

# Coherence of structural visual cues and pictorial gravity paves the way for interceptive actions

**Myrka Zago**

Laboratory of Neuromotor Physiology,  
Santa Lucia Foundation, Rome, Italy



**Barbara La Scaleia**

Laboratory of Neuromotor Physiology,  
Santa Lucia Foundation, Rome, Italy



**William L. Miller**

National Science Foundation, Arlington, VA, USA



**Francesco Lacquaniti**

Laboratory of Neuromotor Physiology,  
Santa Lucia Foundation, Rome, Italy,  
Department of Neuroscience, University of Rome Tor Vergata,  
Rome, Italy, &  
Centre of Space BioMedicine, University of Rome Tor Vergata,  
Rome, Italy



Dealing with upside-down objects is difficult and takes time. Among the cues that are critical for defining object orientation, the visible influence of gravity on the object's motion has received limited attention. Here, we manipulated the alignment of visible gravity and structural visual cues between each other and relative to the orientation of the observer and physical gravity. Participants pressed a button triggering a hitter to intercept a target accelerated by a virtual gravity. A factorial design assessed the effects of scene orientation (normal or inverted) and target gravity (normal or inverted). We found that interception was significantly more successful when scene direction was concordant with target gravity direction, irrespective of whether both were upright or inverted. This was so independent of the hitter type and when performance feedback to the participants was either available ([Experiment 1](#)) or unavailable ([Experiment 2](#)). These results show that the combined influence of visible gravity and structural visual cues can outweigh both physical gravity and viewer-centered cues, leading to rely instead on the congruence of the apparent physical forces acting on people and objects in the scene.

Keywords: interception, scene inversion, pictorial

Citation: Zago, M., La Scaleia, B., Miller, W. L., & Lacquaniti, F. (2011). Coherence of structural visual cues and pictorial gravity paves the way for interceptive actions. *Journal of Vision*, 11(10):13, 1–10, <http://www.journalofvision.org/content/11/10/13>, doi:10.1167/11.10.13.

## Introduction

“I wonder if I shall fall right through the earth! How funny it'll seem to come out among the people that walk with their heads downwards! The Antipathies, I think” (L. Carroll, *Alice's Adventures in Wonderland*). How would Alice's brain deal with the visual dynamics of an upside-down world? This apparently odd question is in fact relevant vis-à-vis the issue of the neural constraints inherent in coding the visual effects of biological and gravitational forces, two key classes of events in our daily experience.

It is easier and faster to deal with objects when they are “right-side up.” “Right-side up” can be defined relative to a variety of biologically relevant reference frames (Howard, 1982; Lacquaniti, 1997). The influence of each such frame can be revealed experimentally by manipulating their relative alignment. The viewer-centered frame is linked to the observer. The gravity-centered frame is

linked to physical gravity. The visual frame is linked to oriented, structural features present in the scene (Runeson, 1988). In outdoor scenes, orientation cues are provided, for instance, by the visible horizon, trees, and buildings. In indoor scenes, orientation can be indicated by the floor, ceiling, lateral walls, objects, and people inside the room. Another potentially important orientation cue is given by the visible influence of gravity on the trajectory of objects moving in the scene.

Most studies dealing with the role of different reference frames have concentrated on perceptual discrimination tasks. Thus, recognition of scenes, people, and actions tends to be faster and more accurate when they are aligned with the observer, whether both the scene and the observer are upright or they are both tilted (e.g., Chang, Harris, & Troje, 2010; Dyde, Jenkin, & Harris, 2006; Epstein, Higgins, Parker, Aguirre, & Cooperman, 2006; Köhler, 1940; Kushiro, Taga, & Watanabe, 2007; McMullen & Jolicoeur, 1992; Reed, Stone, Bozova, & Tanaka, 2003;

Rock, 1988; Troje, 2003; Yin, 1969). For instance, when a digitally edited photograph of a face is presented upside down relative to the observer, the ability to detect gross distortions and abnormalities is strongly impaired and the responses are slowed down (Lobmaier & Mast, 2007; Thompson, 1980), mainly because of a deficit in coding configural information (Freire, Lee, & Symons, 2000). Both the discrimination impairment and the increase in response time are compatible with the hypothesis that the recognition of an inverted, manipulated face requires that featural and configural cues are rotated mentally to a retinal upright orientation, overtaxing the capacity of the underlying mechanisms (Lobmaier & Mast, 2007; Rock, 1988). Similar viewer-centered inversion effects have been described for the discrimination of static whole-body postures (Reed, Stone, Bozova, & Tanaka, 2003) and of biological motion in point-light walk stimuli (e.g., Chang, Harris, & Troje, 2010; Pavlova & Sokolov, 2000; Sumi, 1984). These latter effects depend on an impaired processing of both the global configural shape of the walker (Bertenthal & Pinto, 1994) and the local motion signals of the distal limb segments (Troje & Westhoff, 2006).

Although a viewer-centered reference frame may allow optimal processing of object properties in the canonical perspectives and although egocentric cues dominate in the encoding of visual scenes, allocentric cues contribute as well (see Palmer, 1999). Thus, observers are sensitive to the artificial inversion of the visual effects of gravity on the motion of inanimate objects (Bingham, Schmidt, & Rosenblum, 1995; Indovina et al., 2005). In particular, Bingham et al. (1995) studied the visual recognition of patch-light displays of inanimate events when either the observer or the stimulus was rotated by 180°. Recognition diminished in both conditions as compared to the normal upright condition, but performance was worse when the stimulus was inverted, particularly in the case of events that involved gravity (falling ball, pendulum motion). In addition, a role of orientation relative to gravity has been documented for processing global and local motion cues of point-light walk stimuli, presumably in connection with expectations about the dynamics of the body due to gravity (Chang & Troje, 2009; Shipley, 2003; Troje & Westhoff, 2006).

