Stereovision for action reflects our perceptual experience of distance and depth

Carlo Campagnoli

Department of Psychology, Princeton University, Princeton, NJ, USA, ✉

Sholei Croom

Department of Cognitive, Linguistic, and Psychological Science, Brown University, Providence, RI, USA ✉

Fulvio Domini

Department of Cognitive, Linguistic, and Psychological Science, Brown University, Providence, RI, USA
Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy ✉

Binocular vision is widely recognized as the most reliable source of 3D information within the peripersonal space, where grasping takes place. Since grasping is normally successful, it is often assumed that stereovision for action is accurate. This claim contradicts psychophysical studies showing that observers cannot estimate the 3D properties of an object veridically from binocular information. In two experiments, we compared a front-to-back grasp with a perceptual depth estimation task and found that in both conditions participants consistently relied on the same distorted 3D representation. The subjects experienced (a) compression of egocentric distances: objects looked closer to each other along the *z*-axis than they were, and (b) underconstancy of relative depth: closer objects looked deeper than farther objects. These biases, which stem from the same mechanism, varied in magnitude across observers, but they equally affected the perceptual and grasping task of each subject. In a third experiment, we found that the visuomotor system compensates for these systematic errors, which are present at planning, through online corrections allowed by visual and haptic feedback of the hand. Furthermore, we hypothesized that the two phenomena would give rise to estimates of the same depth interval that are geometrically inconsistent. Indeed, in a fourth experiment, we show that the landing positions of the grasping digits differ systematically depending on whether they result from absolute distance estimates or relative depth estimates, even when the targeted spatial locations are identical.

Introduction

To pick up an object we must first estimate its position relative to our body and its shape. Stereovision can unambiguously determine these properties, given that ocular vergence specifies the egocentric distance of an object while binocular disparities determine its 3D structure (Howard & Rogers, 1995; Wheatstone, 1838). Since this source of information is most reliable within the personal space of an agent, where reach-to-grasp actions are performed, it has been assumed to be accurate (Bradshaw et al., 2004; Jackson, Jones, Newport, & Pritchard, 1997; Loftus, Servos, Goodale, Mendarozqueta, & Mon-Williams, 2004; Melmoth & Grant, 2006; Mon-Williams & Dijkerman, 1999; Servos & Goodale, 1994; Servos, Goodale, & Jakobson, 1992; Watt & Bradshaw, 2000).

However, this postulated accuracy seems to be at odds with findings in perceptual studies. When only binocular vision is available, observers typically experience two phenomena: (a) *visuospatial compression*, objects at different distances look closer to each other than they really are; and (b) *depth underconstancy*, objects appear increasingly shallower as their distance from the viewer increases (Foley, 1980; Johnston, 1991; Sousa, Brenner, & Smeets, 2011; Tittle, Todd, Perotti, & Norman, 1995; Volcic, Fantoni, Caudek, Assad, & Domini, 2013).

Although recent investigations have revealed that similar patterns of visual distortions affect reach-to-grasp actions (Bozzacchi & Domini, 2015), it is still unclear whether stereovision for action yields the same 3D information as stereovision for perception. The aim

Physical distances   **Perceived distances**   **Egocentric function**

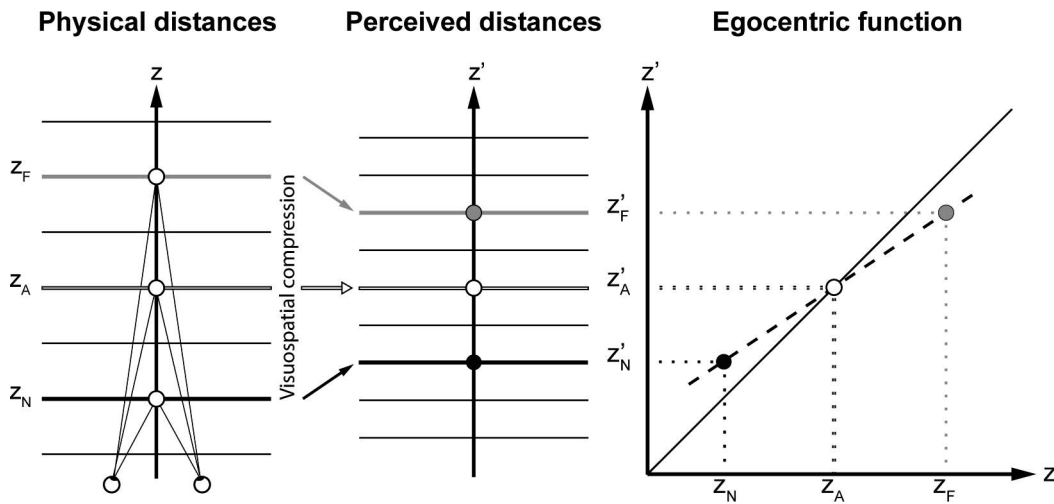
Figure 1. Left, an observer fixates at three distances (z_N, z_A, z_F). Center, perceived distances are less separated than their physical counterparts, resulting in visuospatial compression: near locations (black) appear more distant, whereas far locations (gray) appear closer. Right, the *egocentric function* (dashed line) defines the relationship between physical (z) and estimated (z') distances. Since the range of estimated distances is smaller than the physical range, the slope of this function is smaller than 1.

of this study is to show that the same mechanisms underlying perceptual biases affect the planning of goal-directed actions.

Visuospatial compression and depth underconstancy

Binocular disparities can specify the exact depth of an object only if scaled by an accurate estimate of the viewing distance. Within the peripersonal space, stereovision provides two fundamental sources of information for distance. The first, which is extraretinal, is the vergence angle formed by the eyes when fixating at a specific location (Cormack, 1984; Foley, 1980). The second, of retinal origin, is the pattern of vertical disparities, due to the separation between corresponding projections in the direction orthogonal to the interocular axis (Bishop, 1989; Bradshaw, Glennerster, & Rogers, 1996; Brenner, Smeets, & Landy, 2001; Mayhew & Longuet-Higgins, 1982; Rogers & Bradshaw, 1995). When vertical disparities are negligible and vergence is the main cue to distance, observers experience what we will refer to as visuospatial compression.

Visuospatial compression can be ascribed to a systematic failure in the estimation of an object's egocentric distance. According to Foley (1980), "Perceived distance of near targets exceeds physical distance; perceived distance of far targets is less than physical distance" (p. 411). In other words, objects closer than a specific distance z_A are seen more distant than they are and those farther than z_A are seen closer than they are. The perceptual space of egocentric distances is therefore compressed, since, as indicated in

the diagram of Figure 1, the spacing between egocentric distances in the physical domain maps onto a smaller spacing in the perceptual domain.

If we assume that within a small range of distances this mapping is linear (Foley, 1977), the relation between physical (z) and perceptual (z') distances can be modeled with a line, which we define as the *egocentric function*, having a slope smaller than 1 and passing through the point of veridicality (z_A, $z'_A = z_A$) (Figure 1). The slope of the egocentric function, or *compression factor*, decreases as the magnitude of visuospatial compression increases.

The compression of visual space not only affects the perceived egocentric distance of an object, but also its apparent depth extent Δz (Figure 2). In the domain of stereovision, Δz is accurately specified by retinal disparity information only if scaled by its egocentric distance z (Brenner & van Damme, 1999; Glennerster, Rogers, & Bradshaw, 1996; Rogers & Bradshaw, 1995). However, if the *scaling distance* z_s is smaller than the true distance ($z_s < z$), the object's depth is consequently underestimated and, conversely, $z_s > z$ leads to the overestimation of Δz.

Visuospatial compression therefore gives rise to depth underconstancy (Gregory, 1963; Howard & Rogers, 1995): (a) stimuli close to the body appear deeper, because they are judged as more distant (Figure 2, black triangle), and (b) stimuli far from the body appear shallower, because they are judged as closer than they are (Figure 2, gray triangle). As a result, depth underconstancy can be described as a compression of scaling distances (see Appendix). The relationship between z_s and z will be termed *scaling function*. If the scaling distances coincide with the perceived egocentric distances ($z_s = z'$), then the scaling function
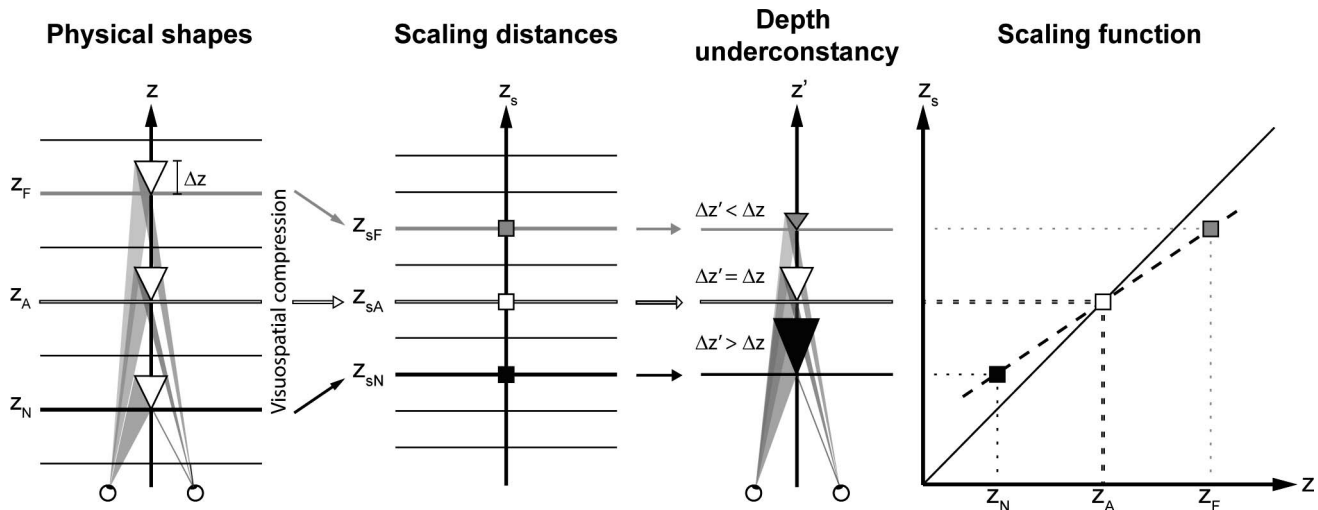
Figure 2. Visuospatial compression leads to depth underconstancy. Left, the relative disparity between two points (the angular difference between the projections of the same segment on the eyes—shaded areas) uniquely specifies their depth separation Δz once the viewing distance z is known. Center, if the object's distance is underestimated (gray square), its depth looks shallower ($\Delta z' < \Delta z$, gray triangle) and vice versa (black square and triangle). Right, the scaling function defines the relationship between physical (z) and scaling (z_s) distances; z_A indicates the distance where judgments are veridical (see Appendix).

will also have a compression factor smaller than 1 (Figure 2, right).

Does vision for action share the same 3D representation as vision for perception?

The previous section illustrates two ways in which visuospatial compression can be analytically described within the domain of stereovision: (a) through an egocentric function, obtained by measuring the perceived position of an object, and (b) through a scaling function, obtained by measuring the object's apparent depth and calculating the corresponding scaling distance.

In previous work, we found that when the vision of the hand is prevented during grasping, the reaching endpoint reveals systematic biases in the estimation of egocentric distance (Bozzacchi, Volcic, & Domini, 2014, 2016; Campagnoli et al., 2012). Moreover, we have also shown that both perceptual judgments and reach-to-grasp actions reveal systematic depth underconstancy (Bozzacchi & Domini, 2015; Foster, Fantoni, Caudek, & Domini, 2011; Volcic et al., 2013). These results suggest that both the egocentric function and the scaling function are compatible with a compression of visual space. However, they do not provide a direct comparison between perception and action.

The main goal of this study was to test two hypotheses: (a) The visual space is characterized by visuospatial compression of egocentric distances, which causes depth underconstancy (in other words, the egocentric and the scaling functions are the same) and

(b) The visual space, specific for each individual, is the same for perception and reach-to-grasp actions (i.e., the individual egocentric and scaling functions are the same across tasks).

These hypotheses were tested by asking participants to judge the relative depth of a three-rod configuration and to grasp it (see Figure 3A). In order to avoid effects of learning, which may reduce or eliminate the very biases we are interested in measuring, subjects grasped virtual objects. They were instructed to shape their hand at the end of the action as if they were holding the object firmly between their index finger and thumb. Since the required action involved a front-to-back grip of the object, the final grip aperture (*FGA*) provided a measure of visual estimate of depth. In addition to the *FGA* we also measured the participants' final hand position (*FHP*). Defined as the position reached by the midpoint between their index finger and thumb at the end of the movement, the *FHP* is a measure of the estimated distance to the object. The perceptual counterpart of the *FGA* was the manual size estimation (*MSE*), where observers matched the sensed separation between their index finger and thumb to the perceived depth of the object. The resemblance of the *MSE* task to a pantomime grasp performed at a specific location raises the question as to how proprioceptive signals from the final position of the hand and the haptic feedback are integrated for the execution of a grasp in depth. In Experiments 2 and 3, we investigated this point.

According to the first hypothesis, visuospatial compression and depth underconstancy stem from the same mechanism. Thus, objects near to the body
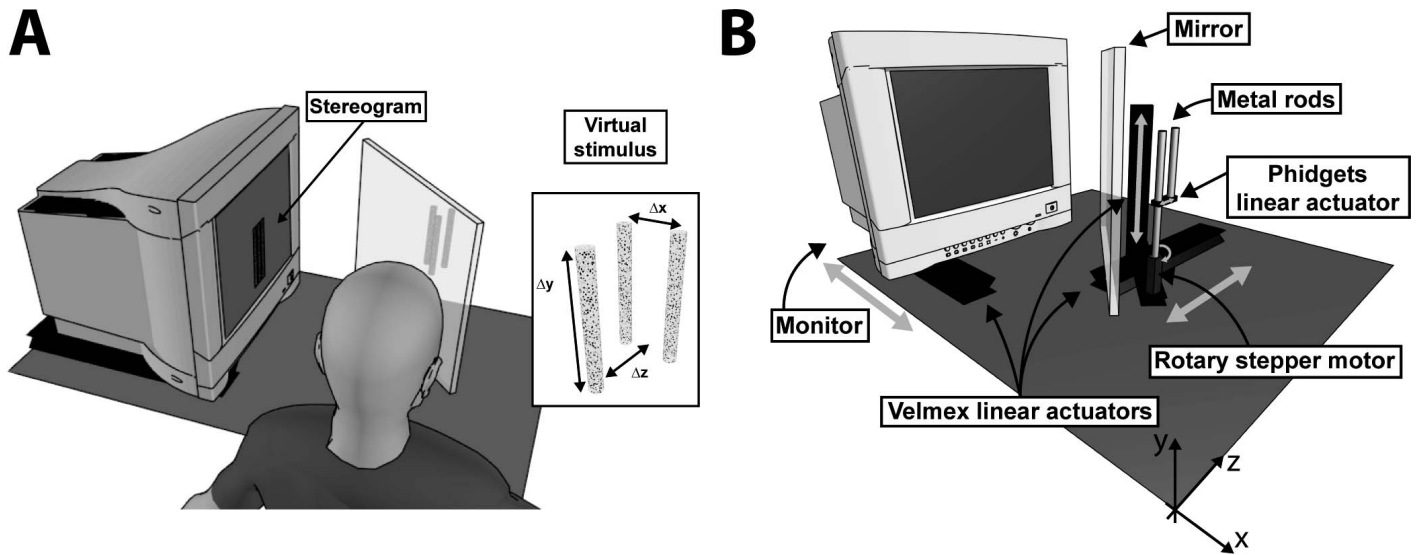
Figure 3. (A) In all experiments, the stereogram of the stimulus rendered on the monitor was reflected by a mirror and viewed by the observer as a 3D virtual object at some distance beyond the mirror's surface, through the use of 3D goggles. The viewing distance of the virtual stimulus was modified by moving the monitor either closer to the mirror or farther away from the mirror. (B) Schematics of the setup. Monitor and physical objects were moved by a series of linear actuators. Gray arrows indicate the direction of motion provided by each motor.

appear farther away and deeper than they are, while objects at a greater distance appear closer and shallower than they are. A direct test of the first hypothesis is only possible in the grasping task, where the estimates of both distance and depth are provided on each trial by the *FHP* and the *FGA,* respectively. Let's say, for instance, that two subjects, A and B, are looking at the same visual scene. Subject A experiences a largely compressed visual space, while Subject B is almost veridical. When Subject A is asked to grasp an object at two distances, z_N and z_F, she reaches instead at the distances FHP_N and FHP_F, where the separation between these two locations ($FHP_F - FHP_N$) is half the actual separation ($z_F - z_N$). Therefore, the egocentric function of this subject has a compression factor of 0.5. Subject B, who reaches almost at the correct locations, will show instead a compression factor of nearly 1.0. Regarding the grip aperture, Subject A should grasp the near object as deeper than the far object, while Subject B should grasp them as almost identical. This is because, according to the first hypothesis, both the scaling function and the egocentric function of Subject A should show a compression factor of 0.5, while both functions of Subject B should show a compression factor close to 1.0.

