

Self-motion and the perception of stationary objects

Mark Wexler Francesco Panerai Ivan Lamouret Jacques Droulez

*Laboratoire de Physiologie de la Perception et de l'Action
Collège de France, 11 pl. Marcelin Berthelot, 75005 Paris, France*

September 2, 2000

One of the ways we perceive shape is through seeing motion [1, 2, 3]. Visual motion may be actively generated (for example, in locomotion), or passively observed. In the study of how we perceive 3D structure from motion (SfM), the non-moving, passive observer in an environment of moving rigid objects has been used as a substitute [1] for an active observer moving in an environment of stationary objects; the ‘rigidity hypothesis’ has played a central role in computational and experimental studies of SfM [4, 5]. Here we demonstrate that this substitution is not fully adequate, because active observers perceive 3D structure differently from passive observers, despite experiencing the same visual stimulus: active observers’ perception of 3D structure depends on extra-visual self-motion information. Moreover, the visual system, making use of the self-motion information treats objects that are stationary (in an allocentric, earth-fixed reference frame) differently from objects that are merely rigid. These results show that action plays a central role in depth perception, and argue for a revision of the rigidity hypothesis to incorporate the special case of stationary objects.

A moving observer’s retinal image of a stationary object undergoes deformation that depends both on the observer’s head motion and on the three-dimensional structure of the object. Rogers and Graham [6, 7] pioneered the study of actively produced motion parallax, and found no effect of self-motion on observers’ performance when compared to that of a non-moving observer receiving similar optic information. Other studies have found that self-motion helps to resolve discrete symmetries in optic flow [8, 9], or to decrease integration times in SfM [10].

In the first experiment we tested extra-retinal contributions to the extraction of depth from motion by means of a cue-conflict paradigm, in which motion parallax cues to 3D structure were weighed against conflicting linear perspective (i.e., the assumption that lines nearly parallel or perpendicular in the image are actually parallel or perpendicular in 3D space). The observer saw a planar 3D grid in motion, and provided an estimate of its ‘tilt’ (i.e., the direction of its normal relative to the frontoparallel plane [11]). Motion parallax could be actively produced or passively observed. In the active case, parallax was due to the observer’s head movements around a virtual object; in the passive case, the observer remained still while watching a replay of the optic flow from a preceding active trial. The

tilt of the plane defined by perspective cues differed from the tilt defined by motion cues by $\Delta T = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ or 180° . Further details and examples of stimuli are shown in Fig. 1.

In their tilt judgments, subjects could always ignore a given cue, switch between cues, or base their responses on a weighted average of the two cues—although theoretical considerations suggest that for conflicts as large as ours, cue averaging would not be optimal [12]. If on a particular trial the observer relied on the motion cue alone, he or she would perceive a spatially irregular structure that contradicts laws of linear perspective [13], but one that undergoes rigid 3D motion. If the observer relied on the perspective cue alone, he or she would perceive a structure that is spatially regular (on the average), undergoing deformation in time and thus violating the rigidity assumption that is supposed to underlie SfM [4, 14, 15, 5] (but see [16, 17, 18, 19])—as in the well-known Ames window phenomenon [20, 21, 22].

Tilt responses in Experiment 1 are shown in Fig. 2. If the response is in accordance with motion cues, we would expect a peak at 0° (all angles are defined with respect to motion tilt); if the response is in accordance with perspective cues, the peak would be at ΔT . Furthermore, in the motion-cue case, an ‘inverted’ response of 180° (or, equivalently, -180°) is possible. This is due to the approximate symmetry of optic flow under the simultaneous transformation of angular speed $\omega \mapsto -\omega$ and tilt $T \mapsto T + 180^\circ$, which is only exact in the case of parallel projection; in our case, the solution with tilt $T + 180^\circ$ is not perfectly rigid [8, 9]. As can be seen from Fig. 2, the responses are based on a multimodal mixture of perspective and motion cues (the multimodality is present in individual subject data), and, as predicted by the model of Landy et al. [12] for large conflicts, not on cue averages.

The main effect of self-motion is on the rel-

ative strength of the motion and perspective cues: *responses in accordance with motion cues are more frequent in the active than in the passive condition*. In order to quantify this self-motion effect, we counted trials with responses based on motion cues, and those with responses based on perspective cues. The effect of self-motion on cue selection can be seen clearly (see Fig. 2b). A similar effect of self-motion was predicted by Gibson [23] in discussing the Ames window (see [24] for related work).