A growing body of experimental work, including a number of the articles cited above, suggests that the observation of the effects of a gravity field apparently acting on people and objects in a distant scene can provide a significant visual “vertical” reference (Howard, 1982). We employ the term “pictorial gravity” (or “visible gravity”) to distinguish the apparent gravity from physical gravity. Pictorial gravity is not always aligned spatially with physical gravity, as it happens, for instance, when we watch a tilted picture or movie. Moreover, while the magnitude of physical gravitational acceleration of an object is constant, the acceleration of the resulting retinal image is inversely related to the apparent viewing distance (assuming an object’s motion in a fronto-parallel plane).

However, pictorial cues (e.g., familiar size of people and objects, linear perspective, shading, texture gradient) built-in to the visual scene may help viewers gauge the spatial scale and orientation of the scene (Bingham, 1993) and estimate the effects of pictorial gravity on the objects of that scene (Miller et al., 2008). Thus, when pictures of static postures of a human standing on a platform and tilted in the roll plane are presented in different orientations, observers are able to judge the stability of this human body relative to pictorial gravity even for directions that are not concordant with the direction of physical gravity (Lopez, Bachofner, Mercier, & Blanke, 2009). In addition, it has been shown that the discrimination precision of a target motion duration is higher when the target accelerates in the downward direction of a virtual room (consistent with pictorial gravity) than when it accelerates in the upward direction (violating gravity), and this is so whether the virtual room and target motion direction are aligned with the observer or they are tilted by 45° (Moscatelli & Lacquaniti, 2011). Finally, it has been shown that viewing a rotated photograph or video clip with strong polarization cues, which indicate relative “up” and “down” directions in the picture, can alter the perceived direction of absolute “up” and “down” directions in the world (Dyde, Jenkin, & Harris, 2006; Jenkin, Jenkin, Dyde, & Harris, 2004; Jenkin, Dyde, Jenkin, Zacher, & Harris, 2011). Thus, perceptual biases have been documented in relation to the orientation of the stimuli relative to gravity (Chang, Harris, & Troje, 2010; Lobmaier & Mast, 2007; Lopez, Bachofner, Mercier, & Blanke, 2009) and to intrinsic visual references (Jenkin, Jenkin, Dyde, & Harris, 2004; Tadin, Lappin, Blake, & Grossman, 2002).

Despite the wealth of studies on scene orientation, to our knowledge no one has yet addressed the relative roles played by the direction of the scene and by the direction of pictorial gravity on viewers’ responses to objects moving in that scene. When scene direction and target gravity direction are changed independently of each other, will the observer remain tuned to the former, to the latter, or to both directions at the same time?

Here, we addressed this issue by asking participants to press a button that triggered the motion of one of two different types of hitter (depending on the participant’s group). If properly timed, the hitter intercepted a target accelerated by a virtual gravity in a strongly polarized scene. Performance feedback was provided (Experiment 1), or it was suppressed (Experiment 2). In both cases, we employed a factorial design to evaluate the effects of scene orientation (normal or inverted) and target gravity (normal or inverted relative to physical gravity).

## Experiment 1

We presented the scene of an arena (see Figure 1), rendered in perspective with computer graphics and

including several pictorial cues about the up–down polarity and the spatial scale. The scene (“s”) was presented either in the normal orientation relative to an upright observer (downward “s” direction in **Figures 1A** and **1B**) or upside down (upward “s” direction, **Figures 1C** and **1D**). The magnitude of target acceleration (“g”) was always  $9.81 \text{ m s}^{-2}$  (typical gravitational acceleration, scaled to the scene), while its direction was either that of normal gravity (downward “g” in **Figures 1A** and **1D**) or that of an inverted gravity (upward “g,” **Figures 1B** and **1C**). The factorial design crossed the direction of the scene and the direction of target gravity, resulting in four blocked conditions (labeled as in **Figure 1**): (A) normal scene and gravity, (B) normal scene and inverted target gravity, (C) inverted scene and gravity, and (D) inverted scene and normal target gravity. Thus, target kinematics consistent with gravity was concordant with the scene orientation only in conditions A and C.

To investigate the influence of different types of cues, we tested two groups of participants for differences in interception performance. In each group, the button press

triggered the motion of a different virtual effector: (1) a bullet shot by a static human character or (2) a sliding piston. A key difference between these two effectors was that the shooting character in the former protocol was located far from the interception point, whereas the sliding piston in the latter protocol was located near the interception point. In this manner, we tested whether the spatial eccentricity of key reference cues affects the performance. Overall, the study was based on a mixed-model design with scene direction and gravity direction as within-subjects factors and interception mode (Bullet and Piston) as between-groups factor.

