



# The stationarity hypothesis: an allocentric criterion in visual perception

Mark Wexler \*, Ivan Lamouret, Jacques Droulez

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

Received 17 July 2001

## Abstract

Having long considered that extraretinal information plays little or no role in spatial vision, the study of structure from motion (SfM) has confounded a moving observer perceiving a stationary object with a non-moving observer perceiving a rigid object undergoing equal and opposite motion. However, recently it has been shown that extraretinal information does play an important role in the extraction of structure from motion by enhancing motion cues for objects that are stationary in an allocentric, world-fixed reference frame (Nature 409 (2001) 85). Here, we test whether stationarity per se is a criterion in SfM by pitting it against rigidity. We have created stimuli that, for a moving observer, offer two interpretations: one that is rigid but non-stationary, another that is more stationary or less rigid. In two experiments, with subjects reporting either structure or motion, we show that stationary, non-rigid solutions are preferred over rigid, non-stationary solutions; and that when no perfectly stationary solutions is available, the visual system prefers the solution that is most stationary. These results demonstrate that allocentric criteria, derived from extra-retinal information, participate in reconstructing the visual scene. © 2001 Elsevier Science Ltd. All rights reserved.

*Keywords:* Structure-from-motion; Self-motion; Head motion; Reference frames; Spatial vision

## 1. Structure from motion and the moving observer

### 1.1. Background

It has long been known that retinal motion can serve as a cue for three-dimensional (3D) structure in human observers (von Helmholtz, 1867; Wallach & O'Connell, 1953; Rogers & Graham, 1979). However, an infinite number of interpretations, coupling 3D structure and movement, are compatible with any two-dimensional optic flow field. To extract a definite interpretation, the observer must therefore make use of a regularizer, an assumption to constrain the degrees of freedom in the problem. In both models of natural vision and in work on artificial vision, the most popular candidate for such a regularizer is the rigidity assumption (RA): the motion of an arbitrary object made up of  $N$  points has  $3N$  degrees of freedom, while that of a rigid object has only

6. It can be shown that this reduction in the number of degrees of freedom allows, in principle, for the extraction of a definite 3D interpretation (up to an arbitrary absolute distance), when a rigid solution exists, even from stimuli that are highly impoverished in space and in time (Ullman, 1979; Koenderink, 1986).

One of the motivations behind the RA is an argument from ecological utility. A significant fraction of optic flow experienced by a normal observer is due to the observer's own movement in an environment where most of the objects are stationary in an external, allocentric, world-fixed reference frame. If retinal stimulation (which carries only egocentric information) is the only input to spatial vision—we shall call this the optic flow hypothesis—an observer moving through an environment of stationary objects should perceive the same 3D structure as a non-moving observer in an environment of rigid objects undergoing equal-and-opposite motion (Wallach & O'Connell, 1953). Thus, the argument goes, 'rigid motion' is ecologically prevalent, arising not only from external motion of rigid objects but also from the observer's locomotion through a station-

\* Corresponding author. Tel. +33-1-44-27-16-24; fax: +33-1-44-27-13-82.

*E-mail address:* wexler@ccr.jussieu.fr (M. Wexler).

ary environment. The two situations are not identical, as with the same optic flow, the moving observer perceives the environment as fixed, and the non-moving observer perceives the environment as moving (Ono & Steinbach, 1990). As far as the perception of structure, rather than motion, is concerned the hypothesis is nevertheless that the two observers will perceive the same 3D shape, as spelled out by Wallach:

The deformations of the retinal image that give rise to the kinetic depth effect are, therefore, the same whether O moves past a stationary object or whether a stationary O sees an object in partial rotation. *No*

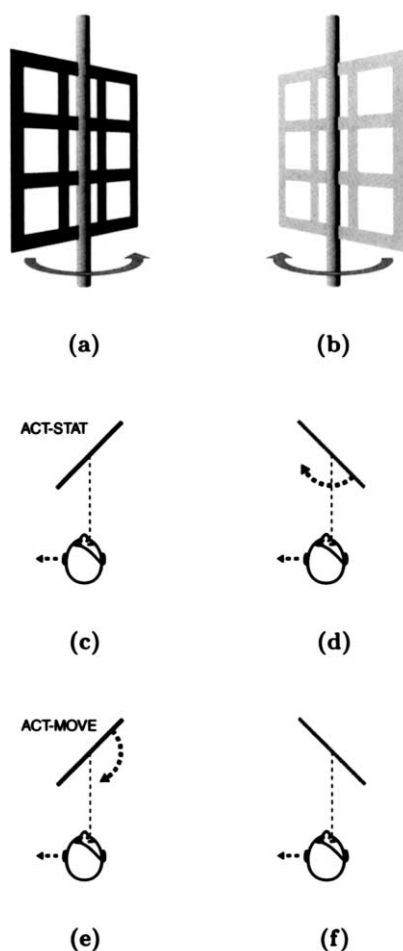


Fig. 1. Geometry of Experiment 1. The optic flow generated by rotating plane (a) is approximately the same as that generated by plane (b), with opposite angular velocity and tilt, as shown in Appendix A. An observer moving past a stationary plane while fixating one of its points (c) may therefore misinterpret the plane as rotating about the fixation point in the same direction as the observer's motion (and twice as fast as the observer's revolution), and having the opposite tilt (d) (condition ACT-STAT). Similarly, an observer moving past a plane that rotates twice as fast as his own revolution about the fixation point (e) may misinterpret the plane as stationary and having the opposite tilt (f) (condition ACT-MOVE). Interpretations (c) and (e) are rigid, while interpretations (c) and (f) are stationary.

*matter how these deformations are produced, the same three-dimensional object should be perceived.* In another respect, however, the two conditions produce different results: in the case where the object is stationary and O moves, the object is not perceived to rotate.

(Wallach, Stanton, & Becker, 1974, p. 339; italics ours)

The optic flow hypothesis makes a strong prediction: if we present a non-moving ('passive') observer with the same optic flow as experienced by a moving ('active') observer<sup>1</sup>, the two should perceive the same 3D structure and, more generally, exhibit the same performance on SfM tasks.<sup>2</sup> Rogers and Graham (1979) found support for the optic flow hypothesis: passive observers had 3D detection thresholds indistinguishable from those of active observers, and this result has been confirmed by subsequent work (Cornilleau-Pérès & Droulez, 1994). The hypothesis was called into doubt, however, by a finding that active observers extract SfM somewhat more efficiently from stimuli with low spatio-temporal coherence than do passive observers (van Damme & van de Grind, 1996).

In contradiction to the optic flow hypothesis, extraretinal information in active observers has been found to disambiguate certain 3D stimuli (Rogers & Rogers, 1992; Dijkstra, Cornilleau-Pérès, Gielen, & Droulez, 1995). It is well known that non-moving observers often experience what are called "motion reversals". This is due to an approximate invariance of optic flow to simultaneous inversions of motion and relative depth (see Fig. 1a–d, Appendix A). In those experiments, when observers actively generated the optic flow through their own motion, they experienced fewer motion reversals than passive observers.

Here, we re-interpret this active/passive difference in motion reversals. It can be noted that in the active conditions of the experiments of Rogers and Rogers (1992) and Dijkstra et al. (1995), the optic flow arises from the projection of stationary objects for moving observers (see Fig. 1c). Thus, in the passive condition, the simulated object and its motion-reversed counterpart move at equal (and opposite) speeds (see Fig. 1a, b), and the non-moving subject experiences a high rate of reversals. In the active condition, however, while the simulated (and, by construction, rigid) object is stationary *in an allocentric reference frame*, its motion-reversed counterpart moves twice as fast as the observer in an

<sup>1</sup> By a 'moving' observer, we mean one in voluntary motion, although his trajectory may not be voluntarily chosen. The distinction between voluntary and involuntary self-motion, which may have psychophysical consequences, will be taken up in future work.

<sup>2</sup> We limit our discussion of observer movement to eye translations, since eye rotations generate very little or no depth information.

allocentric frame (Fig. 1d). In our interpretation (borne out in Experiment 1), a *bias toward perceiving stationary objects* suppresses inversions for the active observer.

We are thus led to conjecture a bias toward perceiving stationary objects, a *stationarity assumption* (SA) for short, in addition to the rigidity assumption (RA), in the extraction of 3D structure and motion from optic flow. By a ‘stationary object’, we mean one that is immobile in an allocentric, *earth-fixed* reference frame. An object is more stationary than another if its points undergo less total motion in this allocentric frame. This parallels the notion of comparative rigidity in the RA (Ullman, 1984).

Most recently, we found strong evidence for the integration of extraretinal information in SfM, and further support for the SA (Wexler, Panerai, Lamouret, & Droulez, 2001). Using stimuli with conflicting perspective and motion cues, we showed that motion cues are used more by active than by passive observers; since a different *structure* was perceived from the same optic flow, this contradicts the optic flow hypothesis. The result only held, however, for stationary objects, showing that the RA coupled to the SA is stronger than the RA alone.