One way to account for the preference for motion cues in the active case is to suppose that the visual system incorporates extra-retinal self-motion information in the process of extracting 3D structure. In the mathematical analysis of the shape-from-motion problem [14], for instance, this would lead to a linearisation of an otherwise (in the passive case) nonlinear system of equations, possibly resulting in greater precision and less variation in perceived 3D structure [25]. Greater reliability of motion depth cues in the active case could enhance their frequency relative to that of perspective cues, as compared to the passive case [12].

Such a ‘precision’ hypothesis would predict that responses in the no-conflict trials of Experiment 1 would be less dispersed in the active than in the passive condition. This is not the case, as can be surmised from the data shown in Fig. 2: the peaks about $T = 0$ are not significantly wider in the passive than in the active conditions. Quantitatively, we fitted the distributions in Fig. 2a with sums of Gaussians centered about $T = 0$ and $T = \Delta T$, and compared the widths of the $T = 0$ peaks using a two-way active/passive $\times \Delta T$ analysis of variance (for $\Delta T > 0^\circ$). Mean widths were 27.3° and 27.1° for the active and passive cases, respectively ($F_{1,7} = 0.01, p > 0.9$). The similarity of the *precision* in the responses to 3D structure in active and passive vision echoes

previous results [6, 9]. Our data, therefore, do not support the precision hypothesis.

A crucial difference between what subjects perceive in the active and passive conditions are *spatial* attributes of the object in an allocentric reference frame. When the observer utilises motion depth cues at the expense of conflicting perspective, he or she perceives a *rigid* 3D object. In the active case, this rigid object is also *stationary* in an allocentric, earth-fixed reference frame, whereas in the passive case, the object specified by motion cues is no less rigid but undergoes movement in space [26]. In principle, stationarity—as opposed to rigidity—is impossible to *reliably* determine from optic information alone. (Physiologically, the visual system may sometimes be fooled into judging stationarity from large-angle optic flow, as in the case of vection, but in general—and in the case of our paradigm—extra-visual information contributes to the perception of object stationarity and self-motion [27].) We propose that motion cues are enhanced in the active condition over the passive condition *because* in the former case they lead to percepts that are not only rigid, but also stationary—the *stationarity hypothesis*. On the other hand, it could simply be that motion cues are enhanced for an observer in motion. There is no way to distinguish the stationarity and the ‘motion-enhancement’ hypotheses in Experiment 1, where moving observers always perceive stationary objects from motion cues.

We tested the stationarity hypothesis in a second experiment, which used the same active/passive cue-conflict paradigm as Experiment 1, but where on some active trials (‘twist’ trials), the virtual object was not stationary but underwent oscillations about a horizontal axis. In the absence of cue conflict, in the twist trials active observers veridically perceived an object undergoing rigid oscillatory motion, in synchrony with their own head movements. The stationarity hypothesis would predict that

in active, non-stationary trials the utilisation of motion cues be reduced relative to active, stationary trials. On the other hand, if the enhancement of motion cues in the active condition of Experiment 1 were due to mere *presence* of self-motion, we would expect no effect of non-stationarity.

As can be seen from Fig. 3a, responses on no-twist trials are similar to those in Experiment 1. In twist trials without cue conflict, there is still a sharp peak about $\Delta T = 0$, showing that subjects were able to do the task even in the presence of a non-stationary stimulus synchronised with their own motion. In twist trials with cue conflict, on the other hand, the dominance of motion cues seen in the stationary case disappears, as predicted by the stationarity hypothesis, and is replaced by approximately equal peaks around motion and perspective cues, reminiscent of the *passive* condition in Experiment 1. To quantify this result, we have calculated the relative strengths of motion and perspective cues; the resulting ratios of motion- to perspective-based trials are shown in Fig. 3b. As predicted by the stationarity hypothesis, the highest ratio was in the one case (active, no-twist) where motion cues yielded a spatially stationary object.

Our two experiments have shown two results. The first, more general result is that extra-visual, movement-related information is incorporated into visual judgments of three-dimensional structure. The visual stimulation in the active and passive conditions of Experiment 1 is the same, yet the active observer responds more frequently based on motion cues than does the passive observer. The second result builds on the first: the relative enhancement of motion cues only occurs when they indicate objects that are stationary in an allocentric reference frame. It seems that the visual system is tuned to react to stationary objects, even when their image deforms due to observer motion. A number of physiologi-

cal findings point to the existence of allocentric coding in mammalian brains [28, 29] could play a role in this process. Our results would point to a revision of the rigidity concept in SfM, and, more generally, the inadequacy of excluding observer motion in the analysis of spatial vision.