According to the second hypothesis, the same distortions of visual space affect perception and action. As a consequence, the *MSE* task should show the same biases as the grasp's *FGA*: Subject A should perceive the near object as deeper than the far one (scaling function with compression factor of 0.5), while Subject B should perceive the two objects as almost identical

(scaling function with compression factor of almost 1.0). As a result, the scaling function derived through inverse geometry from the MSE should match the egocentric function derived from the *FHP* of the grasp.

In summary, the visuospatial compression of egocentric distances revealed by the final hand position in a grasp should coincide, for each individual, with the visuospatial compression of scaling distances derived from depth estimates in both grasping and perception.

Overview of the experiments

We conducted four experiments to investigate whether stereovision for action is the same as stereovision for perception. In Experiment 1, we compared a perceptual task (manually estimate the relative depth of an object presented at one of two distances) with a front-to-back grasp. During the grasp we removed the vision of the hand and used virtual objects to avoid final touch, in order to ensure that the movement was based purely on planning information, without in-flight corrections due to online feedback. We tested whether: (a) the final position of the hand in the grasp was consistent with visuospatial compression, (b) the final grip aperture in both tasks was consistent with depth underconstancy, and (c) the responses of each subject in both tasks revealed a unique representation of the visual space.

We also reasoned that depth underconstancy can be counteracted by simulating objects with different depth extents at two distances, so that they are perceived and

grasped as if they were identical. This hypothesis was tested in Experiment 2.

In Experiment 3, we asked whether adding visual and haptic feedback from the hand helps to overcome the biases of perception while executing a grasp.

In Experiment 4, we addressed an interesting paradox regarding visuospatial compression and depth underconstancy: intervals defined by distance and depth judgments are mutually incompatible. That is, reaching movements aimed at two points should define a different depth separation than the final grip aperture of a grasp movement directed towards the same locations. To test this hypothesis, participants either reached separately to the locations of an object's front and back surfaces or grasped the object front-to-back.

General methods

Participants

Sixty-four students (38 women, 26 men) participated in the experiments (Experiment 1, $n = 13$; Experiment 2, $n = 10$; Experiment 3, front-to-back grasp, $n = 18$; Experiment 3, along-the-side grasp, $n = 17$; Experiment 4, $n = 6$). All participants self-identified as right-handed with normal or corrected-to-normal vision. Experiments 1 and 4 were conducted at the Center for Neuroscience and Cognitive Systems (CNCS) of the Italian Institute of Technology. The participants received a reimbursement of €8 per hour for their effort. The experiment was approved by the Comitato Etico per la Sperimentazione con l'Essere Vivente of the University of Trento and in compliance with the Declaration of Helsinki. Experiments 2 and 3 were conducted at Brown University, where the subjects were paid $8 per hour or granted course credit for participation. Each participant gave informed consent prior to the experiment, which was approved by the Brown University Institutional Review Board.

Apparatus

All four experiments had the same setup, even though they were conducted in two different laboratories (Figure 3). Subjects sat with their head on a chin rest installed along one of the short sides of a rectangular table, facing a semitransparent mirror that was slanted 45° with respect to the fronto-parallel plane. A monitor was located to the left of the observer's head and oriented such that the images on the screen were reflected on the mirror's surface and appeared in front of the eyes. A system of Velmex linear actuators (Velmex, Inc., Bloomfield, NY) was

installed on the table: one actuator carried the monitor along the rightwards-leftwards direction (with respect to the observer's point of view), thus causing the virtual projection on the mirror to look closer or more distant, respectively; another group of two actuators was installed behind the mirror and carried a platform along the z and y axes (relative to the observer's direction of sight, the x axis corresponded to left–right direction, the y axis to the vertical direction, and the z axis to the forward direction). This latter system was used for different purposes depending on the experiments (see below). In Experiments 1 and 2, the motor noise correlated with the viewing distance of the stimuli, and could potentially be picked up as a cue during motor planning. To control for this possibility, in Experiments 3 and 4, the motors reached a midway distance between trials and then they were brought to viewing distance. The fact that we found the same pattern of biases regardless of such manipulation is consistent with the notion that spatial judgments are heavily weighted in favor of vision (Bertelson & Aschersleben, 1998). In the Experiment 3, a combination of small linear actuators attached to a stepper motor (Phidgets Inc., Calgary, AB, Canada) was also mounted on the platform.

Using shutter glasses (liquid-crystal FE-1 goggles by Cambridge Research Systems [Rochester, UK] at CNCS; NVIDIA 3D Vision® 2 wireless glasses [NVIDIA, Santa Clara, CA] at Brown) synchronized with the refresh rate of the screen, participants saw disparity-defined 3D virtual objects that were presented at various distances with consistent vergence and accommodative cues. The position of the eyes in the virtual reality was adapted to each participant's interocular distance, measured with a digital pupillometer (Reichert Inc., Depew, NY). We used an Optotrak 3020 Certus motion capture system and small infrared emitters (Northern Digital, Waterloo, ON, Canada) to record index and thumb trajectories. A calibration procedure allowed us to track the 3D position of the index finger and thumb's finger pads, which was calculated in real time based on the coordinates of three markers attached to each fingernail. Optotrak recordings and the motion of both Velmex and Phidgets actuators were controlled by custom C++ programs using specific API routines provided by each vendor. The rendering of the stimuli was made using OpenGL (the stimuli are described in detail in each experiment's methods). The width of all stimuli, 4 cm, subtended a small visual angle, to maximize the role of vergence for the estimation of distance, at the expense of vertical disparities (Bradshaw et al., 1996). The accommodative distance, determined by the reflected monitor's surface, bisected the distance between the front and back of the stimuli in Experiments 1, 2, and 4, whereas it coincided with

the front surface in Experiment 3. The maximum distance between the surface of the monitor and the simulated front (or back) of the stimuli was 3 cm, (with the only exception of 5 cm in one stimulus of Experiment 3). This solution made sure that the vergence-accommodation conflict was negligible (Buckley & Frisby, 1993; Frisby, Buckley, & Horsman, 1995; Watt, Akeley, Ernst, & Banks, 2005). In all experiments, the grasping movements started roughly from the same position relative to the body ("home"), located about 20 cm to the right, 30 cm below, and 5 cm ahead of the cyclopean eye.

Data Analysis

Raw movement kinematics of the fingertips of index finger and thumb were processed and analyzed offline, using R (R Core Team, 2017) with custom functions. From the raw 3D positional data of index and thumb we calculated the raw profiles of grip aperture (the Euclidean distance between the fingertips) and hand position (the middle point between the fingertips). The raw data were then smoothed using a cubic spline. Missing data due to sudden invisibility of the markers were interpolated within a maximum window of 17 frames (approximately 200 ms). The first two derivatives of the smoothed trajectories were then calculated to obtain velocity and acceleration profiles of index, thumb, grip aperture, and hand position.

The movement was first segmented into two main sections: (a) from the movement onset (defined as the moment when the hand was further than 5 cm away from the home position along the z-axis) to the time of the maximum grip aperture (MGA), and (b) from the time of the MGA to the end of the movement. In trials without visual and haptic feedback (Experiments 1 and 2, and half of the trials of Experiment 4), we extracted two grasp-related dependent variables from the movement trajectories: Final grip aperture (FGA) and final hand position (FHP). The extraction of the FGA comprised three steps: First, we discarded the portion of the grasp previous to the MGA. Second, from the remaining part, we selected only the portion in which the hand position was no more than 50 mm away from the center of the visual field (which was also the center of the 2D projection of the stimulus) along the x and y dimensions (within a limited set of distances, the positioning of the hand direction-wise is normally highly accurate and precise, unlike the positioning in depth; see for example Messier & Kalaska, 1997). Finally, from the resulting window, we extracted the FGA, which corresponded to the value of the grip aperture that minimized both net velocity and net acceleration. The FHP was extracted next, as the value of the hand position at the time of the FGA. In trials

with feedback (when the hand touched the object and the fingertips were visible; Experiment 3 and half of the trials of Experiment 4), we analyzed the full trajectory by calculating the frame-by-frame Euclidean distance between the thumb and its final position and then subdividing it into 100 equally spaced bins from movement onset to movement end. In order to observe the evolution of the grasp during the reaching, we extracted the average grip aperture in each bin of each trial of each subject. In addition, in Experiment 1 we also extracted the manual size estimation (MSE), defined as the value of the grip aperture during the last frame of the movement recording.

Visual inspection of each individual grasp was performed afterwards by localizing the extracted variables on the trajectories, to check the accuracy of the abovementioned procedure. With the exception of few trials, which were excluded because the trajectory was characterized by many local minima and maxima (overall, 1% of the 5,240 grasps performed over the course of the four experiments), the method was successfully applied to obtain the entire set of dependent variables. All linear fits were done by modeling the dependent variable in question with a linear mixed-effects model using the subject's intercept as random component, with the R package lme4 (Bates, Maechler, Bolker, & Walker, 2015). Significance test on the models' coefficients was performed using the Kenward-Roger approximation for degrees of freedom (Luke, 2017).

Experiment 1

We directly tested the two hypotheses motivating this investigation. First, is there a geometrically consistent relationship between distance estimates and depth estimates? Second, do biases in perception and action stem from the same visual representation?

Methods

Stimuli

The target consisted of patterns of random red dots representing the surface of three cylinders, each 0.5 cm wide in diameter and 6 cm high (Figure 4A). The dots were uniformly distributed on the virtual rods' surfaces, not on the image plane. Their luminance was low enough such that the visual scene was completely dark with the only exception of the stimuli (since the biases we observed did not change after prolonged dark adaptation, we safely assume that all the other elements of the laboratory were not visible enough to provide additional depth information). All the monocular cues were consistent with the stimuli's 3D shape with the
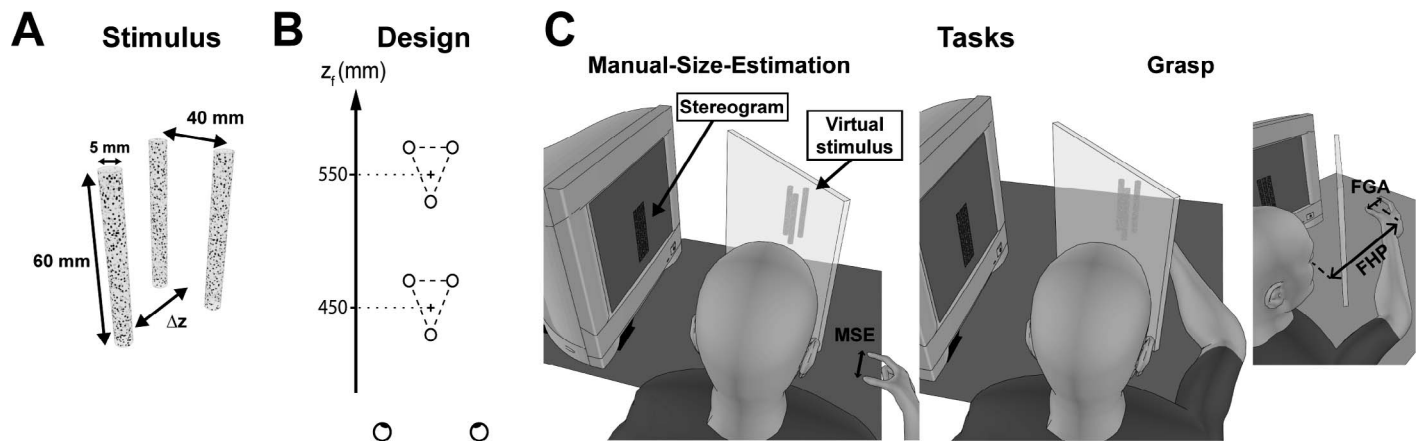
Figure 4. (A) The stimulus consisted of three random-dots cylinders of 5 mm diameter and 60 mm tall. These were positioned at the vertexes of a triangle having a constant base of 40 mm and a variable height (the stimulus depth Δz, which was 30, 40, 50, or 60 mm). (B) Bird's eye view of the experimental design: in both tasks, subjects saw the stimulus at one of two egocentric distances (450 and 550 mm). (C) Tasks and variables. During *MSE*, subjects kept their hand in the same position close to the body. During grasp, participants did not see their hand and limb, nor did they touch any physical object.

exception of the dot size. Given the small diameter of the rods, perceptual distortions of their shape had negligible effects. All three rods appeared vertically oriented, centered at eye height and equally spaced along the horizontal axis, such that the middle rod was at the center of the visual field, whereas the center of the others were 2 cm to its right or left. The middle rod was simulated in front of the two flanking rods at one of four depth separations (30, 40, 50, or 60 mm) and never occluded them. Subjects were instructed to focus on the overall structure of the stimulus: most participants described the target as a "triangle" or a "prism." Here, the term *object* will refer to the three rods as a whole, and the term *viewing distance* to the optical distance of the reflected surface of the monitor.

Procedure

In two separate blocks, participants either judged the depth of the three-rod configuration or grasped it. In each block they viewed a total of 8 stimuli five times. Presented in random order, the stimuli comprised 4 depths (30, 40, 50, and 60 mm) × 2 distances (450 and 550 mm); Figure 4B). The width of the stimulus (the horizontal distance between the rear rods) was always 4 cm (Figure 4A). In each trial, subjects saw only one stimulus at a time, either at the near or at the far distance.

In the *MSE* task, participants rested their hand on a wooden platform, which was located close to their body, and each trial began with the fingers closed together. At a beep, the target appeared automatically and the participants were instructed to open their index finger and thumb along the line of sight without moving the hand, as to indicate the perceived depth of the object (Figure 4C, left). We asked subjects to lift the

fingers from the platform while responding, instead of sliding along the side. This prevented that irregularities on the wooden surface could bias the manual estimation. Each trial lasted 1500 ms from the stimulus onset.

In the grasping task, participants began each trial by having their index and thumb closed together at start position. At a beep, the target appeared. Participants were asked to reach to the object and grasp it front-to-back, pretending to hold it between their fingers until it disappeared 1500 ms later. Since participants had to perform the action without either seeing their hand or receiving haptic feedback, they underwent a short initial training until they could act in the most natural way, as if they interacted with a real object (Figure 4C, right).