## 1.2. Present experiments

Here, we test the stationary assumption directly, by creating stimuli that dissociate stationarity and rigidity. The motion-reversal studies (Rogers & Rogers, 1992; Dijkstra et al., 1995), although in agreement with the SA, do not constitute a stringent test. The active–passive difference in reversals could be explained by subjects’ preferring a stationary solution, when one exists. But the stationary solution was also more rigid than the moving one. If the rigidity difference were undetectable (for small stimuli), the choice of the stationary solution was cost-free, i.e. no other known criterion needed to be sacrificed; if the non-rigidity were detectable, the stationary may actually have been chosen for its rigidity, and not for its stationarity.

In our experiments, we have devised stimuli for the active observer in which two solutions are possible, one that is strictly rigid but moving and the other less rigid but more stationary. We shall call these two solutions the ‘rigid’ and ‘stationary’ solutions, respectively—despite the fact that the second solution may not be perfectly stationary either (but it is always more stationary than the rigid solution). We shall call an object *perfectly* stationary if all of its points are fixed in an earth-fixed reference frame; an object is *more* stationary than another if its points move less in such a frame. For example, an object that rotates about its center is more stationary than the same object undergoing a translation in addition to the above rotation—a comparison

that we used in Experiment 2. In both experiments, we varied the stationary solutions’ non-rigidity (and therefore their ‘cost’, so to speak) by varying the stimulus size: as the stimuli become larger and larger, the non-rigidity of the stationary solution becomes increasingly evident. We also varied the task: in Experiment 1, the task was to report 3D structure, whereas in Experiment 2, the task was to report motion.

How can the stationarity assumption articulate with the rigidity assumption? From parsimony, we consider only models that treat actively generated and passively observed optic flow by the same mechanisms. Three hypotheses are possible.

- *Pure rigidity*: The observer always chooses the most rigid solution, i.e. one in which the perceived 3D object deforms the least.
- *Weak stationarity*: When several solutions are available that are perceived as equally rigid, the observer prefers the most stationary solution. If one solution is perceived as more rigid than all others, it is chosen without regard for its stationarity.
- *Strong stationarity*: An observer will sometimes choose a stationary solution, even if it is detectably non-rigid. Strong non-rigidity, however, may override stationary, i.e. there may be a trade-off between the SA and the RA.

A fourth logical possibility, *pure stationarity*, can be rejected because it yields absurd results for the passive observer. Pure stationarity works for objects that are actually stationary, but breaks down in the presence of observer-independent motion. For an immobile observer, for instance, pure stationarity is maximized when the flow is assumed to be projection of 3D motion perpendicular to the line of sight. Based on this observation, the reader can easily convince himself that a pure SA would fail to account for many common SfM demonstrations, i.e. the case of the passive observer. It is, of course, possible that the active observer uses the pure SA while the passive observer uses the pure RA, but we reject such a model by parsimony. In the two experiments that follow, we find empirical support for the strong stationarity hypothesis.

## 2. Experiment 1

In the first experiment, we compare the stationarity hypothesis to the rigidity hypothesis. In the crucial condition (ACT-MOVE), we present the observer with a stimulus that has two different rigid or nearly rigid interpretations: one that is more stationary and less rigid, and one that is less stationary and completely rigid. (In formal terms, the two stimuli are rigidity maxima.) In this way, we test the strength of the bias towards stationarity, when the choice of stationarity conflicts with the rigidity hypothesis.

Table 1  
Motion (in an allocentric reference frame) of the two solutions in each of the three conditions in Experiment 1

Condition	Rigid solution	Other solution
ACT-MOVE	Moving ( $2\omega$ )	Stationary (0)
PASS	Moving ( $\omega$ )	Moving ( $-\omega$ )
ACT-STAT	Stationary (0)	Moving ( $-2\omega$ )

The observer's eye revolves around the fixation point at angular speed,  $\omega$ .

The existence of these ambiguous stimuli is based on a symmetry of optic flow, illustrated in Fig. 1 and formally derived in Appendix A: simultaneously reversing the motion of a rotating plane and its tilt<sup>3</sup> results in approximately the same optic flow, or exactly the same optic flow in the limit of small stimuli or field of view.<sup>4</sup> If the optic flow resulting from Fig. 1a is misinterpreted as resulting from Fig. 1b, the object will appear to be non-rigid—a non-rigidity that will be greater for larger stimuli. Since our stimuli are drawn as a texture of points whose density on the monitor screen is uniform (in central position), with few or no depth cues other than optic flow, this symmetry will have major perceptual consequences.

Consider the common situation of an observer moving past a stationary surface while fixating one of its points, shown in Fig. 1c. If the center of the observer's eye revolves with angular speed,  $\omega$ , about the fixation point, the optic flow is the same as for a stationary observer and a plane rotating in the opposite direction, that is with angular speed  $-\omega$ , with respect to the observer: the flow resulting from the object rotation in Fig. 1a.<sup>5</sup> This can be confused with the nearly identical flow from Fig. 1b resulting from a plane with opposite tilt that rotates as speed  $+\omega$ , again with respect to the observer, but since the observer's eye already revolves at speed  $\omega$  about the fixation point in an allocentric

reference frame, the second percept rotates at speed  $2\omega$  in the same frame—as shown in Fig. 1d.

The situation just described corresponds to condition ACT-STAT of Experiment 1. The percept corresponding to Fig. 1c is both rigid and stationary, while that corresponding to Fig. 1d is neither stationary nor rigid, with the non-rigidity increasing for larger stimuli.<sup>6</sup> All hypotheses therefore predict that the percept of Fig. 1c will be chosen over that of Fig. 1d.

Consider now the less common, but experimentally more interesting, situation shown in Fig. 1e: an observer moves past a surface while fixating one of its points (the observer's eye revolving with angular speed,  $\omega$ , about the fixation point), while the surface rotates about the same axis, with a speed twice as high as the observer's, i.e.  $2\omega$ . In the observer's egocentric reference frame, the surface rotates with speed  $\omega$ . Applying the same symmetry as that in Fig. 1a–b, this percept can be confused with one that has opposite tilt, and angular speed  $-\omega$  with respect to the observer—and therefore speed 0 in an allocentric frame. This second solution, shown in Fig. 1f, is therefore *almost* stationary.<sup>7</sup>

This situation—which corresponds to condition ACT-MOVE—is interesting because it features conflict between rigidity and stationarity. The percept of Fig. 1e is perfectly rigid, but very non-stationary. The percept of Fig. 1f, however, is much more stationary, but non-rigid—with the non-rigidity increasing as a function of stimulus size. This stimulus allows us to 'put a price' on stationarity, i.e. to compare this bias with the bias towards rigidity.

We test the importance of stationarity, and its weight relative to rigidity, by varying the stimulus size in the ACT-MOVE condition. Since the rigid and stationary solutions differ in tilt by  $180^\circ$ , the responses allow us to calculate the relative frequency of the two solutions in any condition. For very small stimuli, higher-order terms are probably undetectable, so that both the rigid and the stationary solutions appear equally rigid; if the visual system chooses the stationary solution, this choice will therefore be 'free', i.e. bear no cost in rigidity. For larger stimuli, higher-order terms in optic flow will presumably become detectable, and the stationary solution will therefore be seen as non-rigid. Will it continue to be chosen, despite its growing violation of the RA?

<sup>3</sup> The slant of a surface is the angle between its normal and the normal to the frontoparallel plane, while the tilt is the difference of its normal projected into the frontoparallel. If we express the normal of a surface as  $(\sin \sigma \cos \tau, \sin \sigma \sin \tau, \cos \sigma)$ ,  $\sigma$  is the slant,  $\tau$  the tilt. Multiplying the slant by  $-1$  is therefore the same as adding  $180^\circ$  to tilt.

<sup>4</sup> This symmetry is exact in the limit of very small stimuli. Formally, this is because only in this limit do first-order terms dominate higher-order terms. For finite-size stimuli, the rigidity of the inverted solution becomes weaker with increasing size, as shown in Appendix A. Physiologically, the symmetry can be said to be exact when the higher-order terms are undetectable by the visual system (small stimuli), and broken when they are (large stimuli). This is justified by the decrease in inversions for passive observers as the stimulus size increase, as found in Experiment 1.

<sup>5</sup> If the observer's motion is not a pure rotation about the fixation point, the component of revolution may be replaced by an opposite rotation of the plane.

<sup>6</sup> More precisely, rigidity has two maxima: a single global maximum corresponding to the perfectly rigid solution of Fig. 1c and a local maximum, close to the solution of Fig. 1d. The tilt of the local maximum is always opposite that of the global maximum, while the slant may be different. This is demonstrated in Appendix A.

