

Misperception of rigidity from actively generated optic flow

Carlo Fantoni

Department of Life Sciences, University of Trieste,
Trieste, Italy

Center for Neuroscience and Cognitive Systems@UniTn,
Istituto Italiano di Tecnologia, Rovereto, Italy



Corrado Caudek

Department of Psychology, University of Florence,
Firenze, Italy



Fulvio Domini

Center for Neuroscience and Cognitive Systems@UniTn,
Istituto Italiano di Tecnologia, Rovereto, Italy

Department of Cognitive and Linguistic Sciences,
Brown University, Providence, RI, USA



It is conventionally assumed that the goal of the visual system is to derive a perceptual representation that is a veridical reconstruction of the external world: a reconstruction that leads to optimal accuracy and precision of metric estimates, given sensory information. For example, 3-D structure is thought to be veridically recovered from optic flow signals in combination with egocentric motion information and assumptions of the stationarity and rigidity of the external world. This theory predicts veridical perceptual judgments under conditions that mimic natural viewing, while ascribing nonoptimality under laboratory conditions to unreliable or insufficient sensory information—for example, the lack of natural and measurable observer motion. In two experiments, we contrasted this optimal theory with a heuristic theory that predicts the derivation of perceived 3-D structure based on the velocity gradients of the retinal flow field without the use of egomotion signals or a rigidity prior. Observers viewed optic flow patterns generated by their own motions relative to two surfaces and later viewed the same patterns while stationary. When the surfaces were part of a rigid structure, static observers systematically perceived a nonrigid structure, consistent with the predictions of both an optimal and a heuristic model. Contrary to the optimal model, moving observers also perceived nonrigid structures in situations where retinal and extraretinal signals, combined with a rigidity assumption, should have yielded a veridical rigid estimate. The perceptual biases were, however, consistent with a heuristic model which is only based on an analysis of the optic flow.

Introduction

It is commonly assumed that the visual system is optimal insofar as it estimates with maximum precision and accuracy the veridical metric properties of an object's 3-D structure, given sensory signals (Clark & Yuille, 1990; Landy, Maloney, Johnston, & Young, 1995; Ernst, Banks, & Bühlhoff 2000; Ernst & Bühlhoff, 2004; Kersten, Mamassian, & Yuille, 2004; Knill & Pouget, 2004; Knill, 2007; Landy, Banks, & Knill, 2011).

Within such a framework, suboptimal behavior in a laboratory experiment can be ascribed to a lack of sufficient sensory information due to the impoverished stimulus conditions under which the system is forced to operate. In principle, a model that can predict optimal behavior under normal operating conditions should also allow a quantitative prediction of suboptimal performance. Another possibility is that veridical reconstruction is not the goal that the visual system has evolved to pursue (Adelson & Pentland, 1996). Instead, the system merely maintains a calibration between a perceptual space that the observer experiences and the motoric interactions the observer engages in that perceptual space. In this framework, the perceptual space that the brain creates is neither optimal nor veridical in the standard sense. Instead, the visual space that forms the basis for perceptual judgments derives from idiosyncratic and mandatory mappings from sensory information to perceptual experience; in standard computational terms, the system merely follows certain heuristics. Under this model, the

Citation: Fantoni, C., Caudek, C., & Domini, F. (2014). Misperception of rigidity from actively generated optic flow. *Journal of Vision*, 14(3):10, 1–22, <http://www.journalofvision.org/content/14/3/10>, doi:10.1167/14.3.10.

veridicality of perceptual space is never guaranteed, and actual perceptual judgments can exhibit large but predictable biases, in both impoverished and natural stimulus conditions. However, the efficient interaction of the observer with the external world is underwritten by a constant calibration and recalibration between visual space and motoric actions on the basis of constant feedback from both visual and tactile/proprioceptive senses (Vishwanath, 2013; Vishwanath & Hibbard, 2013; Volcic, Fantoni, Caudek, Assad, & Domini, 2013).

In this article we examine the problem of how humans estimate 3-D structure from optic flow information. The standard veridical-reconstruction model (which we refer to as the *optimal-observer model*) makes two major assumptions in order to make this problem computationally tractable; specifically, the ambiguity in sensory signals is resolved through the use of (a) extra retinal information regarding egomotion and (b) an assumption of the rigidity of the external world. We show here that perceived 3-D structure is inconsistent with this model and that neither of the two assumptions appears to be used in deriving 3-D structure from optic flow. Instead, we show that perceptual judgments are consistent with a model in which 3-D structure is directly derived from optic flow following certain simple heuristics, which correctly predict large biases away from veridicality, even under stimulus conditions that mimic natural full-cue conditions.

Structure and motion of a two-plane configuration: A case study

Consider a 3-D structure composed of two planar surfaces slanted about a common vertical axis (Figure 1a) and an observer moving with planar motion (rotations and translations within the horizontal plane) while directing her gaze at the center of the structure. The observer's motion induces a relative rotation ω_r of the two-plane configuration in a viewer-centered coordinate frame, which generates an optic flow pattern carrying information about the 3-D structure (Koenderink & van Doorn, 1975, 1978; Koenderink, 1986; Braunstein & Tittle, 1988). The optic flow information, however, is not sufficient for an accurate estimate of the metric structure of the two-plane configuration. Each surface projects a velocity gradient, termed deformation (*def*)—in this particular case, the rate of change of retinal velocity magnitudes along the horizontal dimension (LITER, Braunstein, & Hoffman, 1993; Domini, Caudek, & Proffitt, 1997; Caudek & Domini, 1998; Domini, Caudek, Turner, & Favretto, 1998; LITER & Braunstein, 1998; Domini & Caudek, 1999). The velocity gradient is ambiguous,

since *def* specifies infinite combinations of surface slant σ and relative rotation ω_r , being $def = \omega_r \tan \sigma$. In Figure 1b, it can be seen how surfaces with different slants viewed by observers moving by different amounts generate the same *def*. In theory, an accurate measurement of the pattern of retinal accelerations would eliminate this ambiguity, but it is known that this measurement is highly unreliable and therefore useless (Norman & Todd, 1993; Domini & Braunstein, 1998; Eagle & Hogervorst, 1999; Hogervorst & Eagle, 2000).

A commonly held assumption is that in order to solve this ambiguity, the visual system combines optic flow information with extraretinal and proprioceptive information about the observer's motion (Ono & Steinbach, 1990; Rogers & Rogers, 1992; Dijkstra, Cornilleau-Pérès, Gielen, & Droulez, 1995; Wexler, Lamouret, & Droulez, 2001; Wexler, Panerai, Lamouret, & Droulez, 2001; Panerai, Cornilleau-Pérès, & Droulez, 2002; Peh, Panerai, Droulez, Cornilleau-Pérès, & Cheong, 2002; van Boxtel, Wexler, & Droulez, 2003; Wexler, 2003; Wexler & van Boxtel, 2005; Jaekl, Jenkin, & Harris, 2005; Colas, Droulez, Wexler, & Bessière, 2007; Dyde & Harris, 2008; Dupin & Wexler, 2013). Since surfaces in the world are mostly stationary, an accurate measurement of the observer's egomotion provides a direct estimate of ω_r and thus of σ from *def*, since $\sigma = \tan^{-1}(def/\omega_r)$. Therefore, an optimal-observer model combining sensory information (retinal and extraretinal) with a *prior for stationarity* is able to accurately derive the metric structure of the two-plane configuration. Note that a prior for stationarity also implies a *prior for rigidity* (Ullman, 1979; Grzywacz & Hildreth, 1987). Since static surfaces do not change their structure during the observer's motion, most 3-D transformations generated by egomotion are necessarily rigid. For example, in the case of the two-plane configuration, if the surfaces are static in the world, then the dihedral angle between the two surfaces does not change either. In the remaining part of this article, we will therefore refer to this prior as to the *stationarity/rigidity prior*.

Nevertheless, empirical data indicate that human observers do not always perceive rigid transformations, even under conditions in which, in principle, they should. For example, Domini and colleagues (1997) showed that a static observer viewing a rotation of the two-plane configuration of Figure 1 typically perceives a nonrigid transformation: The dihedral angle between the two surfaces is seen as changing during the rotation. Specifically, the surface with the larger slant is perceived as rotating faster than the surface with the smaller slant. This result seems to be in contrast with the hypothesis that the human visual system embeds a prior for rigidity in its interpretation

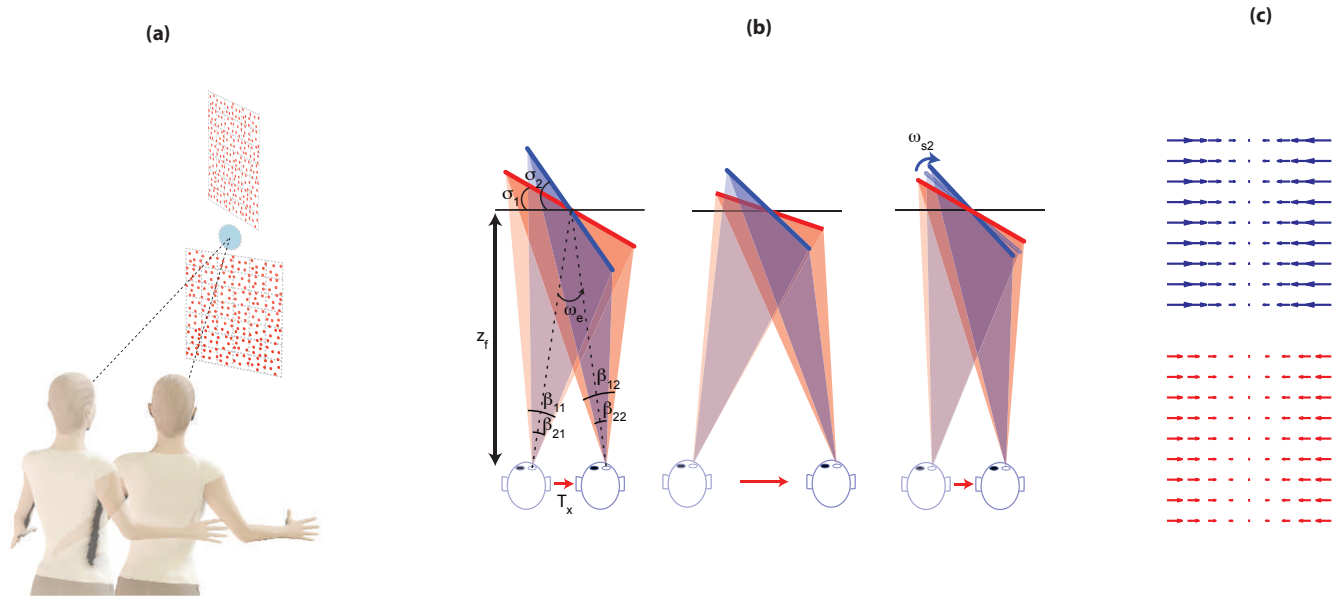


Figure 1. Planar motion of an observer relative to a two-plane configuration. (a) An observer moves rightward while directing her gaze (dashed lines) at the center of a structure composed of two random-dot planar surfaces slanted about the same vertical axis and both tilted in the same direction (Experiment 1). (b, left) Bird's-eye view of the two surfaces, having slants σ_1 and σ_2 , viewed at a distance z_f by an observer moving laterally with speed T_x . The lateral movement induces a rotation of the entire structure, with respect to a viewer-centered reference frame, of angular speed ω_e . Thus the angular speed of the two surfaces is $\omega_{r1} = \omega_{r2} = -\omega_e$. Due to this rotation, the retinal projection of the texture elements of each surface changes in time to generate an optic flow pattern (c). This particular optic flow is entirely defined by the *rate of compression* of the texture pattern, which is the gradient of the optic flow, defined also as *deformation (def)*. At a time t_1 , the surfaces subtend at the eye angles β_{11} and β_{21} , which at a time t_2 , after the lateral translation, become β_{12} and β_{22} . The *deformation* of the two optic flows is given by the ratio between the difference $\beta_{12} - \beta_{11}$ (red) and the time interval for surface 1, and the ratio between $\beta_{22} - \beta_{21}$ (blue) and the time interval for surface 2. The same deformations can be produced by two surfaces with different orientations viewed by an observer moving at a different speed (b, center), or even by a nonrigid configuration, where one surface rotates with respect to the other surface (by an amount ω_{s2}) during the observer's lateral motion (b, right). In this case, the angular speed of this surface in a viewer-centered reference frame is $\omega_{r2} = \omega_{s2} - \omega_e$.

of the optic flow (Glennerster, Tcheang, Gilson, Fitzgibbon, & Parker, 2006). Alternatively, these results could be explained by a model that is optimal but predicts suboptimal performance for stimulus conditions providing insufficient sensory data (Colas et al., 2007).