Results and discussion

Figure 5 shows the average final hand position of the grasp as function of the stimulus's distance, for each relative depth. Subjects grasped the near and the far objects as if they were closer to each other, consistent with visuospatial compression. The egocentric function resulting from fitting the *FHP* data showed a slope smaller than 1, slope$_{FHP\text{-}z}$ = 0.69, 95% CI = [.43, .95]. A repeated measures ANOVA on the *FHP* also revealed a significant effect of the object depth, $F(3, 36) = 7.86$, $p < 0.001$, as shallower objects were grasped farther than deeper objects. Although egocentric distance estimates should only depend on vergence information, this result suggests that allocentric information can influence the transport component of a grasp as well (Sousa, Brenner, & Smeets, 2010; Sousa et al., 2011).
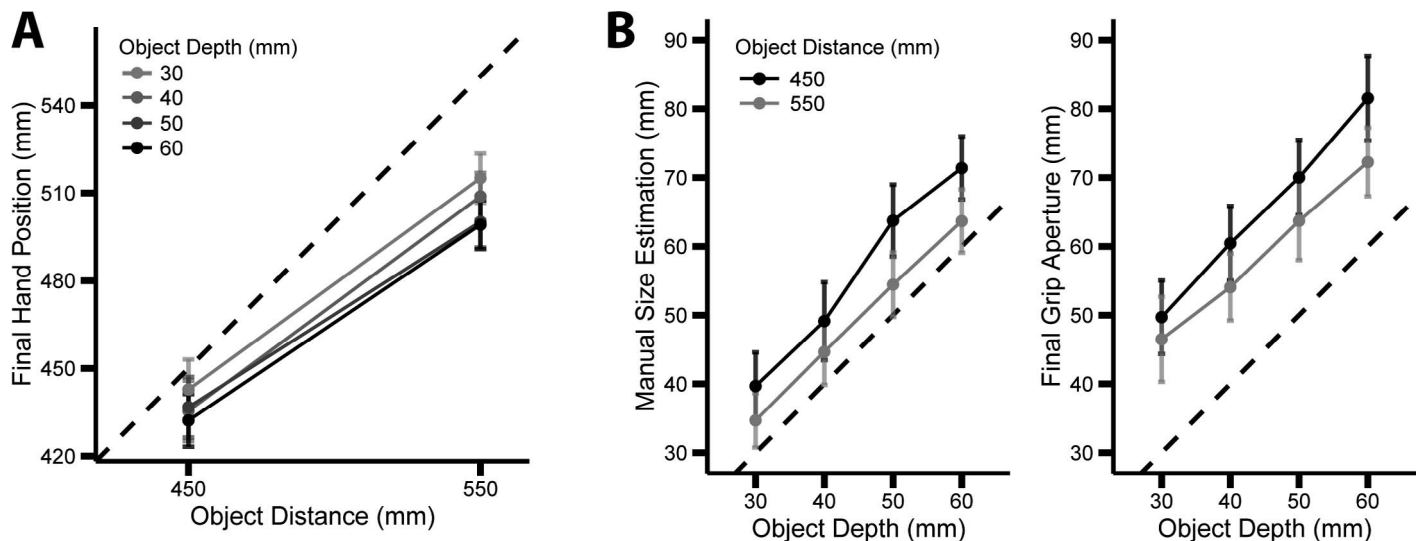
Figure 5. (A) Mean *FHP* of the grasp as function of object distance for each object depth. (B) Mean *MSE* (left) and *FGA* (right) as functions of object depth for each viewing distance. Error bars represent 1 *SEM*. The dashed line indicates accurate performance.

In Figure 5B, the average *MSE* (left) and *FGA* (right) are plotted as function of the object's depth for each viewing distance. In agreement with the hypothesis that the compression of space affected the scaling of binocular disparity during both the perceptual and grasping tasks, both dependent variables were larger for the near stimulus than for the far stimulus. A repeated-measures ANOVA on either response (*FGA* or *MSE*) using distance (450, 550 mm), object's depth (30, 40, 50, 60 mm) and task (grasping vs. perceptual) as within-participant variables revealed three significant main effects: depth: $F(3, 36) = 190.32$, $p < 0.001$; task: $F(1, 12) = 7.66$, $p < 0.02$; and distance: $F(1, 12) = 16.9$, $p < 0.01$. No significant interaction between distance and task was found, $F(1, 12) = 0.05$, $p = 0.82$, indicating that the magnitude of the compression of visual space was not statistically different for perceptual and grasping tasks. Note that these results were obtained by averaging the responses across subjects, to show the general trend in the data. However, doing so may cover individual differences in the scaling of depth information, which were indeed revealed in the following analysis.

To directly test the two main hypotheses, we estimated the egocentric function from the *FHP*, and the scaling functions from the *MSE* and the *FGA*. The egocentric function was determined for each participant through a linear fit of *FHP* as function of viewing distance. The scaling functions were estimated through a nonlinear parameter fit of *MSE* and *FGA* as function of viewing distance and object depth. The nonlinear function is the inverse mapping of the retinal disparity to the observed depth estimates (see Appendix, Equation A7). Figure 6 shows the individual compression factors of the scaling function for the grasping (A) and the perceptual (B) tasks plotted against the

individual compression factors of the egocentric function. Each data point in the graph corresponds to individual subjects.

In agreement with the first hypothesis, the compression of visual space affects the reaching and the grip aperture in the same way (Figure 6A): for every individual, the compression factors of the scaling and egocentric functions derived from the grasp are not statistically different from each other. For example, a subject who exhibited a strong visuospatial compression (Figure 6A, data point enclosed in a square) reached the near and the far stimuli as if they looked much closer to each other than they were, yielding an egocentric function with slope around 0.5. The same subject also grasped the near stimulus as if it looked deeper than the far one, yielding a scaling function with also a slope around 0.5. Conversely, the subject in the encircled data point of Figure 6A exhibited a weak visuospatial compression; thus, she reached to the stimuli at almost their true distance and grasped them with nearly the same *FGA*, resulting in two compression factors close to 1.0.

The second hypothesis, predicting a unique 3D representation for perception and action, is also supported by the data, as can be seen on Figure 6B. Here the slope of the individual scaling functions is derived from the *MSE* data of the perceptual task. Note how the two highlighted subjects exhibit consistent compression of visual space across tasks. The compression factors derived from the *FGA* and the *MSE* were each fitted by a linear model using the *FHP*-derived compression factor as predictor. Consistent with the two hypotheses, the intercepts of both fits were not significantly different from zero: intercept$_{FHP-FGA}$ = −0.25, 95% CI = [−0.61, 0.10]; intercept$_{FHP-MSE}$ = −0.07, 95% CI = [−0.49, 0.34], while the slopes of both
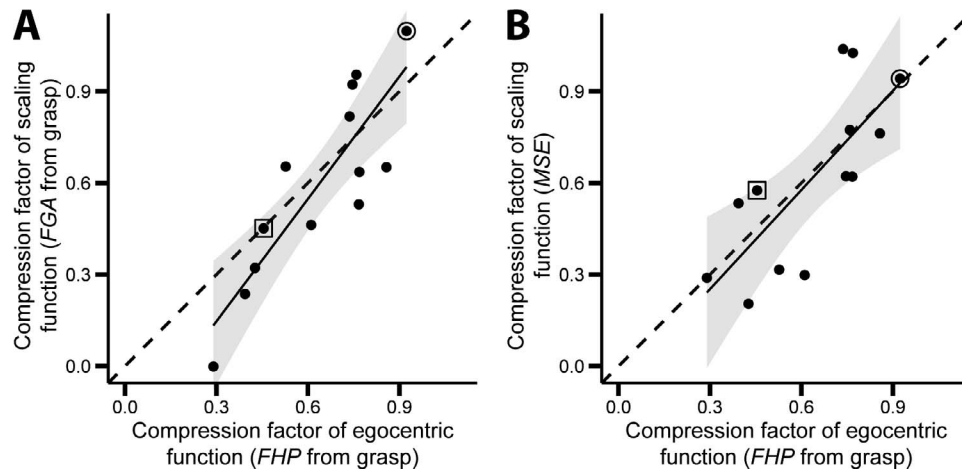
Figure 6. Individual compression factor of the scaling function estimated from the *FGA* of the grasp (A) and from the MSE of the perceptual task (B) as function of the compression factor of the egocentric function estimated from the *FHP*. Each data point is a subject. The dashed line identifies perfect prediction. The black lines correspond to linear fits, with the gray regions specifying the 95% confidence intervals. The framed data points show the results of the same pair of individuals. The participant identified with a square showed a large visuospatial compression, resulting in a shallow egocentric function. Consistently, the slopes of the scaling functions derived from the data of the grasping and perceptual task were also small (compression factor = approx. 0.5). On the other hand, the participant identified with a circle exhibited nearly veridical egocentric and scaling functions (compression factor = approx. 1) in both tasks.

fits were not significantly different from 1: (slope$_{FHP-FGA}$ = 1.34, 95% CI = [0.80, 1.87]; slope$_{FHP-MSE}$ = 1.08, 95% CI = [0.45, 1.71]. This result is particularly interesting, because it shows a strong relationship between parameters generated by fitting data of two entirely different tasks: distance estimates measured in a grasping task predict depth estimates obtained with a perceptual task.

Finally, the average intercepts of the three fits (*FHP*-derived, *FGA*-derived, and *MSE*-derived), corresponding to the point of veridicality, z_A, illustrated in Figures 1 and 2, were not significantly different from each other, $F(2, 24) = 0.2$, $p = 0.82$, remarking the close similarity between spatial representations during grasping and perception.

Experiment 2a

The results of the previous experiment are compatible with the hypothesis that depth perception and reach-to-grasp actions stem from a common representation of visual space (Kopiske, Bruno, Hesse, Schenk, & Franz, 2016). The aim of Experiment 2 was to provide a further test to this hypothesis by showing that physically different, but perceptually matched, stimuli elicit the same motor planning. Each observer performed a perceptual task by setting the simulated depth and width of a three-rod configuration at two egocentric distances to match its structure to that of a standard full-cue stimulus. We predicted that a far

object must be set deeper than a near object to compensate for depth underconstancy.

In a successive task, participants were asked to grasp the two objects they set in the matching procedure. If the hypothesis tested in Experiment 1 is true, then the final grip aperture should be the same for the two objects, in spite of their different physical structure.

Methods

Procedure

Observers performed the perceptual matching task followed by a reach-to-grasp task. In the perceptual task, they matched the perceived structure of a *reference stimulus* to that of a *test stimulus*, identical to the virtual three-rod configuration described in the previous experiment.

The reference stimulus was a physical reproduction of the virtual stimulus, consisting of three wooden cylinders covered with white paper and positioned behind the mirror. The front rod was installed on a movable platform, whereas the two rear rods were attached to a fixed stand. Since the experiment was conducted in the dark, the reference stimulus was made visible by a small black light positioned next to it. To guarantee uniform illumination of the object, a small mirror was positioned at the opposite side of the object from the black light. The reference stimulus was located roughly 25 mm to the left of the participant's cyclopean viewpoint (Figure 7A). While the two physical rear rods always appeared 350 mm away from the observer,
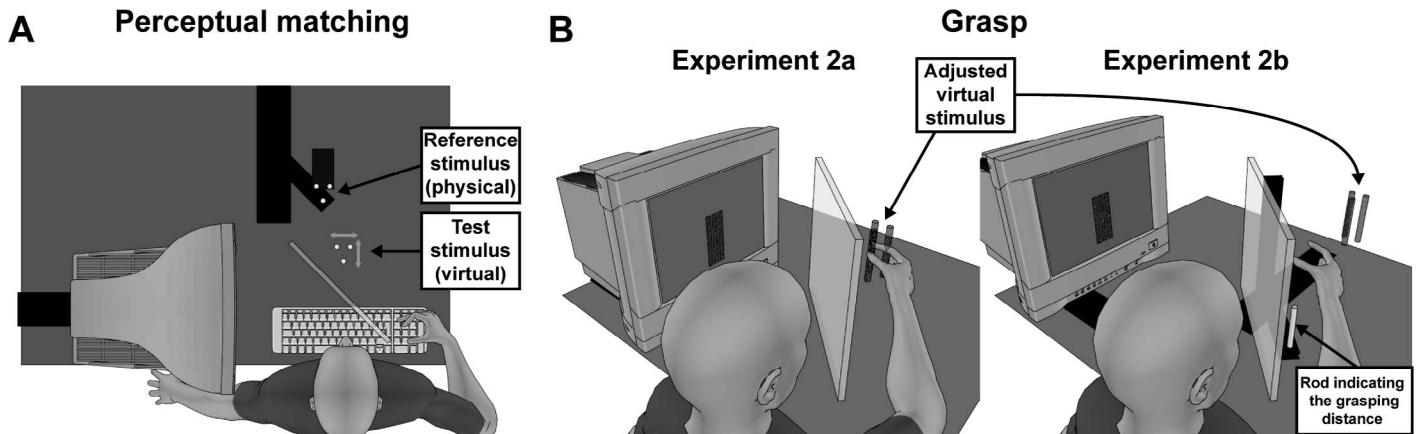
Figure 7. Schematics of the tasks of Experiments 2a and 2b. (A) In the perceptual task subjects viewed a physical (reference) stimulus always at the same distance, and adjusted the width and the depth of a virtual (test) stimulus with the keyboard until the test looked identical to the reference. The test stimulus was viewed either at 270 mm or at 450 mm from the observer. (B) The grasp task of Experiment 2a (left) was identical to that of Experiment 1 (Figure 4C). In the grasp task of Experiment 2b (right), a small rod positioned in the lower visual field indicated the distance, different from that of the virtual stimulus, where the subject had to perform the grasp. When the stimulus was presented at 450 mm the small rod appeared at 270 mm (as in the example), and vice versa.

the front rod was positioned at three absolute distances (300, 310, and 320 mm). This way the depth of the reference stimulus could take on one of three values (30, 40, and 50 mm) randomly chosen at each trial. The width (the distance between the physical back rods) was 40 mm.

In each trial, observers varied the width and depth of the test stimulus via keyboard key press until it appeared to match the structure of the reference stimulus. As in the previous experiment, the test stimulus was presented at one of two distances (270 and 450 mm). The depth of the reference stimulus and the distance of the test stimulus were randomly chosen at each trial. Subjects performed a total of 36 perceptual trials (3 depths × 2 distances × 6 repetitions). The average estimated depths and widths were then calculated to obtain six virtual stimuli. These perceptually equalized stimuli appeared to have the same depth and width at both distances, and were unique for each participant.

In the second part of the experiment, each participant was presented with their own set of perceptually equalized stimuli, one object per trial selected in random order. The task was to reach-to-grasp the stimulus, under the same conditions and procedure used in Experiment 1 (Figure 7B). The grasping task comprised a total of 30 trials: 6 stimuli (3 per distance) × 5 repetitions.

Results and discussion

Figure 8 shows the average adjusted depth (left) and width (right) of the test stimulus as function of the depth of the reference stimulus, for each simulated distance. The results are compatible with the phenomenon of depth underconstancy found in Experiment 1: to achieve a perceptual match, observers adjusted the depth of the far object to be greater than that of the near object, proving that distant objects are perceived shallower than close objects. A repeated-measures ANOVA on the adjusted depth with test stimulus's distance (270 and 450 mm) and reference object's depth (30, 40, 50 mm) as within-participant factors found two significant main effects, test's distance: $F(1, 9) = 15.82$, $p < 0.01$; reference's depth: $F(2, 18) = 43.58$, $p < 0.001$, and a significant interaction, $F(2, 18) = 3.89$, $p < 0.05$. On the contrary, the same analysis on the adjusted width revealed no significant effects: distance: $F(1, 9) = 0.73$, $p = 0.41$; object's depth: $F(2, 18) = 1.98$, $p = 0.17$; interaction: $F(2, 18) = 0.08$, $p = 0.92$.

