posterior into the eight tilt options. We bootstrapped the 95% confidence interval of the prior using the same data ( $N = 1000$  repetitions). Concentration parameters for likelihood functions with  $\kappa$ 's in the upper three quartiles were then fit using the estimated prior.

For each combination of tilt, slant, and distance, the stereoscopic likelihood function,  $L_S(\hat{T}|T)$ , perspective likelihood function,  $L_P(\hat{T}|T)$ , and common prior were likewise used to predict the combined-cue performance following Bayes' rule:

$$p(T|\hat{T}) \propto L_S(\hat{T}|T) \cdot L_P(\hat{T}|T) \cdot p(T). \tag{5}$$

For comparison with the behavior, the predicted performance was fit with a von Mises probability density function.

**Neuronal cue integration models**

We used Bayesian decoding of modeled 3D orientation-selective neurons to assess how neural computation might constrain tilt perception (Ma et al., 2006; Rosenberg et al., 2015). Assuming independent neurons with Poisson spike count statistics, the probability that tilt  $T$  elicits population response ( $\mathbf{r}$ ) is:

$$p(\mathbf{r}|T) = \prod_i \frac{e^{-f_i(T)} \cdot f_i(T)^{r_i}}{r_i!}. \tag{6}$$

Here,  $f_i(T)$  and  $r_i$  are the  $i^{\text{th}}$  neuron's tilt tuning curve and response, respectively. The posterior,  $p(T|\mathbf{r})$ , describing the probability that  $T$  was presented given  $\mathbf{r}$  is proportional to Equation 6 (assuming a uniform prior, a simplification we made to focus on computations specific to cue integration and since the influence of a prior was relatively small in our data). As the number of neurons increases,  $p(T|\mathbf{r})$  converges to a Gaussian. Assuming  $p(T|\mathbf{r})$  guides behavior, the precision of perception ( $1/\sigma^2$  for that Gaussian) is proportional to the gain of the population activity ( $g$ ). The proportionality constant ( $\lambda$ ) depends on the number of neurons and their tuning widths (Ma et al., 2006).

The tuning curves were modeled based on 3D orientation-selective neurons in the caudal intraparietal (CIP) area (Taira et al., 2000; Tsutsui et al., 2002, 2003; Rosenberg et al., 2013). The stimuli used for the neuronal recordings were the same as in the behavioral task, except that only combined-cue stimuli at distances of 37, 57, 97, and 137 cm were shown. The tilt tuning curve at each slant–distance combination was fit with a von Mises function (Rosenberg and Angelaki, 2014a). For each monkey, we calculated the mean response amplitude and tuning width across neurons at each slant–distance combination (Monkey L:  $N = 169$ ; Monkey F:  $N = 180$ ), and linearly interpolated the values for untested distances. These parameters were used to simulate 72 neurons for each monkey, with a 5° spacing between tilt preferences. The model tuning curves were homogeneous in that for a given slant–distance combination, only the preferred tilt differed.

To determine the  $\lambda$  relating gain to precision, we minimized the difference between the observed and simulated perceptual precisions decoded after scaling the population responses by  $\lambda(\kappa_{S,D})$ , which was a function of the slant and distance dependent tuning width ( $\kappa_{S,D}$ ). Four functions were tested: linear, exponential, double exponential, and two-phase exponential. The exponential

function,  $\lambda(\kappa_{S,D}) = DC + G \cdot \exp(-\alpha \cdot \kappa_{S,D})$ , provided the best fit based on Akaike's information criterion (Monkey L:  $DC = 1.3 \times 10^{-3}$ ,  $G = 3.70$ ,  $\alpha = 2.37$ ; Monkey F:  $DC = 1.5 \times 10^{-3}$ ,  $G = 2.03$ ,  $\alpha = 2.50$ ).

Using these  $\lambda(\kappa_{S,D})$  functions, we then fit the response amplitude for each cue-isolated condition to match the monkeys' precisions. Since CIP tilt tuning widths are similar regardless of the defining cue (Tsutsui et al., 2001, 2002), we assumed that the tuning widths were constant across cue conditions. The tilt tuning curves of the model neurons were:

$$f_i(T|S, D, C) = g_{S,D,C} \cdot e^{\kappa_{S,D}[\cos(T-\mu_i)-1]}. \tag{7}$$

Here,  $f_i(T|S, D, C)$  is the  $i^{\text{th}}$  neuron's tilt tuning curve for a given slant ( $S$ ), distance ( $D$ ), and cue condition ( $C$ ). The gain ( $g_{S,D,C}$ ) depended on the slant, distance, and cue condition. The tuning width ( $\kappa_{S,D}$ ) depended on the slant and distance. The preferred tilt is  $\mu_i$ .

Stereoscopic ( $\mathbf{r}_S$ ), left eye perspective ( $\mathbf{r}_{P_L}$ ), and right eye perspective ( $\mathbf{r}_{P_R}$ ) population responses were simulated. We tested three architectures for integrating these responses to create a combined-cue representation ( $\mathbf{r}_C$ ) that was decoded after scaling by  $\lambda(\kappa_{S,D})$ . Since the tuning curves were homogeneous and the simulations did not incorporate a prior, precision was independent of tilt. All simulations were performed with  $T = 180^\circ$ . Since Equation 6 converges to a Gaussian, the decoded posteriors were fit with Gaussian probability functions. To allow for direct comparisons, we refit the monkeys' error distributions with Gaussians. Gaussian and von Mises precision estimates were highly correlated (Monkey L:  $r = 0.99$ ,  $p = 3.8 \times 10^{-74}$ ; Monkey F:  $r = 0.99$ ,  $p = 1.2 \times 10^{-80}$ ).

*Three independent cue-isolated populations*

In this architecture, stereoscopic cues, left eye perspective cues, and right eye perspective cues were represented by three independent neuronal populations. Optimal cue integration was performed by a fourth population that linearly summed the three cue-isolated population responses to create a combined-cue representation (Ma et al., 2006),  $\mathbf{r}_C$ :

$$\mathbf{r}_C = \overbrace{\mathbf{r}_S}^{\text{stereo. pop.}} + \overbrace{\mathbf{r}_{P_L}}^{\text{left eye persp. pop.}} + \overbrace{\mathbf{r}_{P_R}}^{\text{right eye persp. pop.}}. \tag{8}$$

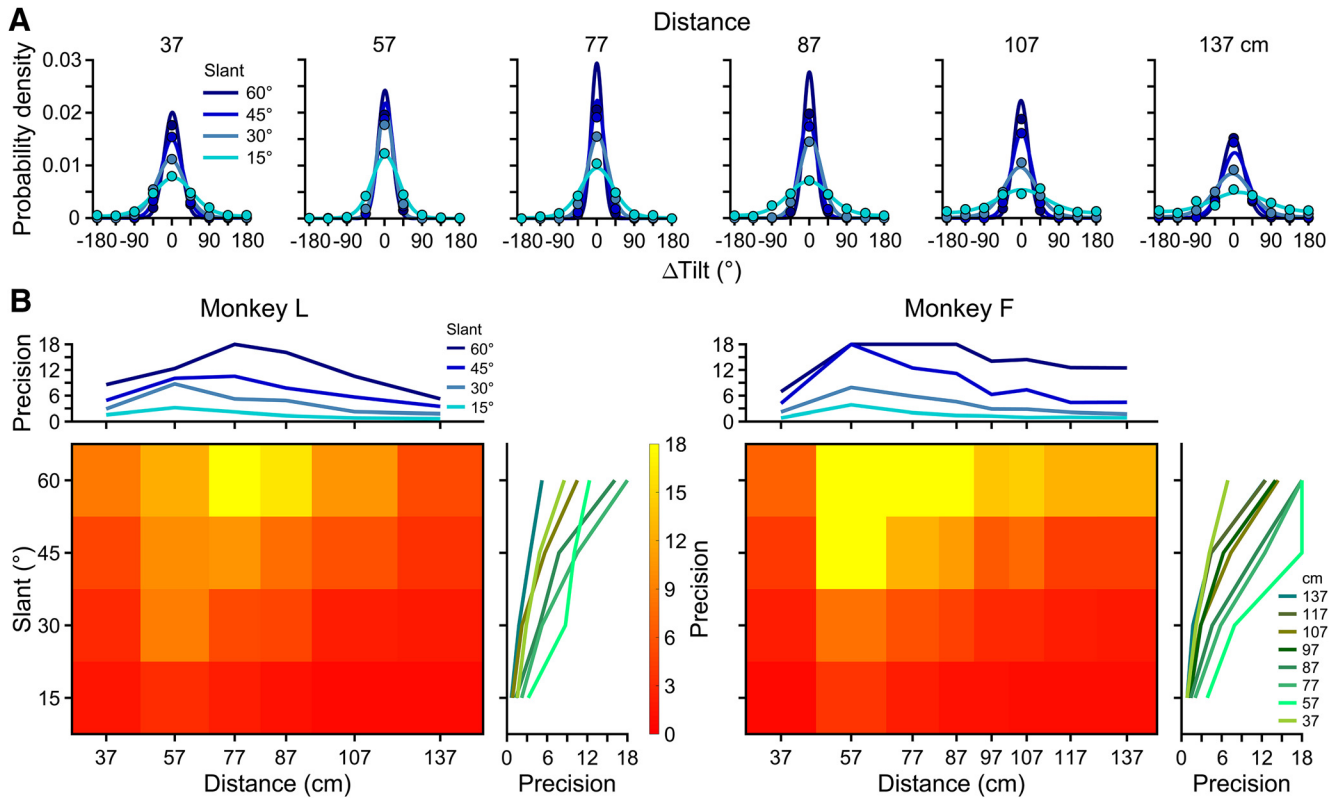
*Two independent cue-isolated populations*