What is critical to our previous study (Domini et al., 1997) is that viewers looked from a static vantage point at projections of rotating surfaces. Thus, they lacked the potential information about the relative motion between them and the distal objects, provided by extraretinal and proprioceptive signals, that is always available when observers move in an otherwise mostly static environment.

In the following section we will show that the optimal-observer model, which utilizes sensory information about the observer's egomotion and embeds a prior for stationarity/rigidity, mistakenly assigns a nonrigid interpretation to rigid stimuli viewed by a static observer and thus predicts our previous findings.

For an active observer who self-generates the optic flow, the model predicts an opposite bias: If the instantaneous optic flow is compatible with a rigid transformation, it assigns a rigid interpretation to the stimuli, even when the actual transformation is nonrigid (i.e., rigid and nonrigid rotations cannot be disambiguated).

The two experiments described here were designed to test the predictions of the optimal-observer model for an active observer. Contrary to its predictions, we show that an active observer is subject to the same biases as a passive observer, casting serious doubts on two widely held assumptions: that a moving observer has access to egomotion information for the interpretation of the optic flow, and that the rigidity and stationarity priors are biologically relevant. Instead, a *heuristic-observer model*, which ignores both egocentric information and a prior for stationarity/rigidity, perfectly predicts the empirical results.

Optimal Bayesian estimation of the relative rotation between two planar surfaces

The two-plane configuration (Figure 1a) undergoes a nonrigid transformation when the angular velocities of the two surfaces with respect to the same arbitrary reference system are different. In an egocentric reference system, this means that $\omega_{r1} \neq \omega_{r2}$. The sensory information available for estimating ω_{r1} and ω_{r2} is the velocity gradients (def_1 and def_2) and an estimate of the observer's egomotion. The latter can also be described in terms of angular velocity, since when the viewpoint shifts along the horizontal axis with linear speed T_x , as in our experiments, it produces a rotation of the surface relative to the observer, in a head-centric coordinate frame, of angular velocity $\omega_e = T_x/z_f$, where z_f is the fixation distance (see Figure 1b). An estimate of the angular velocity $\hat{\omega}_e$ can be provided by extraretinal and proprioceptive signals (Buizza, Leger, Droulez, Bertoz, & Schmid, 1980; Ferman, Collewyn, Jansen, & van den Berg, 1987; Nawrot, 2003; Nawrot & Joyce, 2006; Gu, DeAngelis, & Angelaki, 2007; Bennur & Gold, 2008; Gu, Angelaki, & DeAngelis, 2008; Liu & Angelaki, 2009; Nawrot & Stroyan, 2012).

The main problem with estimating ω_{r1} and ω_{r2} from def_1 , def_2 , and $\hat{\omega}_e$ is that a surface rotation independent of the observer can also contribute to ω_{r1} and ω_{r2} (Figure 1b). If, during a head movement, the surface also rotates (with angular velocity ω_s), then the relative angular velocity between surface and observer is $\omega_r = \omega_s - \omega_e$ (Figure 1b, right panel). The problem of estimating the relative rotation of the two surfaces is therefore ill posed, since the velocity gradients corresponding to each surface (def_1 and def_2) define a set of two equations and four unknowns (ω_{r1} , ω_{r2} , σ_1 , σ_2):

$$\begin{aligned} def_1 &= \omega_{r1} \tan(\sigma_1) = (\omega_{s1} - \omega_e) \tan(\sigma_1) \\ def_2 &= \omega_{r2} \tan(\sigma_2) = (\omega_{s2} - \omega_e) \tan(\sigma_2) \end{aligned} \quad (1)$$

Given its nondeterministic nature, this problem can be defined in probabilistic terms with a normative Bayesian model:

$$\begin{aligned} P(\omega_{r1}, \omega_{r2} | def_1, def_2, \hat{\omega}_e) &= \\ \frac{P(def_1, def_2 | \omega_{r1}, \omega_{r2}) P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)}{P(def_1, def_2)} \end{aligned} \quad (2)$$

where $P(def_1, def_2 | \omega_{r1}, \omega_{r2})$ is the likelihood term, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ the prior, and $P(def_1, def_2)$ a normalizing constant. Since the likelihood term is the product of two likelihoods, then

$$\begin{aligned} P(\omega_{r1}, \omega_{r2} | def_1, def_2, \hat{\omega}_e) &\propto \\ P(def_1 | \omega_{r1}) P(def_2 | \omega_{r2}) P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e) \end{aligned} \quad (3)$$

In summary, according to this model, the probability of observing a specific combination of angular rotations for the two surfaces (ω_{r1} and ω_{r2}) depends on both the gradient of the optic flow produced by each surface (through the likelihood terms) and an additional term that, as we will see shortly, restricts the possible solutions on the basis of egomotion information ($\hat{\omega}_e$) and a priori assumptions about the rotation of the surfaces in the world.

Domini and Caudek (2003) postulated that the perceptual interpretation of the optic flow is solely based on the velocity gradients. According to that proposal, the perceptual solution is the one maximizing the product of the likelihood terms (i.e., maximum likelihood estimate). Since def is ambiguous, this solution is never veridical (see Appendix A for the mathematical proof). No matter what the physical 3-D rotation of a surface is, the most likely rotation is a monotonically increasing function of def . Thus, if the two-plane configuration undergoes a 3-D rigid rotation but each surface projects a different value of def ($def_1 \neq def_2$), the predicted perceptual interpretation is that of a nonrigid structure. Conversely, if the two surfaces rotate by different amounts but produce the same value of def ($def_1 = def_2$), the nonrigid rotation is predicted to be perceived as rigid. Indeed, these predictions are compatible with the findings of Domini et al. (1997) showing that the two-plane configuration is perceived as rigid only if the two surfaces generate the same value of def .

The term $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ of Equation 3 can change the maximum likelihood estimate (MLE), since it incorporates both information about the observer's egomotion and the stationarity/rigidity prior (see Appendix B for the mathematical proof). On the basis of how precisely the motion of the observer is measured and how strong the stationarity/rigidity prior is, we can discriminate between two possible models:

1. The *optimal-observer* (OO) model includes a strong stationarity/rigidity prior and a precise measurement of the observer's egomotion. For an active observer, this model correctly derives a rigid interpretation whenever the optic flow is compatible with such an interpretation. Most importantly, in the unnatural experimental condition in which the observer is static, the model predicts suboptimal performance, incorrectly assigning nonrigid interpretations to rigid stimuli. For a passive observer, this model substantially mimics the MLE interpretation, thus predicting our previous findings [see "Optimal observer (OO)" in Appendix C].
2. The *heuristic-observer* (HO) model predicts the behavior of an observer who has access to noisy egomotion information, and embeds a weak stationarity/rigidity prior. It can be shown that this model makes predictions for both active and passive

observers that are qualitatively identical to those of the MLE model [see “Heuristic observer (HO)” in Appendix C].

Note that the HO model is different from one based on noisy egomotion information but embedding a strong stationarity/rigidity prior—that is, a model with an observer insensitive to egomotion [OIE; see “Observer insensitive to egomotion (OIE)” in Appendix C]. Such a model indeed predicts, for a passive observer, a rigid solution and must therefore be discarded as not biologically plausible.

For passive observers, both the HO and OO models predict the results observed by Domini et al. (1997): Two rigidly rotating surfaces with different slants are perceived as nonrigid. The goal of our study is to distinguish between the HO model and the OO model. *Are biases in a passive observer’s judgments of rigidity due to a visual system that ignores both the stationarity/rigidity prior and the observer’s egomotion (HO model), or do they take place because passive viewing of the optic flow forces the visual system to operate in nonideal conditions (OO model)?*

We addressed this question by studying the perception of rigidity for active observers. If the predictions of the OO model are correct, then an active observer should perceive two surfaces projecting different deformations, like in the example discussed so far, as rigid. Moreover, we will see in the next section that the OO model predicts for an active observer a bias towards rigidity for both rigid and nonrigid transformations. Instead, the HO model predicts that an active observer assigns a rigid interpretation to the optic flow only when the two surfaces project the same deformations.

Experiments 1 and 2

Rationale and predictions

In the two experiments, observers viewed the two-plane configuration in two viewing conditions: active and passive. In the active viewing condition, the observers moved laterally and the optic flow was entirely or partly generated by their movement. In the passive viewing condition, the same optic flow was replayed and viewed from an immobile viewpoint.

During the movement of the active observer, one of the two surfaces, referred to as the *reference surface*, was always stationary in an allocentric reference frame ($\omega_{st} = 0$). The other surface, referred to as the *target surface*, rotated in depth by an amount proportional to the observer’s own motion ω_e . Thus, the amount of rotation of the target surface was defined by a gain

factor g so that $\omega_{st} = g\omega_e$. The factor g was made to vary within the range $[-1, +1]$. At the extreme values, $g = +1$ caused an allocentric rotation ω_{st} of the target surface of equal magnitude but opposite direction to the relative rotation ω_e induced by the observer’s translation; whereas $g = -1$ caused an allocentric rotation ω_{st} of the target surface equal to the relative rotation ω_e . Only $g = 0$ specified a static target surface and, therefore, a rigid structure.

In Experiment 1 we asked the observers to judge which surface rotated faster (the target surface or the reference surface), whereas in Experiment 2 the observers judged whether the dihedral angle between the two surfaces expanded or shrank. For both tasks, a chance-level performance indicated that the observers perceived a rigid structure, since in such case neither surface rotated faster (Experiment 1) and the dihedral angle neither shrank nor expanded (Experiment 2). If observers reliably judged that one surface was rotating faster (Experiment 1) or that the dihedral angle shrank or expanded (Experiment 2), then it meant they perceived a nonrigid transformation.

In both experiments, the gain factor was varied through a staircase procedure devised to find (a) the point of subjective rigidity (PSR)—that is, the value of g giving rise to a chance-level performance—and (b) the just-noticeable difference (JND), that is, the smallest g variation from the PSR which gave rise to a nonrigid percept. The deformations projected by the reference and target surfaces were $def_r = -\omega_e \tan \sigma_r$ and $def_t = (g\omega_e - \omega_e) \tan \sigma_t$, respectively.

We used the model of Equation 3 to predict the perceived rotation difference between the target and reference surfaces for the range of g values tested in our experiments. To this purpose, we considered the viewing parameters used in Experiment 1 (lateral head translation of 160 mm performed at 125 mm/s at a viewing distance of 668 mm), but the results of the simulation generalize to Experiment 2 as well. The input to the HO and OO models were the average values of def and the average observer’s motion ω_e measured during the experimental sessions. For the passive viewing condition it was assumed that the egomotion signal specifies an immobile observer ($\omega_e = 0$).

Figure 2 depicts the results of the simulation for the OO (right column) and HO (left column) models. The results are plotted as a function of the gain g (bottom x-axis) and the corresponding difference between the deformations of the two surfaces, $def_r - def_t$ (top x-axis). The bottom row shows the difference between the maximum a posteriori (MAP) estimates ($\hat{\omega}_{rr} - \hat{\omega}_{rt}$) of the relative angular velocities of the reference ($\hat{\omega}_{rr}$) and target ($\hat{\omega}_{rt}$) surfaces. In the top row, the same data are replotted in terms of the probability of perceiving the target surface as rotating slower than the reference

