Object representation in a gravitational reference frame

Alexandriya M.X. Emonds[1,2], Ramanujan Srinath[2,3,4], Kristina J. Nielsen[2,3,5], Charles E. Connor[2,3,5*]

[1]Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

[2]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, Maryland 21218, USA

[3]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

[4]Current address: Department of Neuroscience, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

[5]Senior authors

[*]Correspondence: connor@jhu.edu

ABSTRACT

When your head tilts laterally, as in sports, reaching, and resting, your eyes counterrotate less than 20%, and thus eye images rotate, over a total range of about 180°. Yet, the world appears stable and vision remains normal. We discovered a neural strategy for rotational stability in anterior inferotemporal cortex (IT), the final stage of object vision in primates. We measured object orientation tuning of IT neurons in macaque monkeys tilted +25 and –25° laterally, producing ~40° difference in retinal image orientation. Among IT neurons with consistent object orientation tuning, 63% remained stable with respect to gravity across tilts. Gravitational tuning depended on vestibular/somatosensory but also visual cues, consistent with previous evidence that IT processes scene cues for gravity's orientation. In addition to stability across image rotations, an internal gravitational reference frame is important for physical understanding of a world where object position, posture, structure, shape, movement, and behavior interact critically with gravity.

INTRODUCTION

Reflexive eye movements compensate for up/down and right/left head movements, but when your head tilts laterally, as during sports, driving[1], social communication[2–5], working in cramped environments, reaching for distant objects, and resting in bed, your eyes compensate less than 20%[6,7], so retinal images rotate around the point of fixation. But the perceptual compensation for this is so automatic and complete that we are usually unaware of the image rotation, and visual abilities are not strongly affected. This perceptual stability is more than just a generalization of recognition across orientations. Critically, our perceptual reference frame for objects remains stable with respect to the environment and gravity. As a result, trees still appear vertical and apples still appear to fall straight to the ground, even though their orientations and trajectories on the retina have changed.

Here, we explored the hypothesis that this perceptual stability is produced by transforming visual objects into a stable, non-retinal reference frame. Our previous work has shown that the primate ventral visual pathway[8] implements an object-centered reference frame[9–18], stabilizing against position and size changes on the retina. But this still leaves open the *orientation* of the ventral pathway reference frame. Our recent work has shown that one channel in anterior ventral pathway processes scene-level visual cues for the orientation of the gravitational reference frame[19,20], raising the possibility that the ventral pathway reference frame is aligned with gravity. Here, we confirmed this hypothesis in anterior IT[8], and found that gravitational alignment depends on both visual and vestibular/somatosensory[21,22] cues. To a lesser extent, we observed tuning aligned with the retinal reference frame, and object orientation in either reference frame was linearly decodable from IT population responses with high accuracy. This is consistent with psychophysical results showing voluntary perceptual access to either

reference frame[23]. The dominant, gravitationally aligned reference frame not only confers stability across image rotations but also enables physical understanding of objects in a world dominated by the force of gravity.

RESULTS

Object tuning in a gravitational reference frame

Monkeys performed a dot fixation task while we flashed object stimuli on a high-resolution LED monitor spanning 100° of the visual field in the horizontal direction. We used evolving stimuli guided by a genetic algorithm[12,13,16–20] to discover 3D objects that drove strong responses from IT neurons. We presented these objects centered at fixation, across a range of screen orientations, with the monkey's head (fixed to a rotating chair) tilted clockwise (–) or counterclockwise (+) by 25° about the axis of gaze (through the fixation point and the interpupillary midpoint) (Fig. 1a,b). Compensatory ocular counter-rolling was measured to be ~6° based on iris landmarks visible in high-resolution photographs, consistent with previous measurements[6,7]. We also found the that 6° compensation produced the closest agreement in the retinal reference frame between object orientation tuning functions across tilts (Fig. S1).

The Fig. 1 example neuron was tested with both full scene stimuli (Fig. 1a), which included a textured ground surface and horizon, providing visual cues for the orientation of gravity, and isolated objects (Fig. 1b), presented on a circular gray background in an otherwise dark room, so that only vestibular and somatosensory cues indicated the orientation of gravity. Object orientation tuning remained stable with respect to gravity across tilts, peaking at orientation 0°, for both full scene (Fig. 1c,d) and isolated object (Fig. 1g,h) stimuli. Correspondingly,

