

# OBJECT SPEED PERCEPTION DURING LATERAL SELF-MOTION

*Björn Jörges<sup>1</sup>, Laurence R. Harris<sup>1\*</sup>*

<sup>1</sup> Center for Vision Research, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada

\* Corresponding Author

## Abstract

Judging the velocity of objects during observer self-motion requires disambiguating retinal stimulation caused by the observer's own movement and retinal stimulation caused by object movement. According to the Flow Parsing hypothesis, observers first estimate their own motion based on visual and other cues. They then subtract the retinal motion corresponding to their own movement from the total retinal stimulation and interpret the remaining stimulation as pertaining to object motion. While the phenomenon has been studied to some extent for motion-in-depth and rotational motion, lateral motion has been largely neglected. The Flow Parsing hypothesis yields predictions both for the precision and the accuracy of target speed estimation during self-motion: firstly, subtracting noisier self-motion information from retinal input should lead to a decrease in precision when the observer is moving during motion observation. Furthermore, when self-motion is only simulated visually, while other cues such as vestibular sensory information and data from efference copies are unavailable, self-motion is likely to be underestimated, which should yield an overestimation of target speed when target and observer move in opposite directions and an underestimation of target speed when target and observer move in the same direction.

## Significance

We rarely stand still while interacting with our inherently dynamic environment. It is crucial for us to obtain accurate and precise estimates of the movement of objects relative to ourselves and relative to the environment even while we are moving. Here, we investigate how we judge the speed of moving objects during visually evoked lateral self-translation. Surprisingly, this measurement has never been attempted before. Estimating the physical speed of objects accurately and precisely is important for a variety of tasks that require an allocentric reference frame.

## Introduction

When observing a moving target while an observer is moving, the same retinal speeds can correspond to vastly different physical velocities. When an observer moves in the same direction, parallel to a moving object, the retinal speed of the object is partially cancelled out, and when they move in the direction opposite to the object, the retinal stimulation due to self-motion may be added to the retinal speed of the object. To obtain an accurate estimate of the object's velocity, observers must therefore obtain an accurate estimate of their own velocity and subtract or add the consequences of this movement to the retinal motion of the target. More specifically, the Flow Parsing Hypothesis (Dupin & Wexler, 2013; Rushton & Warren, 2005; Warren & Rushton, 2008, 2009) posits that, to estimate object motion from ambiguous retinal input representing the sum of object and self-motion, observers first

compute which components of retinal stimulation are caused by their own motion in the environment. Then, they subtract this self-motion information from the overall retinal stimulation and attribute the remaining stimulation to object motion in the scene. When self-motion is experienced only visually while undergoing no physical motion, the visual motion creates a conflict between visual and vestibular inputs as a result of which self-motion is likely to be underestimated, leading to biases in judgments of object velocities although oddly this has never been quantified for horizontal translation. The effect has been shown to some extent for vertical observer and object translation (Dyde & Harris, 2008), as well as for rotating observers (Garzorz, Freeman, Ernst, & MacNeilage, 2018; Hogendoorn, Alais, MacDougall, & Verstraten, 2017; Probst, Loose, Niedeggen, & Wist, 1995) and motion-in-depth (Gray, MacUga, & Regan, 2004). Furthermore, it has been argued that self-motion information is noisier than retinal information concerning object motion, especially when observers have only visual information about their own movement at their disposal (Fetsch, Deangelis, & Angelaki, 2010). Subtracting noisy self-motion information from retinal motion in order to obtain an estimate of target velocity should thus decrease precision (Dokka, MacNeilage, DeAngelis, & Angelaki, 2015). Such a subtraction process is relatively straightforward for the consequences of angular self-motion, but for lateral motion, the geometry requires additional computations involving estimates of the distance of the object to the observer and the direction of object motion relative to the observer's motion. More specifically, observers need to first estimate their own motion in an allocentric world frame by using retinal stimulation attributable to the induced motion of static objects in the environment and other sensory and efferent information such as vestibular activity. Then, this estimate needs to be used to generate an estimate of the retinal stimulation expected to be caused by the observer's motion. This estimated retinal stimulation due to self-motion is then subtracted from the total retinal stimulation, which allows the remaining retinal stimulation to be interpreted as external object motion. The process is known as "flow parsing" in which the different aspects of the total optic flow are attributed to these different causes.

It is important to note that flow parsing is only necessary when humans need to represent the kinetic properties of their environment in an allocentric frame. For computations performed in an egocentric frame, it is generally sufficient to time interceptive actions and avoid collisions according to the velocity of the target relative to the observer. It is true that ecological, optic-flow-based heuristics have successfully explained human performance in paradigmatic cases such as the outfielder problem (Fink, Foo, & Warren, 2009; Wilson & Golonka, 2013). However, humans are able to recover, represent and use the physical parameters of their environment in a variety of tasks (Burr, Tozzi, & Morrone, 2007; Fajen, Parade, & Matthis, 2013; Ilg, Schumann, & Thier, 2004; Wexler, 2003).

