Motion perception by a moving observer in a three-dimensional environment

Lucile Dupin

Laboratoire Psychologie de la Perception, CNRS & Université Paris Descartes, Paris, France

Mark Wexler

Laboratoire Psychologie de la Perception, CNRS & Université Paris Descartes, Paris, France

Perceiving three-dimensional object motion while moving through the world is hard: not only must optic flow be segmented and parallax resolved into shape and motion, but also observer motion needs to be taken into account in order to perceive absolute, rather than observer-relative motion. In order to simplify the last step, it has recently been suggested that if the visual background is stationary, then foreground object motion, computed relative to the background, directly yields absolute motion. A series of studies with immobile observers and optic flow simulating observer movement have provided evidence that observers actually utilize this so-called “flow parsing” strategy (Rushton & Warren, 2005). We test this hypothesis by using mobile observers (as well as immobile ones) who judge the motion in depth of a foreground object in the presence of a stationary or moving background. We find that background movement does influence motion perception but not as much as predicted by the flow-parsing hypothesis. Thus, we find evidence that, in order to perceive absolute motion, observers partly use flow-parsing but also compensate egocentric motion by a global self-motion estimate.

Introduction

When an object moves in the environment, the movement induces translation and deformation of the object’s image on the retina. The retinal image also changes in the case of the observer’s motion: rotation of the eyes in the head or the eyes’ translation in space as the result of head or body movement. Each element of the three-dimensional (3D) scene will move quickly or slowly on the retina depending on its depth. Of course, both object and observer motions may occur at the same time, and therefore changes on the retina are a mixture of these two sources of motion. This ambiguity makes it hard to distinguish and isolate these two sources. For example, how does an observer perceive object motion alone, when this motion is intermixed with parallax flow resulting from the observer’s own movement? This difficulty belongs to a family of problems related to visual stability: how we perceive spatial properties (positions, velocities, sizes, shapes) independent of the deformations in the sensory data due to observer movement, and how we perceive motion in an earth-fixed, rather than egocentric, reference frame. We will use the term absolute motion to refer to motion in an earth-fixed reference frame.

One solution to this problem is to use the visual background with landmarks. When a background is present and is stationary in an earth-fixed reference frame, the earth-fixed and background-centered reference frames are identical. In order to calculate a foreground object’s absolute motion, one can simply calculate its motion relative to the background. We will call this the visual background solution. Evidence for its use is provided by cases in which the background moves with respect to an earth-fixed reference frame, so that the earth-fixed and background-centered reference frames are no longer identical. If the visual background solution is in fact used, then any absolute motion of the background should result in an equal-and-opposite error in the perception of foreground object motion. If the magnitude of the error is less than the background motion, then the usage of the visual background solution can be at most partial. A common example of at least partial use of the visual background solution is the illusory motion of the moon against racing clouds; a laboratory example is the Duncker illusion, in which a small stationary object surrounded by a larger moving background is perceived as moving in the opposite direction to the background (Duncker, 1930). These illusions are special cases of the phenomenon of induced movement (see Reinhardt-Rutland, 1988, for a
call this the self-motion solution. This self-motion estimate may come from extraretinal signals, such as an efference copy of the motor command, or vestibular and proprioceptive information. Efference copy has been most widely—though controversially— theorized to underlie another case of visual stability, that of two-dimensional position and motion during eye movement (see Wurtz, 2008, for a recent review). Extraretinal head movement signals have been shown to modify the perception of 3D motion: if the optic flow simulates a stationary object for a moving observer, the observer usually perceives the object as stationary or nearly so (Rogers & Graham, 1979); if the same optic flow is replayed for a moving observer, the object is perceived to move (Ono & Steinbach, 1990). Extraretinal signals can also modify the perception of 3D shape, favoring shape–motion combinations that are more nearly stationary in an earth-fixed reference frame (Wexler, Panerai, Lamouret, & Droulez, 2001; van Boxtel, Wexler, & Droulez, 2003); moreover, by comparing active and passive observer motion, it has been shown that both efference copy and proprioceptive and vestibular signals contribute to these effects (Wexler, 2003; Dyde & Harris, 2008).

In this study, we will disentangle visual background and self-motion solutions using a task in which an actively moving observer judges the motion of a foreground object, in the presence of a background that also moves. If the observer uses only the visual background solution, the background motion should bias the perception of foreground object motion with a gain of one. In other words, if the background turns with angular speed $\Omega$, the foreground object’s angular velocity should be biased by $-\Omega$, about the same rotation axis. If, on the other hand, the self-motion solution is used, then the background motion shouldn’t affect perceived foreground object motion. A recently published study (MacNeilage, Zhang, DeAngelis, & Angelaki, 2012) has also combined background motion with observer movement in order to test the flow-parsing hypothesis. Using only passive observer movement, MacNeilage et al. (2012) were able to show a significant but small effect of self-motion beyond what is predicted by the flow-parsing hypothesis.

Our study also includes two special cases that serve as control conditions. In one condition, the observer is immobile, while the foreground object and background move. This condition is similar to the experiments of Rushton and Warren (2005), in which observer movement is merely simulated using optic flow. This condition provides an additional check of the effect of background, independent of observer movement. In a second control condition, the observer moves and so does the foreground, but in the absence of any background. This condition provides an independent check on the role of self-motion information.

**Experiment 1**

In this experiment, moving or immobile subjects reported the perceived rotation direction of a planar object, rotating in depth about a vertical axis. In most trials, the object was accompanied by a moving background plane.