<sup>7</sup> Since the percept of Fig. 1f is non-rigid, it is a fortiori non-stationary, but it is nevertheless much more stationary than the percept of Fig. 1e. For convenience, we will refer to it as the stationary solution.

In order to test the detectability of higher-order terms in optic flow, we introduce a passive condition (PASS), in which the subject does not move, but where the rotational component of optic flow is replayed, frame-by-frame, from previously recorded active trials. In PASS, the two solutions are equally non-stationary, rotating with equal speed in opposite directions. In addition, one solution is non-rigid; the detection of this non-rigidity (as borne out by a preference for rigid solution) is used as a test for the detectability of higher-order terms in optic flow. In the limit of small stimuli, the two solutions should be chosen equally often in PASS; as stimuli increase in size, the ‘stationary’ solution should be chosen less often and the rigid solution more often.

In summary, in each of the three conditions (ACT-MOVE, PASS, ACT-STAT), there exists a completely rigid solution, which corresponds to the virtual object actually used to generate the display. In an allocentric reference frame, the rigid solution is stationary in ACT-STAT, but moves in the other conditions. In addition, there is always a second solution, non-rigid but corresponding to a local rigidity maximum. In ACT-MOVE, the second solution is almost stationary. This is summarized in Table 1.

The subject’s task is to indicate the average slant and tilt of the perceived surface. Since, in each condition, the tilts of the two solutions differ by  $180^\circ$  (see Fig. 1a, b), we are able to infer which of the two solutions the subject saw (for example, if the subject saw the stationary solution in conditions ACT-MOVE and ACT-STAT). This technique has a major advantage over simply asking the subject whether he saw a stationary object, although this is of course the point of interest. Had we asked about stationarity, a preference for the stationary solution could have been due to cognitive or response bias, rather than a reflection of perceptual reality: for example, the subject could always *respond* ‘stationary’ without actually experiencing the corresponding percept. With our response method, on the other hand, subjects do not know the one-to-one rela-

tion between tilt and stationary.<sup>8</sup> Therefore, any preference for the stationary solution cannot be due to cognitive factors.

Each of the three hypotheses proposed in Section 1 makes a distinct prediction concerning the outcome of this experiment. We assume that, in the limit of small stimuli, the deformation of the non-rigid solution will become undetectable, and the two solutions will be perceived as equally rigid.

*Pure rigidity:* There is always a perfectly rigid solution, and the difference between the rigid and non-rigid solution should be equally detectable in all conditions. Thus, this hypothesis predicts that the responses will be the same in all conditions. For small stimuli, the difference in rigidity between the two solutions should be undetectable, and in the small-stimulus limit the rigid and non-rigid solutions should be chosen equally often. As stimuli become larger, the rigid solution should be increasingly preferred.

*Weak stationarity:* In the limit of small stimuli, where non rigidity is undetectable,

- in ACT-MOVE the observer will choose the stationary but non-rigid solution;
- in PASS the two solutions will be chosen equally often;
- in ACT-STAT the rigid solution will be chosen, because of its stationarity.

As soon as the rigidity of the two solutions is distinguishable, in PASS the rigid solution should be chosen reliably more often. In this regime, the non-stationary, rigid solution in ACT-MOVE should be chosen as frequently as the rigid solution in PASS.

*Strong stationarity:* As for weak stationarity, in the limit of small stimuli the observer prefers the stationary solution in ACT-MOVE and chooses the two solutions equally often in PASS. When, with larger stimuli, the non-rigidity becomes detectable, and the rigid solution becomes preferred in PASS, the stationary solution is still chosen more often in ACT-MOVE due to its stationarity—even though its non-rigidity is detectable. For even larger stimuli, when non-rigidity becomes very strong, observers choose the rigid solutions in all conditions.

These predictions are summarized in Fig. 2.

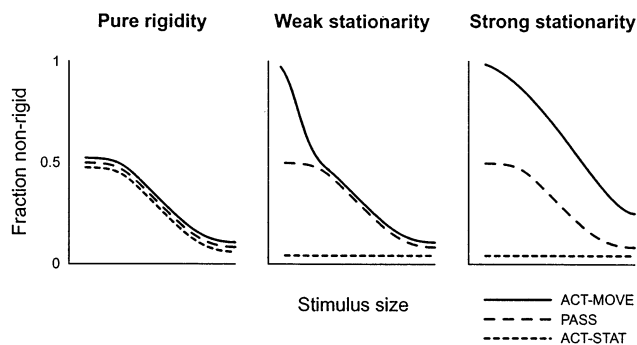


Fig. 2. Schematic predictions of the three hypotheses for Experiment 1, expressed as the frequency that the non-rigid solution perceived.

## 2.1. Methods

### 2.1.1. Apparatus

Subjects’ translational eye displacement was measured by a special-made precision mechanical head tracker (Panerai, Hanneton, Droulez, & Cornilleau-Pérès, 1999), which features submillimeter within-trial precision. The head tracker was sampled at the same

<sup>8</sup> Also, subjects cannot know the motion–structure relation, even if they are not naïve—tilt is varied randomly from trial to trial.

rate as the display monitor, 85 Hz. The tracker latency was lower than the inter-sample interval. The same computer (Pentium II 400 MHz) sampled the head tracker (using a National Instruments PCI-6602 card) and controlled the stimulus display (Sony GDM-F500 CRT monitor driven by a Matrox G400 video card at a resolution of about 5 arcmin/pixel).

### 2.1.2. Stimuli

Trials were in one of the three conditions: ACT-STAT, ACT-MOVE, and PASS. In ACT-STAT, subjects performed a lateral head motion past a virtual stationary object (Fig. 1c). In ACT-MOVE, the virtual object rotated about the fixation point, at a speed twice that of the observer's revolution (Fig. 1e). In the PASS condition, the subject experienced the same optic flow as in a previous active trial (minus the expansion or contraction component), but without performing head motion. What follows is a complete technical description of the generation of these stimuli.

Subjects viewed stimuli with their dominant eye, with the other eye blindfolded. Let  $\mathbf{r}_0 = (x_0, y_0, z_0)$  be the initial position of the dominant eye on a given trial.<sup>9</sup> In ACT-MOVE and ACT-STAT, we generated 200 points in the  $z = 0$  plane, centered about  $\mathbf{R}_0 = (x_0, y_0, 0)$  and with a radius of 1, 2, 4, 6 or 8 cm (about 7°, 14°, 28°, 41° and 53° of visual angle with a typical distance of 16 cm). The points were then projected (in perspective projection, here and throughout) from  $\mathbf{r}_0$  onto a slanted plane passing through  $\mathbf{R}_0$ , yielding a set of points in space,  $\mathcal{P}_0$ , that formed the virtual object. The surface slant was 45°, and tilt varied between 0° and 345°, in steps of 15°.

On each subsequent frame (i.e. interval between monitor retrace, about 12 ms), the following operations were performed: the current eye position ( $\mathbf{r}$ ) was read from the head tracker;  $\mathcal{P}_0$  was transformed to a new set,  $\mathcal{P}$ , depending on the eye position  $\mathbf{r}$ ;  $\mathcal{P}$  was projected onto a set of pixels, drawn white on black. In ACT-MOVE, we first calculated the rotation,  $\Omega$ , that transforms  $(\mathbf{r}_0 - \mathbf{R}_0)/|\mathbf{r}_0 - \mathbf{R}_0|$  to  $(\mathbf{r} - \mathbf{R}_0)/|\mathbf{r} - \mathbf{R}_0|$ .  $\Omega^2$  was then applied to  $\mathcal{P}_0$  to generate  $\mathcal{P}$  (see Fig. 1). In ACT-STAT, the points were stationary:  $\mathcal{P} = \mathcal{P}_0$ .

In each trial of PASS, subjects experienced the same rotational optic flow as in a previous active trial,  $T$ , but without performing head movement. Let  $\bar{\mathbf{r}}_0$  and  $\bar{\mathbf{R}}_0$  be the initial eye position and stimulus center in  $T$  (barred variables refer to  $T$ ). Let  $\mathbf{r}_0 = (x_0, y_0, z_0)$  be the initial position of the dominant eye in the passive trial;  $\mathcal{P}_0$  was the same as in  $T$ , but now centered about  $\mathbf{R}_0 = (x_0, y_0, 0)$  rather than  $\bar{\mathbf{R}}_0$ . On each subsequent frame, the

following steps were performed: the eye position,  $\mathbf{r}$ , was read from the head tracker, and the eye position of the corresponding frame of  $T$ ,  $\bar{\mathbf{r}}$ , was read from a file; a rotation,  $\Omega$ , was calculated that transforms  $(\mathbf{r} - \mathbf{R}_0)/|\mathbf{r} - \mathbf{R}_0|$  to  $(\bar{\mathbf{r}} - \bar{\mathbf{R}}_0)/|\bar{\mathbf{r}} - \bar{\mathbf{R}}_0|$ ;  $\Omega^{-1}$  was applied to  $\mathcal{P}_0$  in order to generate  $\mathcal{P}$ ;  $\mathcal{P}$  was projected for the eye position  $\mathbf{r}$ .