The compression of visual space did not seem to impact the adjustment of the width. This behavior likely resulted from a combination of two factors. First, the range of distances that we tested was perhaps too small: since retinal size only varies linearly with distance, any bias on the scaling of width due to visuospatial compression may have not been statistically detectable in our design (subjects did show a tendency towards setting a greater width at the far distance but it was negligible; but see also van Damme & Brenner, 1997). Second, since the viewing distance (270 or 450 mm) corresponded to the center of the stimulus (i.e., see Figure 4B), subjects fixated at increasingly farther locations when adjusting the back surface of the 30 mm, the 40 mm, and the 50 mm deep object, respectively. Since changes in the fixation angle are more apparent between locations that are close to
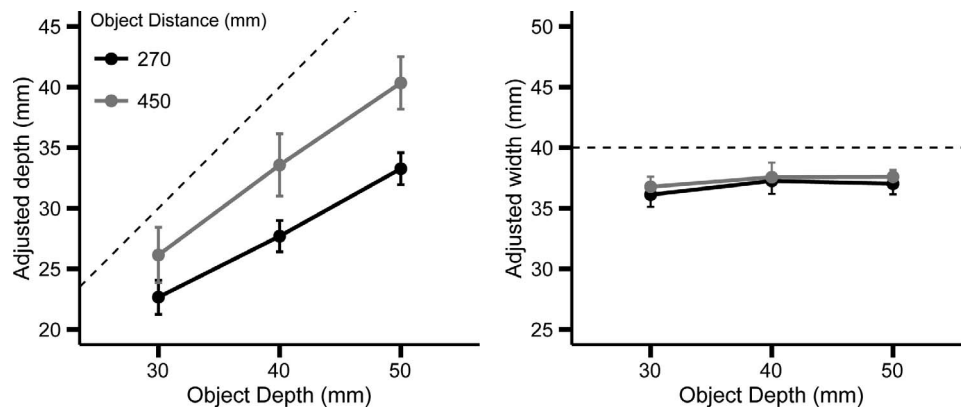
Figure 8. Mean adjusted depth (left) and width (right) of the test stimulus as function of object depth for each distance. Note that these results, like those of Figure 5B, are averaged across participants to show the general trend in the data, even though each individual likely scaled depth information in a different way.

the body, it may be that at 270 mm the subjects did not apply the same scaling distance for all width adjustments but instead greater scaling distances with increasing object's depth. This likely pushed the responses further, closer to those observed at 450 mm.

As a result of the matching task, in the second part of the experiment participants grasped objects that were on average 5.48 mm deeper at the far distance than at the close distance. Figure 9 (left panel) shows, for each distance, the average *FGA* as function of the reference stimulus depth (to which the test stimuli were perceptually matched). Despite the difference in simulated depths, grasps yielded almost the same *FGA* at the two distances. A repeated-measures ANOVA on the *FGA* revealed a significant effect of the object's depth, $F(2, 18) = 9.64$, $p < 0.01$, and a near-to-significant effect of distance, $F(1, 9) = 2.9$, $p = 0.12$: the *FGA* tended to be larger at the far distance. Even though this difference was only about 2.5 mm, which was therefore about half

the difference between the corresponding adjusted depths, we tested whether subjects actually responded to the physical depth in a control experiment.

Experiment 2b

It has been long shown that grip apertures are also planned based on the expected accuracy of the transport component (Bootsma, Marteniuk, MacKenzie, & Zaal, 1994; Jakobson & Goodale, 1991; Wing, Turton, & Fraser, 1986): independent of object size, reaching movements towards far distances yield a larger grasp because they are typically faster and less spatially accurate. This would explain the nearly significant effect of reaching distance on the *FGA* found in Experiment 2a (Figure 9, left): the position where one plans to do a grasp should affect the grip aperture
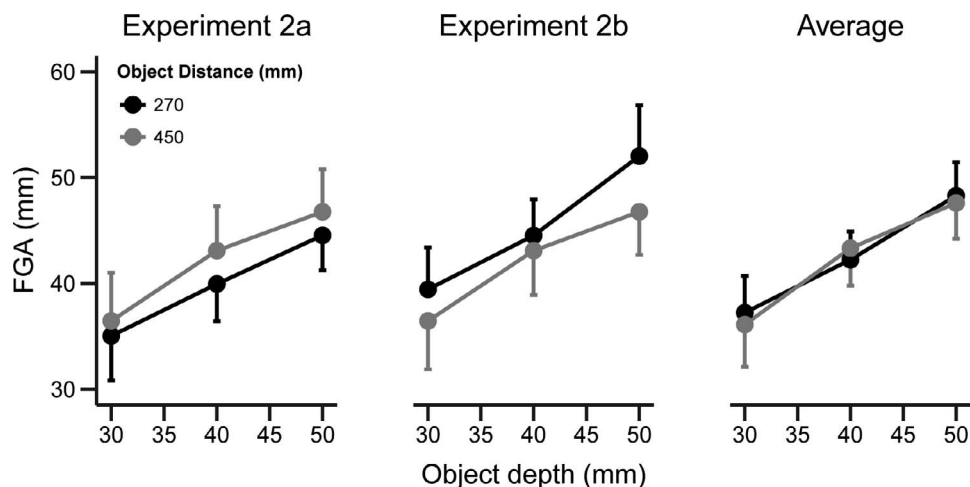

Figure 9. *FGA* as function of the object depth for each viewing distance (270 mm in black; 450 mm in gray). Left, results from Experiment 2a (reaching distance identical to viewing distance). Center, results from Experiment 2b (reaching distance opposite to viewing distance). Right, average *FGA* across Experiments 2a and 2b.

irrespective of the estimated depth of the grasped object. We therefore reasoned that the opposite effect ought to be expected if the subjects were induced to grasp the closer object by reaching at the far distance and the far object by reaching at the close distance.

Methods

Procedure

Apparatus, stimuli, and procedure were the same used for the grasp task of Experiment 2a with one exception. In this case, the same 10 participants grasped at the distance opposite to where the object appeared. When a stimulus appeared at 270 mm, a small rod was positioned at 450 mm on the tabletop and illuminated by a small black light, so that it would be visible through the half-silvered mirror in the lower visual field. Without seeing their hand, subjects reached at the perceived distance of the rod and mimicked a grasp at that position. The opposite happened when the stimulus appeared at 450 mm (Figure 7B, right panel).

Results and discussion

In agreement with the predictions, the results showed the opposite effect of grasping distance found in the previous experiment (Figure 9, center): the *FGA* was now larger for the near object than for the far object, indicating that the hand position was the only factor modulating the scaling of the grip aperture. Indeed, a repeated-measures ANOVA showed a significant effect of object's depth, $F(2, 18) = 7.07$, $p < 0.01$, and object's distance, $F(1, 9) = 8.71$, $p = 0.02$, on the *FGA*. Notably, the effect of distance we found was opposite to that of the previous experiment.

The grasp tasks of Experiments 2a and 2b were identical except for where the hand was brought to grasp the object. Recall that the near and far virtual stimuli appeared identical because of the previous adjustment task. Hence, the comparison between experiments allowed us to test the separate contributions of hand and object's distance on the estimation of depth. We ran a repeated-measures ANOVA on the *FGA* of both experiments using object depth, $F(2, 18) = 10.24$, $p = 0.001$, object distance and hand distance as within-subjects factors: as expected, the hand distance affected the *FGA*, $F(1, 9) = 9.30$, $p = 0.01$, while the object distance did not, ($F(1, 9) = 0.09$, $p = 0.77$). In conclusion, the near and the far objects, which appeared identical in the perceptual task, were grasped with identical grip apertures, after controlling for where the grasp was executed (Figure 9, right panel). This result represents converging evidence that perception

and action share the same representation of the visual space.

Experiment 3

The grasping tasks described previously, executed in the dark without vision of the hand and the physical presence of the object, were designed with the specific purpose of assessing the participants' depth estimates at planning. A potential disadvantage of this experimental methodology is that it may test mechanisms deviating from those that regularly guide action. It may be speculated that after a few grasps, during which the participant does not feel the physical presence of the object, the task diverts from being purely action-based to involving a perceptual strategy (Goodale, Jakobson, & Keillor, 1994; Westwood, Chapman, & Roy, 2000). In other words, a grasp gradually becomes a manual size estimation task.

The flip side of this line of reasoning, which deems this procedure an invalid test of vision for action, is that repeated grasps of a limited set of physical objects could potentially modify the sensorimotor mapping. That is, through an error correction mechanism akin to those observed in sensorimotor adaptation experiments, the recorded kinematics of the grasp do not reflect the visual representation guiding the action. Instead, they identify a modified mapping between a biased 3D representation and the motor program, correcting for inaccuracies in visual information processing.

The main goal of this experiment was to test whether the previous findings can be attributed to the persistent lack of haptic feedback, an interpretation that we will define as the *pantomime grasp* hypothesis. The term pantomime grasp traditionally refers to a grasp toward a remembered object; that is, one that disappears at movement onset. Under such conditions, subjects can see their hand throughout the action but cannot see the stimulus, so they have to mimic the grip aperture corresponding to their best estimate of the object's size. According to this hypothesis, vision-for-action yields veridical estimates of size only if convergent visual and tactile feedback are available. As a consequence, if neither is present, vision-for-action switches to vision-for-perception (Holmes, Lohmus, McKinnon, Mulla, & Heath, 2013; Hosang, Chan, Jazi, & Heath, 2016; Jazi, Yau, Westwood, & Heath, 2015).

Participants executed a series of grasps towards objects identical to those of Experiment 1, with the exception that a physical object was felt at the end of the movement and that the visual feedback of the grasping digits was seen throughout the action. If the systematic biases observed in Experiments 1 and 2 are
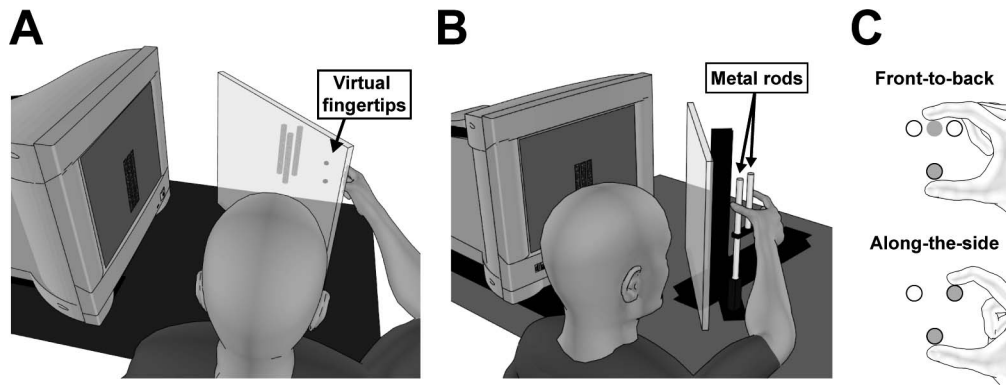
Figure 10. (A) Subjects reached-to-grasp the virtual stimulus while seeing two additional small 3D discs aligned with their index and thumb fingertips. (B) Behind the mirror, not seen, were two metal rods aligned with the virtual stimulus in two different ways, to allow for two different grasps. (C) In the front-to-back grasp, the metal rods (gray filled circles) were positioned such that they matched the front virtual rod and the middle between the two virtual back rods (virtual rods are represented by black open circles). In the along-the-side grasp, the metal rods were aligned with the front and the right back virtual rods. The simulated index fingertip disappeared as soon as the index went behind the right virtual rod. Therefore, the index disappeared in the final moments of the front-to-back grasp, whereas it remained visible until object contact in the along-the-side grasp.

not due to the absence of visuo-haptic feedback, then they should be observed under these conditions too. Moreover, the discrepancy between the estimated depth of the grasped object, which is biased at planning, and the felt size of the object, should produce systematic error corrections, as predicted by what we define as the *sensorimotor correction* hypothesis (Kopiske, Cesanek, Campagnoli, & Domini, 2017).

Methods

Procedure

The procedure was identical to the grasp task of Experiment 1, with the following exceptions. First, while reaching, participants could see two luminous dots that coincided with the centers of their index and thumb fingertips (Figure 10A). Second, we provided haptic feedback by aligning two metal rods with the virtual image (Figure 10B). The metal rods were mounted on a moving platform and brought to position by linear actuators. Subjects contacted the stimulus along the depth dimension, as in the previous experiments (*front-to-back* grasp, Figure 10C). To provide accurate tactile feedback, one metal rod matched the position of the front virtual cylinder, while the other rested at the same depth of the two rear virtual cylinders, positioned midway between them. Since the index finger is normally occluded by an object at the end of a front-to-back grasp, the index virtual fingertip also disappeared as soon as it passed behind the right back rod (i.e., as soon as it was behind the stimulus's back surface). However, it is possible that the sudden lack of visibility of the forefinger could induce participants to complete the action by guiding the finger on the estimated position of the back rod,

effectively allowing depth underconstancy to affect action again. To control for this confound, another group of subjects grasped the same stimuli diagonally, by touching the front and right-side rods (*along-the-side* grasp, Figure 10C). In this condition, one metal rod was again aligned with the front virtual rod, while the other metal rod matched the position of the right rear virtual rod. This ensured that both fingers remained visible throughout the whole movement until object contact. A cover was mounted behind the mirror, so that subjects could not see through it. Therefore, the metal rods were only felt by the subjects but never seen. While grasping, participants only saw the virtual stimulus and the virtual rendering of their fingertips.

A 2 (depth: 30 and 50 mm) $\times$ 2 (distance: 270 and 450 mm) $\times$ 2 (grasp type: front-to-back, along-the–side) design with five repetitions for each combination yielded a total of 40 trials. 18 subjects executed the front-to-back grasps, and 17 subjects grasped the stimulus along the side.