There are two major sources of information about passive self-motion: visual and vestibular cues (Fetsch et al., 2010), which are integrated according to their relative reliability (Fetsch, Turner, DeAngelis, & Angelaki, 2009). How much each sense contributes to the global percept of self-motion seems to depend on different parameters, such as the task and the self-motion profile. Dokka et al. (Dokka et al., 2015), for example, found for direction judgments of a probe presented in the fronto-parallel plane during lateral observer motion that vestibular information in the absence of visual information led to a vast underestimation of self-motion. Visual information only elicited a higher accuracy and having both visual and vestibular cues available increased accuracy only marginally beyond accuracy for visual information only. In a more direct test of perceived self-motion, Harris et al. (Harris et al., 2000) found that vestibular stimulation evoked by moving the observer through the environment was an extremely

potent cue to self-motion which induced a vast overestimation of the distance moved. Visual cues to self-motion were efficient, too, but less so than vestibular cues. However, depending on stimulus parameters such as simulated acceleration, visual cues alone can also lead participants to overestimate their movement (Redlick, Jenkin, & Harris, 2001). For active self-motion (i.e., movements initiated by the observer such as walking through the environment), efference copies and proprioceptive information can serve as further cues. For example, judgments about the distance travelled seem to be more reliable if motion was self-generated as opposed to experienced passively (Becker, Nasios, Raab, & Jürgens, 2002; Frissen, Campos, Souman, & Ernst, 2011; Jürgens & Becker, 2006).

Remarkably, the literature is quite sparse with regards to assessing object motion during lateral, visually simulated observer motion: Warren and Rushton (Warren & Rushton, 2007) found that translational visually evoked self-motion lead observers to perceive the trajectory of a linearly moving probe as tilted towards the direction of the simulated translation. MacNeilage et al. (MacNeilage, Zhang, DeAngelis, & Angelaki, 2012) showed that vestibular cues could help distinguish self-motion from object motion, especially for lateral observer motion. Similarly, Dokka et al. (Dokka et al., 2015) investigated the extent to which observer motion (visual cues only, vestibular cues only, and both visual and vestibular cues) influenced the judged direction of vertical downwards motion with a small lateral component. They found biases in line with insufficient compensation for self-motion in all observer-motion conditions, as well as decreases in sensitivity. Niehorster and Li (Niehorster & Li, 2017) quantified the extent to which flow parsing was complete for straight-ahead self-motion by having participants judge the direction of a probe that moved vertically upwards. Importantly, all these studies used direction judgments as proxies to probe the completeness of flow parsing, while a direct psychophysical investigation of perceived *velocities* is notably missing from the literature. Furthermore, while some of these studies presented their stimuli in compelling stereo 3D, none immersed the observer in a virtual environment. It is not unlikely that flow parsing is facilitated by a more realistic environment. This study aims to shed light on the extent to which visually evoked self-motion influences perceived lateral object *speed* in a naturalistic setting. This is particularly relevant as the visual system has been shown to use velocity information to extrapolate object trajectories to compensate for noisy online information and neural delays (Aguado & López-Moliner, 2019; Aguilar-Lleyda, Tubau, & López-Moliner, 2018; Jörges & López-Moliner, 2019; López-Moliner, Brenner, Louw, & Smeets, 2010). The aim of this project is thus to verify the impact of visually simulated observer motion on accuracy and precision for object speed judgments during lateral translation, which will further our understanding of Flow Parsing and help us understand the conditions under which Flow Parsing is incomplete. More specifically, our hypotheses are:

- When the observer experiences **no visual self-motion** during object motion observation, we expect the **highest accuracy** of speed estimation.
- When visual observer motion is simulated **opposite to the object motion (e. g. observer moves to the right, object moves to the left)** during object motion observation, we expect them to **overestimate the observed speed**.
- When visual observer motion is **simulated in the same direction as the target (e. g. both observer and target move to the right)** during object motion observation, we expect them to **underestimate the observed speed**.
- We expect the **precision to be lower** when the subject experiences visually simulated observer motion during object motion observation relative to when they are static.

## Participants

We tested 30 participants (see power analysis) from the population of PhD and undergrad students at York University with equal numbers of men and women. Due to the culturally independent nature of the phenomenon under study, we do not believe our results are likely to be relevantly skewed by WEIRD people effects (Henrich, Heine, & Norenzayan, 2010). Participants had normal or corrected-to-normal vision and had to achieve a stereoacuity of 63 arc seconds or below on the Fly Stereo Acuity Test. The project has received ethics approval from the Human Participant Ethics Review Sub-Committee at York University. Informed consent was obtained from all subjects and the experiment was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). We continued data collection until we tested 30 subjects who satisfied the criteria laid out below under “Intended interpretation of visually simulated observer motion and task”.

## Apparatus

All the experiments were performed in virtual reality with participants remaining physically static and seated. We programmed the stimuli in Unity (2019.2.11f1), while object motion, visually simulated observer motion and the psychophysical staircases were controlled in C# via its integration with Unity. The Unity project is available on Open Science Foundation (<https://osf.io/m6ukw/>). Stimuli were presented in an Oculus Rift (CV1, firmware version 709/b1ae4f61ae). Participants responded by means of a finger mouse.

## Setup

Our experiment consisted of a Two Interval Forced-Choice Task where participants were asked to indicate which of two intervals contained objects moving at the higher speed.

### *Environment and general layout*

Participants were immersed in a virtual 3D environment that included depth cues from lighting, shadows and the scale of the textures of the floor and the wall backdrop. The ball appeared to the left of the observer if it moved to the right, and to the right of the observer when it moved to the left. The exact position was determined by target speed and visual observer motion (see Equation 1 below). See Figure 1A for a diagram of the visual scene and Figure 1C for a screenshot from the experiment; a short sequence of the experiment can be downloaded [here \(GitHub\)](#).