**Methods**

**Participants**

Eleven unpaid participants volunteered to take part in the experiment (nine men, two women; mean age 29 years). All had normal or corrected-to-normal vision. The subjects were the two authors, seven experienced psychophysical observers (but who were naïve as to the hypotheses of the study), and two inexperienced and naïve observers (SB, RD). Participants were treated in accordance with the relevant aspects of the Declaration of Helsinki.

**Apparatus**

Stimuli were displayed on a large stereoscopic plasma monitor (Panasonic TX-P65VT20E, image size 144 x 81 cm [Panasonic, Osaka, Japan]), set to its native 1920 x 1080 resolution. The side-by-side stereoscopic mode (in which the image for the left eye is compressed in the left half of the frame buffer, and the image for the right eye in the right half) was used, which effectively halved the horizontal resolution for each eye, down to 960 x 1080. The images for the two eyes were decompressed to full width and presented.
sequentially by the monitor at 120 Hz, with the synchronized shutter glasses ensuring that each eye sees only its own image. We used the active LCD shutter glasses delivered with the monitor (Panasonic TY-EW3D10E). The rapid decay time of the plasma monitor ensured a near-absence of “ghost” images (the left-eye image seen by the right eye and vice versa).

The position of the center of each eye in space was measured on-line using a LaserBird optical head tracker (Ascension Technology, Burlington, VT). The sensor (weight 40 g) was mounted on a light but tightly fitting helmet, which the participants wore on their heads. The head tracker provides the 3D position and orientation of the sensor, at a sampling frequency of 240 Hz, in the reference frame of the LaserBird scanner. These data were converted to the reference frame of the monitor by means of a specially developed calibration procedure involving touching the sensor to the monitor surface. The position of the eyes relative to the sensor was measured for each participant using another specially developed procedure, in which the participants optically aligned each eye with two sets of two points on a planar object. The eye positions were normalized so as to be spaced by 6.4 cm. The data were filtered using the LaserBird's DC and AC narrow notch filters, as well as its Kalman predictive filter with prediction interval set to 24 ms, to partly compensate for sensor latency.

**Stimuli**

The stimulus was a dynamic, stereoscopic projection of two virtual objects, a foreground plane and a background plane (on some trials, the background was absent). The geometry of the stimulus will be described in a virtual 3D space prior to projection, and in a subject-independent, earth-fixed reference frame. The foreground plane had size 30 by 30 cm, and was positioned so that its center coincided with the physical center of the monitor; the background plane had size 160 by 80 cm, and was centered at a point 40 cm behind the center of the monitor. Both planes’ edges were vertically and horizontally oriented. The foreground was textured using a green, nonperiodic pattern; the background was textured with a different, gray, nonperiodic pattern (see Figure 1).

A small green fixation point was positioned in the center of the foreground object. The foreground plane, presented on every trial, rotated about a vertical axis passing through its center. Its angular velocity (denoted \( V_o \)) was \( \pm 4, \pm 7, \text{ or } \pm 10 \) deg/s. (In our convention, positive angular velocities correspond to counterclockwise rotations as seen from the top.) The initial orientation of the foreground square was either \( +45 \) or \(-45 \) deg, also rotated about the vertical axis. The background could rotate, be static, or be absent. Its rotation was about the same axis as the foreground object, i.e., the vertical axis passing through the center of the foreground plane (see Figure 2). The reason for this choice of axes was to simulate the relative motion between foreground and background planes during self-motion. The background rotation velocity (denoted \( V_b \)) was 0, \( \pm 8, \pm 12, \text{ or } \pm 16 \) deg/s. The initial orientation of the background was calculated in order to obtain orientations symmetrical about the fronto-parallel at the beginning and the end of the trial. The two stimulus planes were projected onto the darkest possible background on our monitor. Participants wore sunglasses, inside the stereo shutter glasses, to further attenuate this background light, but the “black” background was nevertheless slightly visible (i.e., participants could faintly perceive the monitor’s edges).

The duration of the stimulus was 800 ms. Each trial had fixed values of foreground object and background velocities (\( V_o \) and \( V_b \), respectively). Velocity values within a block were determined using a factorial design with randomized order.

![Figure 1. One monocular frame from the stimulus in Experiment 1.](https://jov.arvojournals.org/)

![Figure 2. Schematic top view of the stimulus in Experiment 1.](https://jov.arvojournals.org/)
In order to generate each eye’s image, data from the motion pair tracker were polled immediately before each stereo pair was drawn (i.e., at 60 Hz), and the approximate positions in 3D space of the two eyes were calculated using the calibration data described above. The image for each eye was calculated using perspective projection, with the center of projection corresponding to the position of the eye. The foreground and background objects appeared simultaneously at the beginning of the trial and disappeared simultaneously at the end of the trial.

Procedure

Following the presentation of the stimulus, the participants’ task was to report the perceived direction of foreground object rotation in an earth-fixed reference frame. Participants were explicitly told to report motion with respect to the (unseen) experimental room and to ignore any motion of the background plane. The direction of rotation was reported as either leftward or rightward motion of the nearest edge of the foreground object (corresponding to clockwise or counter-clockwise rotations, respectively, of the foreground plane as would have been seen from above), and was indicated by clicking on the left or right mouse button.