Results and discussion

Since the fingers contacted a physical object at the end of each grasping movement, we could not use the *FGA* as dependent variable. Instead, we looked at how the grip aperture evolved over the course of the reaching trajectory. Contrary to the pantomime grasp prediction, we hypothesized that the grasp is planned based on distorted estimates of the object's depth, even when visual and tactile feedback are available. Because of depth underconstancy, we expected the grip aperture to be on average smaller when reaching for the far object than when reaching for the near object.
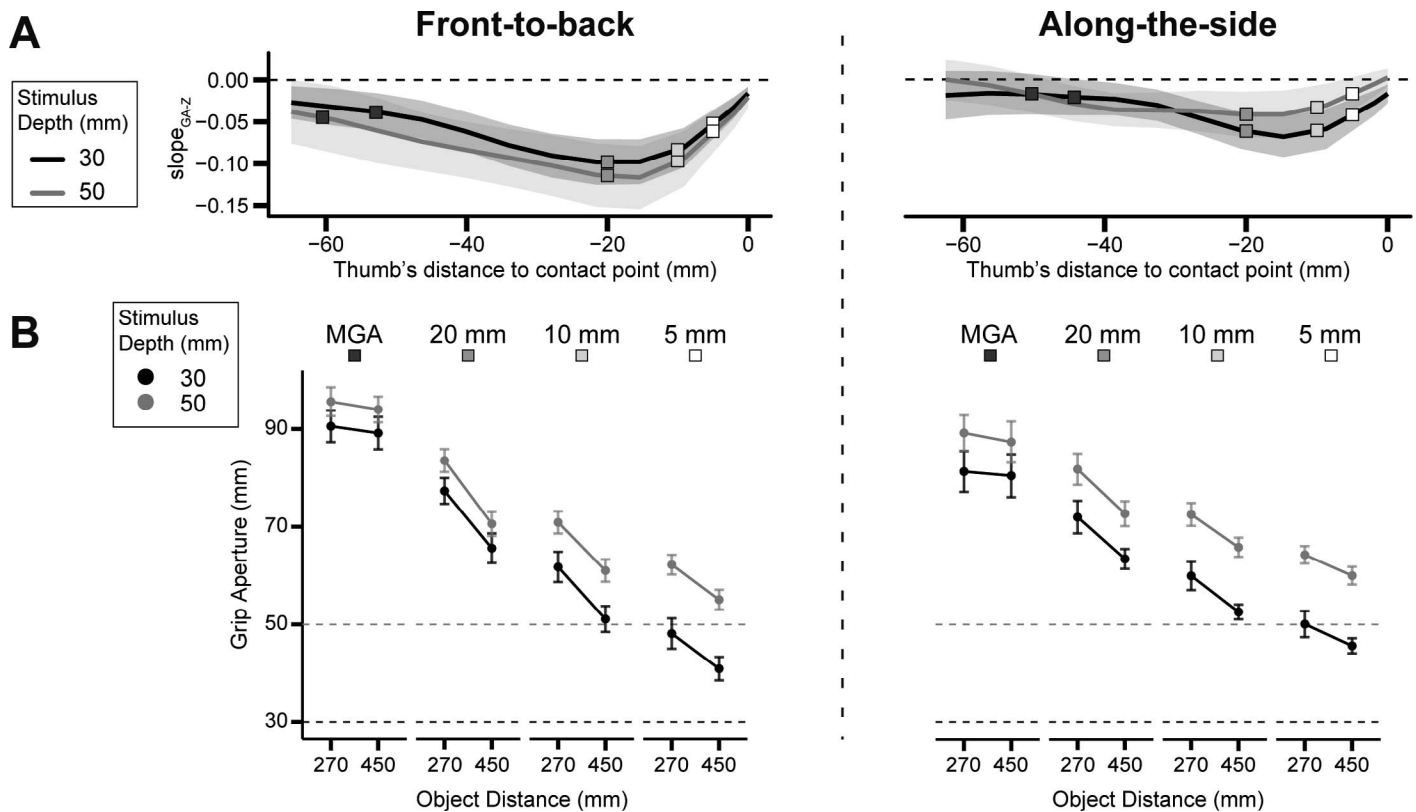
Figure 11. Trajectory analysis from the data of Experiment 3. Left, front-to-back grasp; right, along-the-side grasp. (A) Slope of the function relating the grip aperture to the object distance (slope$_{GA-Z}$) as function of the thumb's distance to contact location, for each object's depth (black, 30 mm; gray, 50 mm). Negative values mean depth underconstancy. In each panel, little squares mark four instants along the trajectory, illustrated in the panels below. (B) Grip aperture plotted as function of the object distance, for each object's depth (black, 30 mm; gray, 50 mm), at four instants along the trajectory (*MGA*, and thumb 20, 10, and 5 mm away from stimulus).

To test this hypothesis, we computed the frame-by-frame Euclidean distance between the thumb (the finger that was always visible in both movements) and its final position. This variable was then divided into 100 equally spaced bins, and in each of them the grip aperture was averaged and analyzed as a linear function of the object's distance and depth. Of particular interest was the slope of the function relating the grip aperture to the object distance (slope$_{GA-Z}$) (Figure 11A). We found that it was significantly negative during the entire final phase of the movement, that is, when the grip was finely adjusted in preparation to grasp. Figure 11B shows four snapshots of the grip aperture—distance function: at *MGA* time, and when the thumb was 20, 10, and 5 mm away from its final position. In the front-to-back grasp, we found a negative slope$_{GA-Z}$, meaning depth underconstancy, already at *MGA* time, although it was not yet significant, $F(1, 52) = 2.08$, $p = 0.15$. The lack of significant effect can be attributed to the fact that the *MGA* of a grasp in depth does not scale sufficiently well with object size, as can be observed on Figure 11B, and consistent with the results of previous studies (Boz-

zacchi et al., 2016; Hibbard & Bradshaw, 2003). This is due to the fact that in order to avoid collision with the object, the subject has to produce an unusually large maximum grip aperture (about 90 mm), which approaches a mechanical upper bound. The effect then progressively increased in magnitude and remained significant until 5 mm before the thumb contacted the object: [20 mm: $F(1, 52) = 35.74$, $p < 0.001$; 10 mm: $F(1, 52) = 21.28$, $p < 0.001$: 5 mm, $F(1, 52) = 14.98$, $p < 0.001$]. A similar pattern of results was found for the along-the-side grasp: [*MGA*, $F(1, 49) = 1.44$, $p = 0.24$; 20 mm: $F(1, 49) = 38.56$, $p < 0.001$; 10 mm: $F(1, 49) = 26.35$, $p < 0.001$; 5 mm: $F(1, 49) = 11.55$, $p = 0.001$].

The presence of depth underconstancy in both grasps suggests that the grip aperture was always planned based on a wrong estimate of the object's 3D structure. The case of the along-the-side grasp is especially remarkable, since the bias emerged even when subjects did not need an accurate estimate of the object depth. In principle, their grasp could have been guided by nulling the relative position of the online visual feedback of the fingertips and corresponding contact locations.
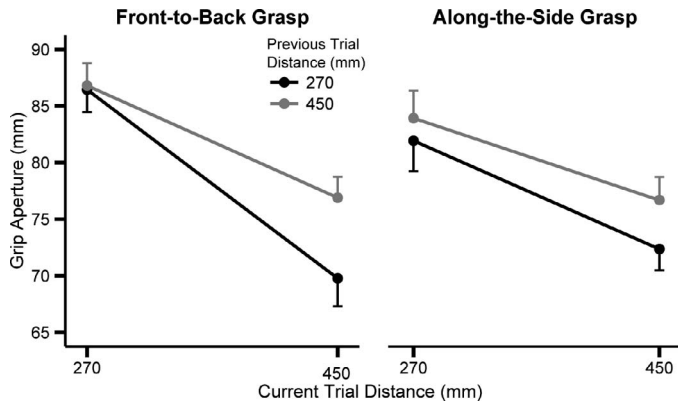
Figure 12. Grip aperture 20 mm prior to the thumb's contact with the object's surface as function of the current trial distance, for each previous trial distance (270 mm in black and 450 mm in gray).

We found converging evidence that these systematic errors affected grasping actions by analyzing the effect of previous trial on the grip aperture. Here we show the analysis relative to the moment when the thumb was 20 mm away from object contact, since at that point of the movement the effect of distance on the grip aperture was the largest. However, results relative to other instants during the final phase of the movement are comparable. We ran a mixed ANOVA using the type of grasp as a between-subjects factor and three within-subject factors (object's depth, current trial's distance, and previous trial's distance) and we found a significant main effect of the previous trial's distance, $F(1, 33) = 13.5$, $p < 0.001$, as the grip aperture was larger immediately after grasping a far object than a near object (Figure 12). This is consistent with the sensorimotor correction hypothesis. Suppose, for example, that the depth of the object at the near distance is overestimated. In this case, the physical contact of the grasping fingers with the object detects a discrepancy between the expected depth, which is biased, and the felt depth. This causes a correction in the sensorimotor mapping leading to a smaller grip aperture in the subsequent grasp. The opposite happens when grasping the far object: after detecting that the stimulus's depth is underestimated at the end of a given grasp, the visuomotor system responds to the error by generating a greater grip aperture in the subsequent grasp.

Experiment 4

Although visuospatial compression and depth underconstancy stem from the same biased estimate of fixation distance, they generate two patterns of visual space distortions that are not geometrically compatible. This is because visuospatial compression is the outcome of absolute distance judgments, whereas depth under-constancy emerges from judgments of depth intervals. In order to clarify this point, consider an observer who is assessing the location of two target objects lying in near space and separated in depth by a distance Δz, such that one is in front of the other (Figure 13, upper left panel, z_{front} and z_{back}). In this task, she estimates the absolute distance of either the front or the back target, in separate trials. Since both objects are in near space, we assume that they will appear as farther than they are. More importantly, the egocentric function (see Figure 1) predicts that the front object's distance will be overestimated to a larger degree than the back object's distance. Thus, the perceived depth interval calculated by subtracting the two judgments ($z'_{front} - z'_{back}$) will be smaller than the actual physical interval (Figure 13, black circles on upper left panel). In a second task, she estimates the depth of an object having the front and back surfaces at the same exact locations as the targets of the previous task (Figure 13, lower left panel, z_{front} and z_{back}). Presumably, this involves scaling the relative disparity between the front and back surfaces by the object fixation distance (in this example, the front surface at z_{front}). Since we assume that z_{front} is overestimated, the perceived depth will be larger than the physical depth (Figure 13, black triangle on lower left panel). Hence, we stumble upon a paradox: the same depth interval Δz is either underestimated or overestimated, depending on whether it results from egocentric or allocentric tasks.

The goal of this experiment is to investigate whether these spatial inconsistencies can also be found with reaching and grasping movements. We hypothesize that (a) a reaching task is based on an absolute distance estimate from ocular vergence and (b) the grip aperture of a grasping task depends on the relative depth estimate resulting from scaled binocular disparities. In order to test this hypothesis, in separate trials participants either pinched a rod positioned in two different spatial locations along the line of sight (Figure 13, upper left), or grasped an object spanning the same depth interval (Figure 13, lower left). If the hypothesis is correct then the depth separation of the reached locations during the pinch is predictably smaller than the final grip aperture of the grasp.

Furthermore, in order to make specific predictions about the landing locations of each digit in the grasp relative to the pinch, we assume that the thumb determines the transport component, as we found in recent work (Volcic & Domini, 2014). Therefore, we expect that the thumb will reach the perceived egocentric location of the closest vertex of the triangle, which is also the location of the front pinch (z'_{front}). The critical prediction is about the landing location of the index. Here, we assume that in order to recover the depth of the object Δz, retinal disparities are scaled by
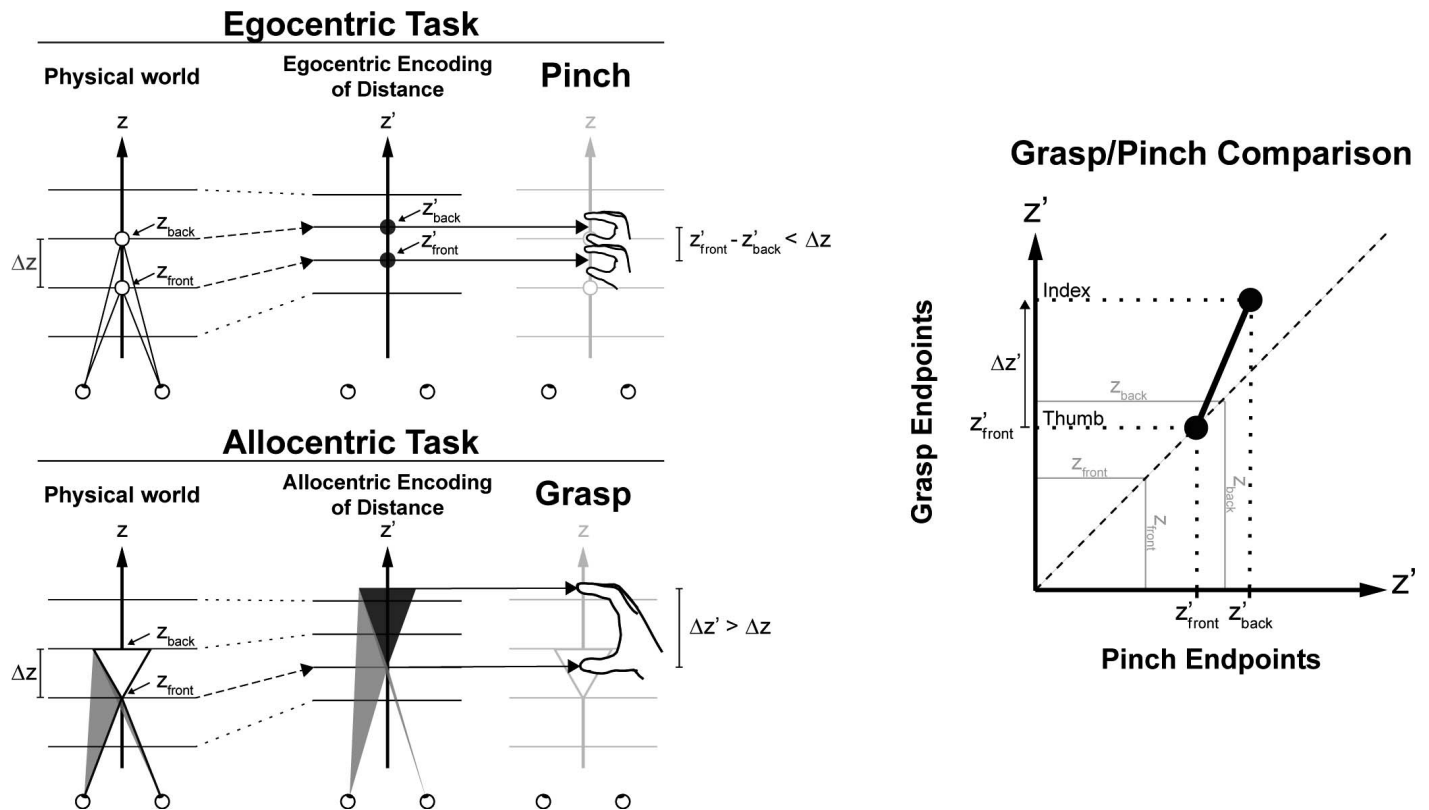
Figure 13. Predictions of the final contact location of thumb and index finger in an egocentric (pinch) and allocentric (grasp) actions directed at the same spatial locations (z_{front} and z_{back}): (1) Upper left, when pinching a target present at one of the two locations, the endpoints can be predicted by visuospatial compression of egocentric distances; (2) Lower left, when grasping front-to-back an object spanning the same spatial extent, the thumb will land at the same location as the front pinch (z'_{front}), but the index location will be determined by the estimated depth of the object ($\Delta z'$). Right, since we assume that the depth of the object is overestimated ($\Delta z' > \Delta z$), we can predict that the final location of the index finger after a grasp ($z'_{front} + \Delta z'$) will be further in depth than the final position of the pinch directed at the same spatial location (z'_{back}).

the estimated fixation distance z'_{front}. The perceived depth $\Delta z'$, added to the predicted location of the thumb at movement completion (z'_{front}), determines the predicted final location of the index finger (Figure 13, right panel).

Methods

Procedure

Participants aimed grasping and reaching actions at a virtual stimulus consisting of a rod in front of a rectangular plane (Figure 14A). The right edge of the plane was positioned such that the index contact point in the grasp was always visible.