In addition to the stimulus dots, there was also a small red fixation point (about 40 arcmin) at  $R_0$ . Apart from the stimuli, the experiment was performed in darkness.

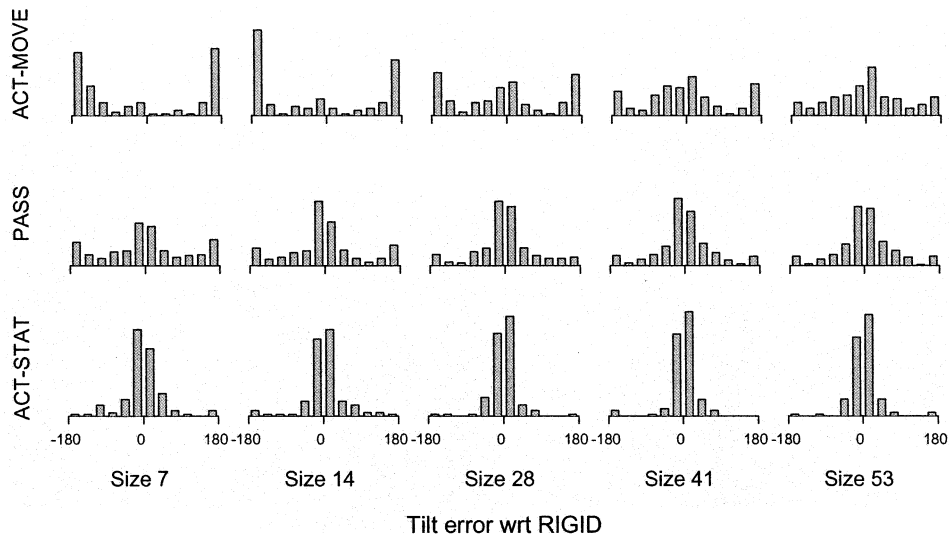
### 2.1.3. Procedure

In the ACT-MOVE and ACT-STAT conditions, the subject viewed the stimulus while making a head translation of 6 cm along the  $x$ -axis (i.e. horizontally and parallel to the monitor). A trial began when the subject's eye was within 5 cm of the ideal starting point, located 16 cm from the monitor. A fixation point then appeared, which the subject was instructed to fixate throughout the trial. The subject then performed a head movement to the right. When  $\Delta x$ , the distance from the starting position along the  $x$ -axis reached 3 cm, a brief tone was sounded, and the subject reversed the direction of movement. When  $\Delta x$  again dropped below 3 cm, the stimulus appeared, with the leftward movement continuing until  $\Delta x$  dropped below  $-3$  cm, at which point, the stimulus disappeared. The stimulus remained visible an average of  $2.5 \pm 0.3$  s. Note that actual trajectories were not arcs about the fixation point; the actual eye position was always used in calculating the projection of the virtual object.

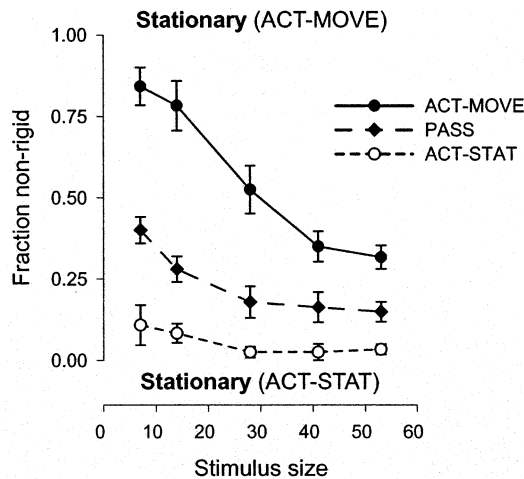
Following the disappearance of the stimulus, the subject returned to the central position, and a probe appeared on the screen. The task was to indicate average surface tilt and slant by adjusting the probe with a joystick. The probe object consisted of a projection of a circle (radius: 4.5°) and a normal to the circle (length: 7°) that extended from the center of the circle toward the subject. The subject adjusted the probe so that the circle appeared to lie in the same plane as the stimulus dots and so that the line appeared perpendicular to the stimulus. In the PASS condition, the subject viewed the stimulus without moving the head, responding as in the active conditions.

In the active block, half the trials were in the ACT-MOVE, and half were in the ACT-STAT condition. The experimental design was factorial; with 24 tilts and 5 stimulus size and two stimulus movement conditions, this yielded 240 trials in the active block. The passive block reproduced trial-for-trial, and in the same order, the previous active block. Five volunteers participated as subjects: three men and two women, ages 20–33 (including two of the authors and three naïve subjects). Following a training session, each subject performed the active block followed by the passive block.

<sup>9</sup> In the coordinate system used, the monitor screen is the plane  $z = 0$ ,  $\hat{x}$  points to the subject's right,  $\hat{y}$  up and  $\hat{z}$  toward the subject.



(a)



(b)

Fig. 3. Results of Experiment 1. (a) Distribution of tilt errors for all subjects, in the three movement conditions, for the five stimulus sizes. Tilt errors are with respect to the tilt of the rigid solution; in ACT-MOVE, tilt of the stationary solution differs by  $180^\circ$ . (b) Fraction of responses that are closer to the non-rigid than to the rigid solution. Error bars represent between-subject standard errors.

## 2.2. Results

To determine whether subjects perceived the normal, rigid solution or the inverted, non-rigid one, we analyzed response tilt, which differed by  $180^\circ$  for the two cases. On each trial, we transformed the response tilt by subtracting the tilt of the rigid solution; henceforth, when we refer to tilt response, we mean the tilt with respect to that of the rigid solution. Therefore, a tilt response of  $0^\circ$  indicates that the subject perceived the rigid solution, while a tilt response of  $180^\circ$  indicates the non-rigid solution. Fig. 3a shows the distribution of tilt responses. As can be seen, the responses cluster around the rigid and non-rigid solutions. The bimodality seen in Fig. 3a is also present in individual subject

data; we never found unimodality about some intermediate tilt.

The stationarity and rigidity properties of the two solutions are summarized in Table 1. In ACT-MOVE, tilt response  $0^\circ$  corresponds to the rigid, non-stationary solution, whereas tilt  $180^\circ$  is the stationary, non-rigid solution. (The solution at  $180^\circ$  is not completely stationary, either (since it is non-rigid), but it is *more* stationary than the solution at  $180^\circ$ .) In ACT-STAT, the solution at  $0^\circ$  is both rigid and stationary; the solution at  $180^\circ$  is non-rigid and non-stationary. In PASS, the two solutions are equally non-stationary; the solution at  $0^\circ$  is rigid, whereas the solution at  $180^\circ$  is non-rigid.

Results depend on both movement condition and stimulus size. In ACT-MOVE, the stationary solution

was seen for the smaller stimuli, while the rigid solution was more often chosen for larger stimuli. In ACT-STAT, the non-stationary, non-rigid solution was almost never chosen, while there was a strong peak around the stationary and rigid solution. In PASS, the stationary and rigid conditions were chosen with almost equal frequency for the smallest stimuli; for larger stimuli, the rigid solution was chosen with higher frequency.

We counted the trials in which tilt response was more than  $90^\circ$  away from the tilt of the rigid solution (i.e.  $0^\circ$ ). Fig. 3b shows fractions of such non-rigid trials, namely their number divided by the total number of trials in each condition. Performing a condition  $\times$  size ANOVA on these fractions, we find the effects of condition and size, as well as their interaction, significant ( $F_{2,8} = 84.8$ ,  $F_{4,16} = 21.5$ ,  $F_{8,32} = 20.0$ , respectively).<sup>10</sup> Even if we exclude ACT-STAT, all three effects remain significant ( $F_{1,4} = 22.0$ ,  $F_{4,16} = 47.6$ ,  $F_{4,16} = 7.11$ , respectively). Performing individual  $t$ -tests (with Bonferroni correction), we find that:

- the non-rigid fraction in ACT-MOVE is greater than in PASS for all values of stimulus size other than  $41^\circ$  (one-tailed, unequal variance);
- the fraction of stationary responses in ACT-STAT is greater than in ACT-MOVE for stimulus sizes  $28^\circ$ ,  $41^\circ$  and  $53^\circ$  (one-tailed, unequal variance);
- the rigid solution is chosen more frequently than the non-rigid solution in PASS, for all stimulus sizes other than the smallest (one-tailed).