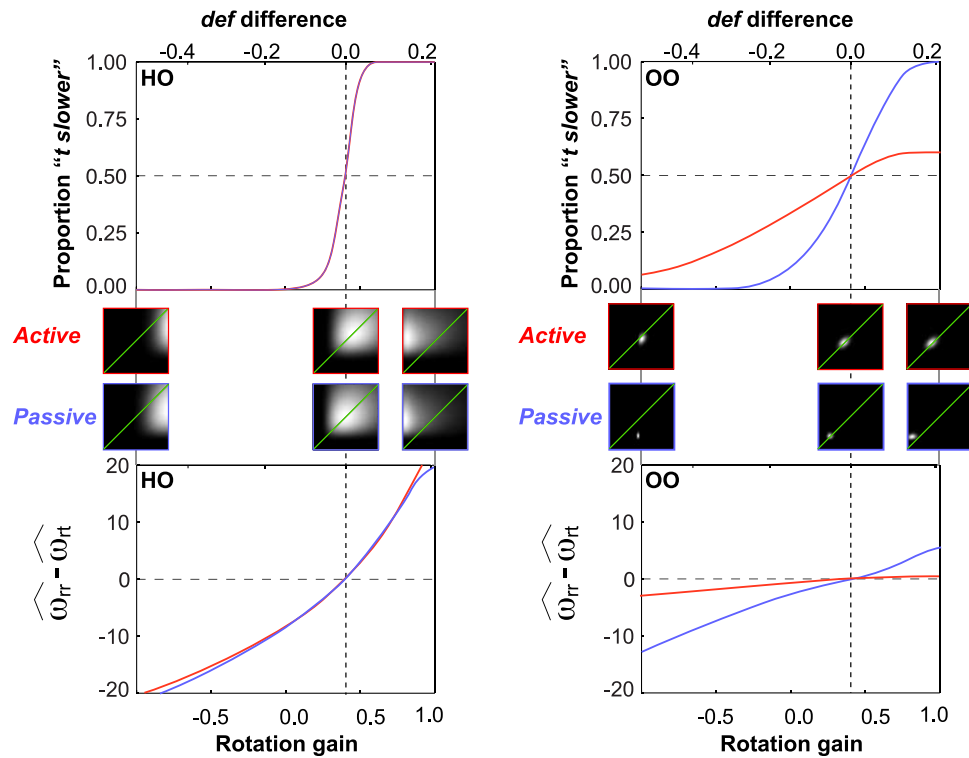


Figure 2. Predicted relative rotation between two surfaces (bottom) and the corresponding probability of perceiving the target surface as rotating slower than the reference surface (top), as a function of the rotation gain g and def difference. The HO model predicts that both a passive (red) and an active (blue) observer will perceive a nonrigid rotation whenever the def difference is detectable (top left). The reason for this behavior is that the posterior of the HO model is peaked at values of relative angular velocities for the reference (ω_{rr}) and target (ω_{rt}) surfaces that solely depend on the values of projected deformations, for both active and passive observers: The larger the def difference, the larger the predicted rotation difference ($\hat{\omega}_{rr} - \hat{\omega}_{rt}$, bottom left). The OO model (right column) makes the same prediction as the HO model for a passive observer (blue), but also predicts that an active observer (red) will mostly perceive rigid transformations ($\hat{\omega}_{rr} - \hat{\omega}_{rt} \approx 0$, bottom right), resulting in an almost flat psychometric function (top right). The center panels represent the posteriors for the HO and OO models as a function of the angular velocities of the two surfaces (ω_{rr} y-axis; ω_{rt} x-axis), with gray levels representing the probability and green lines representing rigid solutions (see Appendix C).

surface, by assuming Gaussian noise on the measurement of $\hat{\omega}_{rr} - \hat{\omega}_{rt}$.

As expected, for the passive viewer the predictions of the two models are the same (blue). The estimated angular velocity of the target surface is smaller than the estimated angular velocity of the reference surface, $\hat{\omega}_{rr} - \hat{\omega}_{rt} > 0$, when the deformation of the target surface is smaller than the deformation of the reference surface ($def_r - def_t > 0$), and vice-versa when the deformation of the target surface is larger than the deformation of the reference surface ($def_r - def_t < 0$). This prediction is in agreement with previous results (Domini et al., 1997). However, for the active observer the predictions of the two models differ substantially (red). For the active observer, the OO model predicts a bias towards a rigid interpretation, whereas the HO model predicts a similar performance for both the active and the passive observer.

General method

Participants

Twenty-one students participated in the experiments. Twelve students of the University of Trieste participated in Experiment 1, in return for course credit. Nine students of the University of Trento were paid to participate in Experiment 2; six of them performed both the active and the passive viewing phases of the experiment. All observers had normal or corrected-to-normal vision and were unaware of the purpose of the experiment. Experiments were undertaken with the understanding and written consent of each subject, with the approval of the Comitato Etico per la Sperimentazione con l'Essere Umano of the University of Trieste (for Experiment 1) and Trento (for Experiment 2), and in compliance with national legislation and the Code of Ethical Principles for

Medical Research Involving Human Subjects of the World Medical Association (Declaration of Helsinki). Each observer participated in all conditions of the factorial within-subjects design.

Apparatus

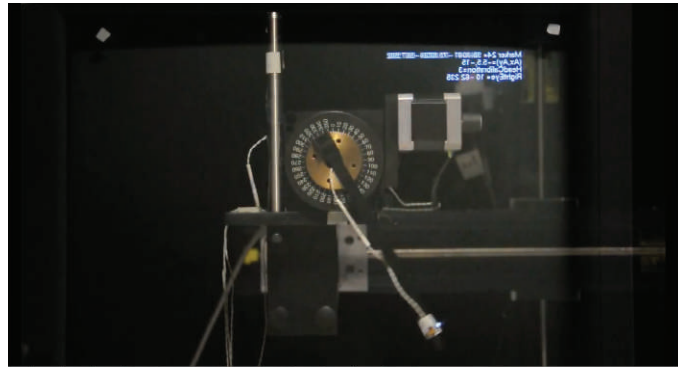
The participants' head motions were tracked in real time by an Optotrak Certus system. A Dell Precision T3400 525W (using an Intel Core 2 Extreme 5252W, QX9650, 3.00 GHz, 1333 MHz FSB, 12 MB L2 Cache) controlled the stimulus display and sampled the tracker. Three sensors on the back of the observer's head were used to calculate the x -, y -, and z -coordinates of the observer's viewpoint in order to update in real time the geometrical projection of each pair of random-dot planar surfaces, through a procedure similar to the one described by Fantoni, Caudek, and Domini (2010, experiment 3). Sampling of the head tracker was set at 360 Hz, so that the tracker latency was lower than the sample interval.

In Experiment 1, the stimuli were displayed on a Sony Trinitron Color Graphic Display GDM-F520 CRT monitor set at a resolution of 1024×768 pixels and a refresh rate of 85 Hz and driven by an nVidia Quadro 5000. In Experiment 2, the same display settings were used (1024×768 screen resolution; 100-Hz refresh rate) on a ViewSonic 9613, 19W CRT monitor driven by an nVidia Quadro FX 4600.

Displays were monocularly viewed through a high-quality front-silvered mirror placed in front of the observer's central viewing position and slanted 45° away from the monitor and the observer's interocular axis (see Fantoni et al., 2010, figure 4). The effective distance from the pupil to the center of the screen was 668 mm in Experiment 1 and 568 mm in Experiment 2.

A custom Visual C++ program supported by OpenGL Libraries and combined with Optotrak API routines was used for stimulus presentation, response recording, and data storage. The same program (a) controlled the slant of the target random-rot plane according to the observer's translation velocity and the selected rotation gain and (b) stored the relative motion between the observer and the planar surfaces in order to replay the same optic flow experienced during active viewing to the passive observer (for further specification, see Fantoni et al., 2010, appendix B).

To control whether our tracking system might artificially induce distortions of the perception of stability during observers' head movements, we measured by how much changes in the depicted virtual stimuli were delayed from changes in the position of the objects being tracked (system lag). In order to do



Movie 1. System lag test of our virtual reality environment.

that, we used an “external” measure similar to the one described by Swindells, Dill, and Booth (2000): A marker was mounted on a rod (10 cm long) sticking out of a circular turntable spinning at a velocity ω of $60^\circ/\text{s}$. A white virtual marker was rendered, ideally superimposed on the physical marker, and 40 measures of instantaneous angular displacement between the physical marker and the virtual marker θ were collected by taking photographs with a Nikon D90 (resolution 3216×2136 , ISO 6400, exposure time = $1/100$, focal length = 38 mm) centered and aligned with the turntable (Movie 1). Two levels of complexity of the graphic scene were also tested: a low level, in which just the virtual marker was displayed, and a high level, in which the marker was displayed together with a complex mesh with more than 10^6 vertices. The estimated system lag (θ/ω) was about 27.9 ± 1.26 ms and was independent of graphical complexity (26.8 ± 1.91 ms vs. 28.9 ± 1.66 ms in the low vs. high complexity condition, $t = 0.82$, $df = 38$, $p = 0.4$). According to the results of Allison, Harris, Jenkin, Jasiobedzka, and Zacher (2001), such a low system lag combined with the low head-translation velocity used in our experiments (about $22^\circ/\text{s}$) is not likely to produce artificial distortions of the perception of stability.

Displays

The displays simulated the perspective projection of two rectangular planar surfaces slanted about the same vertical axis (see Figure 1). The surfaces were defined by randomly positioned small texture elements, which projected as antialiased red dots (1×1 mm) on the image plane (the density was about 8 dots/cm²). The quasirectangular projections were separated by a vertical gap (15 mm) at the center of the screen. In the active condition, the motion of the dots on the screen was calculated in real time by tracking the observers' vantage point. The dots on each simulated planar

surface were projected onto the image plane by using a generalized perspective pinhole model, with the observer's right eye position as the center of projection. In the passive condition the same optic flow was replayed to the passive observer.

We carefully removed cues other than the optic flow that could have specified the 3-D orientation of the planar surfaces. These were the texture gradient and the projected foreshortening of the outline of the rectangular planar surfaces due to perspective projection. To do so, we determined the dot distribution using a back-projection technique so that dots were randomly distributed on the image plane, not on the simulated surface (Banks & Backus, 1998). The height of each plane's projection was 50 mm, while the width randomly varied across trials within a range of 40 to 60 mm.

The display was visible while the observer moved her head rightward. We only kept one direction of head movements, because we found in previous studies that the effect of *def* on perceived surface orientation and motion is exactly the same for the two possible directions of the translatory movement (Fantoni, Caudek, & Domini, 2012). Given the repetitive nature of the observer movement, this also facilitated the execution of the task.

The onset of the test stimulus occurred when the right eye crossed a position 80 mm eccentric to the left of the center of the screen, after the observer reversed her direction of motion. At the average velocity of 230 mm/s, the test stimulus was visible on the screen for about 0.7 s. The stimulus disappeared when the right eye crossed the eccentric position opposite to that of the stimulus onset (i.e., 80 cm eccentric to the right of the screen center). Such a lateral head shift produced a consequent variation of the visual direction of about 13.5° for the viewing distance used in Experiment 1 (668 mm) and 15.7° for the viewing distance used in Experiment 2 (558 mm). The lateral head shift thus produced a rotation ω_e about the vertical axis of the simulated planar surfaces relative to the observer of about $\pm 6.75^\circ$ and $\pm 7.85^\circ$, respectively.

The motion of the dots generated an approximately linear velocity field with velocity vectors mostly parallel to the horizontal axis of the screen. In Experiment 1, the tilt of both surfaces was equal to 180° . This surface tilt coupled with a rightward motion of the head produces an optic flow of pure horizontal compression. In Experiment 2, the tilts of the two surfaces were equal to 0° and 180° . Therefore, one surface (tilt = 180°) produced an optic flow of pure compression, whereas the other (tilt = 0°) produced an optic flow of pure expansion.

During head translation, one of the two surfaces (reference surface) was stationary while the other (target surface) rotated about the vertical axis with

angular velocity proportional to the observer's motion. The vertical position of the two surfaces was randomly selected on each trial, so that the target surface had equal probability of appearing above or below the reference surface.

The simulated slant σ_r of the reference surface was 30° in Experiment 1 and 25° in Experiment 2, and remained constant throughout a trial. The simulated slant σ_t of the target surface was, instead, coupled in real time with the observer's motion through the following equation:

$$\sigma_t = \sigma_{0t} + g\alpha \quad (4)$$

where g is the rotation gain and α the visual direction.

When the gain g was 0, the target surface was stationary with a constant slant $\sigma_t = \sigma_{0t}$. A negative gain ($g < 0$) made the surface rotate in the same direction as the observer's gaze, whereas a positive gain produced the opposite rotation. The time derivative of Equation 4 is the angular velocity of the surface, $\omega_{st} = g\dot{\alpha} = g\omega_e$. Since $\omega_{rt} = \omega_{st} - \omega_e$ is the relative rotation between the target surface and the observer, $\omega_{rt} = g\omega_e - \omega_e$. Therefore, the instantaneous deformation of the target surface is $def_t = (g\omega_e - \omega_e)\tan\sigma_t$, which on average is equal to $(g\omega_e - \omega_e)\tan\sigma_{0t}$. Thus a positive gain reduces the amount of *def* projected by the target surface, whereas a negative gain increases it. Most critically, there is a value of rotation gain for which the target surface generates the same average *def* as the reference surface (Figure 3).

The rotation gain was varied by an adaptive staircase procedure with four randomly interleaved 1-up-2-down, 1-up-1-down, 1-down-1-up, and 2-up-1-down staircases. Staircases were terminated after four reversals. In Experiment 1, the slant σ_{0t} of the target surface could take on two possible values (45° and 55°), whereas only one value (45°) was possible in Experiment 2.

In the passive vision phase, the optic flows were generated by replaying the 2-D transformations generated by the corresponding active vision trials. The stimulus thus included both *translatory* components (inversely proportional to the deviation between the observer's right-eye visual axis and the stimulus center, during the corresponding active vision trial) and *rotational* components (proportional to the three degrees of freedom of head rotations performed by the observer during the corresponding active vision trial). The translatory components of our passive displays (on average, 0.5 cm), together with the display durations (0.7 s) and display size (on average, 5°), were all small enough to prevent vection. This was confirmed by preliminary interviews with the observers.