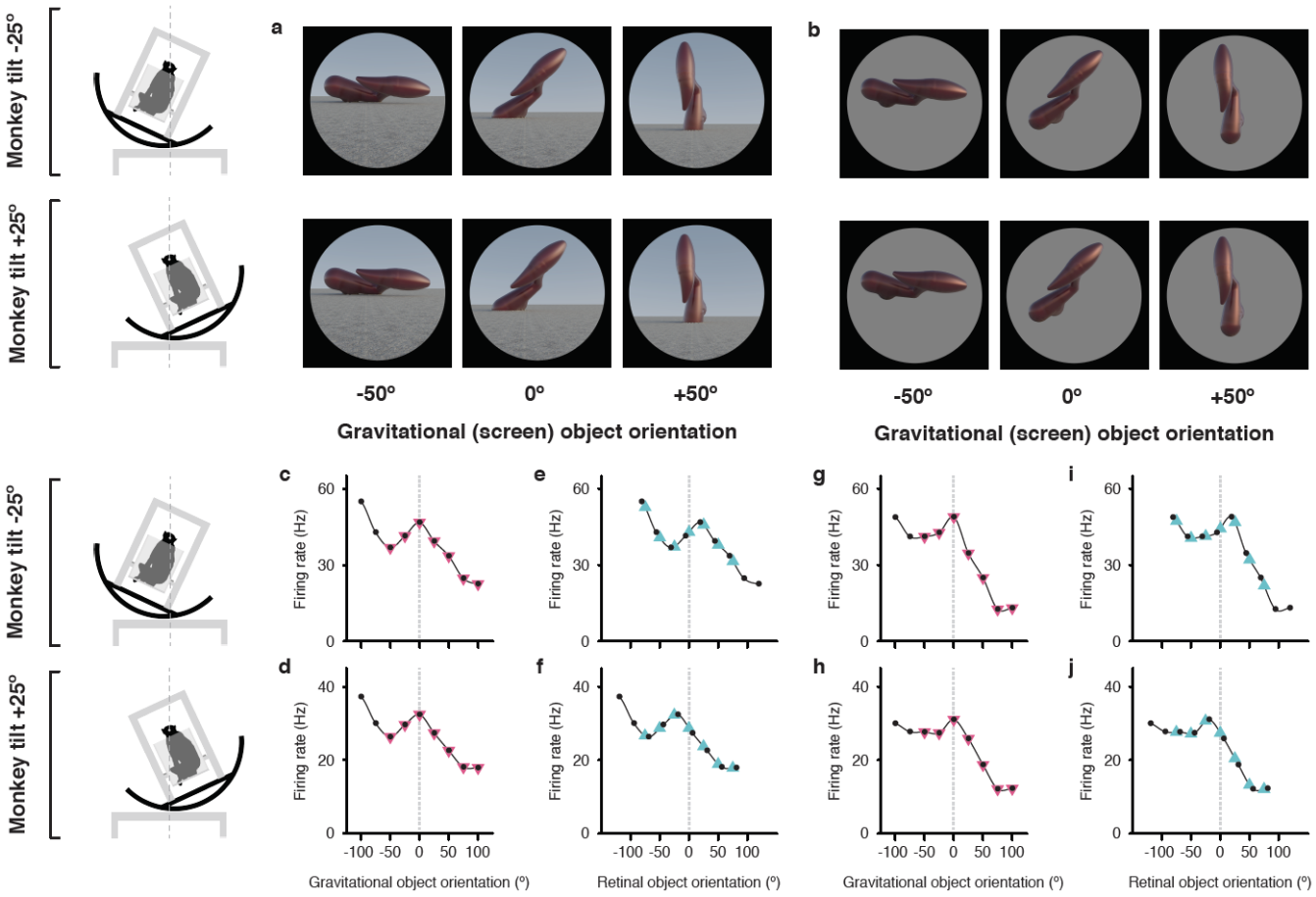


Figure 1. Example neuron tuned for object orientation in a gravitational reference frame. (a,b) Stimuli demonstrating example object orientations in the full scene condition. The orientation discovered in the genetic algorithm experiments is arbitrarily labeled 0°. The two monkey tilt conditions are diagrammed at left. The small *white dots* at the center of the head (connected by vertical *dashed lines*) represent the virtual axis of rotation produced by a circular sled supporting the chair. Stimuli were presented on a 100°-wide display screen for 750 ms (separated by 250 ms blank screen intervals) while the monkey fixated a central dot. Stimuli were presented in random order for a total of 5 repetitions each. (c,d) Responses of an example IT neuron to full scene stimuli, as a function of object orientation on the screen and thus with respect to gravity, across a 100° orientation range, while the monkey was tilted –25° (c) and 25° (d). Response values are averaged across the 750 ms presentation time and across 5 repetitions and smoothed with a boxcar kernel of width 50° (3 orientation values). For this neuron, object orientation tuning remained consistent with respect to gravity across the two tilt conditions, with a peak response centered at 0° (*dashed vertical line*). The *pink triangles* indicate the object orientations directly comparable to the retinal analyses. (e,f) The same data plotted against orientation on the retina, corrected for 6° counter-rolling of the eyes (Fig. S1). The *cyan triangles* indicate the response values directly comparable to gravitational analyses. Due to the shift produced by ocular counter-rolling, these comparison values were interpolated between tested screen orientations using a Catmull-Rom spline. Since orientation tuning was consistent in gravitational space, the peaks are shifted right or left by about 20° each. (g–j) Similar results were obtained for this neuron with isolated object stimuli.

orientation tuning profiles shifted relative to retinal orientation by about 40° between the two tilt conditions (Fig. 1e,f,i,j), shifting the peak to the right and left of 0°. A similar example neuron is presented in Fig. S2a,c–f, along with an example neuron for which tuning aligned with the retina, and thus shifted with respect to gravity (Fig. S2b,g–j).

Distribution of gravity- and retina-aligned tuning

Fig. 2a scatterplots correlation values between object orientation tuning functions in the two tilt conditions calculated with respect to retinal orientation (*x axis*) and gravity (*y axis*), for a sample of 89 IT neurons tested with full scene stimuli. In both the scatterplot and the marginal histograms, color indicates the result of a 1-tailed randomization t-test on each cell for significant positive correlation ($p < 0.05$) in the gravitational reference frame (*pink)*, retinal reference frame (cyan), or both reference frames (*dark gray*) (presumably due to the broad object orientation tuning of some IT neurons[17]).

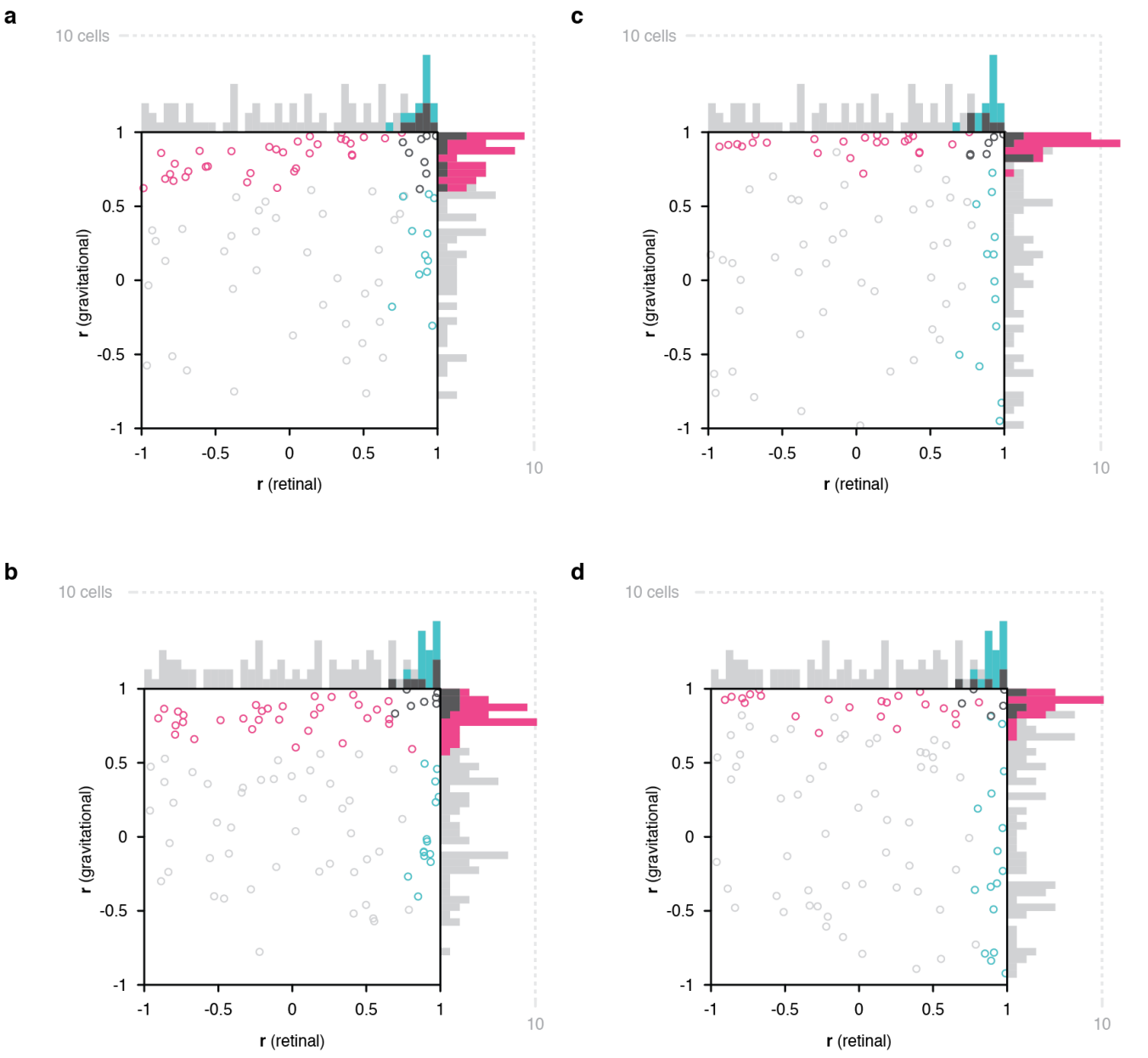


Figure 2. Scatterplots of object orientation tuning function correlations across tilts. (a) Scatterplot of correlations for full scene stimuli. Correlations of tuning in the gravitational reference frame (*y axis*) are plotted against correlations in the retinal reference frame (*x axis*). Marginal distributions are shown as histograms. Neurons with significant correlations with respect to gravity are colored *pink* and neurons with significant correlations with respect to the retinae` are colored *cyan*. Neurons with significant correlations in both dimensions are colored *dark gray*, and neurons with no significant correlation are colored *light gray*. (b) Scatterplot for isolated object stimuli. Conventions the same as in (a). (c) Same scatterplot as in (a), but balanced for number of comparison orientations between gravitational and retinal analysis. (d) Same as (b), but balanced for number of comparison orientations between gravitational and retinal analysis.