### 2.3. Discussion

In light of our results, we now return to the three hypotheses laid out above. The results contradict the pure rigidity hypothesis, which predicts that the allocentric criterion of stationarity plays no role at all in SfM, and therefore that the same solution should be

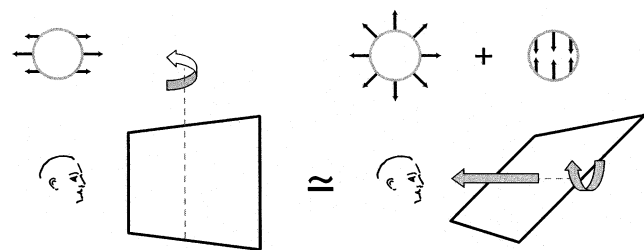


Fig. 4. Geometry of Experiment 2. On the left is a plane with horizontal tilt, rotating about a vertical axis. On the right is a plane with vertical tilt, approaching the observer while rotating about a horizontal axis. Both motions results in the same first-order optic flow, whose components are shown above the corresponding human figure.

<sup>10</sup> We adopt 0.05 as the criterion for statistical significance throughout.

chosen in the ACT-MOVE and PASS conditions. The results also contradict the weak stationary hypothesis, according to which a stationary solution may be preferred, but only if it is no less rigid than the non-stationary solution. This is not the case, since the rigid solution in PASS is always chosen more frequently than the non-rigid solution, and so the higher-order components of optic flow are detectable even for the smallest stimuli; and yet, the rigid solution is always chosen less frequently in the ACT-MOVE than in PASS, in favor of the stationary solution.

Our results do support the strong stationarity hypothesis: the more stationary a solution, the more frequently it is chosen, even at the cost of violating the rigidity assumption. We know that the non-rigidity of the stationary solution is detectable, since it is chosen less and less often in the PASS condition as stimulus size increases.

In this experiment, stationarity is an all-or-nothing criterion; but would this result still hold if the ‘stationary’ object (i.e. one that was *more* stationary) not only deformed, but also clearly moved? To answer this question, we performed a different experiment using a different class of stimuli and a task involving 3D motion rather than structure.

### 3. Experiment 2

Here we exploit another approximate symmetry of optic flow, illustrated in Fig. 4 and derived in Appendix A. Adding a rotation about an axis in the frontal plane to a sagittal translation may cancel the translation and result in a rotation about a different axis—in the limit of small stimuli. For finite-size objects, this invariance results in a second deforming solution, whose non-rigidity grows with stimulus size.

In agreement with this invariance, we previously noticed the following interesting phenomenon. A non-moving observer watches the optic flow from a slanted plane undergoing sagittal translation, while the plane simultaneously rotates about a frontal axis. The rotation of the plane is coupled to its translation in depth, with angular speed  $\omega = V_z / (E_z \tan \sigma)$  (where  $V_z$  is the translation speed,  $E_z$  is the distance to the eye, and  $\sigma$  is the slant). The observer does not perceive this object, however. Instead, he sees the second solution, the one that does not translate, but that is also non-rigid. This perceptual choice is interesting from the point of view of the stationarity assumption, since it implies—as for the stimulus used in Experiment 1—that the visual system prefers to see a more stationary object, even when it is non-rigid. As opposed to Experiment 1, however, it also implies that stationarity is not an all-or-nothing criterion: when no perfectly stationary percept is available (since both solutions undergo rota-



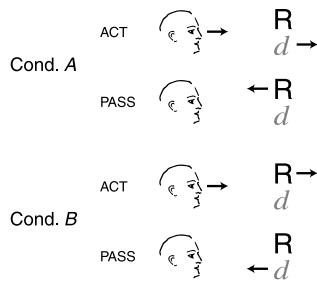


Fig. 5. Summary of conditions in Experiments 2. The rigid object, represented by 'R', and the deforming solution, represented by 'd', are sometimes stationary, and sometimes move in an allocentric reference frame. In active trials, the deforming solution (in condition A) and the rigid solution (in condition B) move with the same phase and amplitude as the observer.

tion at the same angular speed), the visual system does not disregard stationarity altogether but chooses the percept that is *most* stationary.

If the stationarity explanation of the above phenomenon is true, it implies that the result would be different if the translation were generated by the observer. If the object's rotation is coupled to translation (now the observer's) as before, the rigid solution should be perceived instead. Because of the translation being the observer's, rather than the object's, the rigid solution is now more stationary than the deforming solution. We tested this hypothesis in Experiment 2, where the situations just described are the passive (PASS) and active (ACT) trials in condition A, respectively. These are summarized in Fig. 5, where 'R' represents the rigid solution (the virtual rigidly moving object used to generate the stimulus), while 'd' represents the alternate, deforming solution, and the arrows represent movement with respect to an allocentric reference frame. The stationarity hypothesis predicts that—in spite of the same optic flow in active and passive trials—observers should always prefer the more stationary solution: the rigid solution in ACT, and the deforming solution in PASS.

Such a result, however, would be open to an alternative explanation: observers assign more weight to the rigidity criterion when they are moving. In order to distinguish this 'movement/rigidity' explanation from the stationarity hypothesis, we added a second condition, B, to Experiment 2. In the ACT trials of condition B, the center of the rigid object remained at a constant distance from the observer, and therefore underwent a translation in space, with the same amplitude and phase as the observer's sagittal movements (see Fig. 5). The rigid object underwent rotation at angular speed  $\omega = -V_z/(E_z \tan \sigma)$ , so that the center of the second deforming solution would be stationary in space. In the PASS trials, therefore, the rigid solution is more stationary than the deforming solution (Fig. 5). The stationarity hypothesis predicts that in condition B, the

active observer should prefer the deforming solution, while the passive observer should prefer the stationary solution—contrary to condition A. The 'movement/rigidity' hypothesis, however, predicts that the active observer should prefer the rigid solution, as in condition A.

The subject's task was to indicate the axis of rotation of the perceived object—which differed by  $90^\circ$  for the rigid and deforming solutions. As in Experiment 1, we preferred such an indirect task to asking the subject directly about stationarity, which could have opened to the door to cognitive or response bias. As a complement to Experiment 1, where the task concerned 3D shape, in Experiment 2, we asked the subjects about 3D motion.

In order to avoid non-sagittal motion, the center of the stimulus followed the subjects' movements along the  $x$ - and  $y$ -axis, remaining always at the point on the monitor directly opposite each subject's eye. This increased the rigidity of the deforming solution but decreased the stationarity of both solutions by the same amount. In order to deconfound active–passive and learning effects, we performed the experiment in four blocks in the order ACT, PASS, ACT, PASS. In half the subjects, the first two blocks were in condition A and the last two in condition B, while in the rest of the subjects, the order was condition B followed by condition A.

### 3.1. Methods

#### 3.1.1. Apparatus

The apparatus was the same as that in Experiment 1.

#### 3.1.2. Stimuli

The stimulus consisted of 50 points lying on a slanted plane, chosen so that in the observer's initial position, their projections were distributed with a radial distribution proportional to  $r^{-1/3}$ , where  $r$  is the distance from the center. The initial density of points was thus circularly symmetric, with a density proportional to  $r^{-4/3}$ . (This non-uniform distribution was chosen to have relatively few points near the boundaries, in order to give the stimulus an irregular shape.) The plane had a slant  $45^\circ$  and tilt  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  or  $270^\circ$ . Stimuli were of five sizes with the points chosen with a radius no greater than 1, 2, 4, 6, or 8 cm ( $0.95^\circ$ ,  $1.9^\circ$ ,  $3.8^\circ$ ,  $5.7^\circ$  and  $7.6^\circ$  at an average distance of 60 cm).

In the active condition (ACT), the stimulus plane rotated as the observer performed oscillatory, forward-and-back sagittal movements. In order to prevent any movement between the observer and the stimulus other than translations along the  $z$ -axis, the stimulus translated in the image plane, so that its center was always opposite the observer's eye (i.e. if at any given the eye was at point  $(x, y, z)$ , the stimulus center was at  $(x, y,$

0)). In condition *A*, the center of the virtual object always lay on the monitor. In condition *B*, the center of the object was always 60 cm away from the eye along the *z*-axis.

The objects rotated about a frontoparallel axis passing through their center and perpendicular to tilt. The angular speed was  $V_z/(E_z \tan \sigma)$ , where  $V_z$  is the translation speed,  $E_z$  is the distance to the eye, and  $\sigma$  is the slant. In condition *A*, the direction of rotation was such that the deforming solution was always at the same distance to the eye. In condition *B*, the rotation was such that the center of the deforming solution always lay on the monitor. Following rotation, the stimulus dots were projected in perspective for the current eye position and displayed as pixels, as in Experiment 1.

In the passive condition (PASS), the observer experienced identical optic flow as in a corresponding ACT trial, but without performing head movements. The center of the virtual object was always located directly opposite the subject's eye, as in ACT. The virtual object was composed of the same set of dots as in the corresponding ACT trial. On each frame of a PASS trial, the virtual object was at the same distance to the observer's eye and was rotated about its center by the same amount, as in the corresponding frame of the corresponding ACT trial.