## Methods

### Participants

Sixteen subjects participated in this experiment (9 females and 7 males,  $28 \pm 6$  years old, mean  $\pm$  *SD*). They were right-handed (as assessed by a short questionnaire based on the Edinburgh scale), had normal or corrected-

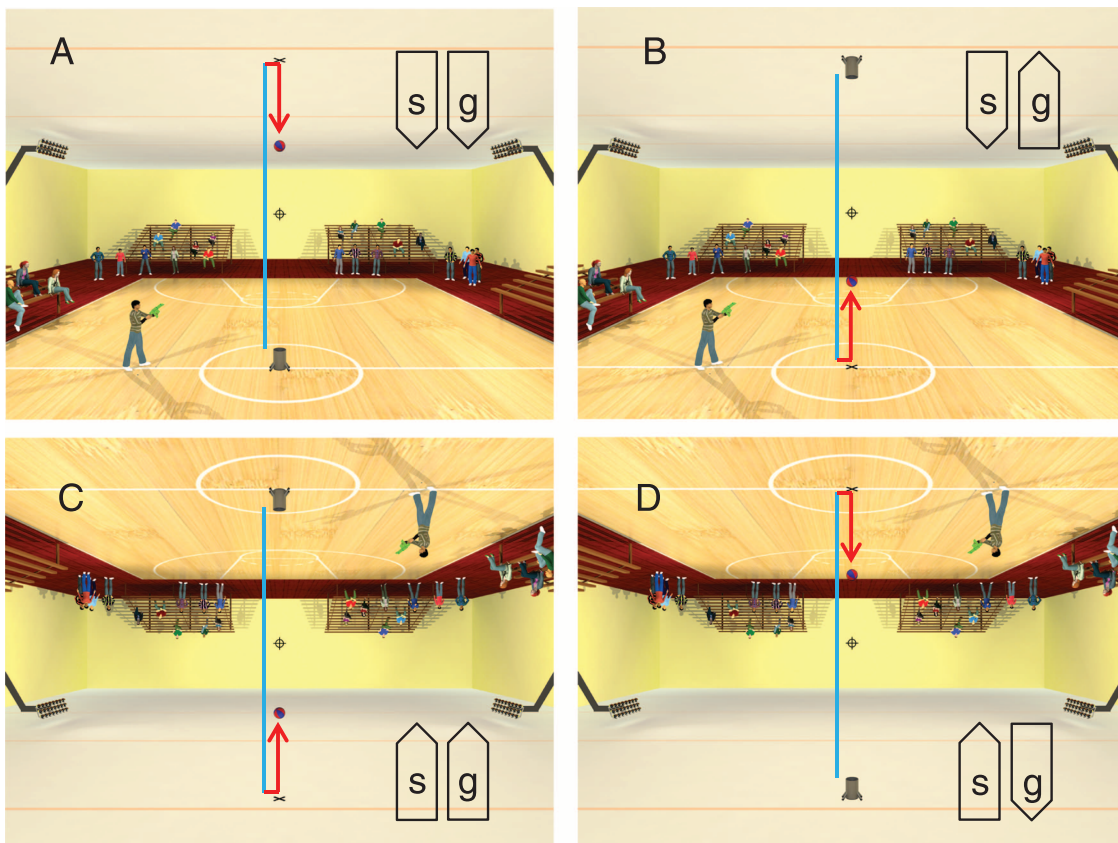


Figure 1. Scenes displayed in the Bullet group. The target ball was launched vertically from the launcher, hit the opposite surface, and bounced back. The target decelerated from launch to bounce (blue trajectory), and it accelerated after bounce (red trajectory). Blue and red segments were not present in the actual movies. When the button was pressed, the standing character shot a bullet toward the interception point (indicated by the crosshair). The direction of the scene (“s”) and the direction of gravity acting on the target (“g”) were varied in different blocks of trials: (A) normal scene and gravity, (B) normal scene and inverted target gravity, (C) inverted scene and gravity, and (D) inverted scene and normal target gravity.

to-normal vision, and gave written informed consent to procedures approved by the Institutional Review Board of Santa Lucia Foundation, in conformity with the Declaration of Helsinki on the use of human subjects in research. Participants were randomly assigned to the Bullet or Piston group, with 8 participants in each group.

### Apparatus, stimuli, and tasks

Participants sat 0.6 m in front of a display (CRT EIZO Flexscan F980, active display size: 402 mm horizontal  $\times$  300 mm vertical, 1600  $\times$  1200 pixel resolution, 100 Hz) in a dimly illuminated room. Visual stimuli were generated with a PNY NVIDIA Quadro FX5600 graphics card. A photodiode counter, placed on an edge of the display, monitored the video output of each image frame at 10 kHz. Participants responded by pressing the button of a computer mouse with their right index finger. Timing of the visual stimuli and motor responses were strictly controlled by means of synchronous acquisition at 10 kHz of the counter and button press via the real-time system PXI-1010 (National Instruments, Austin, TX) under custom LabVIEW (National Instruments, Austin, TX) programs. Moreover, to ensure precise control of timing, all visual stimuli were preprogrammed and downloaded from lookup tables. The stimuli were programmed in C++ using custom software and were rendered using Autodesk Maya 2009 (Autodesk).

Visual stimuli were defined in a right-handed coordinate system with leftward  $X$ -axis and upward  $Y$ -axis in the frontal plane, plus in-depth  $Z$ -axis. Scene projection was computed using on-axis linear perspective, assuming a viewpoint at  $[0, 0, -10 \text{ m}]$  and looking at point  $[0, 0, 0]$ . The horizontal angle of camera view was  $64^\circ$ ; the scene aspect ratio was  $4/3$  (equal to the display aspect ratio). The scene (Figures 1 and 2) subtended  $37^\circ$  by  $28^\circ$ , horizontal and vertical visual angles at 0.6-m viewing distance. The background depicted a portion of an indoor arena (30 m wide  $\times$  56 m long  $\times$  7 m high in the scale of the scene) with a basketball court (15 m wide  $\times$  28 m long) in the center, tiered seating on the sides, and 24 static characters imported from Poser-7 (Smith Micro Software) as spectators placed in different locations to provide an approximate metric reference. The center of the arena was located at the origin  $[0, 0, 0]$  of the  $XYZ$  reference frame. Geometrical cues, textures, directional lights, and shadows were included in the scene to give a sense of depth and up-down. Overhead lighting was provided.