During grasping, participants picked up the object whose front rod was located at one of three distances separated by 40 mm intervals (Figure 14B, grasp panel). The front-to-back extent of the stimulus corresponded to the same interval, such that two distances were the contact point of either the index or the thumb. To test reaching, participants pinched a rod

positioned at the locations of either the front or back surfaces of the grasped objects (Figure 14B, pinch panel). This resulted in three grasps and four distinct pinches. In order to avoid overlearned movements, subjects also grasped 50 mm objects located at the same distances as the 40 mm objects and they pinched their back locations as well. To render the visual stimuli in the grasp and pinch tasks as similar as possible to each other, subjects pinched a rod while always seeing the back surface. Note also that the pinch was adopted so to maximize the similarity between tasks (i.e., both were grasping movements). On the workspace behind the mirror, a vertical stand was mounted on one arm of the Velmex system and supported two physical reproductions of the virtual stimulus (a thin metal cylinder connected to a flat metal plate behind it, one pair for each depth) at different heights. Participants began each trial by resting with their fingers pinched together on a support near their body, while the Velmex motors positioned the physical stimulus at a given distance according to the trial. Once the Velmex system stopped moving, the virtual stimulus appeared
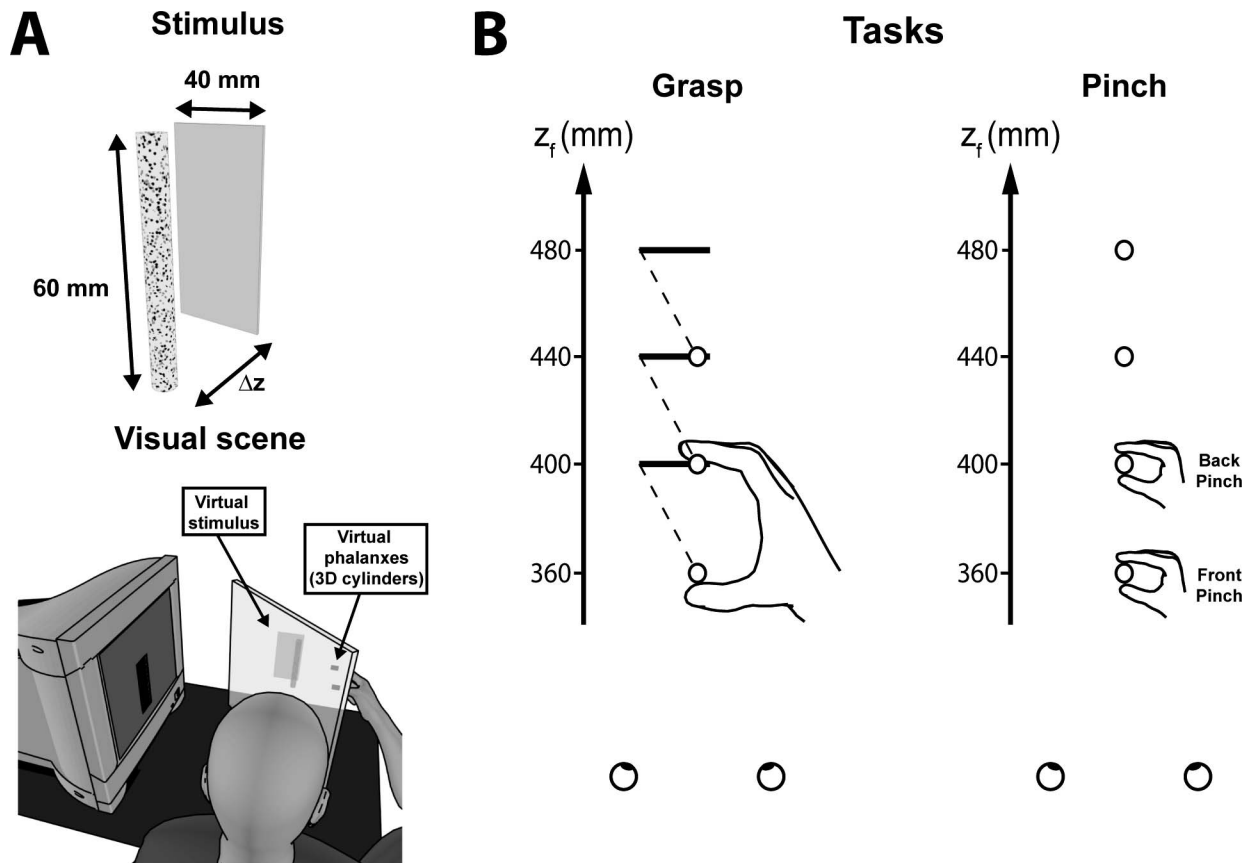
Figure 14. Methods used in Experiment 4. (A) The stimulus comprised of two 3D virtual elements: a rod (front) and a plane (back). In half of the trials, during both grasp and pinch visual and haptic feedback of the fingers' phalanxes were available. (B) In pinch trials, subjects reached for four locations that also defined the front-to-back extent of the grasped objects.

automatically, and participants had 2 s to reach it from the stimulus onset. They were instructed to perform a front-to-back grasp on the stimulus unless a small sphere appeared on top of the front rod, indicating that they were supposed to do a pinch. Participants never saw the physical objects through the mirror, they only saw the virtual stimuli.

Each trial was performed in either a full-feedback condition, or in a no-feedback condition, and at any given trial participants did not know in advance whether they would receive feedback or not. During full-feedback trials, the Velmex system first aligned the physical stimulus with the virtual image, and the participant then started the task. The final phalanxes of their index finger and thumb were represented by two additional 3D cylinders, updated in real time using the coordinates of the Optotrak markers attached to the fingernails. During no-feedback trials, the motors first pushed the stand out of reach, and participants then performed the task without seeing the rendering of their fingers nor touching any physical object. Full- and no-feedback trials were intermixed within the same session, resulting in four equally possible actions: a pinch or a grasp, with or without feedback. This design

was intended to provide ground truth about the size and the position of the stimuli on 50% of the trials, to prevent participants from developing strategies involving pantomime movements.

To summarize, over the course of four blocks, participants performed 13 different actions (6 grasps + 7 pinches) × 2 feedback conditions × 20 repetitions (five repetitions per block), for an overall total of 520 trials. The high number of trials was specifically intended in order to maximize the statistical power for a small sample size ($N = 6$).

Results and discussion

Figure 15A shows the thumb and index endpoints versus the actual physical locations, during no-feedback trials of the pinch and grasp tasks. Consider first the results from the pinch task (Figure 15A, left). The end positions of the thumb and index, which basically coincided, were related to the stimulus distance by a slope smaller than 1, $slope_{Pinch} = 0.77$, 95% CI = [0.7, 0.84], as predicted by visuospatial compression. Turning now at the grasp task (Figure 15A right), it should
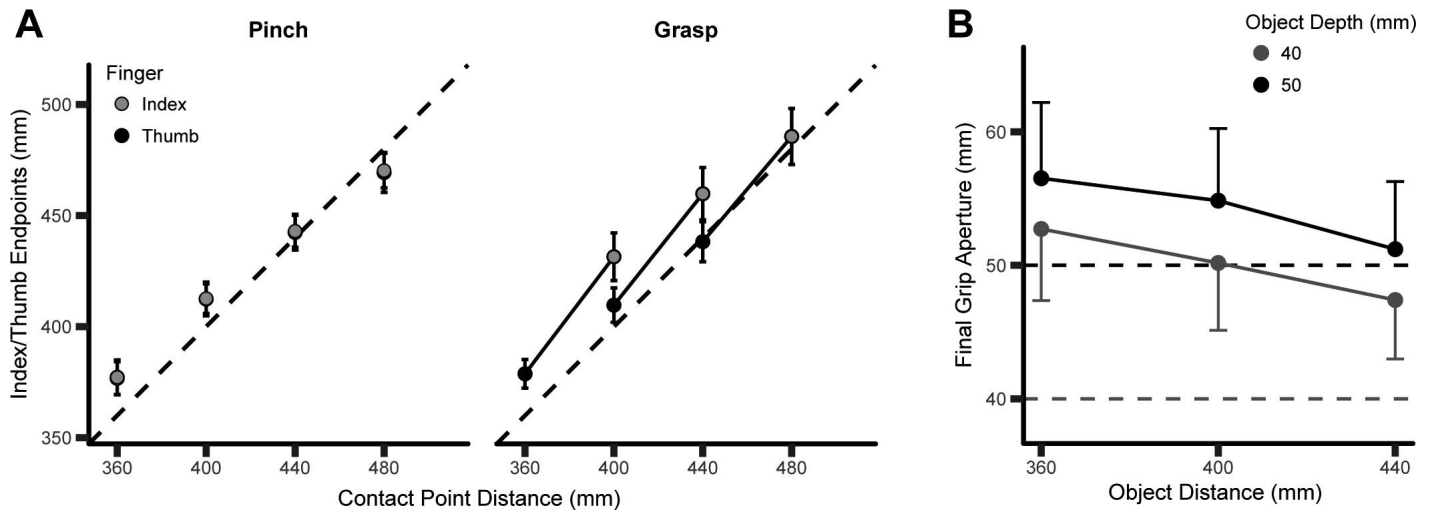
Figure 15. (A) Final positions of index finger and thumb during pinch and grasp tasks (trials without feedback), as function of the actual physical locations of the contact points. (B) *FGA* in the grasp task (trials without feedback) as function of object distance for each simulated object depth.

be noted how remarkably similar the endpoints of the thumb were to those observed in the pinch task aimed at the same contact locations (on Figure 15A, compare the pinch locations at 360, 400, and 440 mm with the corresponding thumb end positions during grasp). This is in agreement with the hypothesis that the thumb guides the transport of the hand during grasp, almost exclusively on the basis of egocentric information. However, the final thumb position was also slightly influenced by allocentric information: the flatter object was grasped as slightly farther than the deeper object, in agreement with the similar result found in Experiment 1. A repeated-measures ANOVA on the final thumb position, using distance and object depth as independent variables, indeed revealed a significant effect of object depth, $F(1, 5) = 50.98$, $p < 0.001$. Whereas the thumb landed at the perceived egocentric distance of the front rod, the index position also depended on the perceived depth of the object. This quantity determined the *FGA*, which in agreement with the previous results, revealed depth underconstancy (Figure 15B). A repeated-measures ANOVA showed in fact that the *FGA* significantly decreased with distance: [object's depth: $F(1, 5) = 33.18$, $p = 0.002$; object's distance: $F(2, 10) = 12.43$, $p = 0.002$].

Consider now the direct test of our predictions, which can be easily visualized by directly comparing the locations reached by index and thumb in the two tasks (Figure 16). Whereas the thumb positions were basically identical (black circles on the pinch and grasp panels), the index positions were systematically farther in depth than the corresponding back pinches (gray circles). Most importantly, the index position can be predicted in two steps by (a) scaling the stimulus disparities with the distance corresponding to the end

location of the front pinch (see Appendix) and (b) adding this depth estimate to the grasping thumb position along the *z* axis (gray areas depict the 95% CI of the prediction). The observed index positions were in good agreement with this prediction, while at the same time they systematically overshot the bisector (Figure 16, grasp/pinch comparison panel). Two linear models were fitted, where the final index position was modeled as a linear function of either the predicted position or the endpoint position of the back pinch (null hypothesis). A goodness-of-fit test confirmed that our predictions fitted the data significantly better, $\chi^2(0) = 14.11$, $p < 0.001$.

These results confirm the distinct role played by egocentric and allocentric information for spatial encoding: absolute distance is represented in egocentric coordinates (i.e., with the observer as the origin), whereas relative depth is defined as an allocentric interval (Foley, 1980; Gogel, 1977, 1990). Previous psychophysical studies have reported cases in which the perceived depth between two points does not match the interval defined by their individual apparent locations (Loomis & Knapp, 2003), suggesting at least a partial dissociation between egocentric and allocentric maps of visual space in both perception and action (Bingham, Crowell, & Todd, 2004; Bingham, Zaal, Robin, & Shull, 2000; Binsted & Heath, 2004; Chen, Byrne, & Crawford, 2011; Eloka & Franz, 2011; Gentilucci, Daprati, Gangitano, & Toni, 1997; Loomis, Philbeck, & Zahorik, 2002; Loomis, Silva, Philbeck, & Fukusima, 1996; Neely, Tessmer, Binsted, & Heath, 2008; Thaler & Goodale, 2010). These data reinforce the idea that different aspects of grasping imply specific 3D information processing (Bingham, 2005; Bingham et al., 2000; Bingham et al., 2004; Brenner & Cornelissen,
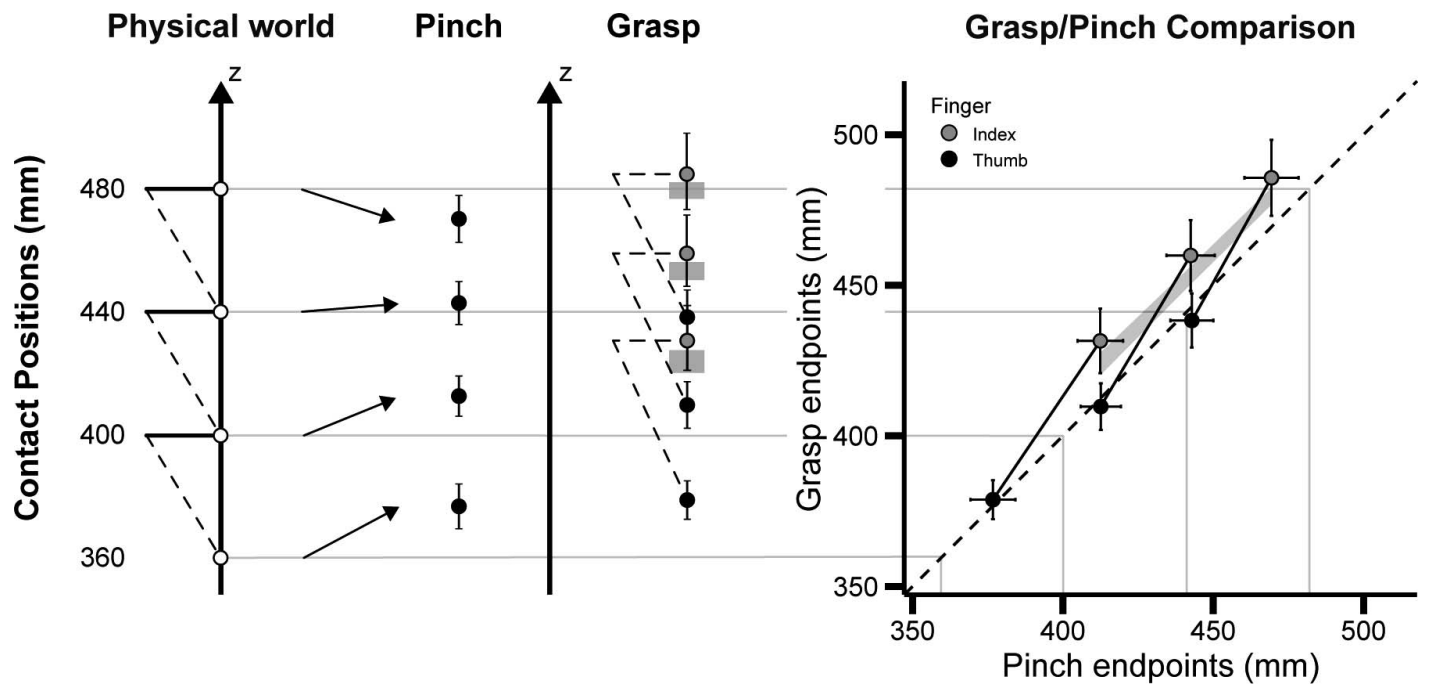
Figure 16. Comparison between grasp and pinch tasks to test the predictions of Figure 13. The filled circles show the average locations of the thumb in the pinch and grasp tasks for movements directed at the physical locations represented by the open circles. A direct comparison of the pinch enpoints with the grasp endpoints indicates that the thumb reached in both tasks the same estimated egocentric location. Errorbars represent 1 *SEM*. The gray bands represents the 95% CI of the predicted final positions of the index finger. The bisector (dashed line) illustrates the null hypothesis (distance and depth estimation share the same distorted space).