In this architecture, stereoscopic cues and both eyes' perspective cues were represented by two independent populations. The response of the perspective population to binocular stimulation was the divisively normalized sum of the squared monocular responses. Optimal cue integration was performed by a third population that linearly summed the stereoscopic cue and single perspective cue population responses to create a combined-cue representation:

$$\mathbf{r}_C = \overbrace{\mathbf{r}_S}^{\text{stereo. pop.}} + \frac{\overbrace{\mathbf{r}_{P_L}^2 + \mathbf{r}_{P_R}^2}^{\text{single persp. pop.}}}{\mathbf{r}_{P_L} + \mathbf{r}_{P_R}}. \tag{9}$$

*One population*

In this architecture, stereoscopic cues and both eyes' perspective cues were represented by a single popula-



**Figure 3.** Tilt perception for combined-cue stimuli. **A**, Probability density functions describing the errors in reported tilts made by Monkey L for each slant–distance combination, calculated using all eight tilts. Columns correspond to distance, and colors to slant. The probability that an error of a given  $\Delta\text{Tilt}$  was made is shown with a point. Correct choices:  $\Delta\text{Tilt} = 0^\circ$ . Solid curves are fitted von Mises probability density functions. At high precisions, there is some deviation between the point representing the probability that the monkey was correct and the probability density function. This deviation reflects discrete versus continuous representations of the area between sampled tilts, and that the sampling interval limits the maximum  $\kappa$  that can be estimated. An upper bound of  $\kappa = 18$  was set based on simulations (see Materials and Methods). **B**, Heat maps showing the precision (von Mises  $\kappa$ ) of tilt perception as a function of slant and distance for both monkeys, calculated using all eight tilts. Red hues indicate lower precision and yellow hues indicate higher precision. Right marginals show  $\kappa$  as a function of slant for each distance. Precision increased monotonically with slant. Upper marginals show  $\kappa$  as a function of distance for each slant. Precision had an inverted U shape as a function of distance. Also see Extended Data Figure 3-1.

tion. The response to binocular stimulation was the diversively normalized sum of the squared cue-isolated responses:

$$r_C = \frac{\overbrace{r_S^2 + r_{P_L}^2 + r_{P_R}^2}^{\text{one pop.}}}{r_S + r_{P_L} + r_{P_R}}. \tag{10}$$

**Code accessibility**

The simulations were performed in MATLAB R2018a on a Windows 10 workstation. The code is available on GitHub: <https://github.com/RosenbergLab/Optimized-but-not-maximized-cue-integration-for-3D-visual-perception>.

**Results**

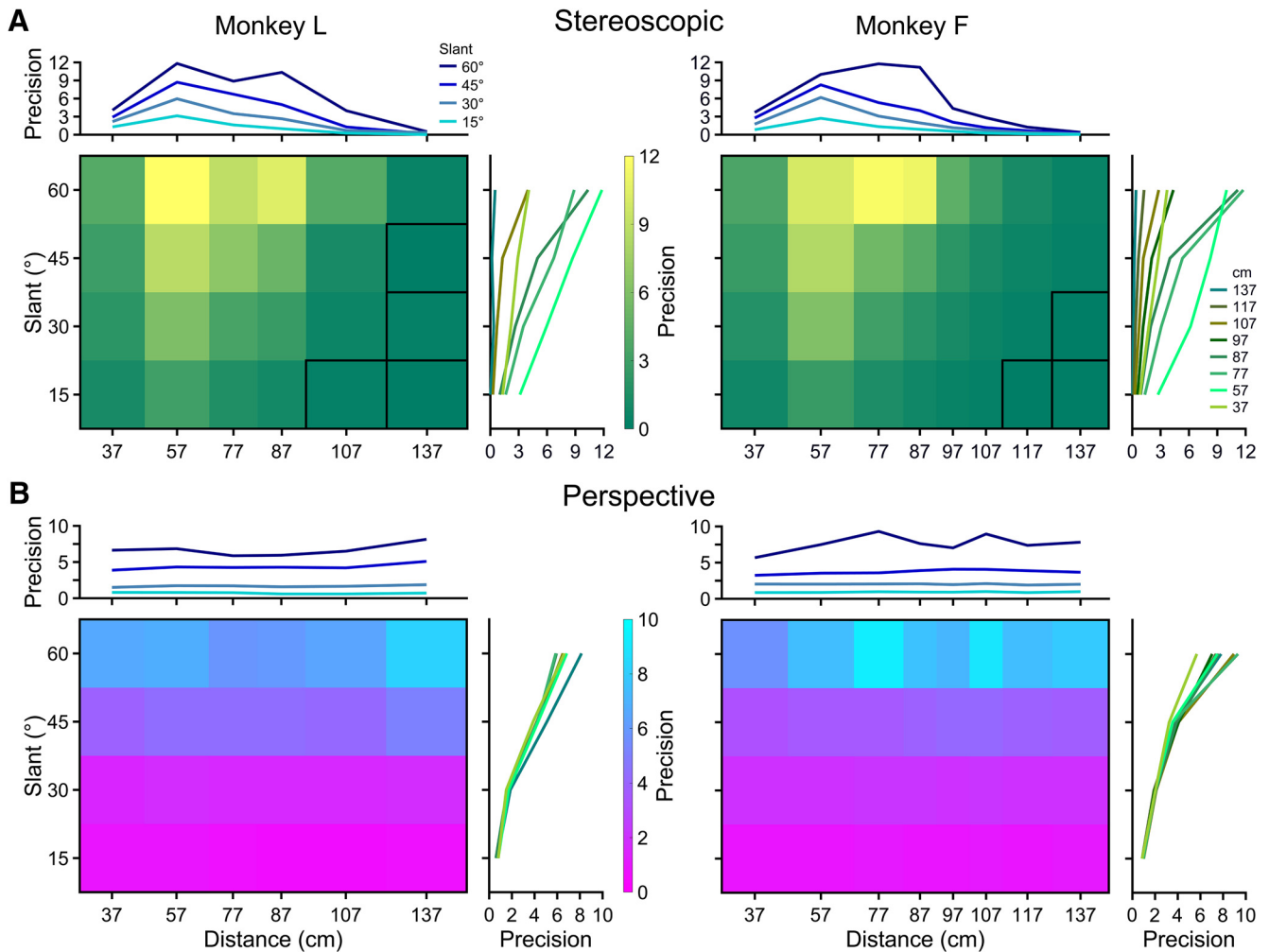
**The precision of tilt perception depends on slant and distance**

Tilt and slant are two angular variables describing 3D planar surface orientation (Stevens, 1983b; Rosenberg et al., 2013). Tilt specifies the direction that the plane is oriented in depth (e.g., right-near = 0°, top-near = 90°),

and slant specifies how much it is oriented in depth (Fig. 2A). For planar surfaces with different slants, distances, and defining cues (Fig. 2B–E), we trained two monkeys to perform an 8AFC tilt discrimination task (Fig. 2F).

We first assessed how the precision of tilt perception depended on slant and distance for combined-cue stimuli defined by stereoscopic and perspective cues. Distributions describing the errors in reported tilts ( $\Delta\text{Tilt} = \text{reported tilt} - \text{presented tilt}$ ) were calculated using all 8 presented tilts. Tilt error distributions for each of the 24 slant–distance combinations tested with Monkey L are shown in Figure 3A. To quantify performance, we fit von Mises probability density functions (solid curves). A fitted density function’s concentration parameter ( $\kappa$ , the circular analog of  $\sigma^{-2}$  for a Gaussian) indicates the precision of perception. Larger  $\kappa$  (taller and narrower densities) indicate greater precision. The joint dependency of precision on slant and distance is shown for both monkeys using heat maps in Figure 3B.

Combined-cue tilt perception depended on slant and distance in distinct ways. At each distance (Fig. 3A, col-



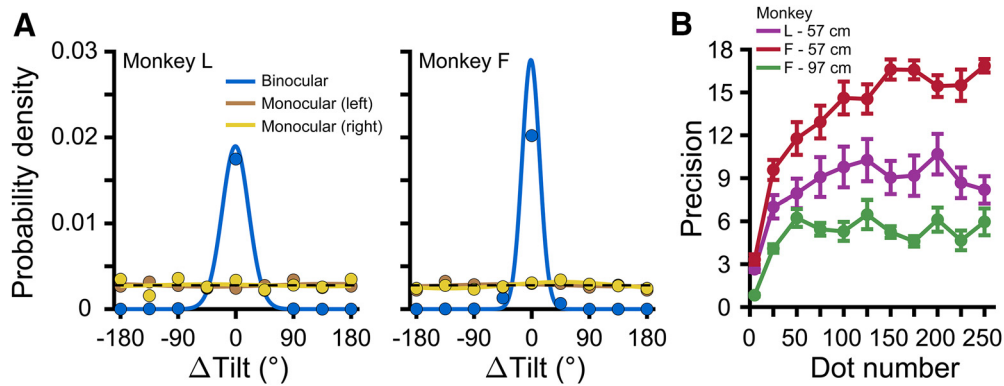
**Figure 4.** Precision of tilt perception for cue-isolated stimuli. **A**, Stereoscopic cues. Precision ( $\kappa$ ) increased monotonically with slant and had an inverted U shape as a function of distance. Performance was at chance level for combinations of small slants and large distances (outlined in black). **B**, Perspective cues. Precision increased monotonically with slant and was independent of distance. Plots follow the format in Figure 3B. Also see Extended Data Figure 3-1.