In the foreground, a cylindrical ball launcher (28-cm diameter, 42-cm height) was placed in the middle of the floor or the ceiling (depending on the block, see below). The point on the opposite surface (ceiling or floor) where the ball would bounce after launch was outlined with a black cross. The nominal location at which the ball had to be intercepted by the participant was indicated by a bead, consisting of a crosshair disk placed at the center of the

far wall, halfway (3.5-m distance) between the floor and the ceiling. In projection, the disk appeared of the same size as the target ball at the interception point ( $0.76^\circ$ ). The point of contact between the hitter and the ball was placed at a ball radius distance from the center of the crosshair, on the left side of the ball (relative to the hitter). The scene (including lights and shadows) could be presented either in the normal upright orientation or upside down in different blocks of trials.

During each test trial, the observer waited until the color of the ball launcher changed from gray to green (random delay of 2–4 s from trial start), and then pressed the mouse button to start the animation. After an additional random delay (0.8–1.2 s), a target ball (24.8-cm diameter) was launched vertically with one of five initial speeds (randomized across trials), bounced on the opposite surface, and returned back toward the launcher. The duration of motion from bounce to arrival at the interception point varied between 615 and 815 ms. To increase realism, the ball was also spun at  $150^\circ \text{ s}^{-1}$  around the vertical axis. Target acceleration was  $9.81 \text{ m s}^{-2}$ , downward or upward. The target decelerated from launch to bounce (blue trajectory in Figure 1), and it accelerated after bounce (red in Figure 1). The ball bounced with a 0.748 restitution coefficient (simulating a typical basketball). Air drag was neglected. Participants were instructed to press the button for interception when the target ball passed in front of the crosshair after the bounce. Button press triggered one of two different effectors (depending on the group): a bullet or a piston. If the effector passed through the interception point within  $\pm 35 \text{ ms}$  relative to the nominal interception time, the response was considered a successful hit and the appropriate feedback signal was delivered (see below). This time window roughly coincided with that necessary for the ball to traverse the



Figure 2. Upright scene displayed in the Piston group of Experiment 1. Button press triggered the piston movement.

crosshair bead. Low-resolution, simplified and noninteractive versions of the animations are shown in [Movies 1](#) and [2](#) for condition A.

**Bullet.** On one side of the scene and relatively far from the interception point, a static character held a toy gun aimed toward the interception point (see [Figure 1](#)). The gun was in the same depth plane as the interception point, at 3.38-m ( $11^\circ$ ) distance from it. After 30 ms from the button press, a bullet was shot from the gun with an initial speed of  $22 \text{ m s}^{-1}$  and reached the nominal contact point 160 ms later with a projectile trajectory dictated by the gravity of the scene. When the target was hit, it exploded in fragments falling on the floor or ceiling under the gravity field acting on the target ball. At the same time, a green rotating star appeared near the interception point to provide an additional reward signal. When the target was missed, it continued its trajectory and finally disappeared into the launcher.

**Piston.** Close to the interception point, a piston placed above a high chair (see [Figure 2](#)) had a red spherical head close to the interception point (at 39-cm,  $1.65^\circ$  distance). After 30 ms following the button press, the piston moved horizontally at a constant speed to reach the contact point 160 ms later. Overall, the spherical head remained within the nominal interception zone for 241 ms. If the ball was successfully intercepted, it was deviated from its original trajectory and moved along a parabolic trajectory dictated by its momentum and the gravity acting on the ball. A green rotating star appeared near the interception point. Total duration of success feedback was 1–1.5 s. If the piston head arrived too early but within 241 ms from the nominal interception time, the ball was displayed as bouncing over the piston, falling with a parabolic trajectory and bouncing on the floor or ceiling. If the piston head arrived too late or if it had already moved away at the time of ball arrival, the ball was displayed as continuing its motion until it finally disappeared into the launcher.

## Procedures

Before the experiment, participants received general instructions and familiarized with the setup and visual scenes by observing the experimenter while he/she performed the task 12 times (3 times for each of the 4 types of scenes A–D). The event sequence during this phase was the same as during the experiment with two major exceptions. First, the target ball always moved at one constant speed (resulting in 700-ms duration from bounce to arrival at the interception point). In this manner, there was no clue to pictorial gravity associated with target motion (because of the lack of target acceleration). Second, the performance feedback was predetermined, independent of actual performance. This latter procedure was aimed at demonstrating all possible types of interaction and feedback (success, early and

late responses) to the participant prior to his/her active performance.

After the familiarization phase, participants were pretested with 20 trials (5 for each of the 4 types of scenes A–D) in which the ball moved at one constant speed (same as in the familiarization phase), and they received feedback based on their actual performance.

The experiment started 5 min after the end of the pretest phase. It consisted of 4 blocks (labeled as in [Figure 1](#)): (A) normal scene and gravity, (B) normal scene and inverted target gravity, (C) inverted scene and gravity, and (D) inverted scene and normal target gravity. The order of blocks was counterbalanced across the participants of each group. Each block included 200 trials consisting of 40 repetitions for each of the 5 initial target speeds, randomized across trials. Target acceleration was  $9.81 \text{ m s}^{-2}$ . The participants of both groups were exposed to identical trial sequences of initial target speed. Each experiment lasted about 1 h 30 min, with 15 min of rest halfway.