### *Targets and visually simulated observer motion*

In one interval participants were presented a ball with a diameter of 0.33 m at a simulated distance of 8 m in front of them, travelling with 6.6 or 8.0 m/s (two target motion profiles). Participants were seated and the camera was simulated 1.3 m above the ground. The targets were simulated 2 m above the ground. The direction in which the targets moved was chosen randomly on each trial. During this interval, participants were either static (“Static with Textured Backdrop”) or experienced visually simulated observer motion to the left or to the right with a Gaussian speed profile (three self-motion profiles), accelerating until reaching peak speed after 0.25 s and then slowing down until coming to a halt at 0.5 s (“Simulated Observer Motion with Textured Backdrop”). The position in time  $x(t)$  was given

by a cumulative Gaussian distribution with a mean of 0.25 s and a standard deviation of 0.08 s divided by 2, multiplied by -1 for trials with visually simulated observer motion to the left. That is, participants were moved visually 0.5 m over the course of 0.5 s, which amounts to a mean speed of 1 m/s. The target's initial position was shifted away from the observer for motion in the same direction, and towards the observer when the observer's visual motion occurred opposite to the target motion, such that observer and target motion were symmetrical, i.e., the distance between observer and target at the beginning was the same as the distance at the end of the trial. To achieve this, the starting position of the target was computed in the following manner:

$$x_{initial} = 0.5 * (Distance_{observer} - v_x * t) \quad [1]$$

$x_{initial}$  denotes the initial position of the target relative to the observer,  $Distance_{observer}$  is the lateral distance the observer in which the observer is simulated to move,  $v_x$  is the horizontal speed of the target, and  $t$  is the duration of the motion interval (0.5 s).

#### Staircase

In the other interval, participants were shown a cloud of smaller moving balls each with a diameter of 0.1 m, as comparison. The balls appeared 1.25 m to the left of the observer (if the big target in the same trial moved to right) or to the right of the observer (if it moved to the left), then moved in the same direction as the big target and disappeared after having travelled 2.5 m. They were spread out vertically over a distance of 1 m. 10 to 15 balls were visible at any given moment. Observers were asked to maintain their gaze on a fixation cross that was continuously displayed straight ahead of them (i.e., also during visually evoked self-motion), 0.8 m under the target (see Fig 1b). The speed of these smaller balls was controlled by a PEST staircase. We employed two staircases for each combination of visually stimulated self-motion (left, right or static) and object motion (6.6 and 8 m/s), one of which started 33 % above the target's speed, and the other one 33 % below target speed (two staircases for each combination of target motion and self-motion). The direction (left-to-right or right-to-left) was chosen randomly for each trials. Thus, there was a total of 2 target speeds x 3 motion conditions x 2 = 12 staircases.. The step sizes were governed by the following rules (Taylor & Creelman, 1967): the initial step size was 1.2 m/s. For the first five trials for each PEST, the step size was maintained. Starting from the eleventh trial, after a reversal (subjects answered "PEST is slower" in the second-to-last trial and "PEST is faster" in the last trial or vice-versa), the step size was halved. After the second same answer, the step size was maintained. After the third same answer, the step size was either maintained, when the step size had been doubled before the last reversal, or doubled when the step size had not been doubled before the last reversal. After four same answers, the step size was always doubled. Each PEST ended when it converged (five consecutive trials with step sizes lower than 0.1) AND participants had judged at least 20 trials of the staircase. If the staircase did not converge, the PEST was terminated after 27 trials. The experiment ended when all PESTs, including the control PESTs (see "A possible confound: Induced Motion" below), had terminated. This took about an hour overall including instructions and breaks, which participants could take every 15 minutes.