2000; Coats, Bingham, & Mon-Williams, 2008; Jeannerod, 1986; Lee, Crabtree, Norman, & Bingham, 2008; Mon-Williams & Bingham, 2007; Smeets, Brenner, de Grave, & Cuijpers, 2002; Thaler & Goodale, 2010) and that reaching distance and grip aperture can be calibrated independently (Bingham, Coats, & Mon-Williams, 2007; Coats et al., 2008; Foster, Januszewski, & Franz, 2015; but see also Smeets & Brenner, 1995, 1999, 2008a, 2008b; Smeets et al., 2002).

These findings are in agreement with those of the previous experiments showing that motor planning based exclusively on stereo information is biased. Therefore, online control and the final contact with a physical object are the only resources that allow the visuomotor system to redeem a wrongly planned movement. To confirm this explanation, it is particularly enlightening to closely inspect the trajectories of the grasps with and without feedback. Figure 17 shows the bird's eye view of these trajectories for the pinch (top) and grasp (bottom) tasks. In both cases, it can be seen how the trajectories of the trials with and without feedback overlapped up to the last moments of the movement; that is, when the visual feedback of the fingers was eventually exploited to correct for the errors at planning. In absence of feedback, close targets were systematically overshot, whereas far targets were systematically undershot, in agreement with visuospatial compression.
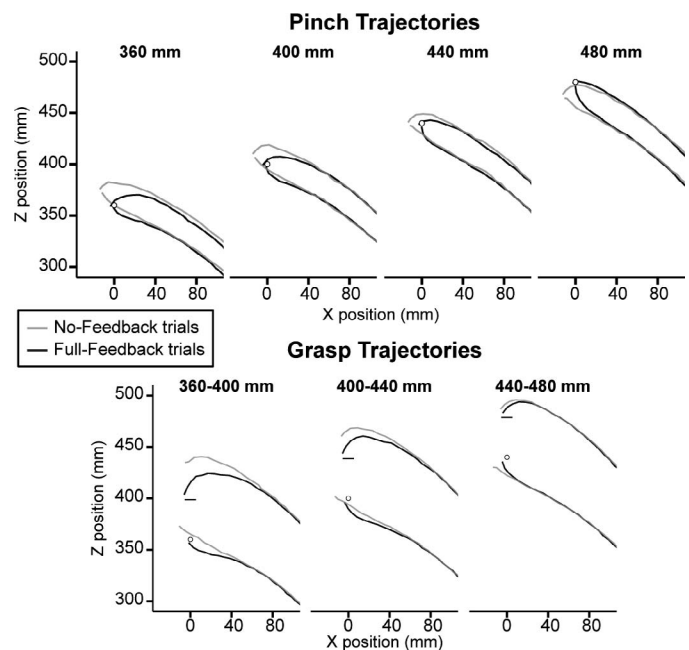

Figure 17. Bird's eye view of the trajectories of the pinching (upper row) and of the grasping (bottom row) in trials with (black) or without (gray) feedback. Each panel shows the data for each individual distance, labeled at the top. The x and z positions of index and thumb are in abscissa and ordinate, respectively.
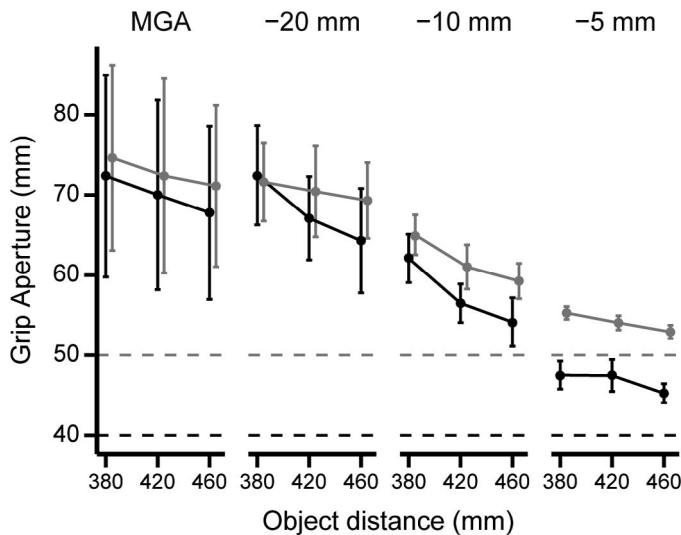
Figure 18. Grip aperture profile taken at four consecutive points along the trajectory of the full-feedback trials in Experiment 4 (at the time of the *MGA*, and when the thumb was 20, 10, and 5 mm away from its final position), as function of the object's distance for each depth (40 mm in black, 50 mm in gray).

Note that the intermixed presence of trials with visual and haptic feedback did not yield any significant improvement on the performance of the trials without feedback. This was revealed by two ANOVAs, with block as independent variable, on both the *FGA*, $F(3, 15) = 1.85$, $p = 0.18$, and the pinch endpoints, $F(3, 15) = 0.41$, $p = 0.74$. Moreover, as can be seen in Figure 18, in the trials with feedback, the analysis of the evolution of the grip aperture over the course of the trajectory (as in Experiment 3) showed depth underconstancy in the final stages of the movement: [*MGA*: $F(1, 28) = 11.64$, $p = 0.002$; 20 mm: $F(1, 28) = 5.05$, $p = 0.03$; 10 mm: $F(1, 28) = 19.09$, $p < 0.001$; 5 mm: $F(1, 28) = 4.98$, $p = 0.03$].

In conclusion, the results of the Experiment 4 are in good agreement with those of the previous experiments, but complement those findings by highlighting the independent roles of egocentric and allocentric information in guiding reach-to-grasp actions. First, egocentric information is almost exclusively responsible for guiding the thumb to the front surface of the target object. This information is biased, as predicted by visuospatial compression. Second, the same biased estimate of egocentric distance scales binocular disparities for computing the object depth extent. This allows prediction of the planned final grip aperture and thus, the end position of the index finger with respect to the thumb. Third, the landing positions of the grasping digits differ systematically depending on whether they result from an egocentric or an allocentric task, even though the targeted spatial locations are identical.

General discussion

The fundamental role of binocular information for guiding action is largely undisputed. Several investigations have shown that stereovision is one of the most powerful cues mediating goal directed actions (Harris, 2004; Servos et al., 1992) and that removing it from the visual input has detrimental effects on both the programming and control of reaching and grasping in humans (Bradshaw & Elliott, 2003; Keefe & Watt, 2009; Marotta, Perrot, Nicolle, Servos, & Goodale, 1995). Vergence has a predominant role in estimating an object's egocentric location, therefore affecting the reaching component of a movement, whereas binocular disparities determine the shaping of the grasp, via computation of an object structure (Melmoth, Storoni, Todd, Finlay, & Grant, 2007; Mon-Williams & Dijkerman, 1999; Watt & Bradshaw, 2000). What remains unclear is whether stereovision provides the action system with accurate measurements of an object 3D shape and distance.

The results of the four experiments yield evidence that reach-to-grasp movements are guided by an inaccurate mapping between the physical space and visual estimates. This mapping is characterized by two interrelated distortions: (a) the range of reaching distances is smaller relative to the same range of physical distances, resulting in visuospatial compression, and (b) the grip aperture of two grasps directed at the same object is greater when reaching near the body than when reaching far from the body, indicating depth underconstancy. In the first two experiments we show that the compression of visual space measured from a reaching task can predict depth underconstancy in a perceptual task on an individual basis: subjects whose reaches span a smaller region of space also show a large depth underconstancy, and vice versa.

The absence of visual and haptic feedback of the grasping hand in these experiments was intentionally chosen in order to avoid online corrections. However, this lack of feedback persisting throughout the experiment may have induced subjects to adopt alternative strategies in action execution, more attuned to a perceptual task (Goodale et al., 1994; Schenk, 2012; Schenk, Franz, & Bruno, 2011; Westwood et al., 2000). To control for this alternative explanation of the findings, in two additional experiments we provided consistent visual and haptic feedback from the fingertips. Even so, the results clearly show that stereo information yields biased estimates of 3D properties at action planning, which persist until the very last instants of a grasp.

Together with recent investigations (Bozzacchi & Domini, 2015; Bozzacchi et al., 2014, 2016; Volcic & Domini, 2016; Volcic et al., 2013), the data from the present study start delineating a clearer picture of the

role of stereovision for the control of reaching and grasping. The phenomena of visuospatial compression and depth underconstancy have in the past been attributed to errors in the translation of the oculomotor vergence into angles, particularly at great distance (*specific distance tendency*; Gogel, 1969) and in absence of other relevant cues to distance (Baird & Biersdorf, 1967; Brenner & van Damme, 1998; Gilinsky, 1951). The present experiments confirm these accounts, since the results can be quantitatively predicted by assuming a common source of error for distance estimates and binocular disparity scaling.

It may be argued, however, that the biases we found in perception and action were caused by the use of virtual instead of real objects. There are at least three main reasons for which we think this is not a likely explanation of our results. First, in our experiments, we prevented the vergence-accommodation conflict, a well-known property of virtual displays that induces flatness. Second, it has been recently shown that visual distortions experienced with our apparatus can be successfully replicated in natural settings during both perceptual and grasping tasks (Bozzacchi & Domini, 2015; Domini, Shah, & Caudek, 2011). Third, several studies demonstrate that the movement kinematics recorded in virtual and physical environments are qualitatively equivalent (Levin, Magdalon, Michaelsen, & Quevedo, 2008, 2015; Magdalon, Michaelsen, Quevedo, & Levin, 2011; but see also Cuijpers, Brenner, & Smeets, 2004, 2008).

However, it is clear that our stimuli, which are specifically designed to isolate the role of vergence and relative disparities, only represent a rarified subset of 3D cues present in the natural environment. Vertical disparities, perspective and ground plane cues are normally available, in addition to vergence and accommodation signals, for distance information. Therefore, it is possible that a better estimate of the object egocentric location not only could reduce or eliminate visuospatial compression, but also fine-tune the scaling of binocular disparities for a more accurate recovery of relative depth. This, in turn, could also be improved by the presence of monocular allocentric information as shading, texture patterns, motion parallax, and object contour. Even though multiple-cue stimuli are also subject to residual, but systematic, perceptual distortions (Norman, Todd, Perotti, & Tittle, 1996), it remains a matter of further investigation whether these biases affect in a similar fashion goal directed actions.

The successful completion of a reach-of-grasp movement may not critically depend on a perfectly accurate metric reconstruction of the scene. The results of Experiments 3 and 4 clearly indicate that wrong estimates at planning are easily sidestepped through the availability of feedback signals allowing online correc-

tions. This can be seen on Figure 17, in which the trajectories of trials with feedback can be directly compared to those without feedback. The kinematics of the movements are basically identical up to the very last phase of the action, during which online control is continuously engaged until object contact. Moreover, in Experiment 3 we show that these late and sudden corrections may influence the execution of successive trials, since they seem to be interpreted by the visuomotor system as indicative of errors in motor planning. After grasping an object whose depth had been overestimated, participants produce a greater grip aperture in the following grasp, and vice versa.

In conclusion, the overall findings of this study show a glaring similarity between the pattern of errors emerging from the phenomenological experience of 3D structure and the kinematics of motor actions. A parsimonious interpretation of these results is, therefore, that visual estimates of object properties constitute a common input to the perceptual and the visuomotor systems (Franz, 2001; Franz & Gegenfurtner, 2008; Kopiske et al., 2016).

Acknowledgments

Commercial relationships: none.
Corresponding author: Carlo Campagnoli.
Email: carlo.campagnoli@princeton.edu.
Address: Department of Psychology, Princeton University, Princeton, NJ, USA.

References

Baird, J. C., & Biersdorf, W. R. (1967). Quantitative functions for size and distance judgments. *Perception & Psychophysics*, *2*(4), 161–166.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, 5(3), 482–489.

Bingham, G., Coats, R., & Mon-Williams, M. (2007). Natural prehension in trials without haptic feedback but only when calibration is allowed. *Neuropsychologia*, 45(2), 288–294.

Bingham, G. P. (2005). Calibration of distance and size does not calibrate shape information: Comparison of dynamic monocular and static and dynamic binocular vision. *Ecological Psychology*, 17(2), 55–74.

Bingham, G. P., Crowell, J. A., & Todd, J. T. (2004). Distortions of distance and shape are not produced by a single continuous transformation of reach space. *Perception & Psychophysics*, 66(1), 152–169.

Bingham, G. P., Zaal, F., Robin, D., & Shull, J. A. (2000). Distortions in definite distance and shape perception as measured by reaching without and with haptic feedback. *Journal of Experimental Psychology: Human Perception and Performance*, 26(4), 1436–1460.

Binsted, G., & Heath, M. (2004). Can the motor system utilize a stored representation to control movement? *Behavioral and Brain Sciences*, 27(1), 25–27.

Bishop, P. O. (1989). Vertical disparity, egocentric distance and stereoscopic depth constancy: A new interpretation. *Proceedings of the Royal Society of London B: Biological Sciences*, 237(1289), 445–469.

Bootsma, R. J., Marteniuk, R. G., MacKenzie, C. L., & Zaal, F. T. (1994). The speed–accuracy trade-off in manual prehension: Effects of movement amplitude, object size, and object width on kinematic characteristics. *Experimental Brain Research*, 98(3), 535–541.

Bozzacchi, C., & Domini, F. (2015). Lack of depth constancy for grasping movements in both virtual and real environments. *Journal of Neurophysiology*, 114(4), 2242–2248.

Bozzacchi, C., Volcic, R., & Domini, F. (2014). Effect of visual and haptic feedback on grasping movements. *Journal of Neurophysiology*, 112(12), 3189–3196.

Bozzacchi, C., Volcic, R., & Domini, F. (2016). Grasping in absence of feedback: Systematic biases endure extensive training. *Experimental Brain Research*, 234(1), 255–265.

Bradshaw, M. F., & Elliott, K. M. (2003). The role of binocular information in the "on-line" control of prehension. *Spatial Vision*, 16(3), 295–309.

Bradshaw, M. F., Elliott, K. M., Watt, S. J., Hibbard, P. B., Davies, I. R., & Simpson, P. J. (2004). Binocular cues and the control of prehension. *Spatial Vision*, 17(1), 95–110.

Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, 36(9), 1255–1264.

Brenner, E., & Cornelissen, F. W. (2000). Separate simultaneous processing of egocentric and relative positions. *Vision Research*, 40(19), 2557–2563.

Brenner, E., Smeets, J. B., & Landy, M. S. (2001). How vertical disparities assist judgements of distance. *Vision Research*, 41(25), 3455–3465.

Brenner, E., & van Damme, W. J. (1998). Judging distance from ocular convergence. *Vision Research*, 38(4), 493–498.

Brenner, E., & van Damme, W. J. (1999). Perceived distance, shape and size. *Vision Research*, 39(5), 975–986.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture, and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, 33, 919–933.

Campagnoli, C., Volcic, R., & Domini, R. (2012). The same object and at least three different grip apertures. *Journal of Vision*, 12(9), 429, doi:10.1167/12.9.429. [Abstract]

Chen, Y., Byrne, P., & Crawford, J. D. (2011). Time course of allocentric decay, egocentric decay, and allocentric-to-egocentric conversion in memory-guided reach. *Neuropsychologia*, 49(1), 49–60.

Coats, R., Bingham, G. P., & Mon-Williams, M. (2008). Calibrating grasp size and reach distance: Interactions reveal integral organization of reaching-to-grasp movements. *Experimental Brain Research*, 189(2), 211–220.

Cormack, R. H. (1984). Stereoscopic depth perception at far viewing distances. *Perception & Psychophysics*, 35(5), 423–428.