## Data analysis

For each trial, we computed the timing error (TE) as the button press time plus 190 ms (time interval between trigger time and arrival time of the effector at the contact point) minus the motion duration of the ball.  $TE = 0$  corresponds to the ideal response, while negative (positive) values of TE correspond to early (late) responses. For each subject, success rate (SR) was computed as the fraction of trials in which the target was intercepted within the time window of  $\pm 35$  ms relative to the nominal interception time, that is,  $-35 \text{ ms} \leq TE \leq 35 \text{ ms}$ . In particular, for the test trials, SR was assessed over all 5 motion durations and the last 20 repetitions of each condition to exclude the initial transient fluctuations of the responses. Success rate was normalized according to  $\arcsin(\text{SR})^{0.5}$ . Repeated-measures mixed ANOVA (split-plot ANOVA) was carried out over the two groups (Bullet and Piston), using “gravity direction” and “scene direction” as within-subjects factors and “interception mode” as between-groups factor. The degrees of freedom for the within-subjects comparisons were corrected in case of deviance from sphericity (Greenhouse–Geisser). An alpha level of 0.05 was used for all statistical tests, and the size of the effect was reported by means of the partial eta squared ( $\eta_p^2$ ). Post-hoc comparisons were conducted using the Fisher’s LSD test.

## Results

To verify that neither interception mode nor group membership *per se* introduced nonspecific biases in the performance, before the actual experiment the participants

of both groups were pretested with a task involving constant speed targets. Here, there was no clue to pictorial gravity associated with target motion (because of the lack of target acceleration). Interception was carried out by means of the same effector (Bullet or Piston) used during the actual experiment. The success rate was low ( $0.33 \pm 0.15$ , mean  $\pm$  *SD*, 2 scene directions  $\times$  2 motion directions  $\times$  8 subjects  $\times$  2 groups,  $n = 64$ ), and it did not vary significantly between the two groups in the pretest trials. Thus, three-way mixed ANOVA (2 [scene direction]  $\times$  2 [motion direction]  $\times$  2 [interception mode]) on the success rate showed no significant effect of either the “interception mode,” “scene,” “motion,” or the interactions.

During the test trials with accelerating targets, the timing errors (TEs) were generally small and without systematic biases toward early or late responses. Mean TE (computed over all 40 repetitions of each condition) was  $2 \pm 14$  ms (2 scene directions  $\times$  2 gravity directions  $\times$  5 motion durations  $\times$  8 subjects  $\times$  2 groups,  $n = 320$ ). However, the performance was not stationary throughout a block of trials but tended to improve with practice due to performance feedback. The changes of TE over successive repetitions of each condition were best fit by means of an exponential function whose time constant (range of 2–8 repetitions) was compatible with a relatively fast learning rate, as it has also been observed in other interception tasks (Zago, Iosa, Maffei, & Lacquaniti, 2010). On average, the absolute value of TE was significantly lower in the last 3 repetitions than in the first 3 repetitions of each condition in both groups (paired *t*-tests, 3 repetitions  $\times$  5 motion durations,  $n = 15$ ,  $P < 0.05$ ).

Given the nonstationarities due to practice effects, the performance was compared across conditions close to steady state. On average, the success rate over the last 20 repetitions was reasonably high ( $0.62 \pm 0.03$ , 2 scene directions  $\times$  2 gravity directions  $\times$  8 subjects  $\times$  2 groups,  $n = 64$ ). However, success rate was not homogenous

across conditions (Figure 3). Indeed, three-way mixed ANOVA (2 [scene direction]  $\times$  2 [gravity direction]  $\times$  2 [interception mode]) showed a significant effect of the interaction between “scene” and “gravity” ( $F_{1,14} = 8.692$ ,  $P = 0.011$ ,  $\eta_p^2 = 0.383$ ), related to the fact that average performance was significantly more successful in the two congruent conditions (A and C) than in the two incongruent conditions (B and D). Thus, mean success rate was  $0.66 (\pm 0.03)$ , 2 scenes  $\times$  8 subjects  $\times$  2 groups,  $n = 32$ ) in the congruent conditions (A and C) and  $0.59 (\pm 0.03)$  in the incongruent conditions (B and D). Post-hoc tests did not reveal any significant difference between the two congruent conditions or between the two incongruent conditions. Instead, both congruent conditions were significantly different ( $P < 0.05$ ) from both incongruent conditions. The effect of the interaction between “scene” and “gravity” was found not only in the average group responses but also at the level of most (6/8) single subjects.

## Experiment 2

Experiment 1 demonstrated that, after learning, accelerating targets are intercepted more successfully when scene direction and gravity direction are concordant, irrespective of the specific effector utilized for interception (Bullet or Piston). Participants were always provided with performance feedback in quasi real time, that is, they operated in a closed-loop mode. In Experiment 2, we tested whether the findings generalize also to open-loop conditions. The visual stimuli and task were similar to those of Experiment 1, but participants were never provided with performance feedback. In addition, to suppress the presence of indirect clues about the relative arrival timing of the ball and the hitter (bullet or piston) in