There were two conditions, one in which the participant was mobile and one in which the participant was immobile. In the mobile condition the participant began each trial near the horizontal center of the monitor. For the purposes of motion control, the participant’s 3D position was taken as the midpoint between the eyes. On each trial there was a direction, left or right, in which the participant was supposed to move; leftward trials alternated with rightward trials, with the direction cued verbally before each trial. The trial began when the participant moved by 5 cm from the center in the corresponding direction. Participants were instructed to move the trunk along with the head, and to keep moving until the stimulus disappeared. They were also instructed to confine their movement to the horizontal axis parallel to the monitor surface as much as possible. Movement along the other axes and speed along the principal axis were not otherwise constrained. In following text, we will report movement in all three axes, showing that movement along the principal axis greatly exceeded that along the other axes. At the end of the trial, a black screen was displayed and the subject reported the perceived direction of rotation of the foreground object. Subjects were instructed to report the direction of object rotation in an earth-fixed frame, not relative to themselves. Average subject velocity, expressed as rotation about the stimulus center, was 17 deg/s (see the following Results and discussion for further details).

In the immobile condition the participant was instructed to remain still (mean velocity 0.08 deg/s), while the foreground and background underwent the same observer-independent motions.

Each condition was run in two separate blocks of 384 trials. The sequence of blocks was immobile, mobile, mobile, and immobile. The entire experiment took about 1 hr per subject.

Participants sat on a stool in front of the monitor, roughly centered horizontally with respect to the screen. The mean vertical position of subjects (calculated as the point midway between the eyes) was 9.9 cm below the vertical center of the monitor, and the mean distance from the monitor was 55.9 cm. The experiment was performed in near darkness.

Results and discussion

Use of visual reference by an immobile observer

In order to determine the role of the background as a visual reference, we first analyzed the data from the condition in which the subject was immobile and the background was present (rotating or stationary).

“Right” responses are coded as +1, “Left” as −1. An example of mean responses for one subject, as a function of object velocity $V_o$, for two extreme values of background velocity $V_b$, is shown in Figure 3a. Both object and background velocities are expressed in an earth-fixed reference frame. For each background velocity $V_b$, we fit responses to a logistic function of $V_o$:

$$R(V_o) \sim \sigma[k(V_o - u)],$$

where $\sigma = 2/(1 + e^{-x}) - 1$ (1)

The regressions were computed using the maximum-likelihood model (Wichmann & Hill, 2001). We used a prior distribution on the slopes $k$, in which we assumed that probability is approximately uniform for $k < k_{\text{max}}$ and then decreases, where $k_{\text{max}}$ is equal to the slope obtained by going from one extreme value of mean response to the other between two adjacent values of $V_o$—this is the largest slope value that can be supported by the discrete nature of our independent variable. In order to obtain a smooth prior, we used a third-order Butterworth function for the prior distribution: $P(k) = 1/[1 + (k/k_{\text{max}})^3]$. The maximization was carried out in Mathematica (ver. 8, Wolfram Research, Champaign, IL), using the FindMaximum function, with initial values for each parameter determined using a linear regression on the data.

An example of the mean data for one subject and the corresponding logistic regressions for different values of $V_b$ is shown in Figure 3b. For a given background velocity, the value of object velocity $V_o$ where the regression curve crosses zero is the point of subjective stationarity (PSS); the point at which the object is
perceived as stationary on the average. As can be seen for one representative subject in Figure 3c, the values of PSS grow in a systematic and roughly linear fashion with background velocity $V_b$. We therefore quantify the relation between PSS and $V_b$ using a linear regression. These two regressions (the logistic one followed by the linear one) are equivalent to a single logistic regression:

$$R(V_o,V_b) \sim \sigma[k(V_o - aV_b - u)]$$  \hspace{1cm} (2)

The coefficient $a$ is equal to the slope of the linear regression of PSS versus $V_b$, the line shown in Figure 3c. It is a measure of the effect of the background reference on velocity perception. If the slope $a$ was 0, then we would conclude that background motion has no effect on perceived foreground object velocity. If the slope was 1, then the PSS would follow background velocity in a one-to-one fashion, as if object motion was perceived only relative to background; in other words, any motion of the background would yield an equal-and-opposite motion of the foreground, consistent with an extreme form of the visual background solution or “flow-parsing.”

Table 1 shows the fitted parameters of this model for individual subjects. All values of coefficient $a$ are between 0 and 1. We calculated confidence intervals for the $a$ parameter using the bootstrap method (Efron & Tibshirani, 1986). We computed standard errors of the mean by calculating the standard deviation on 1,000 bootstrap samples and multiplied these standard errors by the inverse of the cumulative normal distribution evaluated at 0.975, approximately 1.96, to obtain the 95% confidence intervals (also shown in Table 1). All $a$ values are significantly above 0 and significantly below 1. Given these results we can conclude that the background is used as a reference in the perception of object motion, but only partially, with a mean gain of 0.37.

Mobile condition

Participants’ movement: Before describing the results in the mobile condition, we first present data on participants’ movement. Because the movement was