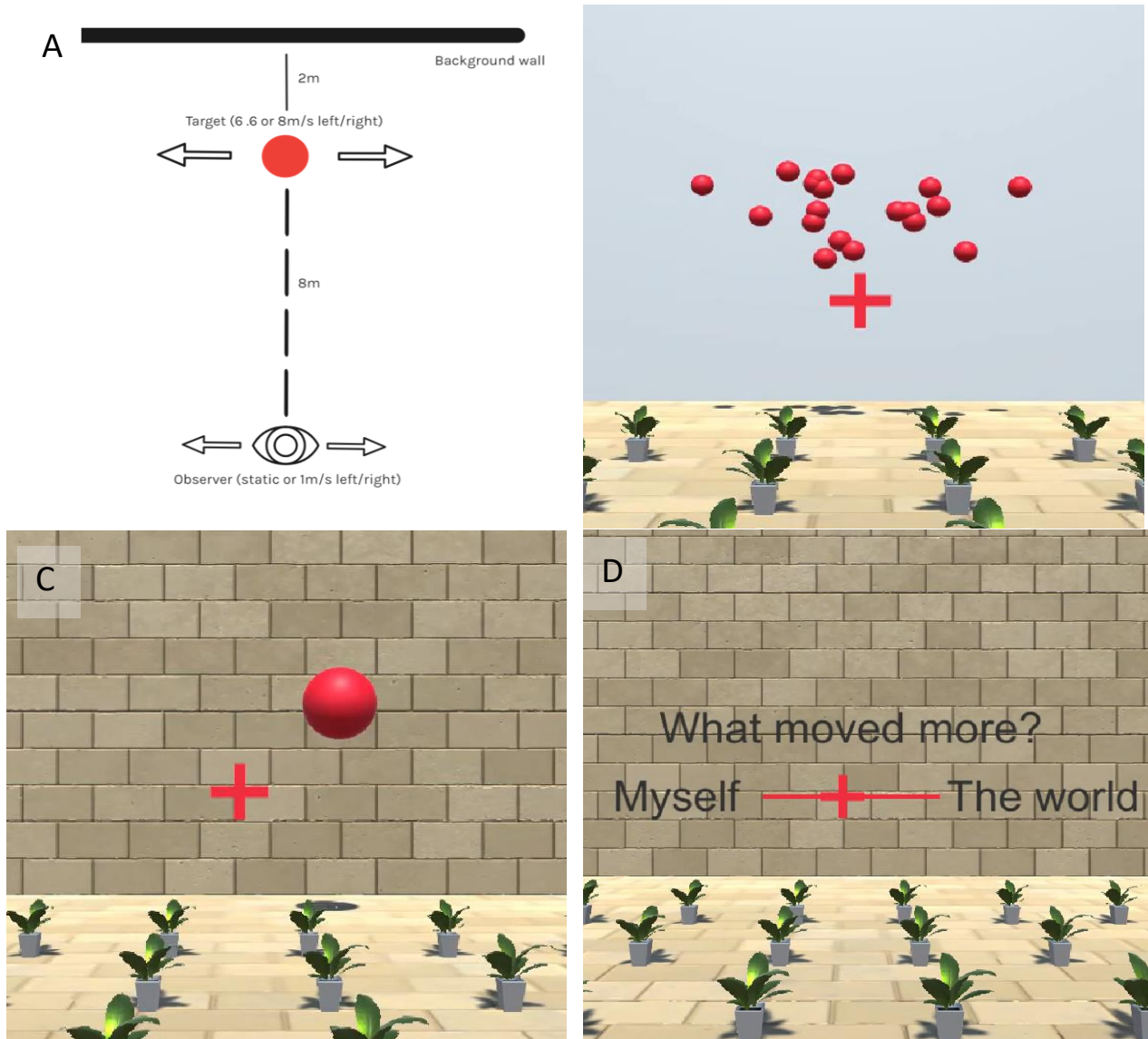


Figure 1: A. Top view of the stimulus scene in one of the test trials. The red circle represents the target, which starts on one side of the midline (see text for details) and moves laterally at 6.6 or 8 m/s for 0.5 s, that is, 3.3 or 4 m. The stylized eye indicates the position of the observer, who can be static or move to the left or to the right for 0.5 s with a Gaussian motion profile and a mean speed of 1 m/s. The target is 8 m away from the observer and 2 m in front of the background wall. B. Screenshot from the program during presentation of the dot cloud in the untextured wall condition ("Blank Backdrop"). C. Screenshot from the program during presentation of the big target in the textured wall condition ("Textured Moving Backdrop"). D. Screenshot from the self-motion judgment conducted after the main body of the experiment. A short sequence of the main experiment can be viewed [here \(Github\)](#). A sequence from the self-motion judgment part of the experiment is available [here \(Github\)](#).

Before starting the actual data collection, participants perform a training session with one PEST where the big target moved at 4 m/s. Subjects were asked to repeat the training if the step size in any of the last five trials was above 0.3 m/s. If they still failed to meet the criterion after a second repetition they were excluded from the experiment.

#### *Intended interpretation of visually simulated observer motion and task*

Our experiment critically depended on our participants perceiving themselves as moving rather than the world as moving. We are furthermore interested in participants making the velocity judgments relative



to the world, not relative to themselves, and we assume that our instruction will make sure of this. However, there is a possibility that either of these assumptions will not hold. This gives rise to five different scenarios during visually simulated observer motion:

- (1) Participants perceive the world as static and themselves as moving and judge object speed relative to the world. This is the intended case.
- (2) Participants perceive themselves as static and the world as moving and judge object speed relative to the world. In this case we would find no effect of visually simulated observer motion at all, that is, there would be no differences between visually simulated observer motion and a (visually) static observer.
- (3) Participants perceive the world as static and themselves as moving and judge object speed relative to themselves. In this case, participants would add the speed of visually simulated observer motion fully onto the target speed, that is, the PSE would be shifted by roughly the mean value of the visually simulated observer motion.
- (4) Participants perceive themselves as static and the world as moving and judge object speed relative to themselves. In this case, participants would also add the speed of visually simulated observer motion fully onto the target speed, that is, the PSE would be shifted by roughly the mean value of the visually simulated observer motion.

To rule out scenarios (2) and (4), we had our participants judge to what extent they felt themselves or the world moving. They were only included into the confirmatory analyses if they had a mean rating between -1 and -0.6 in their judgments about perceived self-motion versus world-motion (see below), indicating that they fully or mostly perceived the world to move, rather than themselves. To achieve that participants made their judgments relative to the world and not to themselves, thus ruling out scenario (3), we gave them very clear instructions to this effect.