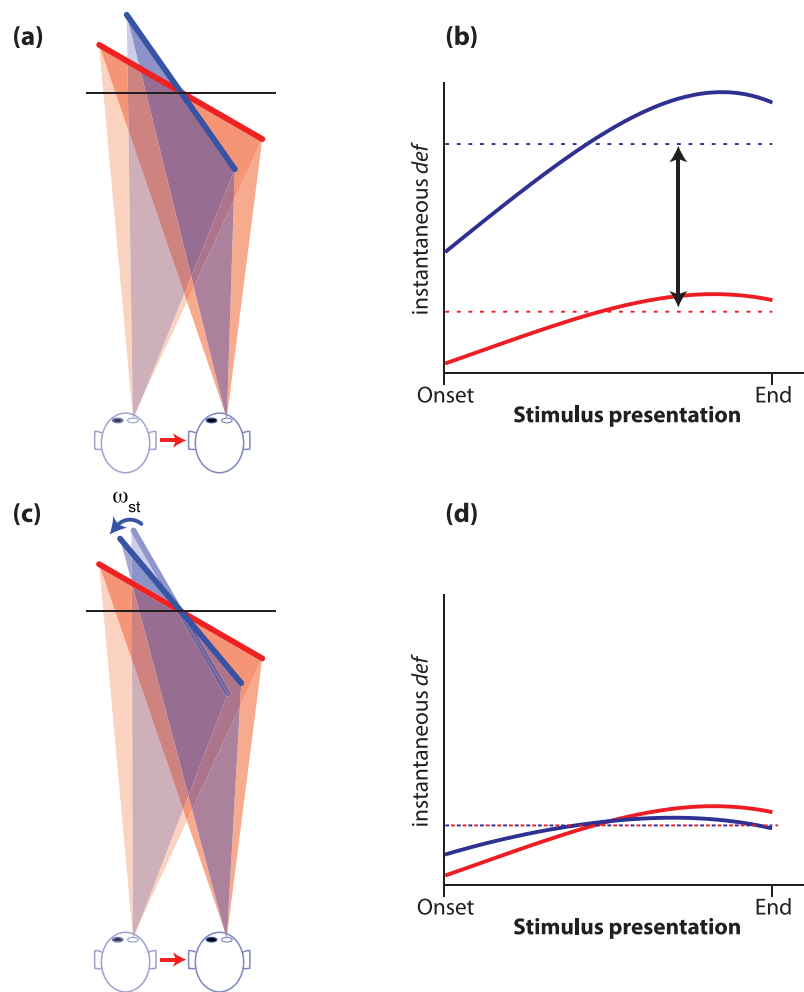


Figure 3. Instantaneous *def* of the target (blue) and reference (red) surface as a function of time. When the observer moves while looking at a rigid two-plane configuration (i.e., $g = 0$, panel a), the *def* of the target surface is always larger than the *def* of the reference surface (b), since the slant of the target surface is larger. When the target surface rotates during the observer translation by a specific amount ($g = 0.43$ in the example, panel c), its instantaneous *def* is on average the same as that of the reference surface (d).

Design

In both experiments, a 2×2 within-subject experimental design was used, with two viewing conditions (active and passive) and with either two slants for the target surface in Experiment 1 (45° and 55°) or two tilt angles for the target surface in Experiment 2 (0° and 180°).

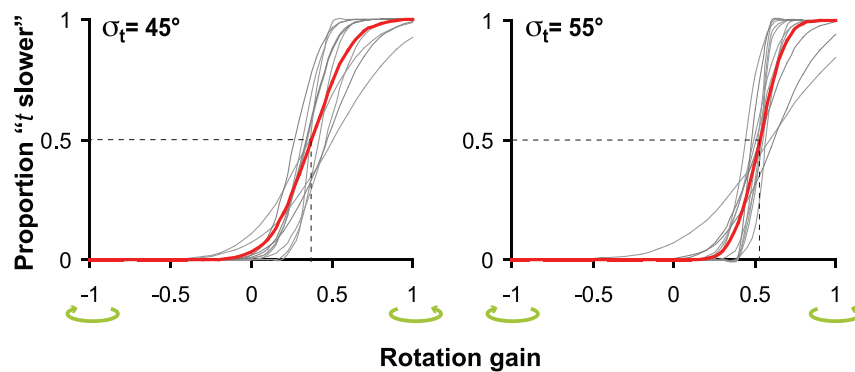
Procedure

Participants were tested individually in complete darkness, so that only the random-dot stimuli were visible during the experiment. To allow for natural head movements, head motion was not restrained. Prior to the experiment, each participant was trained to perform back-and-forth lateral head translations peaking at the required velocity of about 340 mm/s when their gaze crossed the center of the screen.

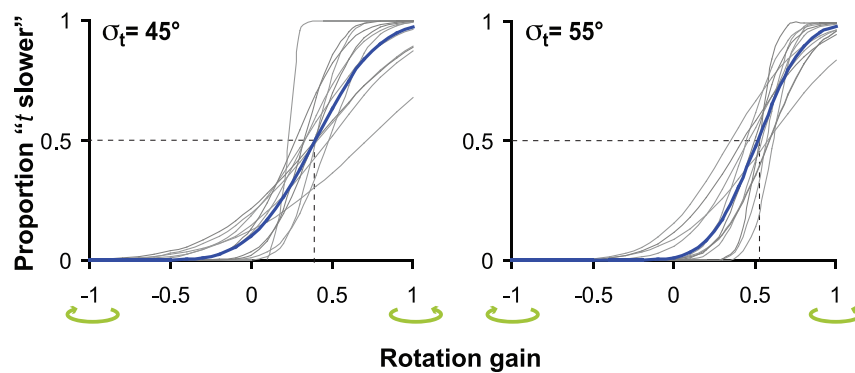
Participants were also instructed to keep fixation at the center of the screen while performing the lateral head translations. At the beginning of each trial, after the observer's gaze was aligned with the depth axis and the center of the screen, a beep was heard and the participant initiated a lateral rightward translation. Once the observer reached a lateral position of 80 mm relative to the center of the screen, a beep signaled that the head's lateral movement had to be reversed. The acoustic signal also provided feedback about the speed of the head translation: A high-pitched sound signaled a speed that was too fast and a low-pitched sound signaled a speed that was too slow.

After one and a half head oscillations at the required speed and the required head orientation (i.e., head pitch and roll were controlled in real time and required to be within the $\pm 5^\circ$ range), the stimulus appeared and remained visible for an entire oscillation cycle (lasting on average 0.7 s). After the stimulus disappeared, the

(a) Active



(b) Passive



(c)

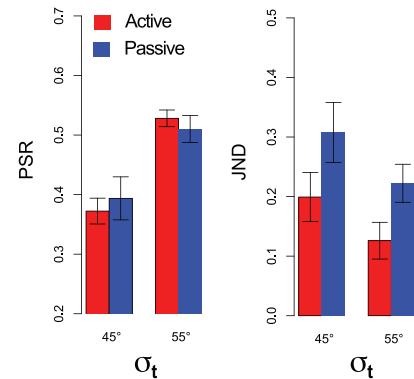


Figure 4. Results of Experiment 1: Which surface rotates faster? (a–b) Individual (gray) and average (red in a, blue in b) cumulative Gaussian fits of the proportion of responses “the target surface rotates slower than the reference surface” as a function of rotation gain g , for active (a) and passive (b) observers. The green curving arrows indicate the direction of rotation of the target surface for negative and positive values of rotation gain. The left and right columns show the results for each level of simulated slant of the target surface (45° and 55° , respectively). (c) Average PSR (left) and JND (right) for passive (blue) and active (red) observers, and for each level of simulated target surface slant. Vertical bars indicate ± 1 standard error of the mean.

observer provided her perceptual judgment through a mouse press.

In Experiment 1, the observer’s task was to judge which of the two surfaces (the upper or the lower one) was rotating faster. In Experiment 2, the observer judged whether the dihedral angle between the two planes appeared to be shrinking or expanding.

Each experimental session lasted about 90 min and consisted of two phases: an active and a passive vision phase. In each phase, eight experimental conditions were tested: 4 staircases \times 2 target-surface slants (45° , 55°) in Experiment 1 and 4 staircases \times 2 target-surface tilt angles (0° , 180°) in Experiment 2.

In order to achieve a stable performance level, both the active and passive vision phases were preceded by a screening session in which observers were presented with 46 trials randomly selected from the four extreme

rotation gain values defining the starting intensity of each staircase (i.e., $g = \pm 0.9$, ± 0.5). Only participants with more than 80% correct responses were admitted to the experimental session.

Experiment 1: Which surface rotates faster?

The results of Experiment 1 are in good agreement with the predictions of the HO model (Figure 4; see Figure 2, left, for the HO model’s predictions). In the active condition, observers were *not* more likely to see a rigid structure (Figure 4a) than in the passive condition (Figure 4b). This result is opposite to the prediction of the OO model, that is, a flattening of the psychometric function in the active condition (Figure 2, right).

Instead, sensitivity to the *def* difference was even larger in the active than in the passive viewing condition. This may indicate that the active observers had an advantage over the passive observers for the measurement of the velocity gradients, for example through a process of retinal stabilization (Oosterhoff, van Damme, & van de Grind, 1993; Cornilleau-Pérès & Droulez, 1994; Aytikin & Rucci, 2012). A similar speculation was made in previous studies (Caudek, Fantoni, & Domini, 2011; Fantoni et al., 2012), but it warrants further research.

Figure 4 illustrates the fitted proportions of “target surface rotating slower” responses as a function of the rotation gain for active (Figure 4a) and passive (Figure 4b) observers. The gray lines are the fitted psychometric functions for each individual observer, while the red and blue lines are the average fitted psychometric functions for active and passive observers, respectively. The fits were based on a cumulative Gaussian model whose parameters were estimated with the constrained maximum likelihood and bootstrap inference method implemented by *psignifit* software (Wichmann & Hill, 2001). The goodness of fit of each best-fit psychometric curve was assessed with the 95% confidence interval criterion based on Monte Carlo simulations of 10,000 data sets.

We defined the PSR as the rotation gain corresponding to the 50% point of each psychometric function (i.e., the rotation gain at which observers were unable to judge which surface was rotating faster). The JND was calculated as the difference between the rotation gain corresponding to the 84% level of the psychometric function and the PSR.

The average PSR and JND for the two slants of the target surface in the active (red bars) and passive (blue bars) viewing conditions are shown in Figure 4c. We analyzed the PSRs and JNDs using linear mixed-effects (LME) models with subjects as random effects, and simulated slant of target surface and viewing condition (passive, active) as fixed effects (Bates & Sarkar, 2007). Two-tailed *p*-values were obtained using Markov chain Monte Carlo simulations (10,000 samples).

For the active observers, a rigid 3-D structure was perceived when the target surface rotated in the counterclockwise direction at about half the speed induced by the observer’s translation (PSR = 0.450 ± 0.020 , $t = 21.9$, $p < 0.001$). A similar bias was found in passive vision (PSR = 0.452 ± 0.024 , $t = 18.71$, $p < 0.001$). Furthermore, the PSR depended on the slant of the target surface: A larger slant corresponded to a larger PSR in both the active (PSR at $45^\circ = 0.37$, PSR at $55^\circ = 0.53$; $t = 6.04$, $p < 0.01$) and passive (PSR at $45^\circ = 0.39$, PSR at $55^\circ = 0.51$; $t = 2.99$, $p < 0.01$) conditions. These effects were accounted for by our LME model,

revealing a significant main effect of the slant of the target surface ($t = 4.41$, $p < 0.001$).

Similar main effects were found on the JNDs (slant: $t = 2.81$, $p < 0.01$; viewing condition: $t = 3.66$, $p < 0.01$). Importantly, the JND in the active condition was significantly smaller than the JND in the passive condition (0.16 and 0.26, respectively, $t = 3.66$, $p < 0.01$), a result that is opposite to the prediction of the OO model. This effect cannot be explained by the unbalanced temporal ordering of the passive and active vision phases. Because the passive vision phase always came *after* the active vision phase, any effect of practice should have produced the opposite result.

These results show that judgments of rigidity were systematically biased. For both passive and active observers, a rigid structure ($g = 0$) was systematically perceived as nonrigid, since the target surface was mostly perceived as rotating faster than the reference surface. Instead, nonrigid structures ($g = \text{PSR}$) were systematically perceived as undergoing a rigid transformation.