### 3.1.3. Procedure

A trial began when the subject's eye was within 5 cm of the ideal starting location, located 60 cm from the screen. In ACT, the subject was instructed to make back-and-forth head movements, with as little left/right and up/down movement as possible. The subject began the movement backwards (away from the screen), while only the frame and fixation point were visible. When the subject's eye position relative to its starting point ( $\Delta z$ ) reached 5 cm, the stimulus dots appeared, and a brief tone was sounded for the subject to reverse movement direction. When  $\Delta z$  attained  $-5$  cm, another tone was sounded, etc. The stimulus disappeared after the subject performed 2.5 such back-and-forth cycles. The trial was restarted if the movement amplitude  $\Delta z$  exceeded  $\pm 10$  cm, or if the duration of any single movement cycle was over 6 s.

The task was to indicate the axis of stimulus rotation, by rotating a probe line in the frontal plane using a joystick. The subject lined up the probe (which appeared immediately following the stimulus dots) with the perceived axis. The axis angle was read off between  $0^\circ$  and  $180^\circ$ .

The optic flow and the task were identical in the passive condition (PASS). The experimental design was factorial: five stimulus sizes and four tilts yielded 20 trials in each block. The block of condition *A* PASS trials immediately followed the condition *A* ACT trials, and similarly for condition *B*. Half the subjects (ran-

domly chosen) performed condition *A* followed by condition *B*, while the other half ran the experiment in the reverse order. Prior to starting the experiment, subjects performed a training block to become familiar with the stimuli and the task. Eight volunteers, all men, aged between 22 and 38 years, participated in the experiment, including one of the authors; the rest of the subjects were naïve concerning the experimental hypotheses.

### 3.2. Results

On ACT trials, subjects performed mostly sagittal motion, as instructed. The mean standard deviation of the *z* coordinate of the eye position was 4.9 cm (corresponding to an amplitude of 7.0 cm), while the mean standard deviations of the *x* and *y* coordinates were 0.7 and 0.6 cm, respectively.

Axis responses will be presented relative to the axis of the rigid solution. The responses run from  $-90^\circ$  to  $+90^\circ$  (since axis *A* +  $180^\circ$  is the same as axis *A*). Since there was no clockwise/counterclockwise asymmetry, we will consider only the absolute value of the responses, which run from  $0^\circ$  to  $90^\circ$ . A response of  $0^\circ$  indicates the rigid solution, a response of  $90^\circ$  the deforming solution. We found no effects of stimulus size, nor of order of conditions (*A*, *B* versus *B*, *A*). We will thus present all data confounded over these variables.

The distribution of responses is shown in Fig. 6a. As can be seen clearly, subjects preferred the more stationary percept in all conditions: the rigid solution in active trials of condition *A* and passive trials of condition *B*, and the deforming solution in passive trials of condition *A* and active trials of condition *B*. Although the optic flow was identical in ACT and PASS trials of condition *A*, the responses were very different. Thus, this ACT/PASS effect clearly demonstrates the contribution of extra-retinal information.

To quantify these effects, we calculated the ratios of 'non-rigid trials' to all trials (where a non-rigid trial was defined as one in which the response was greater than  $\pm 45^\circ$ ). These fractions, as a function of condition and self-motion, are shown in Fig. 6b. A condition (*A*, *B*)  $\times$  self-motion (ACT, PASS) stimulus size ANOVA, with condition order (*A*, *B* versus *B*, *A*) yielded a significant condition  $\times$  self-motion interaction ( $F_{1,5} = 62.1$ ), as well as a significant main effect of condition (more non-rigid responses in *A* than in *B*,  $F_{1,5} = 11.7$ ). The responses in ACT differed significantly from those in PASS in both conditions ( $F_{1,5} = 37.0$  in *A*,  $F_{1,5} = 84.1$  in *B*).

### 3.3. Discussion

The results rule out the pure rigidity hypothesis. In spite of experiencing the same optic flow in the active

and passive trials, subjects showed a very strong bias towards different solutions in ACT and PASS. They tended to choose the more stationary solution in all cases: in condition *A*, the rigid solution in ACT and the deforming solution in PASS, and in condition *B*, the deforming solution in ACT and the rigid solution in PASS. The results cannot be explained by a different weight assigned to rigidity in the moving observer, since this would predict the same responses in condition *B* as in condition *A*, which is clearly not the case.

There is an additional effect that deserves comment. Apart from the presence of self-motion, we see from Fig. 5 that condition *A* ACT is similar to condition *B*

PASS: the center of the rigid solution is stationary, while that of the deforming solution translates. Therefore, it would seem that we should have the same results in these two conditions, and by analogous considerations, in condition *A* PASS and condition *B* ACT. Examining Fig. 6, we see that this is not entirely so: the results of the PASS conditions are more peaked at the extremes of stationarity than those of the ACT conditions. At first glance, this might suggest that non-moving observers assign a higher weight to stationarity.

However, a simpler explanation is possible. In ACT trials, subjects performed some non-sagittal (left–right, up–down) movements, and since the stimulus tracked

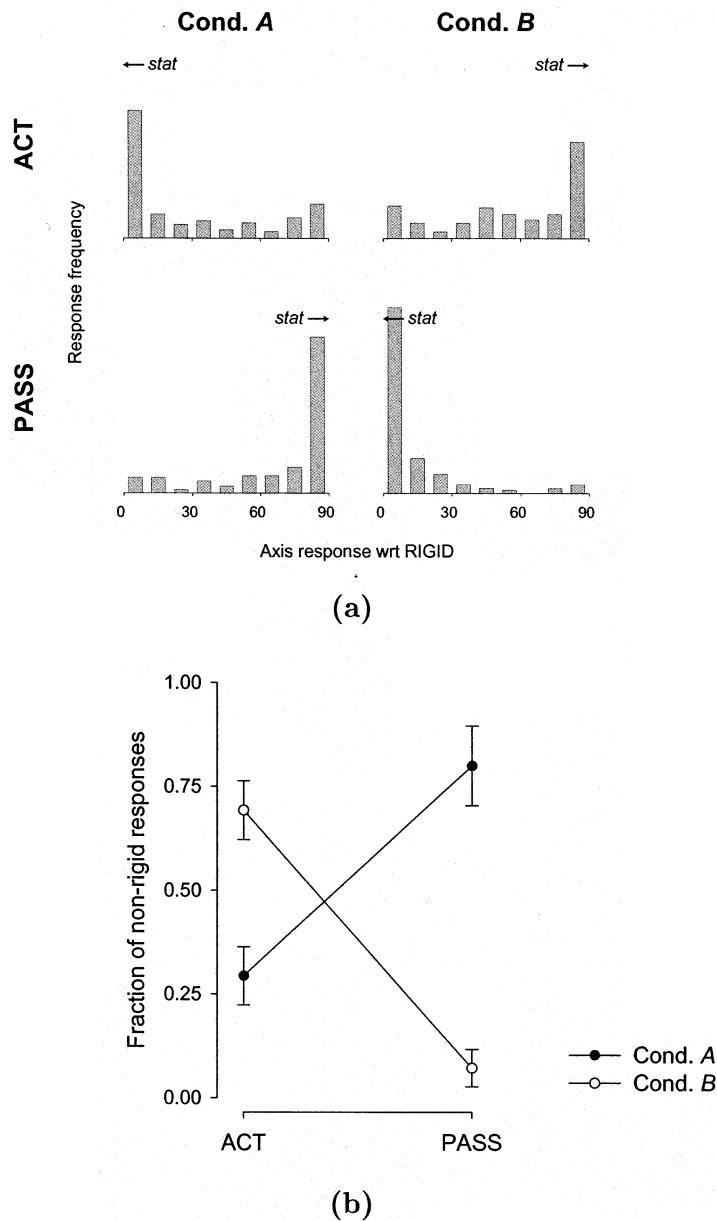


Fig. 6. Results of Experiment 2. (a) Distribution of axis responses for all subjects, in the ACT and PASS conditions, in conditions *A* and *B*. Response 0° indicates the rigid solution, 90° the deforming solution. Predictions of the stationary hypothesis are marked by 'stat'. (b) Fraction of responses closer to the non-rigid solution than to the rigid solution. Error bars represent between-subject standard errors.

these movements, they made all possible percepts less stationary. In PASS trials, however, subjects performed very little movement at all. The greater magnitude of unintended head movements in ACT than in PASS trials therefore explains why PASS responses are more strongly peaked about the stationary solutions.

Contrary to Experiment 1, in this experiment we used a motion, rather than a structure, task. Even though this is a motion task, it is still not open to cognitive bias, since it did not involve reporting whether the object was stationary or moving (Ono & Steinbach, 1990), but rather the axis of rotation. The rotational motions of the rigid and deforming solutions differ relative to the observer, and not in an allocentric reference frame. Thus, for this class of stimuli, perception of stationarity is not mere compensation for self-motion.