Methods

The experiments were performed in monocular viewing conditions in darkness. The 3D position of the subject’s dominant eye was measured by a custom-made mechanical precision eye tracker [30] that was sampled at 85 Hz. In active trials, head-tracker data were read by a PC computer that displayed a polar projection of a virtual object for the current eye position. Data were also stored for use in subsequent passive trials. Stimuli were displayed on a computer monitor with a vertical refresh rate of 85 Hz, and a spatial resolution of 0.3 mm.

In active trials, subjects performed lateral oscillatory head movements about a central point, whose perpendicular distance to the monitor was between 35 and 45 cm. In each trial 3 cycles were performed, with the stimulus appearing after the first half-cycle. Mean maximum displacement amplitudes about stimulus center were $19.7 \pm 7.1^\circ$ and $17.7 \pm 4.3^\circ$, and mean periods were 1.9 ± 0.5 and 2.0 ± 0.5 s in Experiments 1 and 2, respectively.

Stimuli were polar projections of a virtual 3D object, a partial grid of at most 10×10 square cells, each 1 cm in length. The grid was centered about the point on the screen closest to the subject’s initial eye position, with a slant of 45° and a tilt randomly chosen between 0 and 360° . A certain number of distinct cells were randomly chosen and removed from the grid; stimuli consisted of 5, 10, 30, 60 or 100 cells in Experiment 1, and of 10 cells in Experiment 2. Stimuli were drawn as 1-pixel-thick white lines on a black background, with a small red fixation point at the center.

In order to generate conflict stimuli, the initial grids, which had tilt T_p , underwent a projection from the subject’s initial eye position onto a plane passing through the grid center, with a slant of 45° but with a tilt T_m . The resulting virtual object was an irregular grid (similar to the ‘Ames window’ or

‘Ames chair’). (In the analysis of the results, response tilts are measured with respect to T_m , with sign defined so that so that $\Delta T = T_p - T_m \geq 0$.)

Blocks of passive trials were alternated with blocks of active trials. In passive trials, subjects experienced similar optic flow as in active trials, but without head movement. Each passive trial corresponded to a preceding active trial, in that the virtual object used to generate the stimuli was identical, and that rotations of the virtual object about its center with respect to the subject’s eye were identical in passive and active trials. Let the initial eye position be \mathbf{r}_0 and the eye position at a given moment of an active trial be \mathbf{r} (relative to the center of the virtual object used to generate the stimulus); let $\theta = \arccos[(\mathbf{r}_0 \cdot \mathbf{r})/(\|\mathbf{r}_0\|\|\mathbf{r}\|)]$ be the angle between \mathbf{r}_0 and \mathbf{r} , and let \mathbf{a} be the axis generating the rotation from \mathbf{r}_0 to \mathbf{r} (i.e., \mathbf{a} is parallel to $\mathbf{r}_0 \times \mathbf{r}$). At the corresponding moment during the passive trial, the virtual object was rotated by angle $-\theta$ about axis \mathbf{a} before being projected.

In some active trials of Experiment 2 (‘twist trials’), the retinal optic flow underwent a 90° rotation, generated as follows. First, the virtual object underwent the same rotation as the eye about its center (i.e., by angle θ about axis \mathbf{a}). Then \mathbf{r}_0 was rotated by the twist angle about \mathbf{r} to yield \mathbf{r}'_0 , and a new axis $\mathbf{a}' = \mathbf{r}'_0 \times \mathbf{r}$ calculated. Finally, the virtual object was rotated by angle $-\theta$ about the new axis \mathbf{a}' and projected. Thus if the twist angle was zero, the object would remain stationary (i.e., in an egocentric frame it rotated about an approximately vertical axis, as in Experiment 1), while for a 90° twist the object underwent the same rotations in the egocentric frame, but about an approximately horizontal axis.

Following the disappearance of the stimulus, the projection of a virtual probe object was displayed on monitor. The subjects’ task was to adjust the probe, using a joystick, to indicate the tilt of the perceived surface.

In Experiment 1, five active and five passive blocks were performed, with 100 trials in each block. In Experiment 2, one active and one passive block were performed, with 160 trials per block. Eight naive volunteers participated as subjects in each of Experiments 1 and 2.