It is also possible that participants judge motion partially relative to the world and partially relative to themselves. This seems to be roughly equivalent to a scenario where participants judge motion relative to the world but fail to compensate fully for self-motion in their object speed judgments. Shifting from an observer-centered reference frame to a world-centered references has been suggested as the mechanism behind the accuracy-precision trade-off observed by Dokka et al. (Dokka et al., 2015): in this view, transposing a percept into a world-centered reference frame increases accuracy , i.e., enhances compensation for self-motion, at the cost of a decrease in precision.

#### *A possible confound: Induced Motion*

Induced motion occurs when a stimulus is contained within a moving reference frame. Even in a virtual reality 3D presentation, a texture background might be construed as such reference frame and therefore induce motion. Induced motion would bias perceived speed in the opposite direction to the background motion and could therefore mask potential effects of self-motion. That is, any induced motion would tend to cancel out incomplete compensation for self-motion which would otherwise reveal itself as motion in the same direction as the background, leading to an overall accurate speed estimate. The strength of induced motion depends on several factors: The adjacency principle, which states that stimuli that are closer together in space (in all three dimensions) lead to stronger induced motion (Gogel & Koslow, 1972; Gogel & MacCracken, 1979). Furthermore, enclosure seems to be important. Experiments on induced motion without a full rectangular reference frame are rare, but even moving dots have been shown to induce some motion, albeit to a much lesser extent than a full frame.

Brosigle & Whalen (Brosigle & Whalen, 1967), for example, found that induced motion was halved when using a dot as the inducing stimulus moving 0.2° from the induced stimulus, in comparison to the effect of a full rectangular frame at the same distance. Duncker (Duncker, 1929), observed that a horizontal line moving horizontally induced much less motion than a vertical line moving horizontally or a full rectangle, which is arguably a scenario that comes closest to our display.

**Control Conditions: Setup** – To account for this possible confound, we added two additional conditions. In the first condition (“Blank Backdrop”), we minimized possible induced motion effects by using an untextured wall backdrop (Figure 1B). Motion might still be induced by the other objects in the visual scene, but in absence of any traditional frame, induced motion should be minimal. All other experimental parameters were the same as for the main experiment. In the second condition (“Textured Moving Backdrop”), we aimed to minimize perceived self-motion while keeping the induced motion component of the effect intact. We achieved this by moving only the (textured) wall backdrop of the stimuli, while keeping the rest of the visual scene (textured floor, context objects) static. The wall backdrop moved with the same motion profile as the observer was moved visually in the other conditions. This added another 14 staircases (six for the *Textured Moving Backdrop* condition, and eight for the *Blank Backdrop* condition because, for the latter, we also added two PESTs without visually simulated observer motion) that were interleaved in a random order with the staircases described above.

**Probing Perceived Self-Motion** – To assess whether this manipulation worked as intended, after conclusion of the main body of the experiment, we showed the participants the different conditions (*Simulated Observer Motion with Textured Backdrop*, *Blank Backdrop* and *Textured Moving Backdrop*) without the object motion. After they experienced the stimulus (i.e., visually simulated observer motion in the full room, visually simulated observer motion with the untextured wall background or no visually simulated observer motion, but a moving wall backdrop), we asked them to rate on a continuous scale within the virtual environment (see Figure 1D) to which extent they had perceived themselves or the world/wall as moving. Figure 1D shows a screenshot taken from the judgment phase of the task. We repeated this procedure four times for each condition and direction for a total of 24 trials. This took about two minutes. We included only those participants whose mean ratings per condition were within 0.4 of the expected value (i.e., between -1 and -0.6, indicating a high degree of observer motion, for *Blank Backdrop*; between 0.6 and 1, indicating a high degree of wall motion, for *Textured Moving Backdrop*; and between -1 and -0.6, indicating a high degree of observer motion, for the main test condition, *Simulated Observer Motion with Textured Backdrop*). Data collection continued until we achieved the desired number of participants (see power analysis below).

## Analysis of Pilot Data

### *Main Hypotheses: Visually Simulated Observer Motion*

To test our main hypotheses regarding the influence of visually simulated observer motion on precision and accuracy, we performed the following tests over the main conditions, i.e., when the wall backdrop was textured and the participant was moved visually, not the wall backdrop (“Simulated Observer Motion with Textured Backdrop” versus “Static with textured backdrop”).



To assess the **Just Noticeable Difference (JND)** as a measure of precision, we employ General Linear Mixed Modelling, implemented in the R package lme4, according to the recommendations in (Moscatelli, Mezzetti, & Lacquaniti, 2012). We first established a Test Model, in which responses were fitted to a cumulative Gaussian, with subject ID (“Subject”) and horizontal speed ( $Speed_{horizontal}$ , with values -8, -6.6, 6.6 and 8 m/s) as random effects with random intercepts and random slopes per speed difference between target and ball cloud (“Difference”), and self-motion (binary variable “Subject Motion” with the values “Yes” and “No”) and difference in speed between target and ball cloud (“Difference”) and their interaction as fixed effects. In lme4 syntax, this corresponds to:

$$Subject\ Motion * Difference + (Difference | Subject) + (Difference | Speed_{Target}) \quad [2]$$

We then established a Null Model with subject and horizontal speed as random effects with random intercepts, and subject motion profile and difference in speed between target and ball cloud as fixed effects, but not their interaction:

$$Subject\ Motion + Difference + (Difference | Subject) + (Difference | Speed_{Target}) \quad [3]$$

We then use an ANOVA to test whether the test model was significantly better than the null model. If the interaction term improved the model significantly, the subject motion profile has a relevant influence on the slope of the fitted cumulative Gaussian. We expected the interaction parameter to be lower for Motion = “Congruent” and Motion = “Incongruent”, thus putting into evidence that visually simulated observer motion decreases precision in object speed judgments during self-motion.