Apart from confirming the stationarity hypothesis for a different stimulus and a different task than Experiment 1, the results of Experiment 2 show that stationarity is not an all-or-nothing criterion. When no entirely stationary solution is available, and all solutions undergo perceptible motion, the visual system nonetheless shows a strong preference for perceiving the *most stationary* object.

#### 4. General discussion

In two experiments that involved different stimuli and tasks, we have shown that observers do not necessarily perceive the most rigid solution to the SfM problem. Instead, when a second, less rigid but more stationary, solution is available, this solution is often the one perceived. Furthermore, this is so even when it can be demonstrated that the more stationary solution is detectably less rigid (Experiment 1), and when the more stationary solution nevertheless undergoes detectable motion (Experiment 2). Of the three hypotheses discussed in Section 1, our data thus rule out the rigidity and weak stationarity hypotheses, and support the strong stationarity hypothesis.

There are two causes of non-stationarity: deformation and rigid motion in space. In Experiment 1, the stationarity criterion makes the observer perceive the deforming rather than the rigidly moving solution. In Experiment 2, both solutions move, but one solution—the one that deforms—moves less (rotation only) than the other (rotation and translation), and this solution is the one perceived. Therefore, the SA not only operates when a truly stationary solution is present, but also favors the most stationary solution when all solutions undergo motion in space. Thus, in Experiment 2, the stationarity criterion paradoxically leads to the perception of a rotating surface. Furthermore, Experiment 2 shows that the SA is just as operative for the non-moving as for the active observer.

It is important to note that the bias towards stationarity, as uncovered by the present experiments, operates in the absence of any additional visual cues to stationarity, such as a visual background to the scene. It has been suggested that the principle means to distinguish self—from object-motion relies on the presence of such a background (Gibson, 1966). Our experiments, however, are performed in darkness, with the stimulus the only element visible. Therefore, a stationary criterion could only apply provided extra-retinal information (Wexler et al., 2001)

In this work, we have examined stimuli that, amongst their infinite number of structure-from-motion interpretations, admit two local maxima of *rigidity*, and have found that the more *stationary* of the two is often chosen by the visual system. We have thus found that two criteria play a role in the extraction of SfM, rigidity and stationarity. How can these criteria be combined? One possibility is that maxima of rigidity (Ullman, 1984) or some modified form of rigidity (Todd, 1982; Koenderink, 1986) are found first, with the stationarity criterion choosing between the possible solutions. This possibility seems unlikely, since it would require finding all or many of the local maxima of a high-dimensional function. A second possibility is that the criterion of stationarity is applied first, but this also seems unlikely since, in the case of non-stationary rigid objects, this would often lead to extremely non-rigid percepts, which is not the case.

A third possibility is that the visual system simultaneously maximizes rigidity and stationarity. This could happen in at least two different ways. Either the two are independent criteria, of which the visual system maximizes some weighted combination or, more parsimoniously, an assumption of stationarity could be implicit in the SfM system. Any system that solves for SfM must determine both 3D structure and 3D motion simultaneously. If, for example, an iterative procedure is used, the optimization could start from object motion equal-and-opposite to that of the observer, i.e. an object that is stationary. Where no stationary solution exists (either for a non-moving observer or, as in Experiment 2, for an active observer), a moving solution is eventually found. This simple scheme would predict that for identical optic flow, the determination of SfM would be faster for the active than for the non-moving observer. Work using stimuli with short dot lifetimes (van Damme & van de Grind, 1996) confirms this prediction.

Regardless of how the stationarity criterion is applied, and the details of how it articulates with rigidity assumption, its existence implies that 3D structure and motion are coded in the human visual system, implicitly or explicitly, in an allocentric reference frame. Neurophysiological evidence for allocentric coding has been discovered in the hippocampus of rats (O'Keefe &

Nadel, 1978) and, more recently, in the parietal cortex of monkeys (Synder, Grieve, Brotchie, & Anderson, 1998). How we come to have high level, allocentric representations starting with low-level egocentric sensory data has been one of the most compelling questions in psychology and neuroscience. Here, we have presented psychophysical evidence for an allocentric criterion, and therefore for implicit or explicit allocentric coding in visual processes leading to the perception of three-dimensional structure.

## Acknowledgements

We thank Francesco Panerai for technical assistance, and Jeroen van Boxtel for carefully reading the manuscript. M.W. wishes to thank Adélaïde de Chatellus for her hospitality at Saint-Cergue during the writing of this article.

## Appendix A

Consider a plane,  $\mathcal{P}$ , that has normal,  $\hat{\mathbf{N}}$ , passing through a point,  $\mathbf{R}_0$ , and let  $\mathbf{R}$  be a point on the plane  $\mathcal{P}$ .  $\mathcal{P}$  simultaneously undergoes translation with velocity,  $\mathbf{V}$ , and rotation with angular velocity,  $\mathbf{\Omega}$ , about an axis passing through the point  $\mathbf{R}_0$ . Therefore, the three-dimensional motion vector of the point,  $\mathbf{R}$ , is

$$\mathbf{V} + \mathbf{\Omega} \times (\mathbf{R} - \mathbf{R}_0). \quad (1)$$

Now, what is the projection of this three-dimensional motion onto a two-dimensional image plane? Let the image plane also pass through  $\mathbf{R}_0$  and have normal  $\hat{\mathbf{z}}$ . Let  $\mathbf{r}$  be the perspective projection of  $\mathbf{R}$  from the eye located at  $\mathbf{E}$  onto the image plane; our goal is to solve for the two-dimensional optic flow, i.e. the two-dimensional flow field as a function of  $\mathbf{r}$ ,  $\mathbf{u}(\mathbf{r})$ .

Let us express  $\mathbf{R}$  as

$$\mathbf{R} = \mathbf{E} + t(\mathbf{r} - \mathbf{E}). \quad (2)$$

Since  $(\mathbf{R} - \mathbf{R}_0) \cdot \hat{\mathbf{N}} = 0$ , we have

$$t = \frac{(\mathbf{R}_0 - \mathbf{E}) \cdot \hat{\mathbf{N}}}{(\mathbf{r} - \mathbf{E}) \cdot \hat{\mathbf{N}}}. \quad (3)$$

Eqs. (2) and (3) allow us to re-write the three-dimensional motion (Eq. (1)) as a function of position in the two-dimensional image. Projecting this vector onto the image plane, we find the general form of the two-dimensional flow:

$$\mathbf{u}(\mathbf{r}) = \frac{1}{t} \left[ \frac{B_z}{E_z} (\mathbf{r} - \mathbf{E}) + \mathbf{B} \right],$$

$$\mathbf{B} = \mathbf{V} + \mathbf{\Omega} \times [t(\mathbf{r} - \mathbf{E}) - (\mathbf{R}_0 - \mathbf{E})]. \quad (4)$$

The general optic flow (Eq. (4)) is greatly simplified in the limit of very small stimuli. In order to see this, we

set  $\mathbf{r} = (\rho \cos \varphi, \rho \sin \varphi, 0)$  and expand  $\mathbf{u}(\mathbf{r})$  as a Taylor series in powers of  $\rho$  about  $\rho = 0$ , i.e., about small field of view. For our purposes we consider only sagittal translation, i.e.  $\mathbf{V} = (0, 0, V_z)$ , and rotation about axes in the  $xy$  plane, i.e.  $\mathbf{\Omega} = (\omega \cos \alpha, \omega \sin \alpha, 0)$ . If we set  $\hat{\mathbf{N}} = (\sin \sigma \cos \tau, \sin \sigma \sin \tau, \cos \sigma)$  (where  $\sigma$  and  $\tau$  are the surface slant and tilt, respectively) and, without loss of generality, put  $\mathbf{E} = (0, 0, E_z)$ , we find that the lowest order terms in the Taylor series are of first order in  $\rho$ :

$$\begin{aligned} u_x(\rho, \varphi) &= \left[ \frac{V_z}{E_z} \cos \varphi - \omega \tan \sigma \sin \alpha \cos(\varphi - \tau) \right] \rho + \dots \\ u_y(\rho, \varphi) &= \left[ \frac{V_z}{E_z} \sin \varphi + \omega \tan \sigma \cos \alpha \cos(\varphi - \tau) \right] \rho + \dots \end{aligned} \quad (5)$$

where the ellipses represent second-order terms. In the first-order flow field, the term proportional to  $V_z$  is an expansion or contraction due to sagittal translation; the other term is identical to the flow field of a rotating plane under parallel projection.