Of the 52 neurons with consistent object orientation tuning in one or both reference frames, 63% (33/52) were aligned with gravity, 21% (11/52) were aligned with the retinae, and 15% (8/52) were aligned with both. The population tendency toward positive correlation was significant along the gravitational axis (two-tailed randomization t-test for center-of-mass relative to 0; $p = 6.49 \times 10^{-29}$) as well as the retinal axis ($p = 5.76 \times 10^{-10}$). Similar results were obtained for a partially overlapping sample of 99 IT neurons tested with isolated object stimuli in a darkened room to remove visual cues for gravity (Fig. 2b), with 60% (32/53) significant correlation in the gravitational reference frame, 26% (14/53) significant correlation in the retinal reference frame, and 13% (7/53) significant in both reference frames. The population tendency toward positive correlation was again significant in this experiment along both gravitational ($p = 3.63 \times 10^{-22}$) and retinal axes ($p = 1.63 \times 10^{-7}$).

The analyses above were based on the full set of orientation comparisons possible for the gravitational reference frame (7), while the experimental design inevitably produced fewer comparisons for the retinal reference frame (5). Rerunning the analyses based on just 5 comparable object orientations in both reference frames (Fig. 1, *pink* and *cyan triangles*) produced the results shown in Figs. 2c and d. For full scene stimuli, this yielded 56% (23/41) significant gravitational alignment, 27% (11/41) retinal alignment, and 17% (7/41) dual alignment (Fig. 2c). For isolated object stimuli, this reanalysis yielded 58% (28/48) gravitational alignment, 29% (14/48) retinal alignment, and 13% (6/48) dual alignment (Fig. 2d).

Population coding of orientation in both reference frames

Neurons with no significant correlation in either reference frame might actually combine signals from both reference frames, as in other brain systems that interact with multiple reference

frames[24–32]. This would be consistent with human psychophysical results showing mixed influences of retinal and gravitational reference frames, with stronger weight for gravitational[33,34]. For mixed reference frame tuning of this kind, it has been shown that simple linear decoding can extract information in any one reference frame with an appropriate weighting pattern across neurons[24,35,36]. We tested that idea here and found that object orientation information in either gravitational or retinal space could be decoded with high accuracy from the responses of the IT neurons in our sample. The decoding task was to determine whether two population responses (across the 89 neurons tested with full scene stimuli) were derived from same or different orientations, either in gravitational space or retinal space (corrected for counter-rolling). This match/non-match task allowed us to analyze population information about orientation equivalence even though individual neurons were tested using different stimuli with no comparability between orientations. (Across neurons, orientations were aligned according to their order in the tested range, so that each non-match trial involved the same orientation difference, in the same direction, for each neuron.)

The accuracy of linear discriminant models for orientation match/non-match in the gravitational reference frame was 97% (10-fold cross-validation). The accuracy of models for orientation match/non-match in the retinal reference frame was 98%. This easy decoding of information in both reference frames is consistent with psychophysical results showing that humans have voluntary access to either reference frame[23]. High accuracy was obtained even with models based solely on neurons that showed no significant correlation in either gravitational or retinal reference frames (Fig. 2a, *light gray*): 89% for gravitational discrimination and 97% for retinal discrimination. This supports the idea that these neurons carry a mix of retinal and gravitational object orientation signals.

177

178 **Gravity-aligned tuning based on purely visual cues**

179 The results for isolated object stimuli in Fig. 2b,d indicate that alignment of object information
180 with gravity does not require the visual cues present in the full scene stimuli (ground surface
181 and horizon) and can be based purely on vestibular and somatosensory cues for the direction
182 of gravity in a dark room. We also tested the converse question of whether purely visual cues
183 (tilted horizon and ground surface) could produce alignment of object orientation tuning with
184 the visually apparent orientation of gravity, even in the presence of conflicting vestibular and
185 somatosensory cues (i.e. with the monkey in a normal upright orientation). In spite of the
186 conflict, many neurons showed object orientation tuning functions aligned with the visually
187 cued direction of gravity, as exemplified in Fig. 3a,c–f. The 5 object orientations that were
188 comparable in a gravitational reference frame (*pink triangles*) produced consistent responses
189 to object orientations relative to the ground surface and horizon (Fig. 3c,d). For example, the
190 top left stimulus in Fig. 3a (horizon tilt –25°, retinal orientation –25°) has the same orientation
191 with respect to the ground surface as the bottom right stimulus (horizon tilt +25°, retinal
192 orientation +25°). Thus, in the visually-cued gravitational reference frame, these two stimuli
193 line up at 0° orientation in both Fig. 3c and d, and they evoke similar responses. Conversely,
194 the 9 orientations comparable in the retinal reference (*black dots* and *cyan triangles*) produced
195 inconsistent responses (Fig. 3e,f). A different example neuron (Fig. 3b,g–j) exhibited object
196 orientation tuning aligned with the retinae (Fig. 3i,j) and not gravity (Fig. 3g,h).