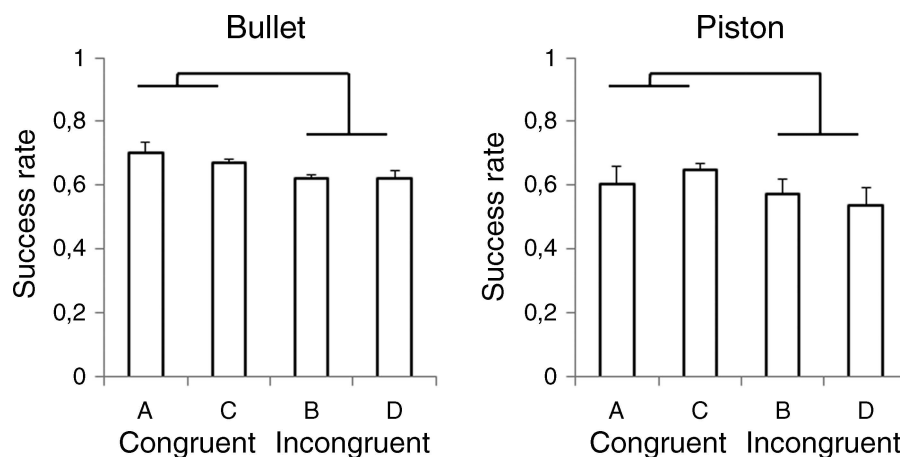


Figure 3. Success rate (over the last 20 repetitions) for each type of scene in the two groups of Experiment 1. Brackets indicate that success rate is significantly ( $P < 0.05$ ) higher for the congruent scenes (A and C) than for the incongruent ones (B and D).

the nominal interception zone, the ball was rendered transiently invisible while it traversed this zone.

## Methods

### Participants

Sixteen right-handed subjects (different from those of [Experiment 1](#)) gave written informed consent to participate in the experiment (8 females and 8 males,  $28 \pm 5$  years old, mean  $\pm$  SD). Participants were randomly assigned to the Bullet or Piston group, with 8 participants in each group.

### Apparatus, stimuli, and tasks

The experimental setup, stimuli, and protocol were identical to those of [Experiment 1](#), except for the following changes. The image of the ball faded out progressively while the ball approached the interception zone (over the last 4 frames prior to arrival), it was fully invisible in the nominal interception zone (1 frame), and it faded in while the ball moved out of the interception zone (first 4 frames after arrival). The ball finally returned into the launcher. In contrast with the ball image, the image of the hitter was never suppressed.

### Procedures and data analysis

Procedures were identical to those described for [Experiment 1](#), except that during all experimental phases (familiarization, pretest, and test trials) the ball image was suppressed around nominal interception time and no performance feedback was provided. Data analysis was the same as for [Experiment 1](#).

## Results

As was found in [Experiment 1](#), the performance did not vary significantly between the two groups in the pretest

trials with constant speed targets. Three-way mixed ANOVA (2 [scene direction]  $\times$  2 [motion direction]  $\times$  2 [interception mode]) on the success rate showed no significant effect for any factor or interaction.

As expected, the lack of performance feedback resulted in much larger timing errors in [Experiment 2](#) than in [Experiment 1](#). Mean timing error (computed over all 40 repetitions of each condition) was  $-43 \pm 42$  ms (5 motion durations  $\times$  2 scene directions  $\times$  2 gravity directions  $\times$  8 subjects  $\times$  2 groups,  $n = 320$ ), indicating a systematic bias toward premature responses.

Success rate tended to be low, due to the substantial timing errors. For the sake of comparison with the results reported for [Experiment 1](#), we computed the success rate over the last 20 repetitions of each condition. On average, this value was  $0.37 (\pm 0.07, 2 \text{ scene directions} \times 2 \text{ gravity directions} \times 8 \text{ subjects} \times 2 \text{ groups}, n = 64)$ , much lower than in [Experiment 1](#). However, as in [Experiment 1](#), success rate was not homogenous across conditions ([Figure 4](#)). Three-way mixed ANOVA (2 [scene direction]  $\times$  2 [gravity direction]  $\times$  2 [interception mode]) showed a significant effect of the interaction between “scene” and “gravity” ( $F_{1,14} = 13.490, P = 0.003, \eta_p^2 = 0.491$ ), related to the fact that average performance was significantly more successful in the two congruent conditions (A and C) than in the two incongruent conditions (B and D). Thus, mean success rate was  $0.41 (\pm 0.06, 2 \text{ scenes} \times 8 \text{ subjects} \times 2 \text{ groups}, n = 32)$  in the congruent conditions (A and C) and  $0.33 (\pm 0.07)$  in the incongruent conditions (B and D). Post-hoc tests did not reveal any significant difference between the two congruent conditions or between the two incongruent conditions. Instead, both congruent conditions were significantly different ( $P < 0.05$ ) from both incongruent conditions. The effect of the interaction between “scene” and “gravity” was found not only in the average group responses but also at the level of most (7/8) single subjects.

In sum, the results of [Experiment 2](#) obtained in the absence of performance feedback confirmed those of [Experiment 1](#).

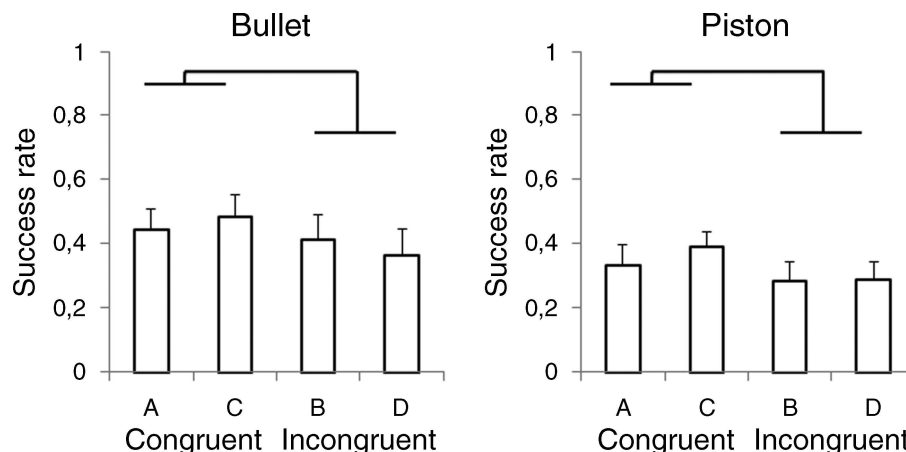


Figure 4. Mean success rate for each type of scene in the two groups of [Experiment 2](#). Same format as in [Figure 3](#).