References

- [1] H. Wallach and D.N. O’Connell. The kinetic depth effect. *Journal of Experimental Psychology*, 45:205–217, 1953.
- [2] M.L. Braunstein. Depth perception in rotating dot patterns. *Journal of Experimental Psychology: Human Perception and Performance*, 72:415–420, 1962.
- [3] G. Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:210–211, 1973.
- [4] S. Ullman. *The interpretation of visual motion*. MIT Press, Cambridge, Mass., 1979.
- [5] J.J. Koenderink. Optic flow. *Vision Research*, 26(1):161–179, 1986.
- [6] B. Rogers and M. Graham. Motion parallax as an independent cue for depth perception. *Perception*, 8(2):125–134, 1979.
- [7] B. Rogers and M. Graham. Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, 22(2):261–270, 1982.
- [8] S. Rogers and B.J. Rogers. Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics*, 52(4):446–452, 1992.
- [9] T.M. Dijkstra, V. Cornilleau-Pérès, C.C. Gielen, and J. Droulez. Perception of three-dimensional shape from ego- and object-motion: comparison between small- and large-field stimuli. *Vision Research*, 35(4):453–462, 1995.
- [10] W.J. van Damme and W.A. van de Grind. Non-visual information in structure-from-motion. *Vision Research*, 36(19):3119–3127, 1996.
- [11] K.A. Stevens. Surface tilt (the direction of slant): a neglected psychophysical variable. *Perception and Psychophysics*, 33(3):241–250, 1983.
- [12] M.S. Landy, L.T. Maloney, E.B. Johnston, and M. Young. Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, 35(3):389–412, 1995.
- [13] F. Attneave and R. Frost. The determination of perceived tridimensional orientation by minimum criteria. *Perception and Psychophysics*, 6:391–396, 1969.
- [14] H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London (B, Biological Sciences)*, 208(1173):385–397, 1980.
- [15] J.T. Todd. Visual information about rigid and non-rigid motion: a geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 8(2):238–252, 1982.
- [16] H. Wallach, A. Weisz, and P.A. Adams. Circles and derived figures in rotation. *American Journal of Psychology*, 69:48–59, 1956.
- [17] E.H. Adelson. Rigid objects that appear highly non-rigid. *Invest. Ophthalm. Visual Sci.*, 26 (Suppl.), 1985.
- [18] P. Sinha and T. Poggio. Role of learning in three-dimensional form perception. *Nature*, 384(6608):460–463, 1996.
- [19] J.E. Sparrow and W.W. Stine. The perceived rigidity of rotating eight-vertex geometric forms: extracting nonrigid struc-

- ture from rigid motion. *Vision Research*, 38(4):541–556, 1998.
- [20] A. Ames. Visual perception and the rotating trapezoidal window. *Psychological Monographs*, 65(7), 1951.
- [21] W.H. Ittelson. *The Ames demonstrations in perception*. Princeton University Press, Princeton, 1952.
- [22] V. Cornilleau-Pérès, M. Wexler, E. Marin, and J. Droulez. The perception of surface orientation in small and wide-field. *Investigative Ophthalmology and Visual Science (Supplement)*, 40, 1999.
- [23] J.J. Gibson. *The ecological approach to visual perception*. Houghton-Mifflin, Boston, 1979.
- [24] A.H. Reinhardt-Rutland. Perceiving surface orientation: Pictorial information based on rectangularity can be overridden during observer motion. *Perception*, 22:335–341, 1993.
- [25] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(333-356), 1988.
- [26] H. Ono and M.J. Steinbach. Monocular stereopsis with and without head movement. *Perception and Psychophysics*, 48(2):179–187, 1990.
- [27] J.A. Crowell, M.S. Banks, K.V. Shenoy, and R.A. Andersen. Visual self-motion perception during head turns. *Nature Neuroscience*, 1:732–737, 1998.
- [28] J. O’Keefe and L. Nadel. *The hippocampus as a cognitive map*. Oxford, 1978.
- [29] L.H. Snyder, K.L. Grieve, P. Brotchie, and R.A. Andersen. Separate body- and world-referenced representations of visual space in parietal cortex. *Nature*, 394:887–891, 1998.
- [30] F. Panerai, S. Hanneton, J. Droulez, and V. Cornilleau-Pérès. A 6-dof device to measure head movements in active vision experiments: Geometric modeling and metric accuracy. *Journal of Neuroscience Methods*, 90(2):97–106, 1999.