197

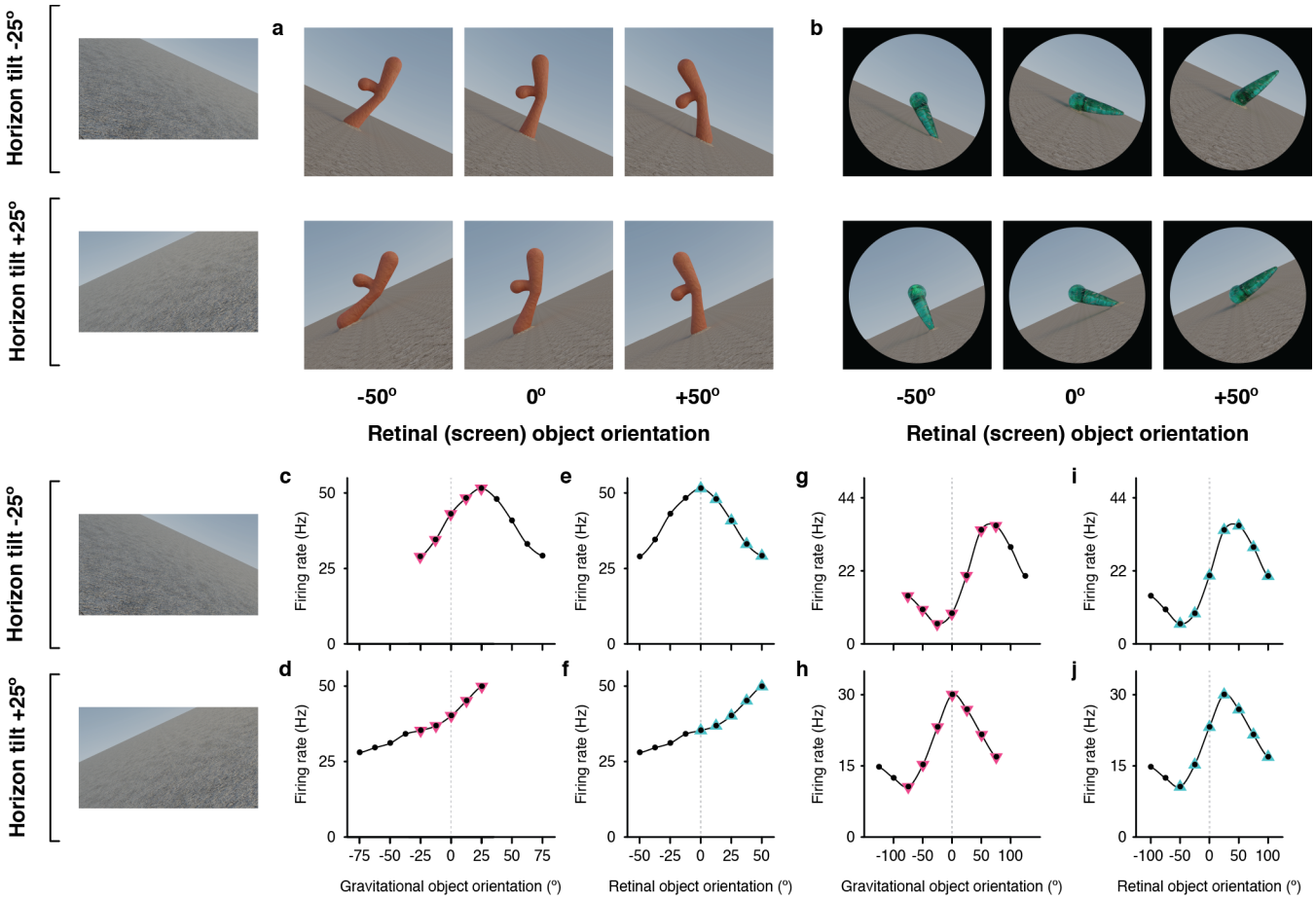


Figure 3. Example neurons tested with tilted horizon stimuli while the monkey remained in an upright orientation. (a,b) Stimuli demonstrating example object orientations in two conditions, with the ground surface, horizon, and sky gradient tilted –25° (clockwise, top row), and with ground surface, etc. tilted +25° (counterclockwise, second row). The monkey was in a normal upright orientation during these experiments, producing conflicting vestibular/somatosensory cues. The retinal orientation discovered in the genetic algorithm experiments is arbitrarily labeled 0°. (c,d) For an example IT neuron tested with the stimuli in (a), object orientation tuning with respect to the visually cued direction of gravity was consistent across the two ground tilts. (e,f) Correspondingly, the neuron gave very different responses to retinal object orientation values between the two ground tilts. (g,h) This different example IT neuron tested with the stimuli in (b) did not exhibit consistent object orientation tuning in visually-cued gravitational space. (i,j) Instead, this neuron maintained consistent tuning for retinal-screen orientation despite changes in ground tilt.

Across a sample of 228 IT neurons studied in this cue conflict experiment, 123 showed significant correlation across visual ground/horizon tilt in one or both reference frames. Of these, 54% (67/123) showed object orientation tuning aligned with the retinal reference frame, 35% (43/123) with the gravitational reference frame, and 11% (13/123) with both (Fig. 4a). The population tendency toward retina-aligned orientation tuning was significant (two-tailed randomization t-test for center-of-mass relative to 0; $p = 8.14 \times 10^{-28}$) as was the tendency toward gravity-aligned orientation tuning ($p = 6.23 \times 10^{-6}$). The experimental design in this case produced more comparisons in the retinal reference frame, and balancing the numbers of comparisons resulted in more equal percentages (Fig. 4b). The main result in this experiment, that many IT neurons exhibit object orientation tuning aligned with visual cues for the direction of gravity, even in the presence of conflicting vestibular/somatosensory cues, argues that visual cues contribute to gravity-aligned tuning under normal circumstances, where they combine with convergent vestibular/somatosensory cues. That would be consistent with our previous discovery that many neurons in IT are strongly tuned for the orientation of large-scale ground surfaces and edges, in the orientation ranges experienced across head tilts[14,15], and more generally with the strong visual representation of scene information in temporal lobe[37–40].

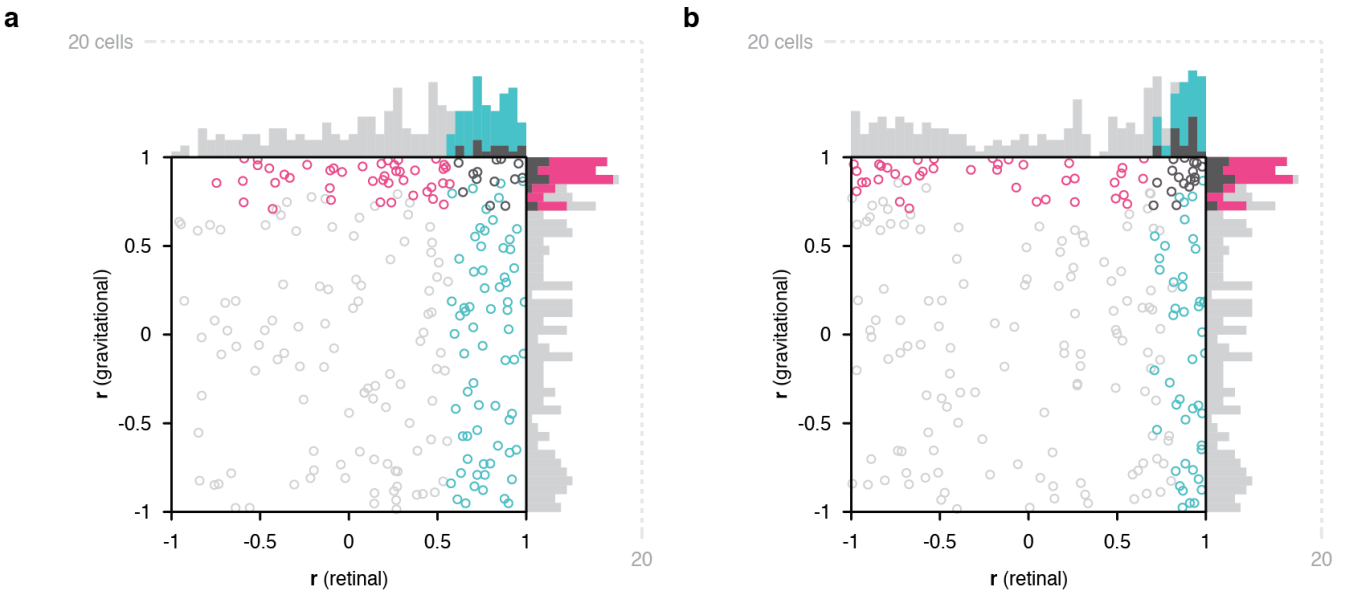


Figure 4. Scatterplots of object orientation tuning function correlations across visual horizon tilts on the screen, with the monkey in an upright orientation. (a) Scatterplot of correlations for full scene stimuli. Correlations of tuning in gravitational space as cued by horizon tilt (*y axis*) are plotted against correlations in retinal space (*x axis*). Marginal distributions are shown as histograms. Neurons with significant correlations in visually-cued gravitational space are colored *pink* and neurons with significant correlations in retinal space are colored *cyan*. Neurons with significant correlations in both dimensions are colored *dark gray*, and neurons with no significant correlation are colored *light gray*. (b) Same scatterplot as in (a), but with correlation values balanced for number of comparison orientations between gravitational and retinal analysis.

DISCUSSION

The fundamental goal of visual processing is to transform photoreceptor sheet responses into usable, essential information—readable, compressed, stable signals, for the specific things we need to understand about the world. In this sense, the transformation described here achieves both stability and specificity of object information. The gravitational reference frame remains stable across retinal image rotations, a critical advantage for vision from the tilting platform of the head and body. And, it enables understanding of object structure, posture, shape, motion, and behavior relative to the strong gravitational force that constrains and interacts with all these factors. It provides information about whether and how objects and object parts are supported and balanced against gravity, how flexible, motoric objects like bodies are interacting energetically with gravity, what postural or locomotive behaviors are possible or likely, and about potential physical interactions with other objects or with the observer under the influence of gravity. In other words, it provides information critical for guiding our mechanistic understanding of and skillful interactions with the world.