umns), the probability density functions grew taller and narrower with increasing slant. This pattern indicates a monotonic relationship between precision and slant, which is visualized in plots of  $\kappa$  versus slant for each distance (Fig. 3B, right marginals). The monotonic relationship between precision and slant is consistent with the slant-dependent reliability of perspective cues (Fig. 1B). Precision had an inverted U shape as a function of distance. This pattern can be seen by comparing probability density functions across columns in Figure 3A. For each slant, the density functions were most concentrated at or just behind the plane of fixation (57 cm). The inverted U-shape relationship between precision and distance is visualized in plots of  $\kappa$  versus distance for each slant (Fig. 3B, top marginals). Decreasing precision with distance is consistent with the distance-dependent reliability of stereoscopic cues (Fig. 1A). The falloff in precision with distance from the fixation plane (both toward and away from the monkey) is consistent with the limited range of horizontal disparities represented by the visual system (Prince et al., 2002). For example, this explains why

greater falloffs in precision occurred between 57 and 37 cm than between 57 and 77 cm, since the magnitude of the disparity pedestal at 37 cm ( $-1.63^\circ$ ) was bigger than at 77 cm ( $0.78^\circ$ ). Asymmetric falloffs were especially evident in the stereoscopic cue data, described in the next section (Fig. 4A).

The interaction of the monotonic relationship between precision and slant and the inverted U-shape relationship between precision and distance resulted in a more gradual falloff in precision with distance at larger slants, giving the heat maps a wedge-shaped appearance (Fig. 3B). From these plots, it is apparent that precision peaked at the largest tested slant ( $60^\circ$ ), and either at or just behind the plane of fixation. The mean precision across all slant-distance combinations was  $\kappa = 6.19 \pm 4.81$  SD for Monkey L ( $N = 24$ ), and  $7.03 \pm 5.92$  for Monkey F ( $N = 32$ ). Although the precision of tilt perception varied substantially as a function of slant and distance, performance was always above chance (Rayleigh test for circular uniformity, all  $p \leq 2.3 \times 10^{-7}$ , Bonferroni corrected for 24 or 32 comparisons for Monkey L or F, respectively). A similar





**Figure 5.** Stereoscopic cue controls. **A**, Stereoscopic cue stimuli were viewed binocularly (blue curves) or monocularly (left eye stimulated: orange; right eye stimulated: yellow). Probability density functions describing the errors in reported tilts made by each monkey, calculated using all eight tilts are plotted. The probability that an error of a given  $\Delta\text{Tilt}$  was made is shown with a point. Correct choices:  $\Delta\text{Tilt} = 0^{\circ}$ . Solid curves are fitted von Mises probability density functions. Chance performance is indicated by dashed black lines. **B**, Precision ( $\kappa$ ) versus dot number for planes at 57 cm (Monkey L: purple; Monkey F: red) and 97 cm (Monkey F: green). Error bars are SEM across sessions.

pattern was found with larger stimuli over a wider range of distances (Extended Data Fig. 3-1A). These results suggest that the precision of tilt perception depended on a combination of physiological factors and the pose-dependent reliabilities of the visual cues.

### Contributions of stereoscopic cues to tilt perception

We next assessed the contributions of stereoscopic cues to tilt perception by analyzing responses to planar surfaces which had no perspective cues that could be used to perform the task (Fig. 2D). Error distributions of reported tilts were calculated using all eight tilts, and fit with von Mises densities to quantify performance. The precision of tilt perception based on stereoscopic cues is plotted as a function of slant and distance in Figure 4A.

The mean precision across all slant–distance combinations was  $\kappa = 3.63 \pm 3.45$  SD for Monkey L ( $N = 24$ ), and  $2.87 \pm 3.29$  for Monkey F ( $N = 32$ ). We found a monotonic relationship between precision and slant, indicating that the monkeys were able to more reliably estimate the direction of the gradient of horizontal disparities at larger slants. There was also an inverted U-shape relationship between precision and distance which, as discussed above, was consistent with a combination of geometric and physiological factors.

The stereoscopic cue and combined-cue precision landscapes were similar in shape (Monkey L:  $r = 0.89$ ,  $p = 4.4 \times 10^{-9}$ ; Monkey F:  $r = 0.82$ ,  $p = 1.1 \times 10^{-8}$ ). However, precision was significantly lower with stereoscopic cue than combined-cue stimuli (Wilcoxon signed-rank test; Monkey L:  $p = 1.8 \times 10^{-5}$ ,  $N = 24$ ; Monkey F:  $p = 8.0 \times 10^{-7}$ ,  $N = 32$ ). At combinations of small slants and large distances, performance with the stereoscopic cue stimuli was not significantly different from chance (Fig. 4A, conditions outlined in black; Rayleigh test and Bonferroni corrected). At greater distances, performance was at chance level even with larger stimuli (Extended Data Fig. 3-1B). These findings indicate that stereoscopic cues did not contribute to tilt perception beyond  $\sim 137$  cm, and imply that perspective cues were entirely respon-

sible for above chance performance with combined-cue stimuli at some poses.

We performed a control experiment to verify that the stereoscopic cue stimuli had no perspective cues that could be used to perform the task (see Materials and Methods). Stereoscopic cue stimuli were presented binocularly (both eyes saw the plane) as well as monocularly (one eye saw the plane). Above chance performance with monocular viewing would indicate that there were usable perspective cues. As expected, the monkeys performed well with binocular viewing (Rayleigh test; Monkey L:  $p = 2.2 \times 10^{-308}$ ; Monkey F:  $p = 2.2 \times 10^{-308}$ ). In contrast, monocular performance was not significantly different from chance (Monkey L: left eye  $p = 0.79$ , right eye  $p = 0.91$ ; Monkey F: left eye  $p = 0.77$ , right eye  $p = 0.11$ ). Thus, the stereoscopic cue stimuli contained no usable perspective information for performing the task (Fig. 5A).

A second control examined whether performance with the stereoscopic cue stimuli was affected by a potential conflict between the stereoscopically defined planar slant (which was always non-zero) and the isotropic shape of the dots which could be interpreted as signaling zero slant. This conflict would increase with greater dot number, causing a decrease in precision (Hillis et al., 2004). We therefore assessed how the precision of stereoscopic cue perception depended on dot number. As expected, precision depended significantly on dot number (Kruskal–Wallis test; Monkey L:  $p = 2.7 \times 10^{-5}$ ; Monkey F:  $p = 9.3 \times 10^{-13}$  at 57 cm; Monkey F:  $p = 1.7 \times 10^{-9}$  at 97 cm). Precision initially increased, reflecting that more dots provide greater signal for performing the task. To test for decreases in precision, we ran pairwise comparisons using Tukey’s HSD test. In each case of a significant difference, the precision at the larger dot number was greater than at the smaller number. There were no significant differences between precisions at dot numbers  $\geq 75$ . These results suggest that precision increased monotonically with dot number, and thus that performance was not affected by this potential conflict (Fig. 5B).

### Contributions of perspective cues to tilt perception

We next assessed the contributions of perspective cues to tilt perception by analyzing responses to planes lacking stereoscopic cues (Fig. 2E). Stereoscopic cues were eliminated by presenting combined-cue stimuli to one eye only. Left and right eye projections of 3D stimuli are not copies of each other, and sensitivity can differ between the eyes, at least for perspective-defined 3D motion (Thompson et al., 2019). We therefore tested whether the precision of tilt perception depended on which eye saw the plane. Error distributions of reported tilts were calculated separately for each eye using all eight tilts, and fit with von Mises densities to compare performance. Precision did not significantly depend on which eye was stimulated (Wilcoxon signed-rank test; Monkey L:  $p = 0.80$ ; Monkey F:  $p = 0.08$ ), so responses to left and right eye presentations were pooled and fit with von Mises densities to quantify performance. The precision of tilt perception based on perspective cues is plotted as a function of slant and distance in Figure 4B.

The mean precision across all slant–distance combinations was  $\kappa = 3.34 \pm 2.42$  SD for Monkey L ( $N = 24$ ), and  $3.59 \pm 2.66$  for Monkey F ( $N = 32$ ). Performance was significantly above chance at all poses (Rayleigh test, all  $p \leq 1.5 \times 10^{-12}$ , Bonferroni corrected). Consistent with the slant-dependent reliability of perspective cues (Fig. 1B), precision increased monotonically with slant. Reflecting that the perspective cues only signaled 3D orientation (see Materials and Methods), precision was independent of distance. A similar pattern was found with larger stimuli over a wider range of distances (Extended Data Fig. 3-1C).

Precision was significantly lower with perspective cue than combined-cue stimuli (Wilcoxon signed-rank test; Monkey L:  $p = 4.9 \times 10^{-4}$ ; Monkey F:  $p = 2.9 \times 10^{-6}$ ). The perspective cue and stereoscopic cue precisions were not significantly different when compared across all slants and distances (Wilcoxon signed-rank test; Monkey L:  $p = 0.53$ ; Monkey F:  $p = 0.13$ ). However, the relative precisions for the two cues were distance dependent. For distances at or just behind the plane of fixation (57, 77, and 87 cm), precision was significantly higher with stereoscopic cues (Wilcoxon signed-rank test; Monkey L:  $p = 4.9 \times 10^{-4}$ ,  $N = 12$ ; Monkey F:  $p = 3.4 \times 10^{-3}$ ,  $N = 12$ ). For nearer (37 cm) and further distances (>87 cm), precision was significantly higher with perspective cues (Wilcoxon signed-rank test; Monkey L:  $p = 4.9 \times 10^{-3}$ ,  $N = 12$ ; Monkey F:  $p = 8.9 \times 10^{-5}$ ,  $N = 20$ ). These findings demonstrate that both stereoscopic cues and perspective cues contribute to 3D perception within peripersonal space, and that perspective cues extend perception beyond the range supported by stereopsis.