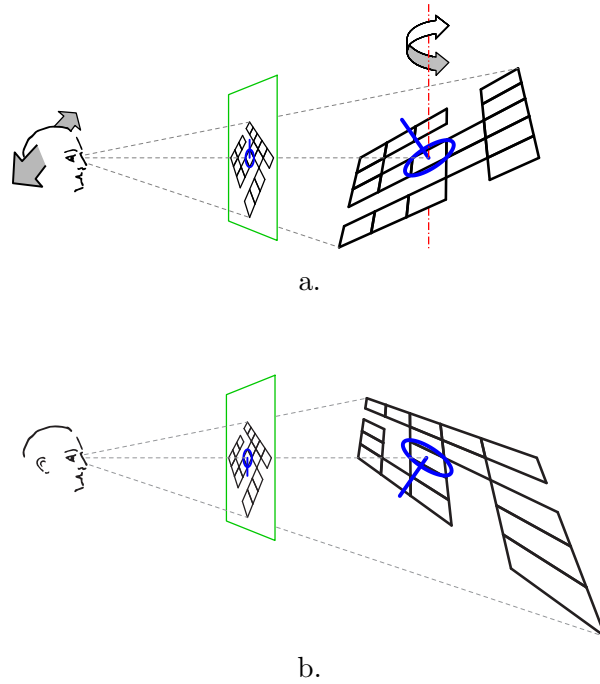
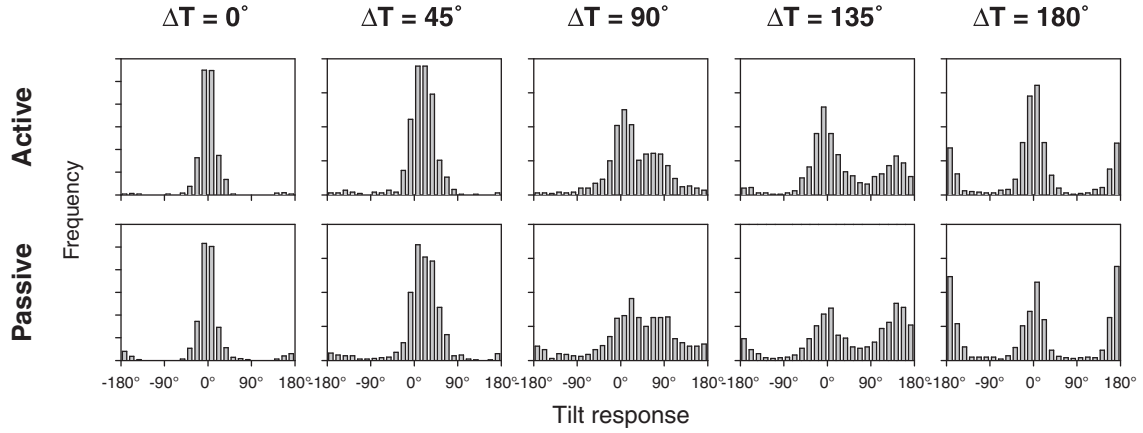
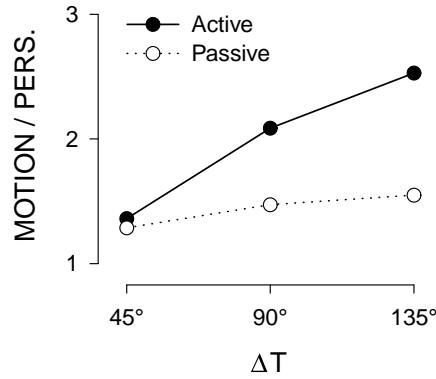


Figure 1: Cue-conflict stimuli used in the experiments. **a.** A no-conflict stimulus, where both perspective and motion cues indicate a surface tilted upwards. Both the virtual object and the proximal stimuli are shown. In the active condition, the subject performed lateral head translations while fixing an object that was stationary (Experiment 1) or oscillating (Experiment 2) in an earth-fixed reference frame; in the passive condition, the subject remained still but experienced the same optic flow as in a previous active trial. The subject's task was to indicate the perceived tilt of the grid by adjusting the virtual probe (shown in blue) so that the ellipse (line) appeared to parallel (perpendicular) to the stimulus. **b.** A conflict stimulus, generated from the no-conflict stimulus by back-projecting its 2D image onto a different plane. Although the virtual object is now tilted downwards, in its central position the projection is identical to the no-conflict stimulus; thus, perspective cues indicate upward tilt, while motion cues indicate downward tilt. In this case, the tilt conflict ΔT would be 180° .



a.



b.