A similar hypothesis about gravity-related tuning for tilted planes has been tested in parietal area CIP (central intraparietal area). Rosenberg and Angelaki[41] measured the responses of 46 CIP neurons with two monkey tilts, right and left 30°, and fit the responses with linear models. They reported significant alignment with eye orientation for 45 of 92 (49%) tilt tests (right and left for each neuron), intermediate between eye and gravity for 26/92 tilt tests (28%), and alignment with gravity for 6/92 tilt tests (7%). However, of the 5 neurons in this last category, only one appeared to show significant alignment with gravity for both tilt directions (Rosenberg and Angelaki[41], Fig. 4D). Thus, while orientation tuning of ~35% of CIP neurons was sensitive to monkey tilt and gravity-aligned information could be extracted with a neural network[41], there

was no explicit tuning in a gravitational reference frame or dominance of gravitational information as found here. There is however compelling human fMRI evidence that parietal and frontal cortex are deeply involved in perceiving and predicting physical events[42], and have unique abstract signals for stability not detected in ventral pathway[43] (though these could reflect decision-making processes[44,45]). Our results and others[46–48] suggest nonetheless that ventral pathway object and scene processing may be a critical source of information about gravity and its effects on objects, especially when detailed object representations are needed to assess precise shape, structure, support, strength, flexibility, compressibility, brittleness, specific gravity, mass distribution, and mechanical features to understand real world physical situations.

Our results raise the interesting question of *how* visual information is transformed into a gravity-aligned reference frame, and how that transformation incorporates vestibular, somatosensory, and visual cues for the direction of gravity. Previous work on reference frame transformation has involved *shifts* in the position of the reference frame. There is substantial evidence that these shifts are causally driven by anticipatory signals for attentional shifts and eye movements from prefrontal cortex, acting on ventral pathway cortex to activate neurons with newly relevant spatial sensitivities[49–54]. Here, the more difficult geometric problem is *rotation* of visual information, such that “up”, “down”, “right” and “left” become associated with signals from different parts of the retina, based on a change in the perceived direction of gravity. This could also involve spatial remapping, but in circular directions, within an object-centered reference frame[10–18]. Humans can perform tasks requiring mental rotation of shapes, but this is time consuming in proportion to the angle of required rotation[55], and seems to rely on an unusual strategy of covert motor simulation[56]. The rotation required here is fast and so

289 automatic as to be unnoticeable. Discovering the underlying transformation mechanism will

290 likely require extensive theoretical, computational, and experimental investigation.

291

REFERENCES

1. Zikovitz, D. C., & Harris, L. R. Head tilt during driving. *Ergonomics*, **42**, 740-746 (1999). https://doi.org/10.1080/001401399185414

2. Halberstadt, A. G., & Saitta, M. B. Gender, nonverbal behavior, and perceived dominance: A test of the theory. *Journal of Personality and Social Psychology*, **53**, 257-272 (1987). https://doi.org/10.1037/0022-3514.53.2.257

3. Mignault, A., & Chaudhuri, A. The many faces of a neutral face: Head tilt and perception of dominance and emotion. *Journal of nonverbal behavior*, **27**, 111-132 (2003).

4. Krumhuber, E., Manstead, A. S., & Kappas, A. Temporal aspects of facial displays in person and expression perception: The effects of smile dynamics, head-tilt, and gender. *Journal of Nonverbal Behavior*, **31**, 39-56 (2007).

5. Mara, M., & Appel, M. Effects of lateral head tilt on user perceptions of humanoid and android robots. *Computers in Human Behavior*, **44**, 326-334 (2015). https://doi.org/10.1016/j.chb.2014.09.025

6. Miller, E. F. Counterrolling of the human eyes produced by head tilt with respect to gravity. *Acta Oto-laryngologica*, **54**, 479-501 (1962). https://doi.org/10.3109/00016486209126967

7. Schworm, H. D., Ygge, J., Pansell, T., & Lennerstrand, G. Assessment of ocular counterroll during head tilt using binocular video oculography. *Investigative Ophthalmology & Visual Science*, **43**, 662-667 (2002).

8. Felleman, D. J., & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, **1**, 1-47 (1991). DOI: 10.1093/cercor/1.1.1-a

9. Pasupathy, A., and Connor, C.E. (1999). Responses to contour features in macaque area

316 V4. J. Neurophysiol. *82*, 2490–2502. https://doi.org/10.1152/jn.1999.82.5.2490

317 10. Pasupathy, A., and Connor, C.E. (2001). Shape representation in area V4: position-specific

318 tuning for boundary conformation. J. Neurophysiol. *86*, 2505–2519.

319 https://doi.org/10.1152/jn.2001.86.5.2505

320 11. Pasupathy, A., and Connor, C.E. (2002). Population coding of shape in area V4. Nat.

321 Neurosci. *5*, 1332–1338. doi:10.1038/nn972

322 12. Carlson, E.T., Rasquinha, R.J., Zhang, K., and Connor, C.E. (2011). A sparse objectcoding

323 scheme in area V4. Curr. Biol. *21*, 288–293. https://doi.org/10.1016/j.cub.2011.01.013

324 13. Srinath, R., Emonds, A., Wang, Q., Lempel, A. A., Dunn-Weiss, E., Connor, C. E., &

325 Nielsen, K. J. Early emergence of solid shape coding in natural and deep network vision. *Curr.*

326 *Biol.*, **31**, 51-65 (2021). https://doi.org/10.1016/j.cub.2020.09.076

327 14. Brincat, S.L. & Connor, C.E. (2004) Underlying principles of visual shape selectivity in

328 posterior inferotemporal cortex. Nat. Neurosci. *7*, 880–886. doi:10.1038/nn1278

329 15. Brincat, S.L. & Connor, C.E. (2006) Dynamic shape synthesis in posterior inferotemporal

330 cortex. Neuron *49*, 17–24. https://doi.org/10.1016/j.neuron.2005.11.026

331 16. Yamane, Y., Carlson, E.T., Bowman, K.C., Wang, Z., and Connor, C.E. A neural code for

332 three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* **11**, 1352–

333 1360 (2008). doi:10.1038/nn.2202

334 17. Hung, C.C., Carlson, E.T., and Connor, C.E. Medial axis shape coding in macaque

335 inferotemporal cortex. *Neuron* **74**, 1099–1113 (2012).

336 https://doi.org/10.1016/j.neuron.2012.04.029

337 18. Connor, C.E., & Knierim, J.J. (2017). Integration of objects and space in perception and

338 memory. *Nature Neuroscience* **20**: 1493–1503. doi:10.1038/nn.4657

19. Vaziri, S., Carlson, E.T., Wang, Z., and Connor, C.E. A channel for 3D environmental shape in anterior inferotemporal cortex. *Neuron* **84**, 55–62 (2014). https://doi.org/10.1016/j.neuron.2014.08.043

20. Vaziri, S., and Connor, C.E. Representation of gravity-aligned scene structure in ventral pathway visual cortex. *Curr. Biol.* **26**, 766–774 (2016). https://doi.org/10.1016/j.cub.2016.01.022

21. Brandt, T., Dieterich, M., & Danek, A. Vestibular cortex lesions affect the perception of verticality. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, **35**, 403-412 (1994). https://doi.org/10.1002/ana.410350406

22. Baier, B., Thömke, F., Wilting, J., Heinze, C., Geber, C., & Dieterich, M. A pathway in the brainstem for roll-tilt of the subjective visual vertical: evidence from a lesion–behavior mapping study. *Journal of Neuroscience*, **32**, 14854-14858 (2012). https://doi.org/10.1523/JNEUROSCI.0770-12.2012

23. Attneave, F., & Reid, K. W. Voluntary control of frame of reference and slope equivalence under head rotation. *Journal of Experimental Psychology*, **78**, 153 (1968). https://doi.org/10.1037/h0026150

24. Stricanne, B., Andersen, R. A., & Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *Journal of Neurophysiology*, **76**, 2071-2076 (1996). https://doi.org/10.1152/jn.1996.76.3.2071

25. Buneo, C. A., Jarvis, M. R., Batista, A. P., & Andersen, R. A. Direct visuomotor transformations for reaching. *Nature*, **416**, 632-636 (2002).