### A 3D analog of the oblique effect and priors over surface tilt

Humans perceive 2D tilt more accurately and precisely at cardinal than oblique tilts (the “oblique effect”; Campbell et al., 1966; Westheimer and Beard, 1998). Similar findings were recently reported for human 3D tilt perception (Kim and Burge, 2018). To test whether the monkeys

showed a 3D analog of the oblique effect, we fit the error distribution of reported tilts for each combination of tilt, slant, distance, and cue condition with a von Mises density function. We then compared the accuracy and precision of perception across all slant–distance combinations as a function of tilt. Analyses were performed separately for each cue condition.

#### Accuracy results

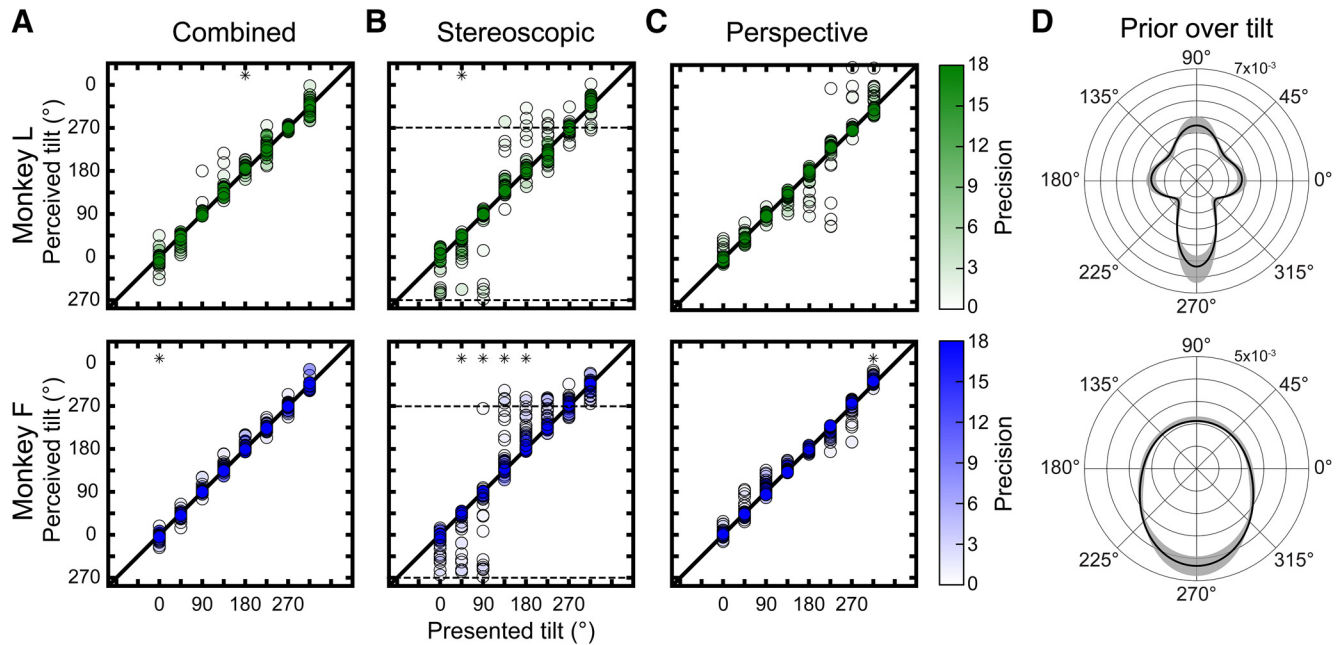
For combined-cue stimuli, the biases at any given tilt were generally small and not significant (Fig. 6A). Across the two monkeys, biases that were much smaller than the tilt sampling interval but statistically significant were found at two tilts (one-sample circular median test,  $p < 0.05$  and Bonferroni corrected for eight tilts). Monkey L had a significant bias at  $180^\circ$  (median  $\mu = 5.31^\circ$ ,  $N = 24$  slant–distance combinations). Monkey F had a significant bias at  $0^\circ$  (median  $\mu = -3.73^\circ$ ,  $N = 32$ ). Both biases were in the direction of  $270^\circ$  (bottom-near). We also compared biases between cardinal and oblique tilts. For Monkey L, the median absolute bias at cardinal tilts was  $3.81^\circ$  ( $N = 24$  slant–distances  $\times$  4 tilts = 96), and at oblique tilts was  $8.79^\circ$  ( $N = 96$ ). Consistent with an oblique effect, the cardinal biases were significantly smaller than the oblique biases (two-sample circular median test,  $p = 1.5 \times 10^{-3}$ ). For Monkey F, the median absolute biases at cardinal ( $3.30^\circ$ ,  $N = 128$ ) and oblique ( $2.89^\circ$ ,  $N = 128$ ) tilts were not significantly different ( $p = 0.62$ ).

For stereoscopic-cue stimuli, the biases at any given tilt were typically small and not significant, but some large biases were evident (Fig. 6B). Across the two monkeys, significant biases were found at five tilts. As indicated by the opacity of the data points, large biases were restricted to low precisions (which occurred at small slants and large distances; Fig. 4A). Thus, perception was accurate so long as sufficient information was available to perform the task well. Between all cardinal and all oblique tilts, Monkey L’s biases were consistent with an oblique effect: the median absolute bias at cardinal tilts ( $7.81^\circ$ ) was significantly smaller (two-sample circular median test,  $p = 1.5 \times 10^{-3}$ ) than at oblique tilts ( $15.58^\circ$ ). For Monkey F, the median absolute biases at cardinal ( $9.36^\circ$ ) and oblique ( $8.31^\circ$ ) tilts were not significantly different ( $p = 0.62$ ). However, when precision was low, the responses of both monkeys were biased toward  $270^\circ$ .

For perspective cue stimuli, the biases at any given tilt were generally small and not significant (Fig. 6C). The only significant bias was for Monkey F at  $315^\circ$  (median  $\mu = 9.28^\circ$ ). Between all cardinal and all oblique tilts, Monkey L’s biases were consistent with an oblique effect: the median absolute bias at cardinal tilts ( $5.5^\circ$ ) was significantly smaller (two-sample circular median test,  $p = 1.8 \times 10^{-4}$ ) than at oblique tilts ( $8.79^\circ$ ). For Monkey F, the median absolute biases at cardinal ( $5.67^\circ$ ) and oblique ( $6.67^\circ$ ) tilts were not significantly different ( $p = 0.45$ ).

#### Precision results

For the combined-cue stimuli, precision was not significantly dependent on tilt for either monkey (Kruskal–Wallis test; Monkey L:  $p = 0.09$ ; Monkey F:  $p = 0.89$ ). After grouping precisions into cardinal and oblique tilts, Mon-



**Figure 6.** Biases in tilt perception occurred at low precisions and were consistent with a prior over 3D tilt. **A–C**, Perceived tilt (presented tilt + mean of the von Mises fit to the error distribution) versus presented tilt (Monkey L:  $N = 24$  slant–distance combinations per tilt; Monkey F:  $N = 32$  per tilt). Diagonals are identity lines. Greater vertical distance from the identity line indicates greater bias. The fill opacity indicates precision ( $\kappa$ ). Asterisks mark biases that were significantly different from 0°. **A**, Combined-cue stimuli. **B**, Stereoscopic cue stimuli. At low precisions, perception was pulled toward 270° (bottom–near), marked by horizontal dashed lines. **C**, Perspective cue stimuli. **D**, Priors over 3D tilt. The angular variable is surface tilt and the radial variable is the probability density value. Shading indicates the bootstrapped 95% confidence interval.

key L showed an oblique effect: the median precision at cardinal tilts (7.38,  $N = 96$ ) was significantly larger (Mann–Whitney  $U$  test,  $p = 0.01$ ) than at oblique tilts (5.28). For Monkey F, the median precisions at cardinal (5.12,  $N = 128$ ) and oblique (5.07,  $N = 128$ ) tilts were not significantly different ( $p = 0.79$ ).

For the stereoscopic cue stimuli, precision was not significantly dependent on tilt for either monkey (Kruskal–Wallis test; Monkey L:  $p = 0.70$ ; Monkey F:  $p = 0.21$ ). Precisions were not significantly different between cardinal (Monkey L: median  $\kappa = 3.34$ ; Monkey F: median  $\kappa = 1.82$ ) and oblique (Monkey L: median  $\kappa = 3.04$ ; Monkey F: median  $\kappa = 1.77$ ) tilts for either monkey (Mann–Whitney  $U$  test; Monkey L:  $p = 0.36$ ; Monkey F:  $p = 0.98$ ).