## Discussion

Two experiments showed that, when intercepting a target that accelerates under gravity along the vertical, the brain can downplay both viewer-centered and gravicentric reference frames and can rely instead on the congruence of the apparent physical forces acting on people and objects in the scene. Indeed, best performance occurred, on average, when the direction of the background scene (and its implied gravity) and the direction of target gravity were concordant, irrespective of whether both were upright or inverted relative to the observer. These results were obtained in both the protocol involving a shooting scene and that involving a sliding piston, despite the wide spatial separation between these two hitters (the shooting character was located far from the interception point, whereas the sliding piston was located near the interception point). Thus, the specific interception mode and the spatial eccentricity of key reference cues were not critical for the effect of scene/gravity congruence.

Strikingly, we found that performance could be superior when both the scene and target gravity were inverted (condition C) than when the scene was upright but target gravity was inverted (condition B). This finding, *prima facie*, conflicts with most previous reports on the effects of scene inversion, which showed that the processing of upside-down scenes is less efficient than that of upright scenes (see [Introduction](#) section). However, it should be remarked that our conclusions are directly applicable only to active interactions with moving objects, that is, to the conditions we tested here. It remains to be seen whether similar conclusions apply to perceptual recognition and discrimination of objects, the tasks typically investigated in studies of scene inversion.

To our knowledge, this is the first study that independently manipulated the direction of the background scene and that of target gravity and, therefore, looked for interactions that go beyond the role of configural cues (mainly involved in the inversion effects). We believe that the critical interaction in our experiments was determined by the congruence of the physical forces apparently acting on the actors and objects that populated the scene. Thus, the layout of the arena we displayed in the background scene implied the static effects of a gravity field pulling people and objects against the support surface.

In [Experiment 1](#), participants received online feedback about their success or failure in intercepting the moving target within the allocated time window. In line of principle, therefore, the effect of scene/gravity congruence may have depended on a trial-by-trial calibration of the performance based on feedback. However, this explanation cannot hold for the results of [Experiment 2](#), in which no performance feedback was ever provided. Here, the effect of scene/gravity congruence must have arisen solely based on intrinsic *a priori* expectations about the most plausible force fields acting globally on objects and people in the scene.

The ability to use congruence in interception timing implies that the brain relies on a self-consistent global representation of different physical forces rather than separately representing the force fields acting locally on subsets of objects. Global congruence represents an ecologically valid solution. It also guarantees accurate decoding of the visual effects of environmental forces, a scheme capable of generalizing across different views of a scene, including the unusual inverted view. We also note that the idea that the brain can take into account congruency of kinetics may be considered germane to the well-established notion that the visual system is able to use contextual information to facilitate perception when the usual spatial relations among objects hold, giving rise to overall contextual congruency (see [Palmer, 1999](#)).

After visual images have been organized into coherent objects and their motions have been interpreted in the context of specific reference frames, further neural processing presumably occurs in order to decode the event that gave rise to the corresponding retinal inputs ([Palmer, 1999](#)). We speculate that event decoding includes coherent and calibrated estimates of the force patterns (kinetics) underlying visual kinematics, although at this stage we cannot determine conclusively whether kinetics is specified quantitatively ([Runeson, Juslin, & Olsson, 2000](#)) or it is based on some simpler, qualitative heuristics ([Gilden & Proffitt, 1989](#); [Zago, McIntyre, Senot, & Lacquaniti, 2008](#)).

Our results may also be relevant to the current debate of whether neural representations for computing time are anchored to retinal coordinates ([Bruno, Ayhan, & Johnston, 2010](#)) or may exhibit a genuine spatial tuning in external space ([Burr, Cicchini, Arrighi, & Morrone, 2011](#)). Clearly, the present finding that timing performance could be equally good irrespective of whether the scene was upright or inverted is compatible with flexible representations in external space.

Returning to the initial question of how Alice would deal with the visual dynamics of an upside-down world, our data suggest that she would do quite well, provided that both people and gravity were inverted in Wonderland (as realized in the scene of the tunnel fall of the charming and insightful cartoon by Walt Disney).

## Acknowledgments

We thank Vincenzo Maffei for early discussions leading to this work, Giuseppe Cotignola and Riccardo De Marco for help with the setup and the experiments, and Alessandro Moscatelli and Paolo Viviani for comments on the manuscript. The work was supported by the Italian Health Ministry, Italian University Ministry (PRIN Project), and Italian Space Agency (DCMC and CRUSOE Grants). W. L. M. was formerly affiliated with the Santa Lucia Foundation and is now affiliated with the National Science



Foundation, but the work described in this manuscript does not necessarily represent the views of the National Science Foundation or the United States Government.

Commercial relationships: none.

Corresponding author: Dr. Myrka Zago.

Email: m.zago@hsantalucia.it.

Address: Santa Lucia Foundation, via Ardeatina 306, 00179 Rome, Italy.