26. Avillac, M., Deneve, S., Olivier, E., Pouget, A., & Duhamel, J. R. Reference frames for representing visual and tactile locations in parietal cortex. *Nature neuroscience*, **8**, 941-949 (2005). doi:10.1038/nn1480

27. Mullette-Gillman, O. D. A., Cohen, Y. E., & Groh, J. M. Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *Journal of Neurophysiology*, **94**, 2331-2352 (2005). https://doi.org/10.1152/jn.00021.2005

28. Mullette-Gillman, O. D. A., Cohen, Y. E., & Groh, J. M. Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered reference frame. *Cerebral Cortex*, **19**, 1761-1775 (2009). https://doi.org/10.1093/cercor/bhn207

29. Caruso, V. C., Pages, D. S., Sommer, M. A., & Groh, J. M. Compensating for a shifting world: evolving reference frames of visual and auditory signals across three multimodal brain areas. *Journal of neurophysiology*, **126**, 82-94 (2021). https://doi.org/10.1152/jn.00385.2020

30. Chang, S. W., & Snyder, L. H. Idiosyncratic and systematic aspects of spatial representations in the macaque parietal cortex. *Proceedings of the National Academy of Sciences*, **107**, 7951-7956 (2010). https://doi.org/10.1073/pnas.0913209107

31. McGuire, L. M., & Sabes, P. N. Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *Journal of Neuroscience*, **31**, 6661-6673 (2011). https://doi.org/10.1523/JNEUROSCI.2921-10.2011

32. Chen, X., DeAngelis, G. C., & Angelaki, D. E. Diverse spatial reference frames of vestibular signals in parietal cortex. *Neuron*, **80**, 1310-1321 (2013). https://doi.org/10.1016/j.neuron.2013.09.006

33. Bock, O. L., & Dalecki, M. Mental rotation of letters, body parts and scenes during whole-body tilt: Role of a body-centered versus a gravitational reference frame. *Human Movement Science*, **40**, 352-358 (2015). https://doi.org/10.1016/j.humov.2015.01.017

34. Corballis, M. C., Nagourney, B. A., Shetzer, L. I., & Stefanatos, G. Mental rotation under head tilt: Factors influencing the location of the subjective reference frame. *Perception & Psychophysics*, **24**, 263-273 (1978).

387 35. Deneve, S., Latham, P. E., & Pouget, A. Efficient computation and cue integration with
388 noisy population codes. *Nature neuroscience*, **4**, 826-831 (2001).
389 36. Pouget, A., Deneve, S., & Duhamel, J. R. A computational perspective on the neural basis
390 of multisensory spatial representations. *Nature Reviews Neuroscience*, **3**, 741-747 (2002).
391 37. Epstein, R., & Kanwisher, N. A cortical representation of the local visual
392 environment. *Nature*, **392**, 598-601 (1998).
393 38. Epstein, R. A. Parahippocampal and retrosplenial contributions to human spatial
394 navigation. *Trends in cognitive sciences*, **12**, 388-396 (2008).
395 https://doi.org/10.1016/j.tics.2008.07.004
396 39. Lescroart, M. D., & Gallant, J. L. Human scene-selective areas represent 3D configurations
397 of surfaces. *Neuron*, **101** 178-192 (2019). https://doi.org/10.1016/j.neuron.2018.11.004
398 40. Kornblith, S., Cheng, X., Ohayon, S., & Tsao, D. Y. A network for scene processing in the
399 macaque temporal lobe. *Neuron*, ***79***, 766-781 (2013).
400 https://doi.org/10.1016/j.neuron.2013.06.015
401 41. Rosenberg, A., & Angelaki, D. E. Gravity influences the visual representation of object tilt in
402 parietal cortex. *Journal of Neuroscience*, **34**, 14170-14180 (2014).
403 https://doi.org/10.1523/JNEUROSCI.2030-14.2014
404 42. Fischer, J., Mikhael, J. G., Tenenbaum, J. B., & Kanwisher, N. Functional neuroanatomy of
405 intuitive physical inference. *PNAS*, **113**, E5072-E5081 (2016).
406 https://doi.org/10.1073/pnas.1610344113
407 43. Pramod, R. T., Cohen, M. A., Tenenbaum, J. B., & Kanwisher, N. Invariant representation
408 of physical stability in the human brain. *eLife*, **11**, e71736 (2022).
409 https://doi.org/10.7554/eLife.71736

410 44. Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the
411 parietal cortex (area LIP) of the rhesus monkey. *Journal of neurophysiology*, *86*(4), 1916-1936.
412 https://doi.org/10.1152/jn.2001.86.4.1916
413 45. Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of*
414 *neuroscience*, *30*(1), 535-574. 10.1146/annurev.neuro.29.051605.113038
415 46. Gallivan, J. P., Cant, J. S., Goodale, M. A., & Flanagan, J. R. Representation of object
416 weight in human ventral visual cortex. *Current Biology*, **24**, 1866-1873 (2014).
417 https://doi.org/10.1016/j.cub.2014.06.046
418 47. Gallivan, J. P., Johnsrude, I. S., & Randall Flanagan, J. Planning ahead: object-directed
419 sequential actions decoded from human frontoparietal and occipitotemporal networks.
420 *Cerebral Cortex*, **26**, 708-730 (2016). https://doi.org/10.1093/cercor/bhu302
421 48. Cesanek, E., Zhang, Z., Ingram, J. N., Wolpert, D. M., & Flanagan, J. R. Motor memories
422 of object dynamics are categorically organized. *Elife*, **10** (2021). 10.7554/eLife.71627
423 49. Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H.
424 Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, **29**, 757-767 (2001).
425 https://doi.org/10.1016/S0896-6273(01)00250-1
426 50. Moore, T., & Armstrong, K. M. Selective gating of visual signals by microstimulation of
427 frontal cortex. *Nature*, **421**, 370-373 (2003).
428 51. Moore, T., Armstrong, K. M., & Fallah, M. Visuomotor origins of covert spatial attention.
429 *Neuron*, **40**, 671-683 (2003). https://doi.org/10.1016/S0896-6273(03)00716-5
430 52. Armstrong, K. M., Fitzgerald, J. K., & Moore, T. Changes in visual receptive fields with
431 microstimulation of frontal cortex. *Neuron*, **50**, 791-798 (2006).
432 https://doi.org/10.1016/j.neuron.2006.05.010

433 53. Schafer, R. J., & Moore, T. Selective attention from voluntary control of neurons in
434 prefrontal cortex. *Science*, **332**, 1568-1571 (2011). DOI: 10.1126/science.1199892
435 54. Noudoost, B., & Moore, T. Control of visual cortical signals by prefrontal dopamine. *Nature*,
436 **474**, 372-375 (2011). doi:10.1038/nature09995
437 55. Shepard, R. N., & Metzler, J. Mental rotation of three-dimensional objects. *Science*, **171**,
438 701-703 (1971).
439 56. Wexler, M., Kosslyn, S. M., & Berthoz, A. Motor processes in mental rotation. *Cognition*,
440 **68**, 77-94 (1998). https://doi.org/10.1016/S0010-0277(98)00032-8
441

MATERIALS AND METHODS

Data and code availability. The data and code that support the findings of this study have been deposited in a public Github repository https://github.com/amxemonds/ObjectGravity. Further information and requests for data or custom MATLAB code should be directed to and will be fulfilled by the corresponding author, Charles E. Connor (connor@jhu.edu).