Cuijpers, R. H., Brenner, E., & Smeets, J. B. (2004). Grasping virtual objects with constant haptic feedback. *Journal of Vision*, 4(8): 409, doi:10.1167/4.8.409. [Abstract]

Cuijpers, R. H., Brenner, E., & Smeets, J. B. (2008). Consistent haptic feedback is required but it is not enough for natural reaching to virtual cylinders. *Human Movement Science*, 27(6), 857–872.

Domini, F., Shah, R., & Caudek, C. (2011). Do we perceive a flattened world on the monitor screen? *Acta Psychologica*, 138(3), 359–366.

Eloka, O., & Franz, V. H. (2011). Effects of object shape on the visual guidance of action. *Vision Research*, 51(8), 925–931.

Foley, J. M. (1977). Effect of distance information and range on two indices of visually perceived distance. *Perception*, *6*(4), 449–460.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*(5), 411–434.

Foster, R., Fantoni, C., Caudek, C., & Domini, F. (2011). Integration of disparity and velocity information for haptic and perceptual judgments of object depth. *Acta Psychologica*, *136*(3), 300–310.

Foster, R., Januszewski, A., & Franz, V. (2015). Incorrect haptic feedback in 50% of trials is sufficient to bias grip aperture. *Journal of Vision*, *15*(12):1150, doi:10.1167/15.12.1150. [Abstract]

Franz, V. H. (2001). Action does not resist visual illusions. *Trends in Cognitive Sciences*, *5*(11), 457–459.

Franz, V. H., & Gegenfurtner, K. R. (2008). Grasping visual illusions: Consistent data and no dissociation. *Cognitive Neuropsychology*, *25*(7–8), 920–950.

Frisby, J. P., Buckley, D., & Horsman, J. M. (1995). Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception*, *24*(2), 181–198.

Gentilucci, M., Daprati, E., Gangitano, M., & Toni, I. (1997). Eye position tunes the contribution of allocentric and egocentric information to target localization in human goal-directed arm movements. *Neuroscience Letters*, *222*(2), 123–126.

Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological Review*, *58*(6), 460–482.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research*, *36*(21), 3441–3456.

Gogel, W. C. (1969). The sensing of retinal size. *Vision Research*, *9*(9), 1079–1094.

Gogel, W. C. (1977). The metric of visual space. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes* (pp. 129–181). New York: Wiley.

Gogel, W. C. (1990). A theory of phenomenal geometry and its applications. *Perception & Psychophysics*, *48*, 105–123.

Goodale, M. A., Jakobson, L. S., & Keillor, J. M. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia*, *32*(10), 1159–1178.

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, *199*, 678–680.

Harris, J. M. (2004). Binocular vision: Moving closer to reality. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical, and Engineering Sciences*, *362*(1825), 2721–2739.

Hibbard, P. B., & Bradshaw, M. F. (2003). Reaching for virtual objects: Binocular disparity and the control of prehension. *Experimental Brain Research*, *148*(2), 196–201.

Holmes, S. A., Lohmus, J., McKinnon, S., Mulla, A., & Heath, M. (2013). Distinct visual cues mediate aperture shaping for grasping and pantomime-grasping tasks. *Journal of Motor Behavior*, *45*(5), 431–439.

Hosang, S., Chan, J., Jazi, S. D., & Heath, M. (2016). Grasping a 2D object: Terminal haptic feedback supports an absolute visuo-haptic calibration. *Experimental Brain Research*, *234*(4), 945–954.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. New York: Oxford University Press.

Jackson, S. R., Jones, C. A., Newport, R., & Pritchard, C. (1997). A kinematic analysis of goal-directed prehension movements executed under binocular, monocular, and memory-guided viewing conditions. *Visual Cognition*, *4*(2), 113–142.

Jakobson, L. S., & Goodale, M. A. (1991). Factors affecting higher-order movement planning: A kinematic analysis of human prehension. *Experimental Brain Research*, *86*(1), 199–208.

Jazi, S. D., Yau, M., Westwood, D. A., & Heath, M. (2015). Pantomime-grasping: The "return" of haptic feedback supports the absolute specification of object size. *Experimental Brain Research*, *233*(7), 2029–2040.

Jeannerod, M. (1986). The formation of finger grip during prehension: A cortically mediated visuo-motor pattern. *Behavioural Brain Research*, *19*(2), 99–116.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*(7), 1351–1360.

Keefe, B. D., & Watt, S. J. (2009). The role of binocular vision in grasping: A small stimulus-set distorts results. *Experimental Brain Research*, *194*(3), 435–444.

Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2016). The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study. *Cortex*, *79*, 130–152.

Kopiske, K. K., Cesanek, E., Campagnoli, C., & Domini, F. (2017). Adaptation effects in grasping the Müller-Lyer illusion. *Vision Research*, *136*, 21–31.

Lee, Y. L., Crabtree, C. E., Norman, J. F., & Bingham,

G. P. (2008). Poor shape perception is the reason reaches-to-grasp are visually guided online. *Perception & Psychophysics*, *70*(6), 1032–1046.

Levin, M. F., Magdalon, E. C., Michaelsen, S. M., & Quevedo, A. A. (2008, August). Comparison of reaching and grasping kinematics in patients with hemiparesis and in healthy controls in virtual and physical environments. In *Proceedings of the 8th international conference for virtual rehabilitation*, (pp. 60). Vancouver, BC, Canada: IEEE.

Levin, M. F., Magdalon, E. C., Michaelsen, S. M., & Quevedo, A. A. (2015). Quality of grasping and the role of haptics in a 3-D immersive virtual reality environment in individuals with stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(6), 1047–1055.

Loftus, A., Servos, P., Goodale, M. A., Mendarozqueta, N., & Mon-Williams, M. (2004). When two eyes are better than one in prehension: Monocular viewing and end-point variance. *Experimental Brain Research*, *158*(3), 317–327.

Loomis, J. M., & Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual environments. *Virtual and Adaptive Environments*, *11*, 21–46.

Loomis, J. M., Philbeck, J. W., & Zahorik, P. (2002). Dissociation between location and shape in visual space. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(5), 1202–1212.

Loomis, J. M., Silva, J. A. D., Philbeck, J. W., & Fukusima, S. S. (1996). Visual perception of location and distance. *Current Directions in Psychological Science*, *5*(3), 72–77.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502.

Magdalon, E. C., Michaelsen, S. M., Quevedo, A. A., & Levin, M. F. (2011). Comparison of grasping movements made by healthy subjects in a 3-dimensional immersive virtual versus physical environment. *Acta Psychologica*, *138*(1), 126–134.

Marotta, J. J., Perrot, T. S., Nicolle, D., Servos, P., & Goodale, M. A. (1995). Adapting to monocular vision: Grasping with one eye. *Experimental Brain Research*, *104*(1), 107–114.

Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, *297*(5865), 376–378.

Melmoth, D. R., & Grant, S. (2006). Advantages of binocular vision for the control of reaching and grasping. *Experimental Brain Research*, *171*(3), 371–388.

Melmoth, D. R., Storoni, M., Todd, G., Finlay, A. L., & Grant, S. (2007). Dissociation between vergence and binocular disparity cues in the control of prehension. *Experimental Brain Research*, *183*(3), 283–298.

Messier, J., & Kalaska, J. F. (1997). Differential effect of task conditions on errors of direction and extent of reaching movements. *Experimental Brain Research*, *115*(3), 469–478.

Mon-Williams, M., & Bingham, G. P. (2007). Calibrating reach distance to visual targets. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(3), 645–656.

Mon-Williams, M., & Dijkerman, H. C. (1999). The use of vergence information in the programming of prehension. *Experimental Brain Research*, *128*(4), 578–582.

Neely, K. A., Tessmer, A., Binsted, G., & Heath, M. (2008). Goal-directed reaching: Movement strategies influence the weighting of allocentric and egocentric visual cues. *Experimental Brain Research*, *186*(3), 375–384.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 173–186.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, *24*(2), 155–179.

Schenk, T. (2012). No dissociation between perception and action in patient DF when haptic feedback is withdrawn. *The Journal of Neuroscience*, *32*(6), 2013–2017.

Schenk, T., Franz, V., & Bruno, N. (2011). Vision-for-perception and vision-for-action: Which model is compatible with the available psychophysical and neuropsychological data? *Vision Research*, *51*(8), 812–818.

Servos, P., & Goodale, M. A. (1994). Binocular vision and the on-line control of human prehension. *Experimental Brain Research*, *98*(1), 119–127.

Servos, P., Goodale, M. A., & Jakobson, L. S. (1992). The role of binocular vision in prehension: A kinematic analysis. *Vision Research*, *32*(8), 1513–1521.

Smeets, J. B. J., & Brenner, E. (1995). Perception and action are based on the same visual information: Distinction between position and velocity. *Journal*

of Experimental Psychology: Human Perception and Performance, *21*, 19–31.

Smeets, J. B. J., & Brenner, E. (1999). A new view on grasping. *Motor Control*, *3*(3), 237–271.

Smeets, J. B. J., & Brenner, E. (2008a). Why we don't mind to be inconsistent. In P. Calvo & T. Gomila (Eds.), *Handbook of cognitive science—An embodied approach* (pp. 207–217). Amsterdam: Elsevier.

Smeets, J. B. J., & Brenner, E. (2008b). Grasping Weber's law. *Current Biology*, *18*(23), R1089–R1090.

Smeets, J. B. J., Brenner, E., de Grave, D. D., & Cuijpers, R. H. (2002). Illusions in action: Consequences of inconsistent processing of spatial attributes. *Experimental Brain Research*, *147*(2), 135–144.

Sousa, R., Brenner, E., & Smeets, J. B. J. (2010). A new binocular cue for absolute distance: Disparity relative to the most distant structure. *Vision Research*, *50*(18), 1786–1792.

Sousa, R., Brenner, E., & Smeets, J. B. J. (2011). Objects can be localized at positions that are inconsistent with the relative disparity between them. *Journal of Vision*, *11*(2):18, 1–6, doi:10.1167/11.2.18. [PubMed] [Article]

Thaler, L., & Goodale, M. A. (2010). Beyond distance and direction: The brain represents target locations non-metrically. *Journal of Vision*, *10*(3):3, 1–27, doi:10.1167/10.3.3. [PubMed] [Article]

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 663–678.

van Damme, W., & Brenner, E. (1997). The distance used for scaling disparities is the same as the one used for scaling retinal size. *Vision Research*, *37*(6), 757–764.

Volcic, R., & Domini, F. (2014). The visibility of contact points influences grasping movements. *Experimental Brain Research*, *232*(9), 2997–3005.

Volcic, R., & Domini, F. (2016). On-line visual control of grasping movements. *Experimental Brain Research*, *234*, 2165–2177.

Volcic, R., Fantoni, C., Caudek, C., Assad, J. A., & Domini, F. (2013). Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination. *The Journal of Neuroscience*, *33*(43), 17081–17088.

Watt, S. J., & Bradshaw, M. F. (2000). Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia*, *38*(11), 1473–1481.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10):7, 834–862, doi:10.1167/5.10.7. [PubMed] [Article]

Westwood, D. A., Chapman, C. D., & Roy, E. A. (2000). Pantomimed actions may be controlled by the ventral visual stream. *Experimental Brain Research*, *130*(4), 545–548.

Wheatstone, C. (1838). Contributions to the physiology of vision—Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *128*, 371–394.

Wing, A. M., Turton, A., & Fraser, C. (1986). Grasp size and accuracy of approach in reaching. *Journal of Motor Behavior*, *18*(3), 245–260.

Appendix

Egocentric function

Consider the scenario depicted in Figure 1: an observer is asked to estimate the egocentric distance of a set of probe locations. Results from previous research (Bozzacchi & Domini, 2015; Foley, 1980) have shown that distances near to the body are overestimated whereas distances far from the body are underestimated (phenomenon that we refer to as visuospatial compression). This relationship between estimated and physical distance can be described by a linear function (which we name egocentric function):

$$z' = z_A + c \cdot (z - z_A) \quad (A1)$$

where z' is the estimated distance, z_A is a hypothetical distance at which the estimation is accurate, c is what we define as *compression factor* (which is less than 1), and z is the actual viewing distance.

Scaling function

The relative disparity projected by two points separated in depth by Δz is approximately equal to

$$d \cong \frac{\Delta z \cdot IOD}{z^2} \quad (A2)$$

where IOD is the inter-ocular distance and z is the viewing distance. Solving Equation A2 by Δz we obtain:

$$\Delta z = \frac{d \cdot z^2}{IOD} \quad (A3)$$

Consider now the scenario depicted in Figure 2: a participant is asked to estimate the relative depth of an object. The estimated depth $\Delta z'$ can be defined by using Equation A3:

$$\Delta z' = \frac{d \cdot z_s^{\,2}}{IOD} \quad (A4)$$

where z_s is the scaling distance; that is, the distance at which an object of depth $\Delta z'$ projects the relative disparity d. By substituting d as specified in Equation A2 on Equation A4 we obtain:

$$\Delta z' = \Delta z \cdot \left(\frac{z_s}{z}\right)^2 \quad (A5)$$

Equation A5 specifies the relationship between estimated depth, physical depth, scaling distance, and physical distance. According to Equation A5, depth estimation is veridical ($\Delta z' = \Delta z$) only when the scaling distance matches the viewing distance; that is, when $\frac{z_s}{z}$ is equal to 1. If, however, the participant scales the binocular disparity by a distance that is greater than the viewing distance ($\frac{z_s}{z} > 1$), the object's depth is overestimated as well. The opposite happens when $\frac{z_s}{z} < 1$.

Equation A5 can be solved by z_s to calculate the scaling distance corresponding to a given estimate of an object's depth. Empirical evidence from previous studies (Volcic et al., 2013) have shown that the scaling distance can be modeled with the same linear function used to model the estimated distance (Equation A1). Therefore, we define the *scaling function* as:

$$z_s = z_A + c \cdot (z_f - z_A) \quad (A6)$$

Comparison of visuospatial compression in perception and action

The present study had two goals: (a) to determine whether the same visuospatial compression affects both distance and depth estimation (in other terms, if egocentric distance and scaling distance are the same), and (b) to determine whether the same visuospatial compression affects both perception and grasping. In order to address both points, in Experiment 1 we concurrently measured z' and $\Delta z'$ in an action task through the FHP and the FGA respectively, and $\Delta z'$ in a perceptual task through the MSE.

We fitted the FHP with Equation A1 and we extracted the compression factor c_{FHP} for each subject (values on the abscissa of both graphs of Figure 6). To estimate the compression factor from the FGA and the MSE, we substituted z_s with the scaling function in Equation A6, and we fitted both variables with the following model:

$$\Delta z' = m + \Delta z \cdot \left(\frac{z_A + c \cdot (z - z_A)}{z}\right)^2 \quad (A7)$$

where the parameter m accounted for the average individual differences in the opening of the grip aperture (individual margin). Using Equation A7 we extracted c_{FGA} and c_{MSE} for each subject (values on the ordinate of Figure 6A and 6B).

Predicted final index position on Experiment 4

We used the thumb's final position from the front pinch (FTP_{pinch}) as an estimate of the apparent distance of the front rod during the planning of the grasp, and used it as a measure of scaling distance (z_s in Equation A5) to obtain the predicted depth estimate:

$$\Delta z' = \Delta z \cdot \left(\frac{FTP_{pinch}}{z}\right)^2 \quad (8)$$

where Δz is the physical depth (40 or 50 mm) and z is the front rod's distance. The predicted final index position of the grasp is therefore $FTP_{pinch} + \Delta z'$.