To assess the **Point of Subjective Equivalence (PSE)**, our Test Model contained the same random effects as above and the self-motion profile (ternary variable “Motion Profile” with the values “Congruent”, “No Motion” and “Incongruent”) and the speed difference between target and ball cloud (“Difference”) as fixed effects (Moscatelli et al., 2012). The lme4 syntax is:

$$Motion\ Profile + Difference + (Difference | Subject) + (Difference | Velocity_{Target}) \quad [4]$$

The Null Model contained the same random effects, and only the difference in speed between target and ball cloud as a fixed effect.

$$Difference + (Difference | Subject) + (Difference | Velocity_{Target}) \quad [5]$$

We compared both models with an ANOVA and expected the Test Model to be significantly better than the Null Model, indicating that visually simulated observer motion had an impact on the PSE. Visually simulated observer motion in the same direction as the target should decrease perceived target speed and visually evoked self-motion in the opposite direction of the target should increase perceived target speed.

### Control conditions

For the control conditions (*Textured Moving Backdrop* and *Blank Backdrop*), we used the model comparison [4]/[5] to assess whether wall motion (*Textured Moving Backdrop*) or visually simulated observer motion (*Blank Backdrop*), respectively, led to any biases in perceived velocity. For *Textured Moving Backdrop*, we expected that same pattern as for the main condition, but a slightly less complete compensation, with the untextured wall backdrop giving fewer cues about visually simulated observer motion. Furthermore, there should be next to no induced motion effects, which should augment the

observed effect further. For *Blank Backdrop*, we expected a small effect of induced motion in the opposite direction of the effect of visually simulated observer motion, i.e., an overestimation of speed when observer and wall moved in opposite directions, and an underestimation of speed when observer and wall moved in the same direction.

### Power Analysis

Based on the analysis plan above, we proceeded to a power analysis via simulation. We computed the power for the main condition (visually simulated observer motion with a textured wall backdrop). The R code used for this power analysis is available online under <https://github.com/b-jorges/Motion-Perception-during-Self-Motion/blob/master/PowerAnalysisMotionEstimation.R>. We first created datasets that would roughly resemble the data we are expecting to collect. At the core of the simulation of these datasets is the assumptions that responses could be described by a cumulative Gaussian function (which approximates what is commonly known as “Psychometric Function”). The mean of the cumulative Gaussian corresponds to the PSE, and its standard deviation is proportional to the JND. We varied the means of the Gaussian according to the self-motion profile. Pilot data (see below) show consistently a bias to interpret the dot cloud as faster; when the observer is static, we thus assume a PSE of 2/3 of the presented speed. When the observer moved opposite to the target, we expected the PSE to be higher than in the static condition, and when the observer moved with the target, we expect the PSE to be lower. We conducted the power analysis assuming a difference of 1/8 of the mean presented speed of the visually evoked self-motion; (Dokka et al., 2015) found biases up to 50 % of the visually simulated observer motion. Their task, directionality judgments about downward motion with a lateral left- or rightward component, bears some similarities to ours, but is different enough to warrant a more conservative estimate for the sake of the power analysis. We also use a more naturalistic environment which may render flow parsing more complete. For the standard deviation, we parted from a Weber fraction of 7 % for the static condition (McKee, 1981), which corresponds roughly to a standard deviation of 10 % of the PSE. Where the observer is moving, we expected increased JNDs and therefore an increased standard deviation. For the sake of this power analysis, we assume that the standard deviation in this case might be 1/4 higher than the standard deviation for a static observer. (Dokka et al., 2015) found increases of up to 200 % in thresholds from no self-motion to visually simulated observer motion. We choose a much more conservative value to account for task differences. Additionally, we varied the PSE and SD per subject by multiplying them with random values drawn from a normal distribution with a mean of 1 and a standard deviation of 0.1. To account for the fact that our staircase leads to a concentration of responses around the PSE, we drew the stimulus strengths from a Cauchy distribution with a location of 1 and a scale of 0.04. We drew 55 stimulus strengths for this distribution (per combination of target speed and self-motion, we use two PESTs with about 27 trials each; see above) and fed them into the cumulative Gaussian we established per condition and subject. This yielded the answer probability per trial. We then used these probabilities to draw binary answers (PEST faster yes/no) from a Bernoulli distribution for each trial.

We simulated 500 of these data sets, conducted the analyses described above over each for 20, 22, 24, 26, 28 and 30 participants. We report the percentage where the Test Model was significantly better than the Null Model in Table 1.

n	Power Accuracy	Power Precision
---	----------------	-----------------

20	1	0.744
22	1	0.800
24	1	0.812
26	1	0.852
28	1	0.876
30	1	0.910

*Table 1: Simulated power values for 20,22,24,26,28 and 30 participants.*

While the effect should be easily detectable for the accuracy-based hypothesis, the precision hypothesis is somewhat harder to detect and requires at least 30 subjects (for a power above 0.9). Note that, as the simulation process involves several sources of uncertainty, some variability is to be expected in the results, which explains why the power difference between 22 and 24 participants is smaller than the difference between 24 and 26 participants.