For the perspective cue stimuli, we found some tilt-dependent differences. For Monkey L, precision was significantly greater at 90° (median  $\kappa = 5.70$ ) than 225° (median  $\kappa = 1.64$ ; Kruskal–Wallis test followed by Tukey’s HSD test,  $p = 1.9 \times 10^{-3}$ ) or 315° (median  $\kappa = 2.53$ ;  $p = 4.7 \times 10^{-2}$ ). For Monkey F, precision was significantly greater at 135° (median  $\kappa = 4.69$ ;  $p = 4.7 \times 10^{-2}$ ) than 45° (median  $\kappa = 2.07$ ). Likewise, precisions were significantly greater at 135° ( $p = 8.5 \times 10^{-4}$ ) and 225° (median  $\kappa = 3.59$ ;  $p = 0.02$ ) than at 270° (median  $\kappa = 1.83$ ). After grouping precisions into cardinal and oblique tilts, Monkey L showed an oblique effect: the median precision at cardinal tilts (4.28) was significantly larger (Mann–Whitney  $U$  test,  $p = 0.02$ ) than at oblique tilts (3.22). For Monkey F, the median precisions at cardinal (2.74) and oblique (3.08) tilts were not significantly different ( $p = 0.11$ ).

These analyses revealed a 3D analog of the oblique effect in Monkey L, but not Monkey F. Individual differences in the 2D oblique effect are also observed in humans (Westheimer and Beard, 1998; Girshick et al., 2011). However, even in Monkey L, the effects were not prominent since: (1) biases were typically small, (2) precision results varied across cue conditions, and (3) differences between cardinal and oblique tilts required pooling to detect. Nonetheless, both monkeys’ biases appeared consistent with the influence of a prior resembling the 3D tilt statistics of natural scenes (Adams et al., 2016; Burge et al., 2016).

#### Priors over surface tilt

Consistent with the effect of a prior, biases were most apparent at low precisions (Fig. 6A–C) and absolute bias was negatively correlated with precision: Spearman  $r = -0.64$ ,  $p = 1.9 \times 10^{-157}$  ( $N = 1344$  tilt  $\times$  slant  $\times$  distance  $\times$  cue condition, both monkeys). At the lowest precisions (stereoscopic cue stimuli with small slants and large distances, both monkeys were biased toward 270°. Monkey L, but not Monkey F, also had larger biases at oblique than cardinal tilts. This monkey-specific difference may reflect different internalizations of a prior over tilt. To estimate the prior that best accounted for each monkey’s biases, the perceptual data were modeled using the product of sensory likelihood functions and a prior over tilt (see Materials and Methods). The model described the data well. The circular correlation between observed and predicted perceived tilts was  $r = 0.92$ ,  $p = 2.2 \times 10^{-308}$  ( $N = 1344$ ). Likewise, the Pearson correlation between ob-



served and predicted precisions was  $r = 0.94$ ,  $p = 2.2 \times 10^{-308}$ .

The estimated priors captured the monkey-specific biases. Monkey L's prior had left/right-near lobes that were smaller than the top-near lobe, and a global peak at the bottom-near lobe (Fig. 6D, top; fitted prior  $\kappa$ 's:  $\kappa_{0^\circ \& 180^\circ} = 2.75$ ,  $\kappa_{90^\circ} = 3.5$ , and  $\kappa_{270^\circ} = 8.5$ ). Net pull due to the lobes' asymmetries accounts for larger oblique than cardinal biases. Monkey F did not show an oblique effect, and correspondingly the prior did not have four distinct cardinal lobes (Fig. 6D, bottom;  $\kappa_{0^\circ \& 180^\circ} = 0$ ,  $\kappa_{90^\circ} = 0$ , and  $\kappa_{270^\circ} = 2$ ). The global peaks at  $270^\circ$  account for the bottom-near biases observed for both monkeys at the lowest precisions. Individual differences in biases were thus consistent with different internalizations of the natural scene statistics of surface tilt, which include more cardinal than oblique tilts and a preponderance of ground planes (Adams et al., 2016; Burge et al., 2016).

### Perceptual cue integration

Perception was more precise with combined-cue than cue-isolated stimuli, and the relative cue-isolated precisions were pose dependent. Given these results, we tested whether the performance with combined-cue stimuli was consistent with an optimal integration strategy. Optimal predictions were derived using the cue-isolated data (see Materials and Methods). We then compared the predictions to the observed combined-cue data to assess whether stereoscopic cues and perspective cues (pooled over the left and right eyes) were dynamically reweighted on a trial-by-trial basis according to their pose-dependent reliabilities.

Example error distributions calculated using all eight tilts are shown for the cue-isolated and combined-cue data (from Figs. 3, 4) along with optimal predictions (derived using Eqs. 2, 3) in Figure 7A–E. The observed combined-cue performance (blue curves) was well described by the model (dotted black curves) over a wide range of precisions. The predictions were strong even if only one cue contributed to perception (Fig. 7E). Across all slant–distance combinations ( $N = 56$ , both monkeys), the mean stereoscopic/perspective  $\kappa$  ratio was 1.05, which is ideal for evaluating cue integration (Fig. 7F). At the same time, the spread of relative cue-isolated precisions allowed us to assess cue integration over a range of naturally occurring cue reliability conditions. The observed and optimal combined-cue precisions were well correlated ( $r = 0.93$ ,  $p = 7.9 \times 10^{-25}$ ) and distributed along the identity line (Fig. 7G), indicating that performance was consistent with an optimal integration strategy. Similar results were found with larger stimuli over a wider range of distances (Extended Data Fig. 3-1D).

We performed a second test of cue integration using the likelihood functions and priors estimated in the previous section. For each combination of tilt, slant, and distance (data in Fig. 6;  $N = 448$ , both monkeys), an optimal prediction of the combined-cue performance was made using the product of the cue-isolated likelihood functions and prior over tilt (Eqs. 4, 5). The observed and optimal perceived tilts were well correlated (circular  $r = 0.98$ ,  $p =$

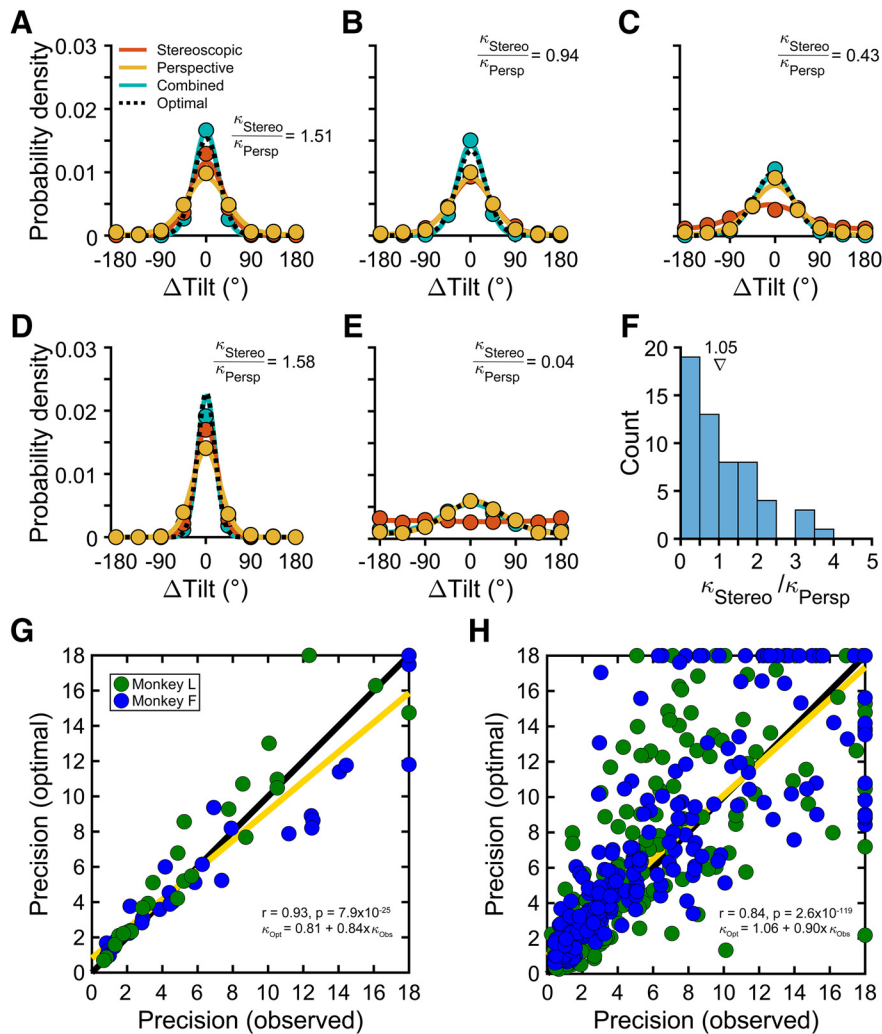
$2.2 \times 10^{-308}$ ), as were the  $\kappa$ 's (Pearson  $r = 0.84$ ,  $p = 2.6 \times 10^{-119}$ ; Fig. 7H). The greater scatter between observed and optimal  $\kappa$ 's when cue integration was assessed at each tilt (Fig. 7H) compared to assessments based on all eight tilts (Fig. 7G) was not surprising since the amount of data in the two analyses differed by  $\sim 8$ -fold. Furthermore, only the second test of cue integration included a prior over tilt. Finding that both optimal precision estimates matched the data well is another indication that the prior had an overall small effect on perception. Together, these findings are consistent with robust 3D perception being achieved through the optimal integration of stereoscopic cues and perspective cues (pooled over the left and right eyes). Since the stimuli were interleaved pseudo-randomly, cue reweighting must have occurred dynamically to match the vagaries of cue reliabilities that occurred with the trial-to-trial changes in surface pose.