<table>
<thead>
<tr>
<th>Subject</th>
<th>$a$</th>
<th>Conf($a$)</th>
<th>$u$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>0.34</td>
<td>0.06</td>
<td>-0.25</td>
<td>0.39</td>
</tr>
<tr>
<td>FP</td>
<td>0.37</td>
<td>0.05</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>LD</td>
<td>0.50</td>
<td>0.03</td>
<td>-0.26</td>
<td>1.00</td>
</tr>
<tr>
<td>LG</td>
<td>0.40</td>
<td>0.06</td>
<td>0.16</td>
<td>0.39</td>
</tr>
<tr>
<td>LN</td>
<td>0.43</td>
<td>0.06</td>
<td>0.08</td>
<td>0.31</td>
</tr>
<tr>
<td>MD</td>
<td>0.35</td>
<td>0.05</td>
<td>0.12</td>
<td>0.62</td>
</tr>
<tr>
<td>MW</td>
<td>0.37</td>
<td>0.05</td>
<td>0.40</td>
<td>0.43</td>
</tr>
<tr>
<td>RD</td>
<td>0.18</td>
<td>0.06</td>
<td>0.22</td>
<td>0.59</td>
</tr>
<tr>
<td>SB</td>
<td>0.53</td>
<td>0.04</td>
<td>0.18</td>
<td>0.52</td>
</tr>
<tr>
<td>SP</td>
<td>0.42</td>
<td>0.06</td>
<td>-0.36</td>
<td>0.40</td>
</tr>
<tr>
<td>TB</td>
<td>0.22</td>
<td>0.06</td>
<td>-0.96</td>
<td>0.40</td>
</tr>
<tr>
<td>All</td>
<td>0.37</td>
<td>-0.02</td>
<td>0.51</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The effect of background (bg) motion on object motion perception in Experiment 1, observer immobile. Fitted parameter values for Equation 2, mean $R^2 = 0.75$. 

Figure 3. Steps of the analysis for one representative participant, SB, in the immobile condition of Experiment 1, showing the effect of background motion. (a) The dots show mean response as function of object angular velocity $V_o$ for two extreme values of background angular velocity $V_b = -16$ deg/s (red) and $V_b = +16$ deg/s (blue). Responses are coded as -1 for perceived rotation with near edge moving to the left (or clockwise seen from above), or +1 for right. Curves show the corresponding logistic fits. (b) Mean response as a function of object velocity $V_o$ for all values of background velocity $V_b$ (coded by color). Curves show the corresponding logistic fits. (c) Points of subjective stationarity (PSS) versus background velocity $V_b$ (same color code as in previous figure), showing a roughly linear relation. The line shows the linear regression.
In order to describe the trajectories, we defined the following axes for the earth-fixed reference frame: the x and y axes were parallel to the monitor and point rightwards and upwards from the subject’s point of view, and the z axis was perpendicular to the monitor, pointing towards the subject. On each frame, we defined the subject’s position as the midpoint between the eyes. First, we calculated the discrete path length of the subject’s movement, the sum of absolute displacements in all three dimensions from frame to frame, while the stimulus was visible. We obtained 12.64 cm for the x axis, 2.91 cm for y, and 1.79 for z, averaged over all trials for all subjects. Thus, total motion in the y and z axes was on the average about 23% and 14% of the motion along the principal x axis. The main dynamical variable of interest is the rotation of the subject’s position in space about the center of the virtual stimulus (see Figure 2, where it is denoted by V_e): it determines the contribution of self-motion to the relative rotations between the subject and the stimulus. To calculate the subject’s movement speed, we first converted the subject’s position to an angle in the horizontal plane about the stimulus center, defined as \( \theta = \arctan \frac{x}{z} \). We then calculated the mean angular speed on each trial, denoted as \( V_s \), as the mean change in \( \theta \) during the display of the stimulus.

Figure 4 shows velocity histograms of four representative subjects. Mean speed for all subjects was 17 deg/s, with between-subject standard deviation 7.4 deg/s, and mean within-subject standard deviation 3.5 deg/s.

**Influence of self-motion without background:** To determine the contribution of self-motion to perceived object movement, we first analyzed data from the condition in which the observer was moving and in which the background was absent.

The sensory consequence of an observer’s movement is to add equal-and-opposite relative motion of the scene in the observer’s egocentric reference frame. In the absence of a background or any other reference object, vision has direct access only to this combination of self-motion and object motion. In other words, the egocentric object velocity, \( V_e \), which is the variable directly driving the optic flow, is equal to \( V_o - V_s \), where \( V_s \) is the velocity of the observer, described previously, and \( V_o \) is absolute velocity, i.e., object velocity in an earth-fixed reference frame. In order to recover absolute motion, the egocentric motion \( V_e \) needs to be compensated for self-motion \( V_s \). However, such compensation (which may be based on motor, vestibular, somatosensory, proprioceptive, or optic flow signals) may be imperfect. We assumed a linear combination of \( V_e \) and the mean value of \( V_s \) over the trial, with the constant \( b \) representing the gain of observer velocity. This gain represented the fraction of the observer’s self-motion used to compensate object motion. While we adopted a linear model for the sake of simplicity—and verified it a posteriori with goodness-of-fit measurements—more complex relations between retinal and extra-retinal signals have been postulated and documented in the compensation for eye movements (Turano & Massof, 2001; Niemeier, Crawford, & Tweed, 2003; Souman & Freeman, 2008; Morvan & Wexler, 2009). We therefore used the following regression:

\[
\mathbf{R}(V_e, V_s) \sim \sigma[k(V_e + bV_s - u)]
\]

In order to combine data with subject motion to the left and right, we transformed leftward trials to equivalent rightward trials by inverting the sign of all velocities (of the foreground, background, and subject) and the response. If \( b \) is equal to 1, then perceived object velocity is, on the average, equal to \( V_o' \); in other words, self-motion is correctly estimated and taken into account, and absolute object motion is perceived. If \( b \) is different from 1, on the other hand, compensation is imperfect: either the observer does not correctly estimate his or her self-motion or does not take it into account correctly. If \( b < 1 \), for example, then self-motion is underestimated (or under-compensated); perceived motion is then in a reference frame intermediate between the egocentric and the earth-fixed.\(^1\)

Finally, if \( b \) is equal to 0, then the observer’s judgments of object motion are entirely based on egocentric motion.