The first-order flow field (Eq. (5)) presents a number of symmetries or ambiguities. In Experiment 1, we exploit the fact that slant and angular speed cannot be separately determined, but occur only in the combination  $\omega \tan \sigma$ . Thus, in the limit of small stimuli, optic flow is invariant under the transformation

$$\left. \begin{aligned} \omega &\mapsto f_\omega \omega \\ \tan \sigma &\mapsto f_{\tan \sigma} \tan \sigma \end{aligned} \right\} \text{provided that } f_\omega f_{\tan \sigma} = 1 \quad (6)$$

In particular, the first-order flow remains unchanged under the simultaneous transformations

$$\omega \mapsto -\omega \quad \sigma \mapsto -\sigma \quad (7)$$

or, equivalently,  $\omega \mapsto -\omega$ ,  $\tau \mapsto \tau + \pi$ . This invariance leads to the ‘motion inversions’ where the observer, mistaking the sign of his or her movement relative to the object, also inverts the surface orientation, leading to the discrete symmetry in tilt  $\tau \mapsto \tau + \pi$ .

What happens to the invariances (Eqs. (6) and (7)) for finite-sized stimuli? To help visualize the answer to this question, we calculated the degree of rigidity of different planar 3D solutions, as a function of the multiplicative  $f_\omega$  and  $f_{\tan \sigma}$  factors, using parameters similar to those in Experiment 1. Our calculation is not complete, as it assumes that all solutions are planar. This restriction is justified by the didactic nature of our computation, by the fact that planes are a good approximation to the real maximum of rigidity, and, empirically, by the subjects’ informal reports of their percepts. The simulations were performed as follows. A circular cloud of dots was generated, then projected onto a surface with slant  $\sigma = 35^\circ$  and tilt  $0^\circ$ . This virtual object was then rotated, by changing  $\sigma$  from  $35^\circ$  to  $55^\circ$ , in steps of  $4^\circ$ . (This simulated the observer’s

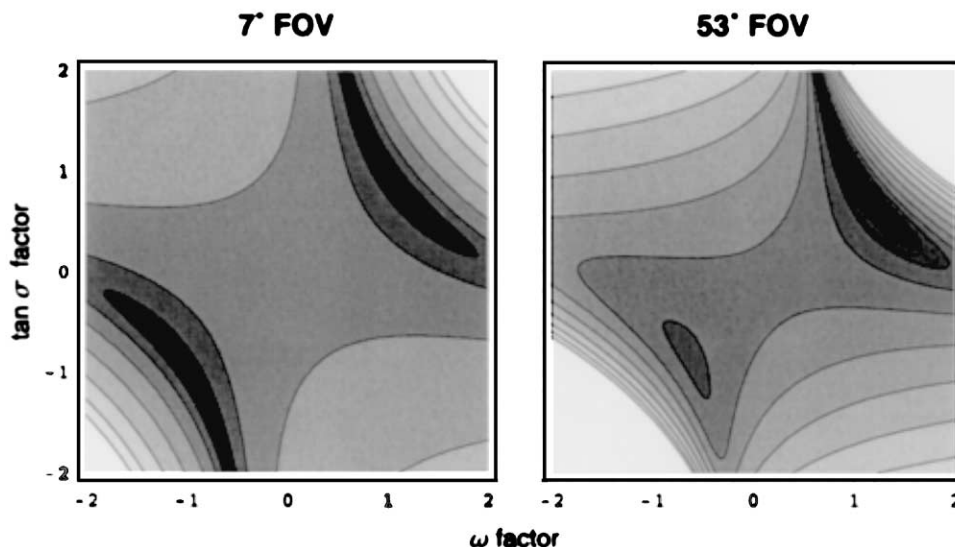


Fig. 7. Similarity of the rigidity landscape of Experiment 1, for small and large stimuli, as a function of the  $\omega$  and  $\tan \sigma$  factors. (Darker shading indicates a higher rigidity.) The scale is arbitrarily chosen.

motion in Experiment 1,  $\pm 3$  cm at 16 cm from the stimulus with slant  $45^\circ$  and tilt  $0^\circ$ .) At the same time, the proximal stimulus was projected onto a second surface with different slant and angular speed, given by  $\tan \sigma' = f_{\tan \sigma} \tan \sigma$  and  $\omega' = f_\omega \omega$ . The degree of rigidity of this second surface was calculated as follows. Let  $r_i^\alpha$  be the 3D position of point  $i$  in frame  $\alpha$ , with a total of  $N$  points and  $F$  frames. Then, we define rigidity as

$$R = - \sum_{\alpha=1}^{F-1} \sum_{i=1}^N \sum_{j=i+1}^N \frac{(|r_i^\alpha - r_j^\alpha| + |r_i^{\alpha+1} - r_j^{\alpha+1}|)^2}{|r_i^\alpha - r_j^\alpha| + |r_i^{\alpha+1} - r_j^{\alpha+1}|}. \quad (8)$$

Eq. (8) is a measure of rigidity because in a rigidly moving object, the distance between any two points is constant over time. For perfectly rigid motion, we will have  $R = 0$ , for non-rigid motion  $R < 0$ . Similar to the definition of rigidity used by Ullman (1984), Eq. (8) sums up the squares of the changes of interpoint distances, weighted by the inverse distance between the two points, so that deformations between nearby points are weighted heavier. Our results, however, would be similar given any sensible measure of non-rigidity.

The results of our simulations are given in Fig. 7, which, for stimuli sizes equal to the smallest and largest used in Experiment 1, shows the rigidity landscape as a function of the  $f_\omega$  and  $f_{\tan \sigma}$  factors. The point  $f_\omega = f_{\tan \sigma} = 1$  corresponds to the surface used to generate the stimulus, which is perfectly rigid by construction, and is therefore a global maximum of rigidity. There is also a local maximum near  $f_\omega = f_{\tan \sigma} = -1$ , corresponding to the inversion symmetry of Eq. (7). The important features of our results are as follows.

- For infinitely small stimuli, the two maxima in Fig. 7 are actually an infinite one-parameter family of solutions, lying on the hyperbolic curves  $f_\omega f_{\tan \sigma} = 1$ . Considering finite-size stimuli, and therefore includ-

ing the second-order terms in optic flow, reduces this infinite family down to two discrete solutions, as assumed in our discussion elsewhere in this article (see, for example, Table 1).

- As stimuli become larger, the local maximum near  $f_\omega = f_{\tan \sigma} = -1$  becomes less sharp compared to the global maximum  $f_\omega = f_{\tan \sigma} = +1$ . Therefore, the non-rigidity of the second solution in Experiment 1, for example, becomes greater for larger stimuli.

In Experiment 2, we consider the special case where the axis of rotation is perpendicular to the surface normal, i.e.  $\alpha = \tau + \pi/2$ . It then follows that the first-order flow (Eq. (5)) is invariant under the simultaneous transformations

$$\begin{aligned} V_z &\mapsto V_z + E_z \omega \tan \sigma \\ \tau &\mapsto \tau + \frac{\pi}{2} \\ \omega &\mapsto -\omega \end{aligned} \quad (9)$$

This invariance is heuristically illustrated in Fig. 4.

## References

- Cornilleau-Pérès, V., & Droulez, J. (1994). The visual perception of three-dimensional shape from self-motion and object-motion. *Vision Research*, 34(18), 2331–2336.
- Dijkstra, T., Cornilleau-Pérès, V., Gielen, C., & Droulez, J. (1995). Perception of three-dimensional shape from ego- and object-motion: comparison between small and large-field stimuli. *Vision Research*, 35(4), 453–462.
- Gibson, J. (1966). *The senses considered as perceptual systems*. London: George Allen and Unwin.
- Koenderink, J. (1986). Optic flow. *Vision Research*, 26(1), 161–179.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. New York: Oxford University Press.

- Ono, H., & Steinbach, M. (1990). Monocular stereopsis with and without head movement. *Perception and Psychophysics*, *48*(2), 179–187.
- Panerai, F., Hanneton, S., Droulez, J., & Cornilleau-Pères, V. (1999). 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.
- Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*(2), 125–134.
- Rogers, S., & Rogers, B. (1992). Visual and nonvisual information disambiguate surfaces specified by motion parallax. *Perception and Psychophysics*, *52*(4), 446–452.
- Snyder, L., Grieve, K., Brotchie, P., & Anderson, R. (1998). Separate body- and world referenced representations of visual space in parietal cortex. *Nature*, *394*, 887–891.
- Todd, J. (1982). Visual information about rigid and non-rigid motion: a geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 238–252.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Ullman, S. (1984). Maximizing rigidity: the incremental recovery of 3-d structure from rigid and nonrigid motion. *Perception*, *13*(3), 255–274.
- van Damme, W., & van de Grind, W. (1996). Non-visual information in structure-from-motion. *Vision Research*, *36*(19), 3119–3127.
- von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*. Hamburg: Voss.
- Wallach, H., & O'Connell, D. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*, 205–217.
- Wallach, H., Stanton, L., & Becker, D. (1974). The compensation for movement-produced changes in object orientation. *Perception and Psychophysics*, *15*, 339–343.
- Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature*, *409*, 85–88.