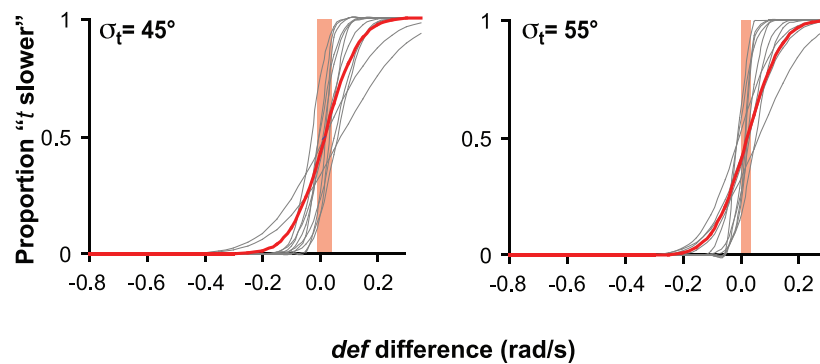
Nevertheless, these biases are consistent with the predictions of the HO model. As can be seen in Figure 5, illustrating individual (gray lines) and average (bold lines) psychometric functions of *def* difference for both active (Figure 5a) and passive (Figure 5b) observers, two surfaces are perceived as undergoing the same rotation in depth only when they project identical velocity gradients (same *def*). When the projected velocity gradients are discernibly different, the two-plane configuration is perceived as undergoing a nonrigid transformation. A statistical analysis on PSR and JND recoded as a function of *def* difference revealed that *def* was the only determinant of the perceptual responses, and the additional contribution of the slant of the target surface was not significant (PSR: $t = 0.42$, not significant; JND: $t = 0.036$, not significant).

In summary, these results are compatible with the HO model, which does not embed either a prior for stationarity/rigidity or a measurement of the observer’s egomotion.

Experiment 2: Is the dihedral angle shrinking or expanding?

In Experiment 1, perception of rigidity was investigated with a task that only indirectly allowed assessment of whether or not observers perceived a rigid transformation of the two-plane configuration. It could be argued that in the event the structure was always perceived as rigid, observers may have adopted

(a) Active



(b) Passive

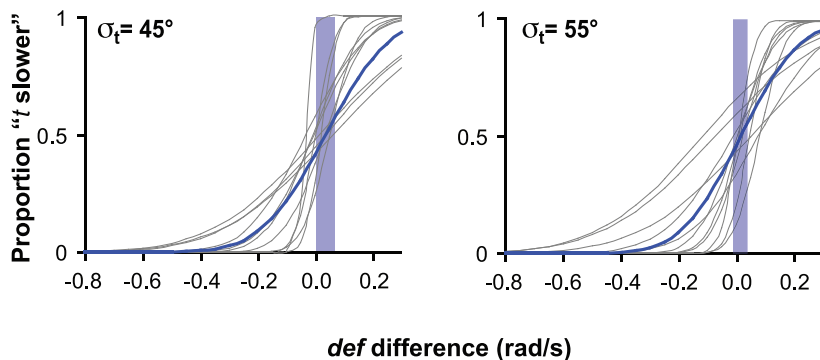


Figure 5. Cumulative Gaussian fits of Figure 4 replotted as function of the *def* difference between the target and reference surfaces. Panel layout and color coding are consistent with those of Figure 4. Shaded bands represent ± 1 standard error of the mean for PSRs.

some ad hoc strategy for deciding which of the two surfaces rotated at a faster rate: maybe deciding to pick the surface projecting a larger velocity gradient as the one rotating faster, effectively performing a 2-D task. Even though this is an unlikely possibility, Experiment 2 was designed to rule out this alternative explanation by adopting a more direct assessment of rigidity.

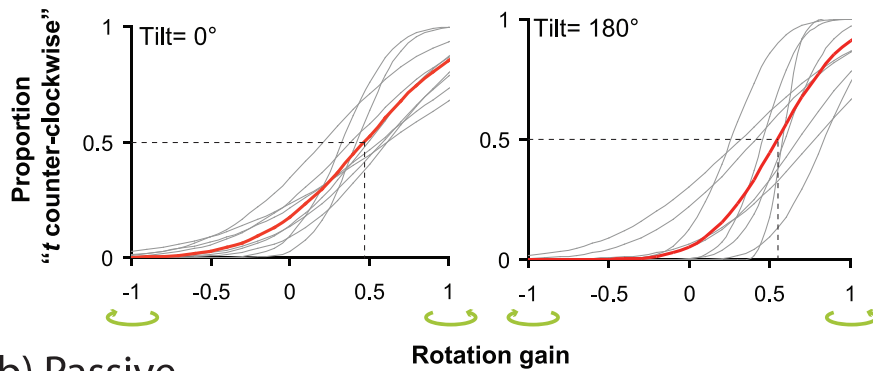
In order to achieve this goal, Experiment 2 differed from Experiment 1 in two important aspects. First, the two surfaces composing the dihedral angle had opposite tilts (0° and 180°). This made it much more difficult to directly compare the velocity gradients of the two surfaces, since the two gradients had different signs. For example, if during the observer translation one surface projected a velocity gradient that was expanding, then the other surface, having opposite tilt, was projecting a velocity gradient that was contracting. Second, the task was to judge whether the dihedral angle was *shrinking* or *expanding*, which required a Euclidean estimate of the 3-D structure.

Results

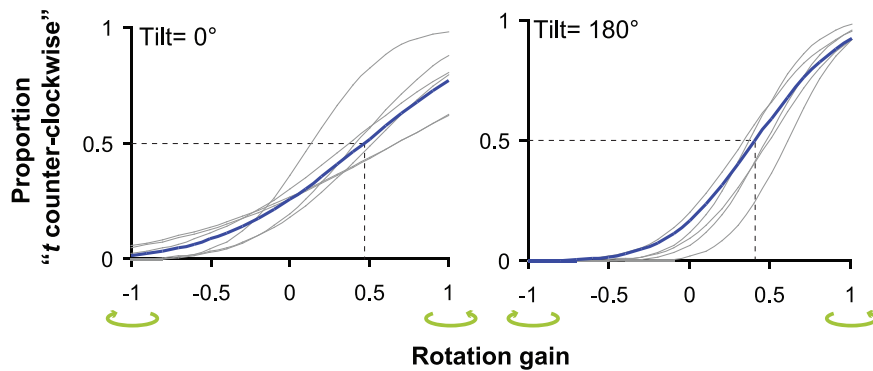
In order to better compare the results of Experiment 2 with those of Experiment 1, we recoded the judgments of shrinkage or expansion in terms of relative rotation of the target surface with respect to the reference surface. For a target surface with a tilt of 180° , a “shrinking” response was recoded as “target surface rotating in a counterclockwise direction relative to the reference surface.” The same coding was also given for a target surface with a tilt of 0° and an “expanding” response.

The results of this experiment showed the same biases as those of the previous experiment, thus ruling out any alternative explanation for the results of Experiment 1. As can be seen in Figure 6, the PSRs corresponded to a rotation gain of about 0.48 ($t = 15.03$, $p < 0.001$), which is close to those found in Experiment 1, and did not depend on the tilt of the target surface ($t = 0.947$, not significant), on whether the observer was active or passive ($t = 0.175$, not significant), or on the interaction of these two variables ($t = 0.954$, not significant). Similar results were

(a) Active



(b) Passive



(c)

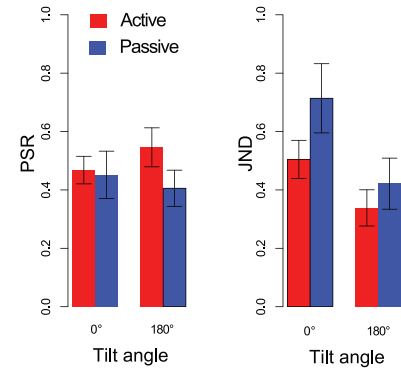


Figure 6. Results of Experiment 2: Is the dihedral angle shrinking or expanding? (a–b) Individual (gray) and average (red in a, blue in b) cumulative Gaussian fits of the proportion of responses “target surface rotating in a counterclockwise direction relative to the reference surface” as a function of rotation gain g , for active (a) and passive (b) observers. The green curving arrows indicate the direction of rotation of the target surface for negative and positive values of rotation gain. The left and right columns show the results for each level of tilt of the target surface (180° and 0° , respectively). (c) Average PSR (left) and JND (right) for passive (blue) and active (red) observers, and for each level of simulated target surface tilt. Vertical bars indicate ± 1 standard error of the mean.

obtained on the JNDs with the same LME model (tilt: $t = 1.87$, not significant; viewing condition: $t = 2.2$, not significant; tilt \times viewing condition: $t = 0.9$, not significant).

As in the Experiment 1, perception of rigidity depended on the difference between the projected velocity gradients (Figure 7), since the two-plane structure was perceived as undergoing a rigid transformation when the *def* difference was 0 (PSR = -0.021 , $t = 1.51$, not significant).

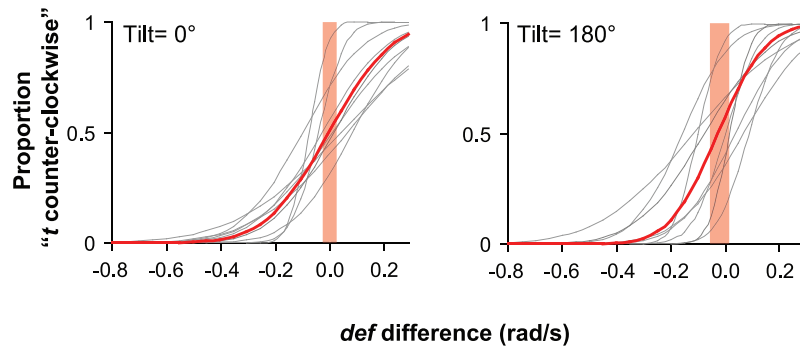
General discussion and conclusions

In two experiments, we studied the perception of rigidity of a two-plane configuration from the information provided by the optic flow. We found that active observers, who self-generated the optic flow, showed systematic biases in the perception of

rigidity. Specifically, a rigid two-plane configuration was inaccurately perceived as nonrigid: The dihedral angle between the two surfaces was judged as changing during the observer’s motion. The biases found for active observers were identical to those of passive observers, who experienced from a static vantage point the same optic flow as the active observers. Whereas perception of rigidity did not depend on the actual rigidity of the distal stimuli, it could be entirely accounted for by the difference between the velocity gradients (deformations, *defs*) projected by the two planar surfaces. When the *def* difference was detectably different from 0, the two surfaces appeared as undergoing different amounts of rotation in depth.

These results replicate for active observers the findings of Domini and colleagues (1997) and are in agreement with the predictions of a model proposed by Domini and Caudek (2003) that derives through a maximum likelihood procedure an estimate of the 3-D

(a) Active



(b) Passive

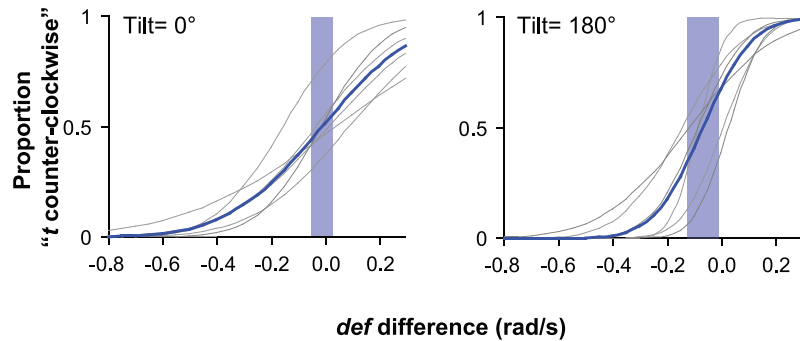


Figure 7. Cumulative Gaussian fits of Figure 6 replotted as a function of the *def* difference between the target and reference surfaces. Panel layout and color coding are consistent with those of Figure 6. Shaded bands represent ± 1 standard error of the mean for PSRs.

angular velocity of a planar surface from the *def* component of the optic flow. Since *def* is inherently ambiguous, the MLE solution is in general inaccurate: It assigns to two different values of *def* two different values of estimated angular velocities, independent of whether or not the two *defs* were generated by surfaces undergoing the same rotational motion. The MLE solution is for a passive observer the best the visual system can do with only optic flow information.