The results of our regression are given for individual subjects in Table 2, which shows the fitted values for the \( b, u, \) and \( k \) parameters, as well as the 95% confidence intervals of \( b \), computed by bootstrap on 1,000 iterations. The mean value of the \( b \) parameter was

![Figure 4. Examples of distributions of subject velocities \( V_s \) in Experiment 1 for four representative subjects, expressed as rotations about the stimulus center in deg/s. (a) AD, (b) FP, (c) LD, and (d) LG.](https://jov.arvojournals.org/)
We observed that, for 10 subjects out of 11, the fitted parameter was significantly greater than 0; for 8 subjects out of 11, b was not significantly less than 1. For one subject, however, computed b was negative, but not significantly different from 0. A t test across subjects showed that the mean value $b = 0.65$ was significantly greater than 0 and less than 1, respectively, $t(10) = 4.92, p < 0.05; t(10) = 2.67, p < 0.05$. Thus, the population mean gain of self-motion is likely to be below 1, although there seem to be large individual differences in this parameter.

Table 2. The effect of self-motion on object motion perception in Experiment 1, no background. Fitted parameter values for Equation 3. Mean $R^2 = 0.66$.

Table 3. The influence of background (a) and self-motion (b) in Experiment 1 when both observer and background are moving. Parameter values for Equation 4, mean $R^2 = 0.61$.

Moving observer and background

We have already shown that both a visual reference (for a stationary observer) and self-motion (in the absence of visual reference) can influence motion perception. We next analyzed data from trials with both a visual reference and observer motion, to check for a possible interaction between these two factors. The regression formula is a generalization of the previous regression models, including both the $V_b$ and $V_s$ terms:

$$R(V_e, V_b, V_s) \sim \sigma[k(V_e - aV_b + bV_s - u)]$$

The fitted values of the $a$ and $b$ parameters and their confidence intervals are shown in Table 3. As in the immobile condition, individual $a$ values were significantly different from both 0 and 1 for every subject. A t test across subjects likewise showed that the mean value $a = 0.43$ was significantly different from both 0 and 1. For 6 out of 11 subjects, individual $b$ values were significantly lower than 1. For one subject, however, computed $b$ was negative, but not significantly different from 0. A t test across subjects showed that the mean value $b = 0.65$ was significantly greater than 0 and less than 1, respectively, $t(10) = 4.92, p < 0.05; t(10) = 2.67, p < 0.05$. Thus, the population mean gain of self-motion is likely to be below 1, although there seem to be large individual differences in this parameter.

Use of visual reference by immobile and mobile observers: Observers could have relied on the visual background differently when moving than when immobile. We therefore compared the $a$ values calculated in the two conditions (immobile in the section Use of visual reference by an immobile observer, mobile in the section Moving observer and background). Mean values of $a$ were 0.37 and 0.43, for observer-immobile and -mobile conditions, respectively; a paired t test showed that these means were not significantly different, $t(10) = 1.78, p = 0.10$. A linear regression showed a correlation between these data: a bootstrap analysis showed that the slope (shown as dotted lines in Figure 5) was significantly different from both 0 and 1, with correlation coefficient $R^2 = 0.25$. There seemed to be less variation in a coefficients in the observer-mobile than in the observer-immobile conditions, although we are at a loss to explain why.

Influence of self-motion with or without a visual reference: The use of self-motion signals may be different if the background is present or absent. For example, self-motion signals may be relied on more (in other words, the $b$ parameter would be greater) in the dark than with a visual background. Figure 6 shows the comparison of $b$ values in these two conditions.
The mean value of $b$ is 0.75 with the background absent and 0.65 with the background present. A paired $t$ test shows that the means in the two conditions are not significantly different ($p = 0.63$), but both the within- and between-subject variability is so large that it is hard to conclude much from this result. The larger variability in $b$ as compared to $a$ coefficients (compare Figure 6 to Figure 5) may be due to the complex cross-modal combination necessary to estimate object motion based on self-motion, and to the generally lower precision of performing absolute as compared to relative judgments (Snowden, 1992).

**Precision with and without a background**

Another way of evaluating the effect of background on motion judgments is to compare task precision—the $k$ parameter in our models—in cases of moving subjects with and without a background. For example, if precision is lower without a background, then observers rely on a visual reference to make their judgments.

The fitted values of $k$ for the no-background, stationary-background ($V_b = 0$), and moving-background conditions ($V_b \neq 0$) are shown in Table 4. The larger the value of slope $k$, the more precise the subject’s judgment. The mean value of $k$ over all subjects with no background was 0.64 ($SD$ 0.27), 0.45 ($SD$ 0.30) with stationary background, and 0.40 ($SD$ 0.16) for moving background. A paired $t$ test showed that precision in the moving-background condition was significantly different and lower than in the condition with no background ($p < 0.05$). Other paired $t$ tests between $k$ where the background was moving compared to stationary and no background compared to stationary showed that $k$ values were not significantly different (respectively, $p = 0.42$ and $p = 0.07$).