### Neuronal models of 3D visual cue integration

We found that combined-cue tilt perception was consistent with the optimal integration of stereoscopic cues and perspective cues (pooled over the left and right eyes). Similar conclusions can be drawn from human slant perception studies (Knill and Saunders, 2003; Hillis et al., 2004). These findings are somewhat surprising since there are actually three cues available for making 3D inferences: stereoscopic cues, left eye perspective cues, and right eye perspective cues. If all three cues were represented independently and optimally integrated, perception would be more precise than observed (Oruç et al., 2003). This observation motivated us to test how the neural architecture and computations responsible for cue integration might constrain the precision of combined-cue tilt perception. By comparing the perceptual data to simulation results, we generate predictions about where and how neural computation might curb 3D perception.

Performance on the task was simulated using Bayesian decoding of modeled neuronal population responses. Responses to stereoscopic cues, left eye perspective cues, and right eye perspective cues were simulated based on CIP area neurons (see Materials and Methods). Neurons in CIP are selective for 3D surface orientation (Taira et al., 2000; Rosenberg et al., 2013), and respond to both stereoscopic and perspective cues (Tsutsui et al., 2001; Rosenberg and Angelaki, 2014b). Furthermore, their inactivation impairs 3D perception (Tsutsui et al., 2001; Van Dromme et al., 2016), and their activity correlates with surface orientation reports (Elmore et al., 2019). We tested three architectures for combining cue-isolated responses to create a combined-cue representation (Fig. 8A). To focus on computations underlying cue integration, and since the effects of a prior were relatively small, we simplified the model by not incorporating a prior over tilt. Thus, perceptual biases were not modeled. For each architecture, we decoded the combined-cue representation to simulate perceptual data (Fig. 8B). Decoded model precisions were compared to the monkey precisions (estimated using all eight tilts; Fig. 3 data) for each slant–distance combination (Fig. 8C).



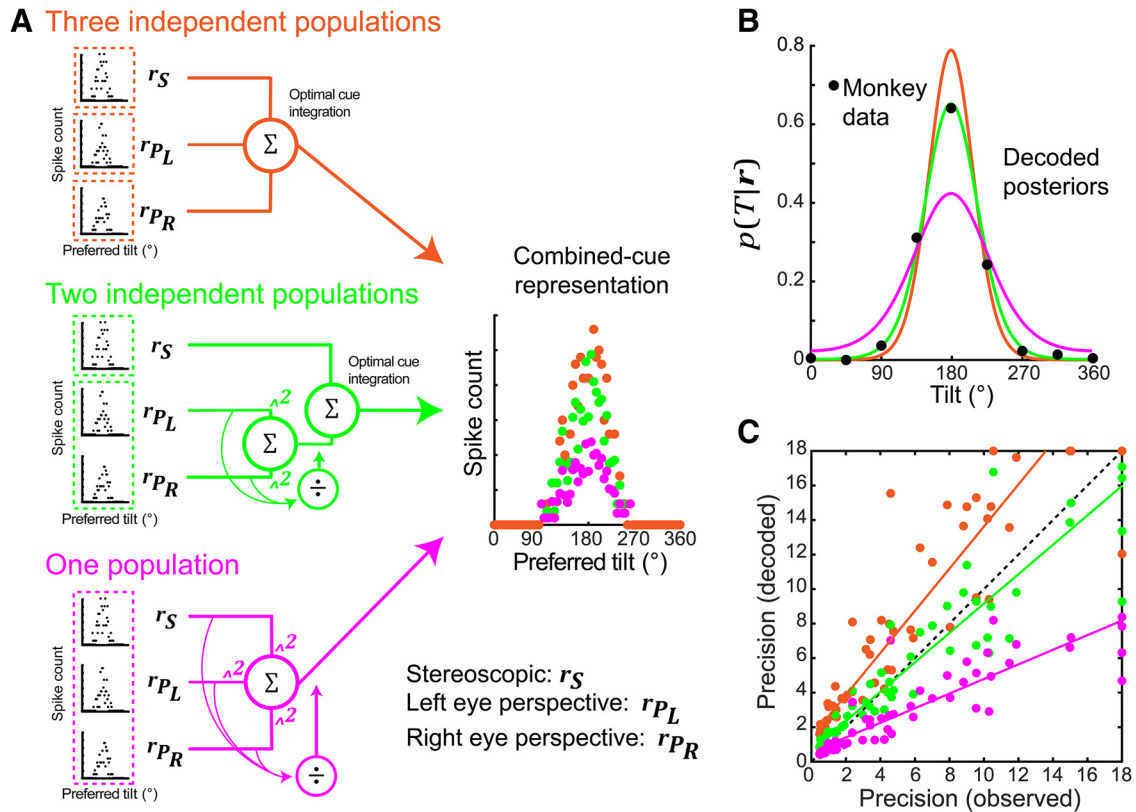


**Figure 7.** Perceptual cue integration. **A–E**, Example densities for each cue condition, calculated using all eight tilts. Solid curves are von Mises fits. Dotted black curves are optimal predictions. Insets show cue-isolated  $\kappa$  ratios. **A**, Slant = 30°, distance = 77 cm (Monkey F). **B**, Slant = 30°, distance = 87 cm (Monkey F). **C**, Slant = 30°, distance = 107 cm (Monkey L). **D**, Slant = 45°, distance = 77 cm (Monkey L). **E**, Slant = 15°, distance = 137 cm (Monkey F). Combined-cue perception depended entirely on perspective cues. **F**, Distribution of cue-isolated  $\kappa$  ratios ( $N = 56$  slant–distance combinations, both monkeys). The triangle marks the mean ratio. **G**, Optimal versus observed combined-cue precision calculated using all eight tilts for each slant–distance combination ( $N = 56$ ). **H**, Optimal versus observed combined-cue precision calculated for each tilt  $\times$  slant  $\times$  distance combination ( $N = 448$ ). Type-II regression lines are shown in yellow ( $\kappa = 18$  excluded). Insets show correlations and regression line equations. Also see Extended Data Figure 3-1.

The first architecture assumed that each of the three cues was represented by an independent neuronal population. A fourth population linearly summed the three cue-isolated population responses to perform optimal cue integration (Ma et al., 2006), thereby creating a combined-cue representation (Fig. 8A, top). This architecture sets an upper bound on combined-cue precision given the cue-isolated precisions. As anticipated, since the perceptual data were consistent with the integration of stereoscopic cues and pooled left and right eye perspective cues, this architecture produced significantly greater precisions than shown by the monkeys: Wilcoxon signed-rank test,  $p = 5.2 \times 10^{-9}$  (Fig. 8C, orange points). Importantly, this simulation showed that if the neural network responsible for tilt perception represented the three cues indepen-

dently and integrated them optimally, then the precision of combined-cue perception would be 1.97 times greater than observed (average decoded/observed precision ratio). Although tilt perception was less precise than theoretically possible, this result does not imply that perception was suboptimal since the monkeys' performance could reflect the optimal integration of non-independent neuronal representations (Oruç et al., 2003).

The second architecture explored this possibility. Given the perceptual data, stereoscopic cues and both eyes' perspective cues were assumed to be represented by two independent populations and optimally integrated. Thus, the critical factor was how the perspective cue population combined the response components contributed from each eye, since this determines the level of dependency



**Figure 8.** Optimized but not maximized cue integration. **A**, Schematics of three architectures for combining responses to stereoscopic cues ( $r_S$ ), left eye perspective cues ( $r_{PL}$ ), and right eye perspective cues ( $r_{PR}$ ). Top, Three independent populations represent each cue (orange). Middle, Two independent populations represent stereoscopic cues and perspective cues from both eyes (green). Bottom, One population represents all cues (magenta). Right, Combined-cue representations for each architecture. Points show the response of each model neuron, ordered along the x-axis by preferred tilt, to a single stimulus presentation (slant = 30°, distance = 37 cm). **B**, Tilt posteriors,  $p(T|r)$ , decoded from the combined-cue representations. Black dots show corresponding data from Monkey L. Given the same cue-isolated responses, precision was greatest for the three independent populations model and lowest for the one population model. The posterior of the two independent populations model matched the monkey’s data. **C**, Comparisons of decoded model precisions and observed monkey precisions. Each point corresponds to one slant–distance combination ( $N = 56$ , both monkeys). The three independent populations model was more precise than the monkeys (nearly all points are above the dashed black identity line). The precisions from the two independent populations model matched the monkeys’ precisions (points are distributed along the identity line). The one population model was less precise than the monkeys (nearly all points are below the identity line). Solid lines are Type-II regressions ( $\kappa = 18$  excluded).

between them. We found that it was possible to account for the perceptual data if the perspective cue population combined the left and right eye response components using quadratic nonlinearities and divisive normalization (Fig. 8A, middle). A similar model describes V1 responses to compound stimuli, and the combination operations are widely implicated in neural processing (Busse et al., 2009; Beck et al., 2011; Carandini and Heeger, 2012; Ni et al., 2012; Louie et al., 2013; Rosenberg et al., 2015). Importantly, these computations made the left and right eye perspective cue representations fully dependent, precluding any improvement in perceptual precision that could have resulted from having two perspective cues. Nevertheless, the stereoscopic cue and perspective cue representations were optimally integrated since the two cue-isolated population responses were linearly summed to create a combined-cue representation (Ma et al., 2006). This architecture produced precisions that were not significantly different from the monkeys’ precisions: Wil-

coxon signed-rank test,  $p = 0.78$  (Fig. 8C, green points). This result suggests that the combination of left and right eye perspective cues using nonlinear canonical computations may be a key factor limiting the precision of tilt perception.