Behavioral task, stimulus presentation, and electrophysiological recording. Two head-restrained male rhesus monkeys (*Macaca mulatta*) were trained to maintain fixation within 1° (radius) of a 0.3° diameter spot for 4 s to obtain a juice reward. Eye position was monitored with an infrared eye tracker (EyeLink). Image stimuli were displayed on a 3840 x 2160 resolution, 80.11 DPI television screen placed directly in front of the monkey, centered at eye level at a distance of 60 cm. The screen subtended 70° vertically and 100° horizontally. Monkeys were seated in a primate chair attached to a +/– 25º full-body rotation mechanism with the center of rotation at the midpoint between the eyes, so that the angle of gaze toward the fixation point remained constant across rotations. The rotation mechanism locked at body orientations of –25º (tilted clockwise), 0º, and +25º (counterclockwise). After fixation was initiated by the monkey, 4 stimuli were presented sequentially, for 750 ms each, separated by 250 ms intervals with a blank, gray background. All stimuli in a given generation were tested in random order for a total of 5 repetitions. The electrical activity of well-isolated single neurons was recorded with epoxy-coated tungsten electrodes (FHC Microsystems). Action potentials of individual neurons were amplified and electrically isolated using a Tucker-Davis Technologies recording system. Recording positions ranged from 5 to 25 mm anterior to stereotaxic 0 within the inferior temporal lobe, including the ventral bank of the superior temporal sulcus, lateral convexity, and basal surface. Positions were determined on the basis

of structural magnetic resonance images and the sequence of sulci and response characteristics observed while lowering the electrode. A total of 368 object-selective IT neurons were studied with different combinations of experiments All animal procedures were approved by the Johns Hopkins Animal Care and Use Committee and conformed to US National Institutes of Health and US Department of Agriculture guidelines.

Stimulus generation. Initially random 3D stimuli evolved through multiple generations under control of a genetic algorithm[12–16], leading to high-response stimuli used to test object orientation tuning as a function of eye/head/body rotation. Random shapes were created by defining 3D mesh surfaces surrounding medial axis skeletons[13]. These shapes were assigned random or evolved optical properties including color, surface roughness, specularity/reflection, translucency/transparency, and subsurface scattering. They were depicted as projecting from (partially submerged in) planar ground surfaces covered with a random scrub grass texture extending toward a distant horizon meeting a blue, featureless sky, with variable ambient light color and lighting direction consistent with random or evolved virtual times of day. Ground surface tilt and object orientation were independent variables of interest as described in the main text. These varied across ranges of 100–200° at intervals of 12.5 or 25°. Ground surface slant, texture gradient, and horizon level, as well as object size and appearance, varied with random or evolved virtual viewing distances. The entire scenes were rendered with multi-step ray tracing using Blender Cycles running on a cluster of GPU-based machines.

Data analysis and statistics. Response rates for each stimulus were calculated by counting action potentials during the presentation window and averaging across 5 repetitions. Orientation tuning functions were smoothed with boxcar averaging across 3 neighboring

values. Pearson correlation coefficients between object orientation tuning functions in different conditions (in some cases corrected for ocular counter-rolling) were calculated for the averaged, smoothed values. Significance of positive correlations were measured with a 1-tailed randomization t-test, $p = 0.05$. (There was no a priori reason to predict or test for negative correlations between orientation tuning functions.) A null distribution was created by randomly assigning response values across the tested orientations within each of the two tuning functions and recalculating the t-statistic 10,000 times. Significant biases of population correlation distributions toward positive or negative values were measured with 2-tailed randomization t-tests, with exact p-values reported. A null distribution was created by randomly assigning response values across the tested orientations within each of the two tuning functions for each neuron, recalculating the t-statistic for each neuron, and recalculating the correlation distribution center of mass on the correlation domain 10,000 times.

Population decoding analysis. We pooled data across 89 neurons tested with full scene stimuli at the two monkey tilts and used cross-validated linear discriminant analysis to discriminate matching from non-matching orientations in both the retinal and gravitational reference frames. Ground truth matches were identical (in either gravitational or counter-rolling corrected retinal coordinates, depending on which reference frame was being tested). Ground truth non-matches differed by more than 25°. We equalized the numbers of retinal and gravitational match and non-match conditions by subsampling. This yielded five potential pairs of matches and 20 potential pairs of non-matches for each reference frame. For each member of a test pair, we randomly selected one raw response value for each neuron from among the 5 individual repetitions for that object orientation. We generated a dataset for all possible test pairs under these conditions. We used the Matlab function fitcdiscr to build optimal models for

514 linear discrimination of matches from non-matches based on response patterns across the 89
515 neurons. We built separate models for retinal and gravitational reference frame match/non-
516 match discrimination. We report the accuracy of the models as 1 – misclassification rate using
517 10-fold cross validation.
518

519 **Acknowledgements:** The authors acknowledge the design and manufacturing contributions of
520 William Nash, William Quinlan, and James Garmon, the software and hardware engineering
521 contributions of Justin Killibrew, and the animal care, handling, training, and surgery
522 contributions of Ofelia Garalde. Dr. Amy Bastian commented on the manuscript.

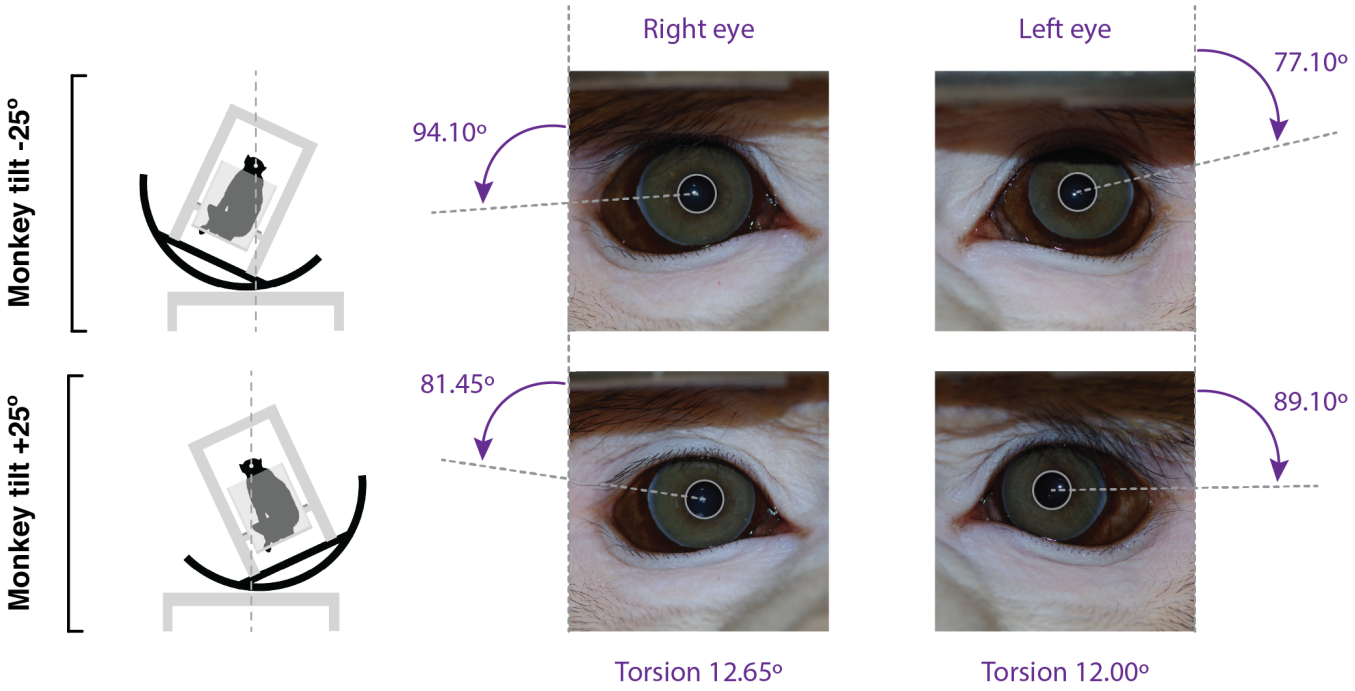
523

524 **Author Contributions:** A.M.X.E., K.J.N., and C.E.C. designed the experiments and analyses,
525 all of which were performed by A.M.X.E., except for the population coding analyses, which
526 were designed and performed by R.S.