<table>
<thead>
<tr>
<th>Subject</th>
<th>$k$ no bg</th>
<th>$k$ stationary bg</th>
<th>$k$ moving bg</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>0.39</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>FP</td>
<td>0.97</td>
<td>1.24</td>
<td>0.62</td>
</tr>
<tr>
<td>LD</td>
<td>0.76</td>
<td>0.35</td>
<td>0.45</td>
</tr>
<tr>
<td>LG</td>
<td>0.47</td>
<td>0.28</td>
<td>0.34</td>
</tr>
<tr>
<td>LN</td>
<td>0.18</td>
<td>0.16</td>
<td>0.19</td>
</tr>
<tr>
<td>MD</td>
<td>0.65</td>
<td>0.82</td>
<td>0.49</td>
</tr>
<tr>
<td>MW</td>
<td>0.87</td>
<td>0.39</td>
<td>0.41</td>
</tr>
<tr>
<td>RD</td>
<td>1.17</td>
<td>0.38</td>
<td>0.29</td>
</tr>
<tr>
<td>SB</td>
<td>0.51</td>
<td>0.45</td>
<td>0.56</td>
</tr>
<tr>
<td>SP</td>
<td>0.62</td>
<td>0.29</td>
<td>0.31</td>
</tr>
<tr>
<td>TB</td>
<td>0.38</td>
<td>0.43</td>
<td>0.37</td>
</tr>
<tr>
<td>All</td>
<td>0.64</td>
<td>0.45</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Table 4. Values of precision, as measured by the $k$ parameter, when the background is absent, when the background is present and stationary and when the background is present and moving.
Discussion

In this experiment, we found that the motion of the background does have an effect on the perception of foreground object motion. However, this effect is far from complete: a background rotating at a certain velocity does not make the foreground appear to rotate at an equal-and-opposite velocity. Background motion seems to have a very similar effect in moving and immobile observers. Thus, neither the “flow-parsing” hypothesis of Rushton and Warren (2005) nor any other visual background solution can fully account for the perception of motion in moving observers. We have also found a significant contribution of self-motion estimates to the perception of object motion. This contribution slightly underestimates the true self-motion, and isn’t significantly different in the presence and absence of visual background.

In this experiment, all stimuli were planar surfaces (both the object and background). In the perception of 3D structure and motion from optic flow, planes give rise to specific ambiguities (Longuet-Higgins, 1984). These ambiguities may account for failure to extract motion from certain types of optic flow in environments composed of planes (Warren & Hannon, 1990; Grigo & Lappe, 1999). Although our stimuli had binocular cues that could disambiguate planar optic flow, we wanted to be sure that our results could be generalized to more complex 3D scenes. We therefore performed a second experiment, based on the immobile condition of Experiment 1, in which the stimulus and background had more complex 3D shapes.

Another possible issue in Experiment 1 is the faintly visible monitor edges. If subjects perceived these stationary edges, they could have provided a conflicting signal to the background, moving at a different velocity with respect to the observer than our background stimulus. In Experiment 2, we therefore covered the monitor with a filter that drastically decreased the visibility of its edges.

Experiment 2

Methods

Unless stated otherwise, the methods were the same as in Experiment 1.

Participants

Five subjects participated (three men, two women, mean age: 28 years). Four had been subjects in Experiment 1 (FP, LD [first author], LN, MD) and all were experienced psychophysical observers. The fifth subject (RR) was new and naive.

Apparatus

The apparatus was the same as in Experiment 1 except that sunglasses that had been worn by the subjects were replaced by a filter on the monitor. To achieve lower luminance levels, the monitor was covered with two layers of 1.2 log unit neutral density filters (Lee Filters, Andover, UK), which drastically decreased the visibility of monitor edges. When questioned following the experiment, no subjects reported being able to see monitor edges.

Stimuli

The object in Experiment 1 was replaced by a cube and the background by a set of smaller cubes at different depths, with the dimensions of the virtual objects chosen so as to yield approximately the same size and shape stimuli when projected (see Figure 7).

The object was initially rotated around the vertical axis by ±22.5 deg. This cube had edges 26 cm long and was centered about the physical center of the monitor, as was the stimulus in Experiment 1. All faces of this cube were textured as the foreground object in Experiment 1.

The background was composed of 12 cubes (four horizontally by three vertically) rotating as a rigid object around the same vertical axis as the background in Experiment 1. Each cube had edges 15 cm long. The texture used was the same as the background object texture in Experiment 1, but it was cropped by a square of size equal to half the height of the initial texture, and then shrunk by a factor of 0.375. In order to add more depth information to the stimulus, the background cubes were positioned at different depths, chosen randomly from a uniform distribution between 20 and

Figure 7. Monocular frame capture of the stimulus in Experiment 2. 3D structure was easier to perceive in the actual stimulus due to motion and binocular disparity.
80 cm behind the origin, independently for each cube and each trial.

Object and background angular velocities were the same as in Experiment 1.

**Procedure**

The procedure was nearly the same as in the immobile condition of Experiment 1. In order to avoid a visual reference on the object, the fixation point was removed but the subject was asked to fixate the foreground cube.

**Results and discussion**

**Results**

We analyzed the data using the same model as in the immobile condition of Experiment 1. We extracted the parameter of interest, \( a \), the gain of background movement in subjects’ motion judgments. For each subject, as in Experiment 1, we found that \( a \) values were significantly between 1 and 0 (see Table 5). Values of \( a \) above 0 mean that the object velocity was perceived relative to background velocity; \( a \) values below 1 indicate that the use of the background was partial. The mean value of \( a \) in Experiment 2 was slightly higher than in Experiment 1, but the difference was not significant.

**Discussion**

In the immobile condition of Experiment 1, using planar stimuli, we showed that background movement had an effect—but only a partial one—on motion perception. In this experiment, we have shown that this result generalizes to a stimulus with a richer depth structure. Even if we cannot be sure that \( a \) values are significantly different between the two experiments, the mean value of \( a \) is slightly higher in the second experiment (0.50 versus 0.41 for the subjects who participated in both).