Figure 2: Results of Experiment 1, averaged over subjects. **a.** Frequencies of tilt responses, in the active and passive conditions, for no tilt conflict ($\Delta T = 0^\circ$) and tilt conflicts $\Delta T = 45^\circ, 90^\circ, 135^\circ$ and 180° . Responses are adjusted so that motion tilt is always at 0° , and so that perspective tilt is positive. **b.** The ratio of motion- to perspective-based responses in the active and passive conditions, as a function of tilt conflict. Motion responses are defined as those falling within $\pm 22.5^\circ$ bins of 0° and 180° , perspective responses as those around ΔT . (Conditions $\Delta T = 0^\circ$ and 180° , where there is a partial confound between the two types of responses, are excluded.) A 3-way analysis of variance on ΔT , cue and self-motion variables showed a significant cue \times self-motion interaction ($F_{1,7} = 6.53, p < 0.05$). For conditions $0^\circ \leq \Delta T \leq 135^\circ$, 7.1% of the motion trials are inverted (i.e., cluster around $T = 180^\circ$) in the active condition, while 20.3% are inverted in the passive condition. A self-motion \times ΔT analysis of variance shows that this difference is significant ($F_{1,7} = 25.4, p < 0.01$) [8, 9].

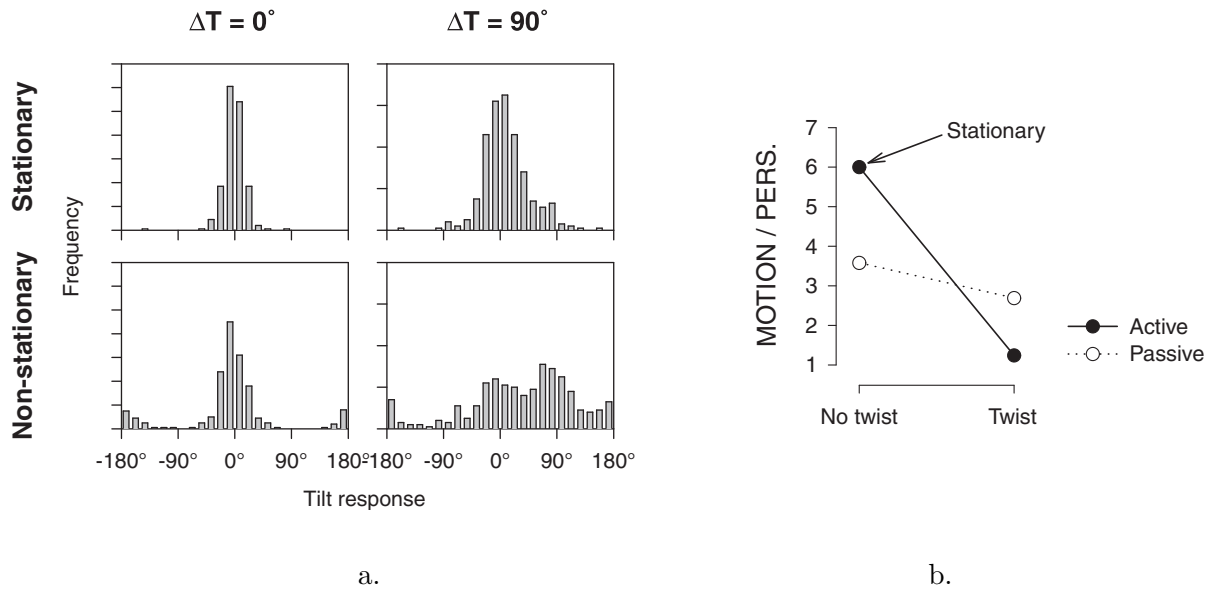


Figure 3: Data from Experiment 2, averaged over subjects. **a.** Frequencies of tilt errors in active stationary and active non-stationary conditions, with and without tilt conflict. Passive results are similar to those in Experiment 1, and, as expected, showed no effect of rotation axis (twist angle). **b.** Ratio of motion- to perspective-cue-based responses with and without twist in active and passive conditions, with tilt conflict. Responses are defined as motion-based when tilt falls within $\pm 45^\circ$ bins about 0° and 180° , as perspective-based when tilt falls within $\pm 45^\circ$ of 90° . The effect of stationarity on the active case was significant, as shown by an ANOVA on the bin counts: in the $\Delta T = 90^\circ$ case, there was an interaction between cue and twist variables ($F_{1,7} = 44.0, p < 0.001$). This effect is not due simply to the change in the axis of rotation that is introduced by the twist, as seen from a cue \times twist \times self-motion interaction ($F_{1,7} = 18.0, p < 0.01$).