We lastly considered a single neuronal population that estimated tilt from stereoscopic cues as well as both eyes’ perspective cues. For consistency with the two independent populations model, response components driven by each of the three cues were combined with quadratic nonlinearities and divisively normalized (Fig. 8A, bottom). Thus, none of the cues were represented independently and no integration of cue-isolated population responses was required to create a combined-cue representation. As anticipated based on the perceptual data and the results from the two independent populations model, this architecture produced significantly lower precisions than shown by the monkeys: Wilcoxon signed-rank test,  $p = 2.5 \times 10^{-9}$  (Fig. 8C, magenta points).

These simulations revealed a neural architecture and computations that could account for the 3D perceptual cue integration results. In particular, they pinpointed the processing of left and right eye perspective cues by a single neuronal population as a factor that could substantially limit the precision of tilt perception. They further suggest that tilt perception is consistent with an optimal integration strategy but not maximized due to the concurrence of several canonical computations.

## Discussion

We evaluated the contributions of stereoscopic cues and perspective cues to tilt perception in macaque monkeys. Tilt discrimination was used to study how 3D perception depends on surface pose (position and orientation) since stereopsis and projective geometry dictate that the reliability of these cues depends on distance and slant (Fig. 1). To date, few studies have investigated tilt perception in humans (Stevens, 1983a; Kim and Burge, 2018) or monkeys (Tsutsui et al., 2001, 2003). We found that tilt perception was accurate across a wide range of poses. Instances of poor accuracy were associated with low precision, and presumably reflected low certainty. Precision was highly pose dependent. With stereoscopic cues, precision increased with slant and decreased with distance from the fixation plane. With perspective cues, precision increased with slant. Perspective cues were the sole contributor to tilt perception at small slant and large distance combinations, indicating that they extend 3D perception beyond the range supported by stereopsis. Notably, these results characterize performance on a task that can be used to study the impact of nuisance variables (slant, distance, and visual cues) on neural coding during abstracted sensory (tilt) discriminations.

### Natural scene statistics may influence 3D perception

The 2D oblique effect is characterized by smaller biases and larger precisions at cardinal than oblique tilts (Campbell et al., 1966; Westheimer and Beard, 1998; Girshick et al., 2011). It is thought to be associated with a prior over orientation that reflects natural scene statistics (Coppola et al., 1998). An oblique effect for 3D tilt was recently found in humans (Kim and Burge, 2018). We assessed whether the monkeys showed a 3D oblique effect, while testing for individual differences and cue-specific dependencies. Monkey L had smaller biases at cardinal than oblique tilts in all three cue conditions, and higher precisions at cardinal than oblique tilts in the combined-cue and perspective cue conditions. These tilt dependencies were not found in Monkey F. This suggests that, like the 2D oblique effect (Westheimer and Beard, 1998; Girshick et al., 2011), the 3D oblique effect is subject to individual differences. Given the extensive training with cardinal and oblique tilts, it is unlikely that training induced the oblique effect in Monkey L.

Both monkeys' reports were biased toward 270° (bottom-near) when precision was especially low. Correspondingly, the estimated priors for both monkeys had global peaks at 270°. This finding is consistent with the influence of a prior for ground planes, which are prominent in natural scenes (Adams et al., 2016; Burge et al.,

2016). It is unlikely that a bottom-near bias reflected a preference for making downward saccades since horizontal eye movements are more accurate than vertical eye movements in humans (Kapoula et al., 2010; Ke et al., 2013), and the oculomotor systems of macaques and humans are similar (Fuchs, 1967). The estimated priors also captured the monkey-specific biases. In particular, only Monkey L's prior had four prominent but asymmetric cardinal lobes, accounting for larger oblique than cardinal biases. These findings are consistent with perception being influenced by a prior resembling the 3D tilt statistics of natural scenes, and further suggest that the internalization of the prior differs across individuals.

### Potential conflicts in stereoscopic cue stimuli

For stereoscopic cue stimuli, there is a potential conflict at non-zero slants since isotropic dots are consistent with zero slant. Supporting this possibility, a previous study found that the precision of slant perception based on stereoscopic cues decreased as the stimulus dot number increased beyond 64 for one observer (Hillis et al., 2004). However, we did not find this result. What may account for the difference? One possibility is that the previously observed decrease in precision may have been unrelated to the potential conflict. The same participant showed a similar pattern with stimuli whose perspective cues signaled the stereoscopically defined reference slant, thus eliminating the relevant conflict. Consistent with this possibility, we found that if performance with stereoscopic cue stimuli was at chance level, performance with combined-cue stimuli was no better than with perspective cue stimuli (Fig. 7E). Had a conflict caused the stereoscopic cue precision to be underestimated, the combined-cue precision would have exceeded the perspective cue precision. There may also be individual differences in the perception of such conflicts, perhaps reflecting differences in the strength of a prior for isotropic visual elements (Knill, 1998b). Our results suggest that, at least in monkeys, an isotropy prior is weak or non-existent. Cognitive processes could also be at play if observers reason that circular dots imply zero slant. Future work can evaluate the extent to which such conflicts (when/if they occur) reflect true perceptual conflicts or cognitive reasoning (likely task, training, and instruction dependent).

### Optimal cue integration

We found that the monkeys achieved robust 3D perception by integrating distinct visual cues. Specifically, the precision of tilt perception was consistent with the optimal integration of independently represented stereoscopic cues and perspective cues (left and right eyes pooled). This result is consistent with previous human slant perception findings (Knill and Saunders, 2003; Hillis et al., 2004). Given that fully independent representations are unlikely (e.g., due to common retinal processing), this body of work suggests that the major sources of noise in 3D orientation estimation based on stereoscopic cues and perspective cues are largely independent. Otherwise, the precision of combined-cue perception would have been worse than observed due to correlated errors, even



if the cues were optimally integrated (Oruç et al., 2003). Perception could also have been more precise than observed since there were actually three cues that could contribute to 3D orientation estimates in these studies: stereoscopic cues, left eye perspective cues, and right eye perspective cues. One explanation for the level of precision observed is that two distinct neuronal populations create 3D orientation estimates: one using stereoscopic cues and one using perspective cues from both eyes (with left and right eye signals combined in a way that makes them fully dependent). Our simulations, discussed in the next section, explored how this might occur.

Since we interleaved the stimuli pseudo-randomly, our findings imply that stereoscopic cues and perspective cues are dynamically reweighted to match the vagaries of cue reliabilities that occur with moment-to-moment changes in viewing conditions, such as happens every time the eyes move. Together with previous human studies (Knill and Saunders, 2003; Oruç et al., 2003; Hillis et al., 2004; Welchman et al., 2005; Preston et al., 2009; Murphy et al., 2013; Rideaux and Welchman, 2018), our results suggest that reliability-dependent cue integration is a conserved computation underlying robust 3D vision in primates.

### Canonical computations might optimize but do not maximize 3D perception

Using simulations, we examined how neural network architectures and computations can constrain perception. Comparisons between the perceptual data and the simulation results produced testable predictions regarding where and how perception might be curbed by neuronal processing. The perceptual data suggested an architecture in which two independent populations representing stereoscopic cues and perspective cues (from both eyes) were optimally integrated. Thus, a critical question was how left and right eye perspective signals were combined by the perspective cue population. We found that the perceptual data could be explained if the two perspective signals were combined with quadratic nonlinearities and divisively normalized. These computations made the two representations fully dependent, precluding any improvement in perceptual precision that could have occurred from having both perspective cues. Humans show individual differences in the extent to which visual cues are dependent on each other (Oruç et al., 2003). It would be straightforward to incorporate such differences into the model. For example, down-weighting the normalization would result in partially dependent perspective cues, and thus increased perceptual precision. Since our focus was on computations specific to cue integration, and evidence for the influence of a prior was relatively small, we did not incorporate a prior over tilt into the model. If future work demonstrated sufficient need to incorporate a prior, there are several potential approaches to consider including heterogeneous tuning curves (Ganguli and Simoncelli, 2014) and tuned divisive normalization (Rosenberg et al., 2015). However, it is currently unclear how priors are physiologically implemented. Future work aimed at an-

swering this question would be valuable for understanding neural computation.

Because of divisive normalization in the two independent populations model that accounted for our findings, the perceptual contributions of left and right eye perspective cues can range from averaging (both signals contribute equally) when they are equally reliable to winner-take-all (only the more reliable signal contributes) when they differ substantially (Busse et al., 2009). As such, this architecture can also explain human cue integration findings that show cue averaging with balanced perspective cues in slant perception (Knill and Saunders, 2003; Hillis et al., 2004), as well as winner-take-all behavior with imbalanced perspective cues in motion-in-depth perception (Thompson et al., 2019). Importantly, this model provides testable predictions about where and how neural computation limits tilt perception. One prediction is that if perspective cues were presented to both eyes without stereoscopic cues, then the response would be equal to that elicited from single-eye stimulation. Future neuronal recording experiments can test this by presenting random dot stimuli binocularly with appropriate perspective rendering for each eye but uncorrelated dot positions across the eyes.