### General discussion

We will first discuss the contribution of the visual background, then that of self-motion. For an immobile observer with a moving background, we find that the perception of object velocity partly depends on the motion of the background. The dependence is a kind of contrast: turning the background one way makes the object appear to turn more in the opposite direction. Moreover, the dependence on background motion appears to be quite linear (Figure 3c). We therefore modeled the effect of background as a linear difference: \( V_p = V_o - a V_b \), where \( V_o \) is foreground object velocity, \( V_b \) background velocity, and \( V_p \) the perceived velocity of the foreground object. If we had found that \( a = 0 \) then perceived motion of the foreground object would have been independent of the background motion, at least in the linear model; \( a = 1 \), on the other hand, would mean that foreground object motion is perceived exclusively relative to the background.

For immobile subjects, we found a mean value of \( a = 0.37 \), and this value is significantly greater than 0 and less than 1. This means that the contrast effect with the background is significant, but also only partial: if the background rotates at the same velocity as the foreground object, the object is not perceived as stationary—although it is stationary with respect to the background—but it is perceived as moving slower than its velocity in a world reference frame, at about 63% of its speed.

There are several different ways, not necessarily mutually exclusive, in which the background could influence perceived object motion. First, the effect could be an example of the phenomenon of induced motion, but in depth. In induced motion, object velocity is perceived in partial contrast to the velocity of its visual context. In our case, the visual background rotating in depth could have created an opposite bias in the estimation of object velocity, similar to the 3D motion contrast effects demonstrated by Warren and Rushton (2009). In this experiment, our results indicate that the perceived velocity of the object is reduced relative to the velocity of the background, which is consistent with the notion of induced motion in depth. However, the exact nature of this effect and its mechanisms require further investigation.

### Table 5

<table>
<thead>
<tr>
<th>Subject</th>
<th>( a ) Experiment 2</th>
<th>( \text{Conf}(a) )</th>
<th>( a ) Experiment 1</th>
<th>( \text{Conf}(a) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP</td>
<td>0.60</td>
<td>0.05</td>
<td>0.37</td>
<td>0.05</td>
</tr>
<tr>
<td>LD</td>
<td>0.41</td>
<td>0.04</td>
<td>0.50</td>
<td>0.03</td>
</tr>
<tr>
<td>LN</td>
<td>0.55</td>
<td>0.05</td>
<td>0.43</td>
<td>0.06</td>
</tr>
<tr>
<td>MD</td>
<td>0.42</td>
<td>0.04</td>
<td>0.35</td>
<td>0.05</td>
</tr>
<tr>
<td>All 4</td>
<td>0.50</td>
<td></td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td>RR</td>
<td>0.35</td>
<td>0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All subjects for each experiment</td>
<td>0.47</td>
<td>0.04</td>
<td>All subjects for each experiment</td>
<td>0.37</td>
</tr>
</tbody>
</table>
case, the background could be used as an anchor for the visual reference frame. Consequently, object velocities (and positions) are computed relative to this frame, and moving the background would alter perceived velocity of foreground objects.

A second way in which background could influence object motion is through the effect of its optic flow on the subject’s estimation of his or her own motion (Lappe, Bremmer, & van den Berg, 1999). This erroneous self-motion estimation, in the case of a moving background, could be incorrectly used to compensate object motion (Wertheim, 1994). In addition, the estimate of self-motion from background optic flow could be modulated by extraretinal signals (Royden, Banks, & Crowell, 1992; Crowell, Banks, Shenoy, & Andersen, 1998). The (mis-)estimation of self-motion from background optic flow may be accompanied by an explicit perception of self-motion in an immobile observer—or not. If so accompanied, it is referred to asvection (Fischer & Kornmüller, 1930; Nakamura & Shimojo, 1999). However,vection takes several seconds to build up (Palmisano, 1996), and since our stimuli only lasted for only 800 ms, they were probably too brief forvection to contribute to our results.

When we calculated the gain of background motion in the mobile observer condition and compared it to the immobile condition, we found no significant difference (mean $a = 0.43$ for mobile observers, $0.37$ for immobile). Thus, although it could have been possible that mobile observers relied more (or less) on the visual background (Wertheim, 1994), the background seems to have exerted the same influence on mobile as on immobile observers.

In the condition with no background and the observer moving, it was obviously impossible to compute relative motion with respect to the background; in order to compute absolute motion, the observer must compensate egocentric object velocity by using an estimate of self-motion. Similar to the case of the moving background, we assumed that perceived velocity $V_P$ was a linear combination of egocentric velocity $V_e$ and self-motion $V_S$, but that self-motion could be estimated with a gain factor, $b$: $V_P = V_e + bV_S$ (see, for example, Freeman & Banks, 1998). Perfect compensation would mean $b = 1$: in this case, perceived object velocity would equal its “real” velocity, $V_e$, in an earth-fixed reference frame; $b = 0$, on the other hand, would mean that the observer perceives object motion in an egocentric reference frame, i.e., only relative to himself. Without a background, we found a mean value of $b = 0.75$, significantly below 1 and above 0. Thus, mobile observers did compensate object motion by an estimate of self-motion, but this compensation was less than its ideal value.

When we calculated the degree of compensation for self-motion in the presence of a visual background, we found a mean value of $b = 0.65$. This value is significantly above 0 and below 1, and not significantly different from the corresponding parameter without background. Thus, the presence of a visual reference does not seem to radically change the way the visual system compensates the perception of object motion by estimated self-motion (Wertheim, 1994). The values that we have obtained for the self-motion gain factor are consistent with previously published results (Wertheim, 1987; Freeman & Banks, 1998; Wexler, 2003).