Our predictions for the effect of induced motion, which we want to probe for in the control conditions, are about accuracy. Considering that it is generally easier to detect accuracy differences than precision differences with the above method, we are confident that the participant number that allows us to detect the precision main effect will also allow us to detect any relevant effect of induced motion.

#### Pre-existing Data

We collected data from seven pilot participants in the main condition (visually simulated observer motion with a texture wall backdrop). One (s07) was excluded because some of her PESTs did not converge. Two participants (s01 and s02) had previously done the task in 2D, but only their 3D data were included in the analysis. Pilot results are largely in line with our predictions: In terms of JNDs, we found that our Test Model was significantly better than the Null Model ( $p = 0.02$ ), and effects trended in the direction of our hypothesis (regression coefficients of -0.078,  $SE = 0.034$ , for the interaction between visually simulated observer motion present and the difference in speed, which corresponds to a lower precision). For the PSEs, we found that our Test Model was significantly better than the Null Model ( $p < 0.001$ ), and the effects go largely in the expected direction (regression coefficients of 0.072,  $SE = 0.05$ , for the main effect of congruent motion, and -0.25,  $SE = 0.053$ , for the main effect of incongruent motion; which corresponds to a lower perceived speed for congruent motion and visually evoked self-motion, and a higher perceived speed for incongruent motion and visually evoked self-motion). The code used for this analysis as well as the pilot data are available [on GitHub](#).

The pilot data were not included into the final analysis; we recruited 30 new subjects.

#### Open Practices

All raw data collected during this project are published in the GitHub repository <https://github.com/b-jorges/Motion-Perception-during-Self-Motion/>, as well as all the code used for analysis. The Unity projects used to present the stimuli and collect data is available on OSF under <https://osf.io/ugtji/>.

## References

- Aguado, B., & López-Moliner, J. (2019). Perceived speed of motion in depth modulates misjudgements of approaching trajectories consistently with a slow prior. *Vision Research*, 159, 1–9. <https://doi.org/10.1016/j.visres.2019.03.009>
- Aguilar-Lleyda, D., Tubau, E., & López-Moliner, J. (2018). An object-tracking model that combines position and speed explains spatial and temporal responses in a timing task. *Journal of Vision*, 18(12), 12. <https://doi.org/10.1167/18.12.12>
- Becker, W., Nasios, G., Raab, S., & Jürgens, R. (2002). Fusion of vestibular and podokinesthetic information during self-turning towards instructed targets. *Experimental Brain Research*, 144(4), 458–474. <https://doi.org/10.1007/s00221-002-1053-5>
- Brosigale, L., & Whalen, P. M. (1967). The effect of enclosure on the allocation of visually induced movement. *Psychonomic Science*, 8(2), 69–70. <https://doi.org/10.3758/BF03330671>
- Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience*, 10(4), 423–425. <https://doi.org/10.1038/nn1874>
- Dokka, K., MacNeilage, P. R., DeAngelis, G. C., & Angelaki, D. E. (2015). Multisensory self-motion compensation during object trajectory judgments. *Cerebral Cortex*, 25(3), 619–630. <https://doi.org/10.1093/cercor/bht247>
- Duncker, K. (1929). Über induzierte Bewegung - Ein Beitrag zur Theorie optisch wahrgenommener Bewegung. *Psychologische Forschung*, 12(1), 180–259. <https://doi.org/10.1007/BF02409210>
- Dupin, L., & Wexler, M. (2013). Motion perception by a moving observer in a threedimensional environment. *Journal of Vision*, 13(2), 1–14. <https://doi.org/10.1167/13.2.15>
- Dyde, R. T., & Harris, L. R. (2008). The influence of retinal and extra-retinal motion cues on perceived object motion during self-motion. *Journal of Vision*, 8(14), 1–10. <https://doi.org/10.1167/8.14.5>
- Fajen, B. R., Parade, M. S., & Matthis, J. S. (2013). Humans Perceive Object Motion In World Coordinates During Obstacle Avoidance. *Journal of Vision*, 13(8), 1–13. <https://doi.org/10.1167/13.8.25>
- Fetsch, C. R., Deangelis, G. C., & Angelaki, D. E. (2010). Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory. *European Journal of Neuroscience*, 31(10), 1721–1729. <https://doi.org/10.1111/j.1460-9568.2010.07207.x>
- Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience*, 29(49), 15601–15612. <https://doi.org/10.1523/JNEUROSCI.2574-09.2009>
- Fink, P. W., Foo, P. S., & Warren, W. H. (2009). Catching fly balls in virtual reality: A critical test of the outfielder problem. *Journal of Vision*, 9(13), 1–8. <https://doi.org/10.1167/9.13.1>
- Frissen, I., Campos, J. L., Souman, J. L., & Ernst, M. O. (2011). Integration of vestibular and proprioceptive signals for spatial updating. *Experimental Brain Research*, 212(2), 163–176. <https://doi.org/10.1007/s00221-011-2717-9>
- Garzorz, I. T., Freeman, T. C. A., Ernst, M. O., & MacNeilage, P. R. (2018). Insufficient compensation for self-motion during perception of object speed: The vestibular Aubert-Fleischl phenomenon.