What is puzzling about these findings is that the performance of active observers in the rigidity discrimination task is basically identical to that of passive observers, even though active observers have access to extraretinal and proprioceptive signals, which, in principle, are strong enough to be informative about the observer's egomotion (Wei & Angelaki, 2004; Jaekl et al., 2005; Caudek et al., 2011; Aytakin & Rucci, 2012). In the section "Optimal Bayesian estimation of the relative rotation between two planar surfaces," we showed that an optimal-observer model, which optimally combines retinal and extraretinal sensory information and assumes a static and rigid environment, is most likely to assign a rigid interpretation to an optic flow compatible with a rigid

transformation. Instead, in both experiments observers mostly perceived nonrigid transformations, although the *instantaneous* optic flow was *always* compatible with a rigid transformation, even for the nonrigid displays.

Note that these results cannot be explained solely on the basis of a noisy estimate of egocentric translation. Whether this estimate is very noisy or simply ignored by the visual system constitutes only part of the story. A fundamental explanatory role is played by the absence of a reliable rigidity prior. In fact, such a prior would bias the observer's judgments toward a rigid interpretation of the two-plane configuration in *both* the passive and active conditions, as predicted by the OIE model described in Appendix C.

Instead, these results are compatible with an HO model, in which *both* the prior for rigidity is uninformative and egomotion information is disregarded. This model is therefore equivalent to the MLE model proposed by Domini and Caudek (2003) for perception of 3-D structure from the passively viewed optic flow, which only relies on the information provided by the local velocity gradients (*def*). The predictions of this model were also confirmed in a series of studies showing systematic biases in both the



Movie 2. A demonstration of the loss of perceptual rigidity to which we are unconsciously exposed during the natural viewing of an everyday object.

perception of motion and slant of actively viewed planar surfaces (Fantoni et al., 2010, 2012; Caudek et al., 2011).

The question of why the visual system relies on ambiguous information provided by *def*, while both ignoring statistically plausible priors, like stationarity or rigidity, and potentially available egocentric signals, remains open. Two possible explanations could be attempted.

A first possibility is that a monocularly viewed optic flow, generated by the motion of a random-dot pattern in an otherwise dark environment, is still a very impoverished viewing condition, which forces the visual system to operate with suboptimal performance, even for an active observer (Wallach, Stanton, & Becker, 1974). The presence of multiple cues, characterizing normal viewing, could provide optical information that is sufficient for an accurate estimate of 3-D properties without the need for additional information carried by extraretinal signals or prior knowledge about world properties. However, even in a natural setting, systematic errors in the perception of rigidity are found—for example, if we walk along a street and look around, two road signs pointing at different street directions from a common pole are perceived as independently rotating as we approach them (Movie 2). Even though monocular richer stimuli still induce the same perceptual distortions, as shown in Movie 2, the absence of binocular disparities constitutes a critical departure from what can be considered ecologically valid stimuli. Indeed, binocular disparities in conjunction with the optic flow could in principle uniquely specify the 3-D structure and motion of an object (Richards, 1985).

A second possibility is that the goal of the visual system is not that of recovering accurate metric information of distal objects (Domini & Braunstein, 1998). Optimal behavior is such insofar it allows an organism to have a successful interaction with the environment for survival. It is possible that this behavior takes place without detailed information about the physical structure of the world (e.g., metric properties). For example, it has been speculated that local affine information, which encodes nonmetric aspects of 3-D structure like the depth order of feature points, determines both our conscious perception of the world and our motor actions (Domini & Caudek, 2011). If this is the case, then perceptual processes are likely to ignore statistical properties of the world that are irrelevant to affine information, like stationarity or rigidity (Jain & Zaidi, 2011).

Keywords: 3-D structure from motion, rigidity, active vision, optic flow, optimal integration

Acknowledgments

We thank Dhanraj Vishwanath, Robert Volcic, and Carlo Nicolini for useful comments on a former version of the manuscript.

Commercial relationships: none.

Corresponding author: Carlo Fantoni

Email: cfantoni@units.it.

Address: Department of Life Sciences, Psychology Unit “Gaetano Kanizsa”, University of Trieste, Trieste, Italy.

References

- Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance. In D. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 409–423). New York: Cambridge University Press.
- Allison, R. S., Harris, L. R., Jenkin, M., Jasiobedzka, U., & Zacher, J. E. (2001). Tolerance of temporal delay in virtual environments. *IEEE Virtual Reality, 2001*, 3, 247–254, <http://dx.doi.org/10.1109/VR.2001.913793>.
- Aytekin, M., & Rucci, M. (2012). Motion parallax from microscopic head movements during visual fixation. *Vision Research, 70*, 7–17.
- Banks, M. S., & Backus, B. T. (1998). Extra-retinal and

- perspective cues cause the small range of the induced effect. *Vision Research*, *38*, 187–194.
- Bates, D., & Sarkar, D. (2007). *lme4. Linear mixed effects models using S4 classes* (R package Version 0.9975) [Computer software]. Retrieved from <http://cran.rproject.org/web/packages/lme4/index.html>.
- Bennur, S., & Gold, J. I. (2008). Right way neurons. *Nature Neuroscience*, *11*, 1121–1122. [PubMed]
- Braunstein, M. L., & Tittle, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 582–590. [PubMed]
- Buizza, A., Leger, A., Droulez, J., Bertoz, A., & Schmid, R. (1980). Influence of otolithic stimulation by horizontal linear head acceleration in optokinetic nystagmus and visual motion perception. *Experimental Brain Research*, *71*, 406–410.
- Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 609–621. [PubMed]
- Caudek, C., Fantoni, C., & Domini, F. (2011). Bayesian modeling of perceived surface slant from actively-generated and passively-observed optic flow. *PLoS ONE*, *6*(4), e18731, doi:10.1371/journal.pone.0018731.
- Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Norwell, MA: Kluwer Academic Publishers.
- Colas, F., Droulez, J., Wexler, M., & Bessière, P. (2007). A unified probabilistic model of the perception of three-dimensional structure from optic flow. *Biological Cybernetics*, *97*, 461–477. [PubMed]
- Cornilleau-Pérès, V., & Droulez, J. (1994). The visual perception of three-dimensional shape from self-motion and object motion. *Vision Research*, *34*, 2331–2336. [PubMed]
- Dijkstra, T. M., Cornilleau-Pérès, V., Gielen, C. C., & Droulez, J. (1995). Perception of three-dimensional shape from ego- and object-motion: Comparison between small- and large-field stimuli. *Vision Research*, *35*, 453–462. [PubMed]
- Domini, F., & Braunstein, M. L. (1998). Recovery of 3d structure from motion is neither euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1273–1295.
- Domini, F., & Caudek, C. (2011). Combining image signals before 3D reconstruction: The Intrinsic Constraint Model of cue integration. In J. Trommershäuser, M. S. Landy, & K. Körding, (Eds.), *Sensory cue integration* (pp. 120–143). New York: Oxford University Press.
- Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 426–444. [PubMed]
- Domini, F., & Caudek, C. (2003). 3D structure perceived from dynamic information: A new theory. *Trends in Cognitive Sciences*, *7*, 444–449.
- Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1111–1129. [PubMed]
- Domini, F., Caudek, C., Turner, J., & Favretto, A. (1998). Discriminating constant from variable angular velocities in structure from motion. *Perception and Psychophysics*, *60*, 747–760. [PubMed]
- Dupin, L., & Wexler, M. (2013). Motion perception by a moving observer in a three-dimensional environment. *Journal of Vision*, *13*(2):15, 1–14, <http://www.journalofvision.org/content/13/2/15>, doi:10.1167/13.2.15. [PubMed] [Article]
- Dyde, R. T., & Harris, L. R. (2008). The influence of retinal and extra-retinal motion cues on perceived object motion during self-motion. *Journal of Vision*, *8*(14):5, 1–10, <http://journalofvision.org/8/14/5/>, doi:10.1167/8.14.5. [PubMed] [Article]
- Eagle, R. A., & Hogervorst, M.A. (1999). The role of perspective information in the recovery of 3D structure-from-motion. *Vision Research*, *39*, 1713–1722. [PubMed]
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69–73.
- Ernst, M., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169.
- Fantoni, C., Caudek, C., & Domini, F. (2012). Perceived surface slant is systematically biased in the actively-generated optic flow. *PLoS ONE*, *7*(3), e33911, doi:10.1371/journal.pone.0033911.
- Fantoni, C., Caudek, C., & Domini, F. (2010). Systematic distortions of perceived planar surface motion in active vision. *Journal of Vision*, *10*(5):12, 1–20, <http://www.journalofvision.org/>

- content/10/5/12, doi:10.1167/10.5.12. [PubMed] [Article]
- Ferman, L., Collewyn, H., Jansen, T. C., & van den Berg, B. (1987). Human gaze stability in the horizontal, vertical and torsional direction during voluntary head movements, evaluated with a three dimensional scleral induction coil technique. *Vision Research*, 27, 811–828. [PubMed]
- Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology*, 16(4), 428–432, doi:10.1016/j.cub.2006.01.019.
- Grzywacz, N. M., & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. *Journal of the Optical Society of America, A*, 4, 503–518.
- Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature Neuroscience*, 11, 1201–1210. [PubMed]
- Gu, Y., DeAngelis, G. C., & Angelaki, D. E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. *Nature Neuroscience*, 10, 1038–1047. [PubMed]
- Hogervorst, M. A., & Eagle, R. A. (2000). The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 934–955. [PubMed]
- Jaekl, P. M., Jenkin, M. R., & Harris, L. R. (2005). Perceiving a stable world during active rotational and translational head movements. *Experimental Brain Research*, 163, 388–399. [PubMed]
- Jain, A., & Zaidi, Q. (2011). Discerning nonrigid 3D shapes from motion cues. *Proceedings of the National Academy of Sciences, USA*, 2010, 1–6, doi:10.1073/pnas.1016211108.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55, 271–304.
- Knill, D. C. (2007). Robust cue integration: A Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *Journal of Vision*, 7(7):5, 1–24, <http://journalofvision.org/content/7/7/5/>, doi:10.1167/7.7.5. [PubMed] [Article]
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712–719, doi:10.1016/j.tins.2004.10.007.
- Koenderink, J. J. (1986). Optic flow. *Vision Research*, 26, 161–180. [PubMed]
- Koenderink, J. J., & van Doorn, A. J. (1978). How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow. In G. Hauske & E. Butenandt (Eds.), *Kybernetik* (pp. 224–247). Munich: Oldenberg.
- Koenderink, J. J., & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22, 773–791.
- Landy, M. S., Banks, M., & Knill, D. (2011). Ideal-observer models of cue integration. In J. Trommershäuser, M. S. Landy, & K. Körding (Eds.), *Sensory cue integration* (pp. 5–29). New York: Oxford University Press.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389–412. [PubMed]
- Liter, J. C., & Braunstein, M. L. (1998). The relationship of vertical and horizontal velocity gradients in the perception of shape, rotation, and rigidity. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1257–1272.
- Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception*, 22, 1441–1465.
- Liu, S., & Angelaki, D. E. (2009). Vestibular signals in macaque extrastriate visual cortex are functionally appropriate for heading perception. *Journal of Neuroscience*, 29, 8936–8945. [PubMed]
- Nawrot, M. (2003). Depth from motion parallax scales with eye movement gain. *Journal of Vision*, 3(11): 17, 841–851, <http://www.journalofvision.org/content/3/11/17>, doi:10.1167/3.11.17. [PubMed] [Article]
- Nawrot, M., & Joyce, L. (2006). The pursuit theory of motion parallax. *Vision Research*, 46, 4709–4725.
- Nawrot, M., & Stroyan, K. (2012). Integration time for the perception of depth from motion parallax. *Vision Research*, 59, 64–71, doi:10.1016/j.visres.2012.02.007.
- Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception and Psychophysics*, 53, 279–291.