## References

- Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, *5*, 221–225.
- Bingham, G. P. (1993). Perceiving the size of trees: Form as information about scale. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1139–1161.
- Bingham, G. P., Schmidt, R. C., & Rosenblum, L. D. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1473–1493.
- Bruno, A., Ayhan, I., & Johnston, A. (2010). Retinotopic adaptation-based visual duration compression. *Journal of Vision*, *10*(10):30, 1–18, <http://www.journalofvision.org/content/10/10/30>, doi:10.1167/10.10.30. [PubMed] [Article]
- Burr, D., Cicchini, G. M., Arrighi, R., & Morrone, M. C. (2011). Spatiotopic selectivity of adaptation-based compression of event duration. *Journal of Vision*, *11*(2):21, 1–9, <http://www.journalofvision.org/content/11/2/21>, doi:10.1167/11.2.21. [PubMed] [Article]
- Chang, D. H., Harris, L. R., & Troje, N. F. (2010). Frames of reference for biological motion and face perception. *Journal of Vision*, *10*(6):22, 1–11, <http://www.journalofvision.org/content/10/6/22>, doi:10.1167/10.6.22. [PubMed] [Article]
- Chang, D. H. F., & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision*, *9*(1):19, 1–17, <http://www.journalofvision.org/content/9/1/19>, doi:10.1167/9.1.19. [PubMed] [Article]
- Dyde, R. T., Jenkin, M. R., & Harris, L. R. (2006). The subjective visual vertical and the perceptual upright. *Experimental Brain Research*, *173*, 612–622.
- Epstein, R. A., Higgins, J. S., Parker, W., Aguirre, G. K., & Cooperman, S. (2006). Cortical correlates of face and scene inversion: A comparison. *Neuropsychologia*, *44*, 1145–1158.
- Freire, A., Lee, K., & Symons, L. A. (2000). The face inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*, 159–170.
- Gilden, D. L., & Proffitt, D. R. (1989). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 372–383.
- Howard, I. P. (1982). *Human visual orientation*. New York: Wiley.
- Indovina, I., Maffei, V., Bosco, G., Zago, M., Macaluso, E., & Lacquaniti, F. (2005). Representation of visual gravitational motion in the human vestibular cortex. *Science*, *308*, 416–419.
- Jenkin, H. L., Jenkin, M. R., Dyde, R. T., & Harris, L. R. (2004). Shape-from-shading depends on visual, gravitational, and body-orientation cues. *Perception*, *33*, 1453–1461.
- Jenkin, M. R., Dyde, R. T., Jenkin, H. L., Zacher, J. E., & Harris, L. R. (2011). Perceptual upright: The relative effectiveness of dynamic and static images under different gravity states. *Seeing Perceiving*, *24*, 53–64.
- Köhler, W. (1940). *Dynamics in Psychology*. New York: Liveright.
- Kushiro, K., Taga, G., & Watanabe, H. (2007). Frame of reference for visual perception in young infants during change of body position. *Experimental Brain Research*, *183*, 523–529.
- Lacquaniti, F. (1997). Frames of reference in sensorimotor coordination. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (vol. 11, pp. 27–64). Amsterdam: Elsevier.
- Lobmaier, J. S., & Mast, F. W. (2007). The Thatcher illusion: Rotating the viewer instead of the picture. *Perception*, *36*, 537–546.
- Lopez, C., Bachofner, C., Mercier, M., & Blanke, O. (2009). Gravity and observer's body orientation influence the visual perception of human body postures. *Journal of Vision*, *9*(5):1, 11–14, <http://www.journalofvision.org/content/9/5/1>, doi:10.1167/9.5.1. [PubMed] [Article]
- McMullen, P. A., & Jolicoeur, P. (1992). Reference frame and effects of orientation on finding the tops of rotated objects. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 807–820.
- Miller, W. L., Maffei, V., Bosco, G., Iosa, M., Zago, M., Macaluso, E., & Lacquaniti, F. (2008). Vestibular nuclei and cerebellum put visual gravitational motion in context. *Journal of Neurophysiology*, *99*, 1969–1982.
- Moscattelli, A., & Lacquaniti, F. (2011). The weight of time: Gravitational force enhances discrimination of visual motion duration. *Journal of Vision*, *11*(4):5, 1–17, <http://www.journalofvision.org/content/11/4/5>, doi:10.1167/11.4.5. [PubMed] [Article]

- Palmer, S. E. (1999). *Vision science. Photons to phenomenology*. Cambridge, MA: MIT Press.
- Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics*, *62*, 889–899.
- Reed, C. L., Stone, V., Bozova, S., & Tanaka, J. (2003). The body inversion effect. *Psychological Science*, *14*, 302–308.
- Rock, I. (1988). On Thompson's inverted-face phenomenon. *Perception*, *17*, 815–817.
- Runeson, S. (1988). The distorted room illusion, equivalent configurations, and the specificity of static optic arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 295–304.
- Runeson, S., Juslin, P., & Olsson, H. (2000). Visual perception of dynamic properties: Cue heuristics versus direct-perceptual competence. *Psychological Review*, *107*, 525–555.
- Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science*, *14*, 377–380.
- Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception*, *13*, 283–286.
- Tadin, D., Lappin, J. S., Blake, R., & Grossman, E. D. (2002). What constitutes an efficient reference frame for vision? *Nature Neuroscience*, *5*, 1010–1015.
- Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception*, *9*, 483–484.
- Troje, N. F. (2003). Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception*, *32*, 201–210.
- Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a “life detector?” *Current Biology*, *16*, 821–824.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.
- Zago, M., Iosa, M., Maffei, V., & Lacquaniti, F. (2010). Extrapolation of vertical target motion through a brief visual occlusion. *Experimental Brain Research*, *201*, 365–384.
- Zago, M., McIntyre, J., Senot, P., & Lacquaniti, F. (2008). Internal models and prediction of visual gravitational motion. *Vision Research*, *48*, 1532–1538.