What are the possible sources of self-motion information that contribute to this (partial) compensation? One is the efference copy of the motor command to move the trunk and head, which—in comparisons between active and passive movement—has been shown to contribute directly to the visual reference frame in 3D vision (Crowell et al., 1998; Wexler, 2003).

Other signals that provide feedback about self-motion—vestibular signals, proprioception, somatosensory information—probably also contribute to the estimate of self-motion (Crowell et al., 1998) that might modify the perception of object motion (MacNeilage et al., 2012). Another extraretinal source of self-motion information could be eye movements. The role of eye movement extraretinal information has been shown in the case of heading perception (Royden et al., 1992; Royden, Crowell, & Banks, 1994). In the case of our experiment, if the subject fixated the same point on the stimulus throughout his or her movement, using the translational vestibulo-ocular reflex, then the counter-rotation of the eye in the head provided self-motion information—provided that the visual system could access and use this signal (Nawrot, 2003). Although this cannot be excluded, it seems unlikely that this signal made a large contribution, for the following reason. Wexler (2003) measured the contribution of extraretinal signals to 3D motion perception, in a case in which subjects moved backwards and forwards, rather than laterally, with the stimulus also moving backwards and forwards. In this case, eye movements could not have played any significant role, but Wexler (2003) nevertheless found a quantitatively similar contribution of extraretinal signals as found here. In addition to extraretinal sources of self-motion information, optic flow provides self-motion information based implicitly on an assumption of visual stationarity (see Lappe et al., 1999, for a review), and this visual self-motion signal may in turn lead to a modification of the perception of object motion (Wertheim, 1994).

Our results show that the precision of object motion judgments, as measured by the slopes of the psychometric curves, deteriorates when a moving background is present as opposed to when the background is
absent. The precision is not significantly different when the background is moving than when it is stationary. Thus, the addition of visual information does not improve the precision of the observer even if the background is stationary. If optic flow contributes to self-motion estimation, but with a different gain than extraretinal signals, one explanation could be that the retinal and extraretinal cues to self-motion are in conflict when a background is present.

We have shown that 3D background motion does have an effect on the perception of motion of foreground objects, and that this effect is a contrast: when the background moves in the same direction as the foreground motion, the foreground object is perceived to move slower, and when the background moves in the opposite direction, the foreground object is perceived to move faster. This is further documentation of induced movement in depth, evoked by Rushton and Warren in support of their flow-parsing hypothesis, according to which the perception of motion relative to the background alone is sufficient for the moving observer to perceive absolute object motion (Rushton & Warren, 2005; Warren & Rushton, 2008, 2009).

However, we have measured the “gain” of the background effect, \( a \), defined as follows: if the background velocity changes by \( \Delta \), then the perceived velocity of the foreground object changes by \( -a\Delta \). We have found that this gain is roughly equal to 0.4–0.5, with a between-subject standard deviation of about 0.1. This gain is significantly below 1, in all individual subjects and in our subject population as a whole. Moreover, the background gain is not significantly higher in the moving observer (mean 0.43) than in the immobile (0.38). This is in contradiction to a strong form of the flow-parsing hypothesis, according to which motion relative to the background is the only way in which a moving observer can perceive absolute object motion, and which would seem to require a gain of 1. Instead, some other factor must also contribute to the perception of absolute object motion.

This other factor is, of course, an estimate of self-motion. As already discussed, this estimate can come from a variety of extraretinal signals: efference copies of trunk, head, and eye motor commands; vestibular signals; and proprioception. These signals, converted into an estimate of the observer’s own rotation about the stimulus, are added to the egocentric stimulus velocity, with a gain of roughly 0.7. Although the self-motion gain is not significantly less than 1, other studies have also found a self-motion gain below 1 (Wexler, 2003; Dyde & Harris, 2008) (but see Redlick, Jenkin, & Harris, 2001). This is as if the observer was either underestimating his or her self-motion or deliberately using less than the full estimate for the purposes of estimating object velocity. In the domain of eye movements, similar underestimation can account for the Filehne illusion and the Aubert-Fleischl effects (Mack & Herman, 1973; Wertheim, 1994, 1987; Freeman & Banks, 1998; Morvan & Wexler, 2009).

Another possibility is that observers are overestimating object motion rather than underestimating self-motion (Freeman & Banks, 1998). However, it is difficult to test this possibility in the absence of well known stimulus factors affecting the perceived speed gain of motion in depth.

In sum, we have shown that the two possible methods for perceiving absolute object motion—compensating for self-motion and computing motion with respect to the background—are used by the actively moving observer. Moreover, both methods most likely have a gain below 1, although there do seem to be large individual differences. It is interesting to note that the sum of the two gains, roughly 0.7 and 0.4, happens to be quite close to 1. This could be a clue that the two methods, which are usually simultaneously available in real-life situations, work in tandem.

Keywords: motion perception, self-motion, three-dimensional vision

Acknowledgments

We thank Ivan Lamouret and Jacques Droulez for collaborating in the preliminary version of these studies, and Simon Rushton, Paul Warren, and Brett Fajen for useful conversations. This work was partly funded by the ERC PATCH project.

Commercial relationships: none.
Corresponding author: Lucile Dupin.
Email: lucile.dupin@parisdescartes.fr.
Address: Laboratoire Psychologie de la Perception, CNRS & Université Paris Descartes, Paris, France.

Footnote

1 The effect could also be due the overestimation of retinal velocity (Freeman & Banks, 1998) instead of underestimation of self-motion.

References