- Ono, H., & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception and Psychophysics*, *48*, 179–187. [PubMed]
- Oosterhoff, F. H., van Damme, W. M., & van de Grind, W. A. (1993). Active exploration of three-dimensional objects is more reliable than passive observation. *Perception*, *22*, 99.
- Panerai, F., Cornilleau-Pérès, V., & Droulez, J. (2002). Contribution of extraretinal signals to the scaling of object distance during self-motion. *Perception and Psychophysics*, *64*, 717–731. [PubMed]
- Peh, C.-H., Panerai, F., Droulez, J., Cornilleau-Pérès, V., & Cheong, L.-F. (2002). Absolute distance perception during in-depth head movement: Calibrating optic flow with extra-retinal information. *Vision Research*, *42*, 1991–2003. [PubMed]
- Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America*, *2*, 343–349.
- Rogers, S., & Rogers, B. J. (1992). Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics*, *52*, 446–452. [PubMed]
- Swindells, C., Dill, J. C., & Booth, K. S. (2000). System lag tests for augmented and virtual environments. In M. S. Ackerman & K. Edwards (Eds.), *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology* (pp. 161–170). New York: ACM.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- van Boxtel, J. J. A., Wexler, M., & Droulez, J. (2003). Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision*, *3*(5):1, 318–332, <http://www.journalofvision.org/content/3/5/1>, doi:10.1167/3.5.1. [PubMed] [Article]
- Vishwanath, D. (2013). Experimental phenomenology of visual 3D space: Considerations from evolution, perception, and philosophy. In L. Albertazzi (Ed.), *Handbook of experimental phenomenology: Visual perception of shape, space and appearance* (pp. 181–204). Hoboken, NJ: Wiley-Blackwell.
- Vishwanath, D., & Hibbard, P. B. (2013). Seeing in 3-D with just one eye: Stereopsis without binocular vision. *Psychological Science*, *24*(9), 1673–1685, doi:10.1177/0956797613477867.
- Volcic, R., Fantoni, C., Caudek, C., Assad, J., & Domini, F. (2013). Visuomotor adaptation changes the processing of binocular disparities and tactile discrimination. *Journal of Neuroscience*, *33*, 17081–17088.
- Wallach, H., Stanton, L., & Becker, D. (1974). The compensation for movement-produced changes in object orientation. *Perception and Psychophysics*, *15*, 339–343.
- Wei, M., & Angelaki, D. E. (2004). Does head rotation contribute to gaze stability during passive translations? *Journal of Neurophysiology*, *91*(4), 1913–1918, doi:10.1152/jn.01044.2003.
- Wexler, M. (2003). Voluntary head movement and allocentric perception of space. *Psychological Science*, *14*, 340–346. [PubMed]
- Wexler, M., Lamouret, I., & Droulez, J. (2001). The stationarity hypothesis: An allocentric criterion in visual perception. *Vision Research*, *41*, 3023–3037. [PubMed]
- Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature*, *409*, 85–88. [PubMed]
- Wexler, M., & van Boxtel, J. J. (2005). Depth perception by the active observer. *Trends in Cognitive Science*, *9*, 431–438. [PubMed]
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception and Psychophysics*, *63*, 1293–1313.

Appendix A

Maximum Likelihood Estimate of angular velocities from *defs*

In Equation 3, the likelihoods $P(def_i|\omega_{ri})$, $i = 1, 2$, can be calculated by integrating over the nuisance variable σ the image formation model $P(def_i|\omega_{ri}, \sigma)$ multiplied by the prior distribution $P(\sigma)$:

$$P(def_i|\omega_{ri}) = \int_{\sigma} P(def_i|\omega_{ri}, \sigma)P(\sigma)d\sigma \quad (A1)$$

Assuming Gaussian noise in the measurement of *def*, $P(def_i|\omega_{ri}, \sigma)$ is a Gaussian distribution with mean at the measured value of the velocity gradient (*def*) and variance s_{def}^2 . The prior distribution of surface slant $P(\sigma)$ is considered to be an uninformative Gaussian centered at 0 with variance s_{σ}^2 (Domini & Caudek, 2003; Colas et al., 2007).

In Figure A1 we show $P(def_i|\omega_{ri}, \sigma)$, $P(\sigma)$, and $P(def_i|\omega_{ri})$ for $def_1 = 0.2$ rad/s (Figure A1a) and $def_2 =$

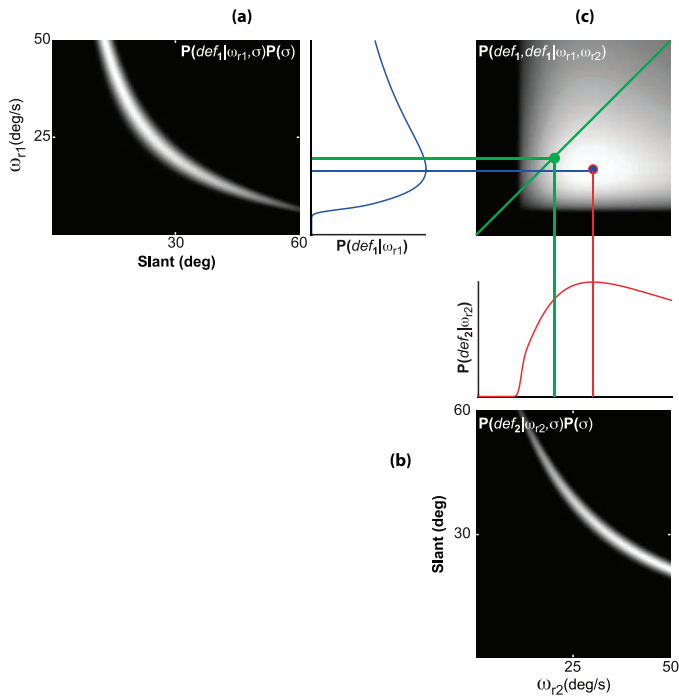


Figure A1. MLE of relative angular velocities from *defs*. The Likelihood function $P(def_1, def_2|\omega_{r1}, \omega_{r2})$ (c) for two surfaces projecting velocity gradients $def_1 = 0.2$ rad/s and $def_2 = 0.35$ rad/s is the product of two likelihood functions $P(def_1|\omega_{r1})$ (a) and $P(def_2|\omega_{r2})$ (b) calculated by integrating over the nuisance variable σ the products $P(def_1|\omega_{r1}, \sigma)P(\sigma)$ (a) and $P(def_2|\omega_{r2}, \sigma)P(\sigma)$ (b). The MLE corresponds to two values of relative angular velocities ω_{r1} (blue lines) and ω_{r2} (red lines) that are different from the actual angular velocity (green lines) and are also different from each other (blue circle with red rim). Therefore, the MLE is that of a nonrigid transformation (the diagonal green line indicates rigid solutions, i.e., $\omega_{r1} = \omega_{r2}$). Gray levels correspond to probability.

0.35 rad/s (Figure A1b). In this simulation, we considered an observer moving at $\omega_e = 19.35^\circ/s$ viewing a static two-plane configuration for which $\sigma_1 = 30^\circ$ and $\sigma_2 = 45^\circ$. We now consider the Maximum Likelihood Estimate (MLE) of ω_{r1} (blue lines) and ω_{r2} (red lines), which are the values of relative angular velocities maximizing $P(def_1, def_2|\omega_{r1}, \omega_{r2})$. These values are different from the real value of $\omega_r = \omega_e = 19.35^\circ/s$ (Figure A1c, green circle) and also different from each other (Figure A1c, x- and y-coordinates of the red outlined circle). Therefore, the MLE is that of a nonrigid structure, since the estimated rotation of the surface producing a larger value of *def* is larger. The result of this simulation is compatible with the findings of Domini and colleagues (1997) showing that perceived angular velocity is a monotonically increasing function of *def* and that the two-plane configuration is perceived as rigid only if the two surfaces generate the same value of *def*.

Appendix B

The role of egomotion information and a prior for stationarity/rigidity

The term $P(\omega_{r1}, \omega_{r2}|\hat{\omega}_e)$ of Equation 3 can change the MLE interpretation (see Appendix A), since it incorporates information about the observer’s egomotion and the stationarity/rigidity prior. It can be shown that

$$P(\omega_{r1}, \omega_{r2}|\hat{\omega}_e) = \int_{\omega_e} P(\omega_{r1}|\omega_e)P(\omega_{r2}|\omega_e)P(\omega_e|\hat{\omega}_e)d\omega_e \quad (B1)$$

where $P(\omega_{r1}|\omega_e) = P_{\omega_s}(\omega_{s1} = \omega_{r1} + \omega_e)$ and $P(\omega_{r2}|\omega_e) = P_{\omega_s}(\omega_{s2} = \omega_{r2} + \omega_e)$, with $P_{\omega_s}(\omega_{s1})$ and $P_{\omega_s}(\omega_{s2})$ indicating the a priori distributions over the surface angular velocities ω_{s1} and ω_{s2} . These a priori distributions (modeled as Gaussians centered at 0 with standard deviation s_{ω_s}) are sharply peaked at 0 if the surfaces are assumed to be stationary in the world, since a stationary surface is defined by $\omega_s = 0$. If both a priori distributions are narrowly peaked at 0, then both surfaces are stationary and as a consequence the structure is rigid.

The term $P(\omega_e|\hat{\omega}_e)$ —a Gaussian distribution centered at $\hat{\omega}_e$ with standard deviation s_{ω_e} —defines the precision of the measurement of the egomotion angular velocity $\hat{\omega}_e$. This distribution is narrowly peaked at $\hat{\omega}_e$ if the egomotion estimate is very precise.

In summary, $P(\omega_{r1}, \omega_{r2}|\hat{\omega}_e)$ and its influence on the posterior critically depends on the precision of the measurement of the egomotion angular velocity $\hat{\omega}_e$, and the strength of the prior for stationarity/rigidity.

Appendix C

The extent to which the Maximum A Posteriori (MAP) estimate—that is, the pair $(\omega_{r1}, \omega_{r2})$ that maximizes the posterior—differs from the MLE can be seen in Figures C1 through C3. Depending on the values of s_{ω_e} (precision of egocentric motion estimate) and s_{ω_s} (strength of the stationarity/rigidity prior), we can foresee the following three qualitatively different models.

Optimal observer (OO)

This model, similar to the one proposed by Colas et al. (2007), includes a strong stationarity/rigidity prior and a precise measurement of the observer’s egomotion (Figure C1). First consider the active observer (Figure

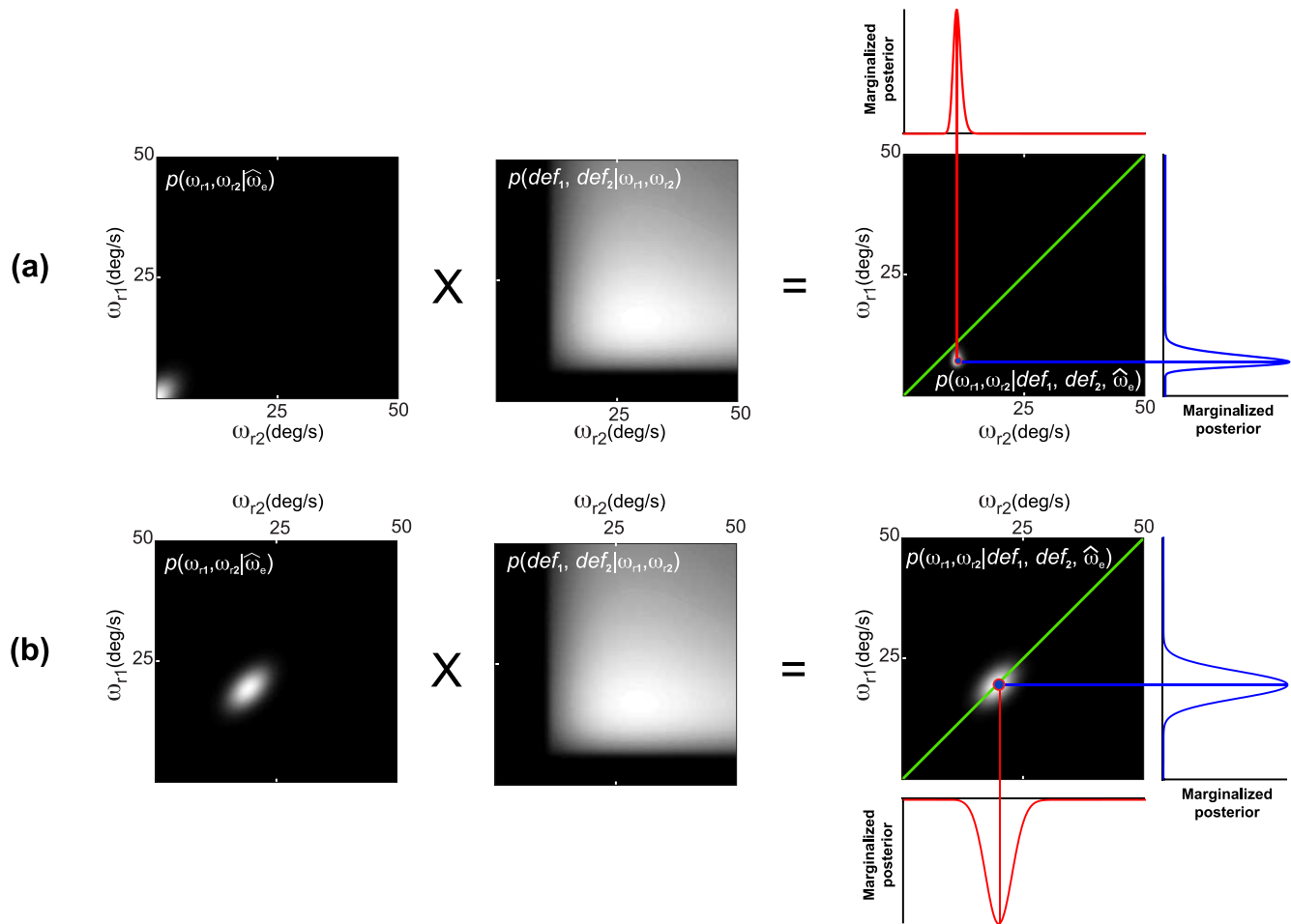


Figure C1. MAP estimate of relative angular velocities according to the OO model. The posterior distribution $P(\omega_{r1}, \omega_{r2} | def_1, def_2, \hat{\omega}_e)$ is obtained by multiplying the likelihood $P(def_1, def_2 | \omega_{r1}, \omega_{r2})$ by $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$. If the observer’s egomotion is measured with precision and it is assumed that surfaces in the world are stationary, then $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ is sharply peaked at the value of angular velocity equal to the observer’s egomotion ($\hat{\omega}_e$). For a passive observer, who is static, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ is peaked at 0 (a). For an active observer, moving with relative angular velocity $\hat{\omega}_e = 19.35^\circ/s$, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ is peaked at $19.35^\circ/s$ (b). The posterior distribution for the passive observer is still peaked, like the likelihood, at two values of relative angular velocities ω_{r1} (blue lines) and ω_{r2} (red lines) that are different from each other (a, right). Instead, the posterior distribution for the active observer is peaked at two values of relative angular velocities ω_{r1} (blue lines) and ω_{r2} (red lines) that are equal to $\hat{\omega}_e$ (b, right). Therefore, the MAP for a passive observer specifies a nonrigid solution, whereas the MAP for an active observer specifies a rigid solution.