Journal of Vision, 18(13), 1–9. <https://doi.org/10.1167/18.13.9>

Gogel, W. C., & Koslow, M. (1972). The adjacency principle and induced movement. *Perception & Psychophysics*, 11(4), 309–314. <https://doi.org/10.3758/BF03210385>

Gogel, W. C., & MacCracken, P. J. (1979). Depth adjacency and induced motion. *Perceptual and Motor Skills*, 48(2), 343–350. <https://doi.org/10.2466/pms.1979.48.2.343>

Gray, R., MacUga, K., & Regan, D. (2004). Long range interactions between object-motion and self-motion in the perception of movement in depth. *Vision Research*, 44(2), 179–195. <https://doi.org/10.1016/j.visres.2003.09.001>

Harris, L. R., Jenkin, M., & Zikovitz, D. C. (2000). Visual and non-visual cues in the perception of linear self motion. *Experimental Brain Research*, 135(1), 12–21. <https://doi.org/10.1007/s002210000504>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2–3), 61–83; discussion 83–135. <https://doi.org/10.1017/S0140525X0999152X>

Hogendoorn, H., Alais, D., MacDougall, H., & Verstraten, F. A. J. (2017). Velocity perception in a moving observer. *Vision Research*, 138, 12–17. <https://doi.org/10.1016/j.visres.2017.06.001>

Ilg, U. J., Schumann, S., & Thier, P. (2004). Posterior parietal cortex neurons encode target motion in world-centered coordinates. *Neuron*, 43(1), 145–151. <https://doi.org/10.1016/j.neuron.2004.06.006>

Jörges, B., & López-Moliner, J. (2019). Earth-Gravity Congruent Motion Facilitates Ocular Control for Pursuit of Parabolic Trajectories. *Scientific Reports*, 9(1), 1–13. <https://doi.org/10.1038/s41598-019-50512-6>

Jürgens, R., & Becker, W. (2006). Perception of angular displacement without landmarks: Evidence for Bayesian fusion of vestibular, optokinetic, podokinesthetic, and cognitive information. *Experimental Brain Research*, 174(3), 528–543. <https://doi.org/10.1007/s00221-006-0486-7>

López-Moliner, J., Brenner, E., Louw, S., & Smeets, J. B. J. (2010). Catching a gently thrown ball. *Experimental Brain Research*, 206(4), 409–417. <https://doi.org/10.1007/s00221-010-2421-1>

MacNeilage, P. R., Zhang, Z., DeAngelis, G. C., & Angelaki, D. E. (2012). Vestibular facilitation of optic flow parsing. *PLoS ONE*, 7(7). <https://doi.org/10.1371/journal.pone.0040264>

McKee, S. P. (1981). A local mechanism for differential velocity detection. *Vision Research*, 21(4), 491–500. [https://doi.org/10.1016/0042-6989\(81\)90095-X](https://doi.org/10.1016/0042-6989(81)90095-X)

Moscattelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision*, 12(11), 1–17. <https://doi.org/10.1167/12.11.26>

Niehorster, D. C., & Li, L. (2017). Accuracy and tuning of flow parsing for visual perception of object motion during self-motion. *I-Perception*, 8(3), 1–18. <https://doi.org/10.1177/2041669517708206>

Probst, T., Loose, R., Niedeggen, M., & Wist, E. R. (1995). Processing of visual motion direction in the fronto-parallel plane in the stationary or moving observer. *Behavioural Brain Research*, 70(2), 133–144. [https://doi.org/10.1016/0166-4328\(95\)80003-4](https://doi.org/10.1016/0166-4328(95)80003-4)

Redlick, F. P., Jenkin, M., & Harris, L. R. (2001). Humans can use optic flow to estimate distance of travel.

483        *Vision Research*, 41(2), 213–219. [https://doi.org/10.1016/S0042-6989\(00\)00243-1](https://doi.org/10.1016/S0042-6989(00)00243-1)

484        Rushton, S. K., & Warren, P. A. (2005). Moving observers, relative retinal motion and the detection of  
485        object movement [2]. *Current Biology*, Vol. 15, pp. 542–543.  
486        <https://doi.org/10.1016/j.cub.2005.07.020>

487        Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient Estimates on Probability Functions. *The Journal*  
488        *of the Acoustical Society of America*, 41(4A), 782–787. <https://doi.org/10.1121/1.1910407>

489        Warren, P. A., & Rushton, S. K. (2007). Perception of object trajectory: Parsing retinal motion into self  
490        and object movement components. *Journal of Vision*, 7(11), 1–11. <https://doi.org/10.1167/7.11.2>

491        Warren, P. A., & Rushton, S. K. (2008). Evidence for flow-parsing in radial flow displays. *Vision Research*,  
492        48(5), 655–663. <https://doi.org/10.1016/j.visres.2007.10.023>

493        Warren, P. A., & Rushton, S. K. (2009). Perception of scene-relative object movement: Optic flow parsing  
494        and the contribution of monocular depth cues. *Vision Research*, 49(11), 1406–1419.  
495        <https://doi.org/10.1016/j.visres.2009.01.016>

496        Wexler, M. (2003). Voluntary head movement and allocentric perception of space. *Psychological*  
497        *Science*, 14(4), 340–346. <https://doi.org/10.1111/1467-9280.14491>

498        Wilson, A. D., & Golonka, S. (2013). Embodied Cognition is Not What you Think it is. *Frontiers in*  
499        *Psychology*, 4(February), 1–13. <https://doi.org/10.3389/fpsyg.2013.00058>

500