527

528 **Conflict of Interest Declaration:** The authors declare that they have no competing interests.

SUPPLEMENTARY INFORMATION



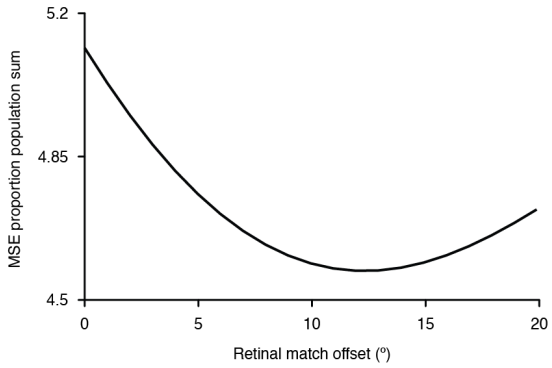


Figure S1. Analysis of eye counter-rotation during tilt. Eye orientation was estimated with lines connecting the center of the pupil with visualizable features at the edge of the iris. For the right and left eyes in the same monkey, the measured difference in eye orientation was 12.65° for the right eye and 12.00° for the left eye (*upper right*). For all the data, from both monkeys, we normalized and summed the mean squared error (MSE) between responses at corresponding retinal positions, as a function of the offset value between retinal positions used to compensate for counter-rotation (*lower left*). The minimum MSE was measured at 12° offset, corresponding to 6° rotation from normal in each of the tilt conditions.

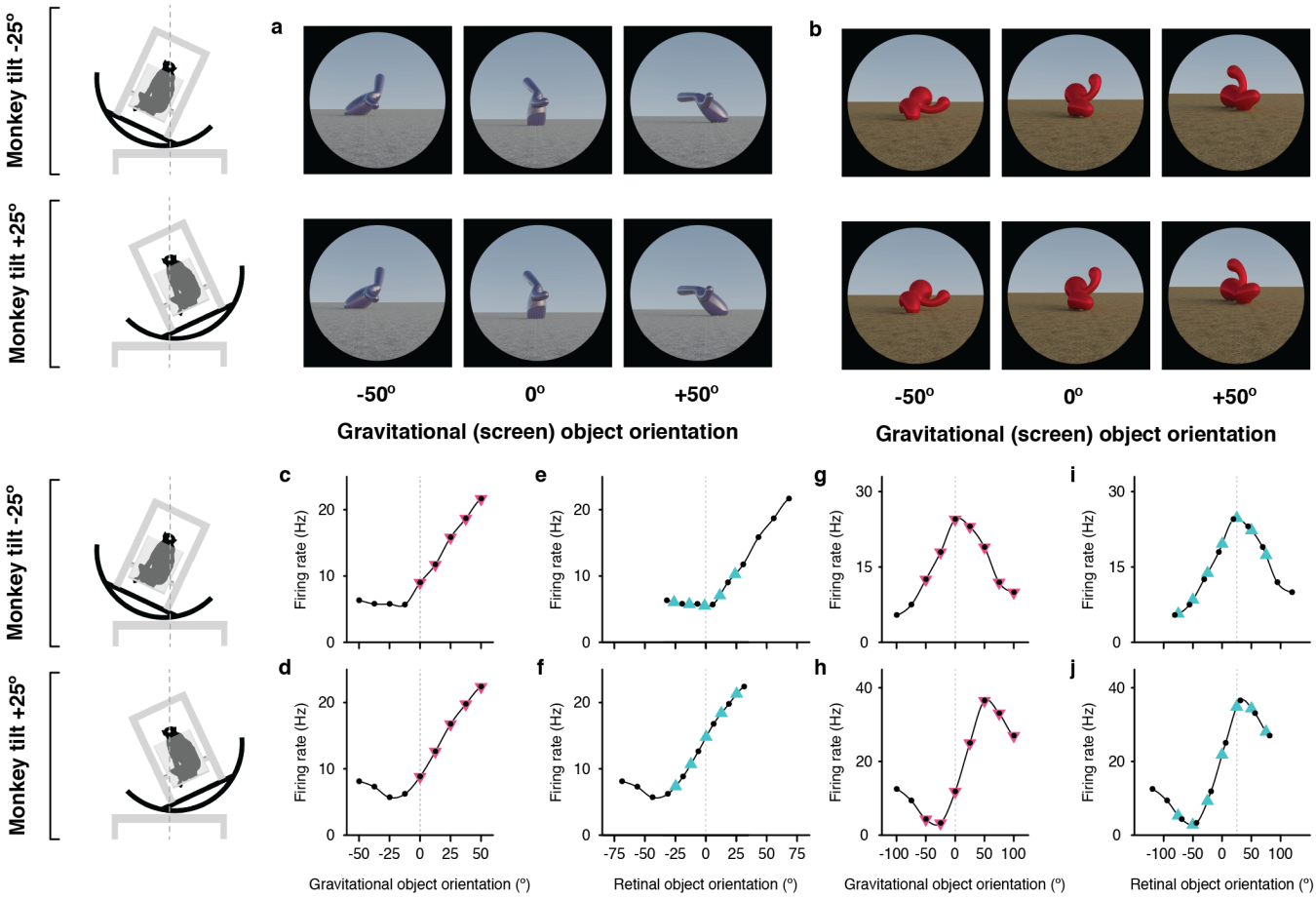


Figure S2. Example neurons tuned in gravitational space and retinal space. (a,b) Stimuli demonstrating example object orientations used to study the two IT neurons. The orientation discovered in the genetic algorithm experiments is arbitrarily labeled 0°. The two monkey tilt conditions are diagrammed at left. (c,d) Responses of the example IT neuron studied with the stimuli shown in (a), as a function of object orientation on the screen and thus with respect to gravity, across a 100° orientation range, while the monkey was tilted –25° (c) and 25° (d). Response values are averaged across the 750 ms presentation time and across 5 repetitions and smoothed with a boxcar kernel of width 50° (3 orientation values). For this neuron, object orientation tuning remained consistent in screen/gravity space across the two tilt conditions. Other details as in Fig. 1. (e,f) The same data plotted against orientation on the retina, corrected for 6° counter-rolling of the eyes in each tilt condition. Due to the shift produced by ocular counter-rolling, these comparison values were interpolated between tested screen orientations using a Catmull-Rom spline. Since orientation tuning was consistent in gravitational space, the tuning functions are shifted right or left by about 20° each. (g,h) Responses of a different example IT neuron studied with the stimuli shown in (b), as a function of object orientation on the screen and thus with respect to gravity, across a 100° orientation range, while the monkey was tilted –25° (c) and 25° (d). In this case, the tuning peak was shifted about 40°, in the direction expected for orientation tuning in retinal space. (j,k) The same data plotted against orientation on the retina, corrected for 6° counter-rolling of the eyes in each tilt condition. The correspondence between curves in (j) and (k), with peaks at near 0°, is consistent with orientation tuning in retinal space.