C1b). Since s_{ω_e} and s_{ω_s} are very small, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ is sharply peaked at the veridical solution $\omega_{r1} = \omega_{r2} = \hat{\omega}_e$ (Figure C1b, left panel). In this case, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ has a strong influence on the widely spread likelihood function (Figure C1b, middle panel), therefore producing MAP estimates (Figure C1b, right panel) defining a rigid interpretation ($\omega_{r1} = \omega_{r2}$).

Consider now the passive observer (Figure C1a). $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ is sharply peaked at the solution $\omega_{r1} = \omega_{r2} = 0$, since the sensed egocentric motion is zero (Figure C1a, left panel). In this case, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ only pulls the MAP solution towards small rotations, but it effectively constitutes a noninformative prior,

since it does not favor any particular solution that is different from zero.

Therefore, for a passive observer the MAP solution is qualitatively similar to the MLE solution, assigning a nonrigid interpretation to the two-plane rotation (Figure C1a, right panel).

Heuristic observer (HO)

This model includes a weak stationarity prior and a noisy measurement of the observer’s egomotion, Figure C2 (Fantoni et al., 2010; Caudek et al., 2011; Fantoni et

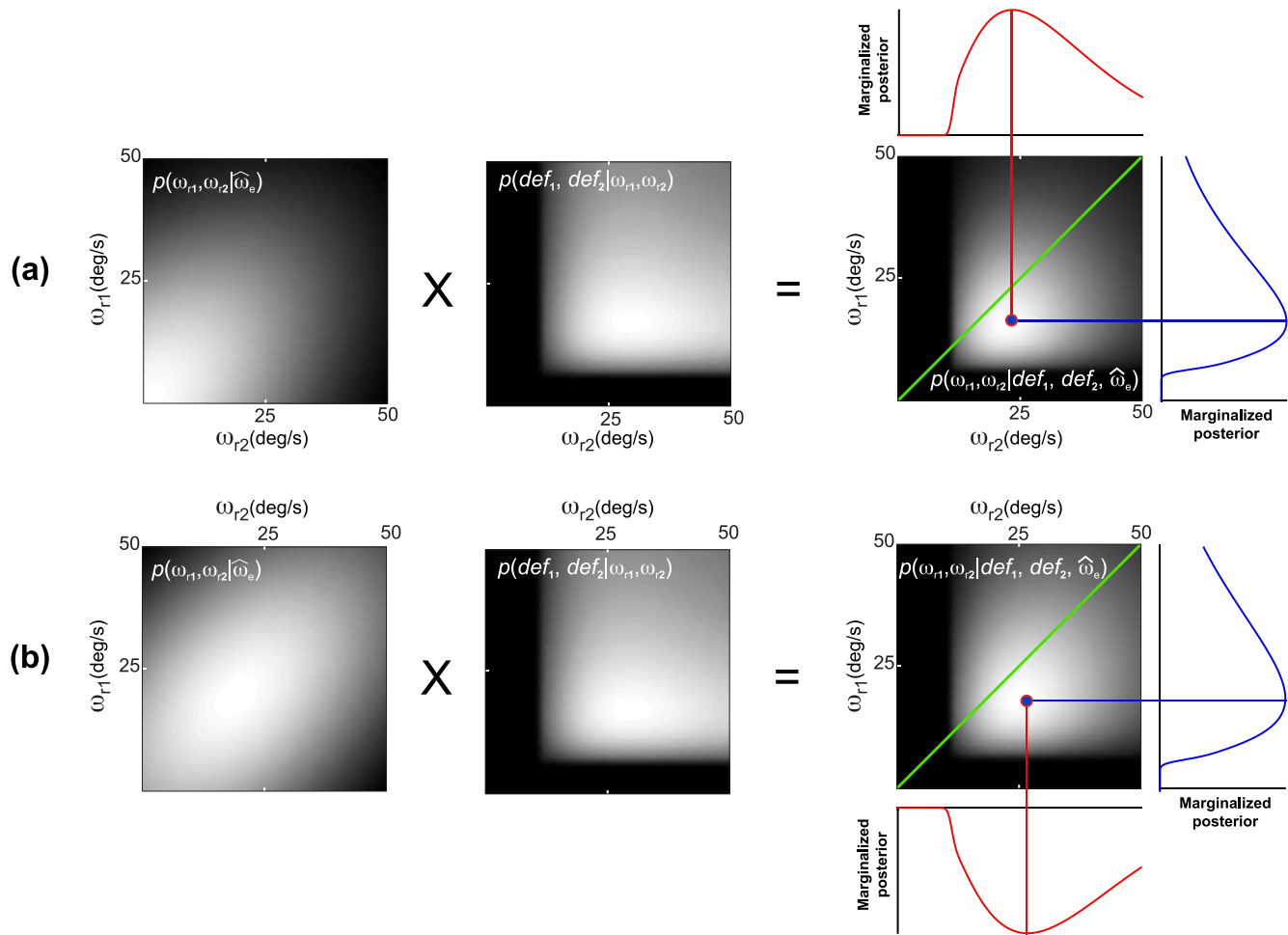


Figure C2. MAP estimate of relative angular velocities according to the HO model. The posterior distribution $P(\omega_{r1}, \omega_{r2} | def_1, def_2, \hat{\omega}_e)$ is obtained by multiplying the likelihood $P(def_1, def_2 | \omega_{r1}, \omega_{r2})$ by $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$. If the observer’s egomotion is measured with very low precision and the prior for stationarity/rigidity is uninformative, then $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ is widely distributed, peaked at 0 for a passive observer (a) and at $\hat{\omega}_e$ for an active observer (b). Given the very weak influence of $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ over the likelihood for both an active and a passive observer, the posterior is peaked, like the likelihood, at two values of relative angular velocities ω_{r1} (blue lines) and ω_{r2} (red lines) that are different from each other. Therefore, the MAP for both a passive and an active observer favors a nonrigid solution.

al., 2012). For both passive (Figure C2a) and active (Figure C2b) observers, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ is widely spread, since the system has no access to reliable information about the observer’s egomotion (large s_{ω_e}) and does not assume that surfaces in the world are stationary (large s_{ω_s}). In this case, $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ constitutes a non-informative prior which does not change the MLE interpretation: For both passive and active observers, the MAP estimate defines a nonrigid transformation.

Observer insensitive to egomotion (OIE)

This model includes a strong stationarity prior and a noisy measurement of the observer’s egomotion (Figure C3). For both passive (Figure C3a) and active (Figure C3b) observers, the estimates of $(\omega_{r1}, \omega_{r2})$ are the same. In this case, the measurement of ω_e is uncertain, but the strong stationarity prior imposes the condition that $\omega_{r1} = \omega_{r2}$. The MAP estimate defines a rigid transformation for both active and passive observers.

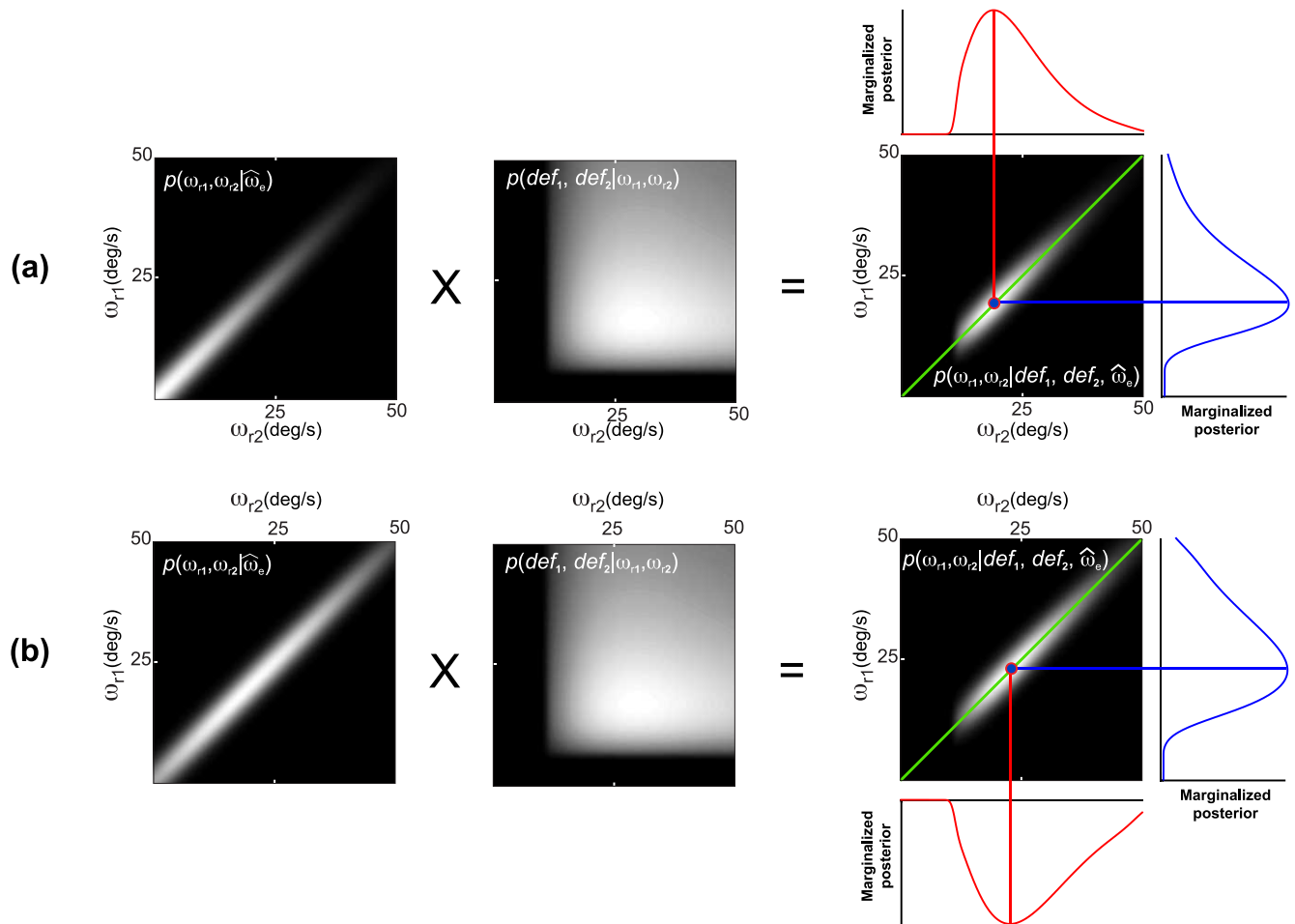


Figure C3. MAP estimate of relative angular velocities according to the IEO model. The posterior distribution $P(\omega_{r1}, \omega_{r2} | def_1, def_2, \hat{\omega}_e)$ is obtained by multiplying the likelihood $P(def_1, def_2 | \omega_{r1}, \omega_{r2})$ by $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$. If the observer’s egomotion is measured with very low precision but the prior for stationarity/rigidity is highly informative, then $P(\omega_{r1}, \omega_{r2} | \hat{\omega}_e)$ specifies a family of rigid solutions ($\omega_{r1} = \omega_{r2}$). In this case also, the posterior favors rigid interpretations for both a passive (a) and an active (b) observer.