Our simulations showed that independent representations of stereoscopic cues, left eye perspective cues, and right eye perspective cues could double the precision of tilt perception. Why then has the visual system not implemented this network and integration strategy? One possibility is the biological inefficiency associated with the number of neurons required to maintain three separate representations, and the duplication of computational units to separately estimate 3D information from left and right eye perspective signals. Another possibility is that 3D estimates derived from perspective cues are noisy, and pooling the two signals is necessary for noise reduction.

The current results are consistent with optimized (through the linear summation of independent stereoscopic cue and perspective cue population responses) but not maximized (due to combining left and right eye perspective cue representations using quadratic nonlinearities and divisive normalization) 3D perception. Analogous limitations may exist in other sensory systems which have multiple inputs sensitive to the same signals, as occurs in audition, vestibular processing, and bimanual touch. Lastly, our results highlight that “optimal” is in the eye of the beholder, and most meaningful in the context of a specific neural architecture.

## References

- Adams WJ, Elder JH, Graf EW, Leyland J, Lutigheid AJ, Murry A (2016) The Southampton-York natural scenes (SYNS) dataset: statistics of surface attitude. *Sci Rep* 6:35805.
- Alberts BB, de Brouwer AJ, Selen LP, Medendorp WP (2016) A Bayesian account of visual-vestibular interactions in the rod-and-frame task. *eNeuro* 3:e0093.
- Alizadeh AM, Van Dromme I, Verhoef BE, Janssen P (2018) Caudal intraparietal sulcus and three-dimensional vision: a combined functional magnetic resonance imaging and single-cell study. *Neuroimage* 166:46–59.



- Ban H, Welchman AE (2015) fMRI analysis-by-synthesis reveals a dorsal hierarchy that extracts surface slant. *J Neurosci* 35:9823–9835.
- Banks MS, Hooge IT, Backus BT (2001) Perceiving slant about a horizontal axis from stereopsis. *J Vis* 1:55–79.
- Barton RA (2004) Binocularity and brain evolution in primates. *Proc Natl Acad Sci USA* 101:10113–10115.
- Beck JM, Latham PE, Pouget A (2011) Marginalization in neural circuits with divisive normalization. *J Neurosci* 31:15310–15319.
- Burge J, McCann BC, Geisler WS (2016) Estimating 3D tilt from local image cues in natural scenes. *J Vis* 16:2.
- Busse L, Wade AR, Carandini M (2009) Representation of concurrent stimuli by population activity in visual cortex. *Neuron* 64:931–942.
- Campbell FW, Kulikowski JJ, Levinson J (1966) The effect of orientation on the visual resolution of gratings. *J Physiol* 187:427–436.
- Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. *Nat Rev Neurosci* 13:51–62.
- Coppola DM, Purves HR, McCoy AN, Purves D (1998) The distribution of oriented contours in the real world. *Proc Natl Acad Sci USA* 95:4002–4006.
- Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. *Annu Rev Neurosci* 24:203–238.
- Dakin CJ, Rosenburg A (2018) Gravity estimation and verticality perception. *Handb Clin Neurol* 159:43–59.
- Elmore LC, Rosenberg A, DeAngelis GC, Angelaki DE (2019) Choice-related activity during visual slant discrimination in macaque CIP but not V3A. *eNeuro* 6:e0248.
- Fuchs AF (1967) Saccadic and smooth pursuit eye movements in the monkey. *J Physiol* 191:609–631.
- Ganguli D, Simoncelli EP (2014) Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Comput* 26:2103–2134.
- Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* 14:926–932.
- Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge: Cambridge University Press.
- Hegd e J, Van Essen DC (2005) Role of primate visual area V4 in the processing of 3-D shape characteristics defined by disparity. *J Neurophysiol* 94:2856–2866.
- Henderson M, Vo V, Chunharas C, Sprague T, Serences J (2019) Multivariate analysis of BOLD activation patterns recovers graded depth representations in human visual and parietal cortex. *eNeuro* 6:e0362.
- Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: optimal cue combination. *J Vis* 4:967–992.
- Howard IP, Rogers BJ (1995) Binocular vision and stereopsis. New York: Oxford University Press.
- Kapoula Z, Yang Q, Vernet M, Dieudonn e B, Greffard S, Verny M (2010) Spread deficits in initiation, speed and accuracy of horizontal and vertical automatic saccades in dementia with Lewy bodies. *Front Neurol* 1:138.
- Ke SR, Lam J, Pai DK, Spering M (2013) Directional asymmetries in human smooth pursuit eye movements. *Invest Ophthalmol Vis Sci* 54:4409–4421.
- Kim B, Kenchappa SC, Sunkara A, Chang TY, Thompson L, Doudlah R, Rosenberg A (2019) Real-time experimental control using network-based parallel processing. *eLIFE* 8:e40231.
- Kim S, Burge J (2018) The lawful imprecision of human surface tilt estimation in natural scenes. *eLIFE* 7:e31448.
- Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C (2007) What's new in Psychtoolbox-3. *Perception* 36:1.
- Knill DC (1998a) Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Res* 38:1655–1682.
- Knill DC (1998b) Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Res* 38:2635–2656.
- Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res* 43:2539–2558.
- Landy MS, Banks MS, Knill DC (2011) Ideal-observer models of cue integration. In: *Sensory cue integration*, pp 5–29. New York: Oxford University Press.
- Louie K, Khaw MW, Glimcher PW (2013) Normalization is a general neural mechanism for context-dependent decision making. *Proc Natl Acad Sci USA* 110:6139–6144.
- Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9:1432–1438.
- Murphy AP, Ban H, Welchman AE (2013) Integration of texture and disparity cues to surface slant in dorsal visual cortex. *J Neurophysiol* 110:190–203.
- Murray RF, Morgenstern Y (2010) Cue combination on the circle and the sphere. *J Vis* 10:15.
- Nguyenkim JD, DeAngelis GC (2003) Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *J Neurosci* 23:7117–7128.
- Ni AM, Ray S, Maunsell JH (2012) Tuned normalization explains the size of attention modulations. *Neuron* 73:803–813.
- Oru c I, Maloney LT, Landy MS (2003) Weighted linear cue combination with possibly correlated error. *Vision Res* 43:2451–2468.
- Preston TJ, Kourtzi Z, Welchman AE (2009) Adaptive estimation of three-dimensional structure in the human brain. *J Neurosci* 29:1688–1698.
- Prince SJ, Cumming BG, Parker AJ (2002) Range and mechanism of encoding of horizontal disparity in macaque V1. *J Neurophysiol* 87:209–221.
- Qamar AT, Cotton RJ, George RG, Beck JM, Prezhdo E, Laudano A, Tolia AS, Ma WJ (2013) Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proc Natl Acad Sci USA* 110:20332–20337.
- Rideaux R, Welchman AE (2018) Proscription supports robust perceptual integration by suppression in human visual cortex. *Nat Commun* 9:1502.
- Rosenberg A, Angelaki DE (2014a) Gravity influences the visual representation of object tilt in parietal cortex. *J Neurosci* 34:14170–14180.
- Rosenberg A, Angelaki DE (2014b) Reliability-dependent contributions of visual orientation cues in parietal cortex. *Proc Natl Acad Sci USA* 111:18043–18048.
- Rosenberg A, Cowan NJ, Angelaki DE (2013) The visual representation of 3D object orientation in parietal cortex. *J Neurosci* 33:19352–19361.
- Rosenberg A, Patterson JS, Angelaki DE (2015) A computational perspective on autism. *Proc Natl Acad Sci USA* 112:9158–9165.
- Sanada TM, DeAngelis GC (2014) Neural representation of motion-in-depth in area MT. *J Neurosci* 34:15508–15521.
- Seilheimer RL, Rosenberg A, Angelaki DE (2014) Models and processes of multisensory cue combination. *Curr Opin Neurobiol* 25:38–46.
- Stevens KA (1981) The information content of texture gradients. *Biol Cybern* 42:95–105.
- Stevens KA (1983a) Surface tilt (the direction of slant): a neglected psychophysical variable. *Percept Psychophys* 33:241–250.
- Stevens KA (1983b) Slant-tilt: the visual encoding of surface orientation. *Biol Cybern* 46:183–195.
- Taira M, Tsutsui KI, Jiang M, Yara K, Sakata H (2000) Parietal neurons represent surface orientation from the gradient of binocular disparity. *J Neurophysiol* 83:3140–3146.
- Thompson L, Ji M, Rokers B, Rosenberg A (2019) Contributions of binocular and monocular cues to motion-in-depth perception. *J Vis* 19:2.
- Tsutsui K, Jiang M, Yara K, Sakata H, Taira M (2001) Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *J Neurophysiol* 86:2856–2867.
- Tsutsui K, Sakata H, Naganuma T, Taira M (2002) Neural correlates for perception of 3D surface orientation from texture gradient. *Science* 298:409–412.

- Tsutsui K, Jiang M, Sakata H, Taira M (2003) Short-term memory and perceptual decision for three-dimensional visual features in the caudal intraparietal sulcus (area CIP). *J Neurosci* 23:5486–5495.
- Van Dromme IC, Premereur E, Verhoef B-E, Vanduffel W, Janssen P (2016) Posterior parietal cortex drives inferotemporal activations during three-dimensional object vision. *PLoS Biol* 14:e1002445.
- Welchman AE, Deubelius A, Conrad V, Bülthoff HH, Kourtzi Z (2005) 3D shape perception from combined depth cues in human visual cortex. *Nat Neurosci* 8:820–827.
- Westheimer G, Beard BL (1998) Orientation dependency for foveal line stimuli: detection and intensity discrimination, resolution, orientation discrimination and vernier acuity. *Vision Res* 38:1097–1103.