Multiple distance cues do not prevent systematic biases in reach to grasp movements

Karl K. Kopiske*[1, 5], Chiara Bozzacchi*[1,4], Robert Volcic[1,2], & Fulvio Domini[1,3]

1. *Center for Neuroscience and Cognitive Systems@UniTN, Istituto Italiano di Tecnologia, Rovereto, Italy*

2. *Department of Psychology, New York University Abu Dhabi, Abu Dhabi, UAE;*

3. *Department of Cognitive, Linguistic & Psychological Sciences, Brown University, RI, USA*

4. *Social Brain Laboratory, Netherlands Institute for Neuroscience, Netherlands Academy for Arts and Sciences (KNAW), Amsterdam, NL*

5. *Cognitive Systems Lab, Institute of Physics, Chemnitz University of Technology, D-09126 Chemnitz, Germany*

* Chiara Bozzacchi and Karl K. Kopiske contributed equally

Corresponding Author:

Karl K. Kopiske

Current Address: Cognitive Systems Lab, Institute of Physics,

Chemnitz University of Technology,

D-09126 Chemnitz, Germany.

Email address: karl.kopiske@physik.tu-chemnitz.de

Abstract

The perceived distance of objects is biased depending on the distance from the observer at which objects are presented, such that the egocentric distance tends to be overestimated for closer objects but underestimated for objects further away. This leads to the perceived depth of an object (i.e., the perceived distance from the front to the back of the object) also being biased, decreasing with object distance. Several studies have found the same pattern of biases in grasping tasks. However, in most of those studies, object distance and depth were solely specified by ocular vergence and binocular disparities. Here we asked whether grasping objects viewed from above would eliminate distance-dependent depth biases, since this vantage point introduces additional information about the object's distance, given by the vertical gaze angle, and its depth, given by contour information. Participants grasped objects presented at different distances (i) at eye-height and (ii) 130 mm below-eye-height, along their depth axes. In both cases, grip aperture was systematically biased by the object distance along most of the trajectory. The same bias was found whether the objects were seen in isolation or above a ground plane to provide additional depth cues. In two additional experiments, we verified that a consistent bias occurs in a perceptual task. These findings suggest that grasping actions are not immune to biases typically found in perceptual tasks, even when additional cues are available. However, online visual control can counteract these biases when direct vision of both digits and final contact points is available.

Introduction

Picking up objects is a goal-directed action that we perform frequently and with no apparent effort. This may suggest that to do this the brain computes with high accuracy and precision the visuomotor transformations that map location and shape of objects. Among all the visual cues available for the programming and control of reaching and grasping in humans, researchers have argued that "binocular cues emerge as being the most important" (Goodale 2011), and it has been put forward that stereo vision for action is veridical (Goodale, 2011), given sufficient information about object position (Chen, Sperandio, & Goodale, 2018).

In theory, the visual stereo information provided by vergence, accommodation and binocular disparity is sufficient to accurately measure the egocentric and allocentric information allowing the correct grasp of an object (Foley, 1980; Melmoth et al, 2006, 2007; Servos, 2000; Goodale 2011). However, it has been shown that binocular information alone produces systematically distorted depth estimates and grasps (Bozzacchi, Volcic, & Domini, 2014, 2016; Brenner & Van Damme, 1999; Foley, 1980; Hibbard & Bradshaw, 2003; Johnston, 1991; Keefe & Watt, 2017, Campagnoli, Croom, & Domini, 2017). These biases are consistent with a negative relationship between egocentric distance and perceived front-to-back depth in both perception and motor actions, and are at odds with the notion of a veridical stereo vision for action. The goal of the current study was to extend these findings by investigating if grasping kinematics remain biased when further depth cues are present.

The negative relationship between egocentric distance and perceived depth results from incorrect egocentric distance estimates of the target object. Consider an object located at eye-height in front of a viewer: the relative disparity between two points, respectively in the front and in the back of the object, does not uniquely determine its depth. The egocentric distance to the object, sensed by ocular convergence, must be accurately assessed for a veridical computation of the object

depth (Fig.1 *left*). If the distance (Z_f) is overestimated then the depth of the object is overestimated as well, since the distance scaling binocular disparities (Z_s) is larger than Z_f. The opposite takes place if the distance is underestimated ($Z_s<Z_f$) (Fig.1 *center*). Empirical evidence from perceptual studies indicates that the egocentric map relating physical distances to scaling distances is compressed, since objects close to the observer appear as being farther away and objects that are far from the observer as being closer (Foley, 1980; Johnson, 1991; Servos, 2000). This visual space compression yields corresponding distortions of depth estimates: when an object is estimated as closer to the observer than it really is ($Z_{s2}<Z_{f2}$), its depth is estimated as smaller ($\Delta Z_{s2}<\Delta Z_{f2}$), and vice versa (Mon-Williams et al., 2001; Domini and Caudek, 2013; Volcic et al., 2013), see Fig. 1 *inset*. Reach-to-grasp studies reveal the same negative relationship between egocentric distance and grip aperture when grasping objects front-to-back. That is, the grip aperture diminishes with the distance to the object (Bozzacchi et al, 2014; 2015; Campagnoli, Croom, & Domini, 2017).

This finding has been demonstrated repeatedly with observers grasping an object located at eye-height and along the line of sight (Bozzacchi et al., 2014; 2015; 2016; Campagnoli, Croom, & Domini, 2017; Volcic and Domini, 2016). Such a setup, however, leaves open the possibility that systematic biases affecting this task can be ascribed to the unusual combination of viewing condition and type of grasp. In this situation, binocular disparities and ocular convergence are the only information available specifying the front-to-back depth of the object and its egocentric distance. Conversely, in a natural setting objects are generally viewed from above while grasped. This condition allows additional allocentric information, such as object contour, and egocentric distance information, given by the vertical gaze angle (Wallach and O'Leary, 1982; Mon-Williams et al., 2001; Gardner and Mon-Williams, 2001; Marotta and Goodale, 1998, Keefe & Watt, 2017). Moreover, viewing an object from above in general allows a clear visibility of the grasping locations and a constant monitoring of their relative positions with respect to the grasping digits (Brenner & Smeets, 1997; Fukui & Inui, 2006).

The goal of the three experiments presented here was to investigate whether this additional information, present in a more typical viewing condition, reduces or eliminates the previously found systematic biases. We compared grasps aimed at objects located at eye level with grasps toward objects vertically displaced, so to be viewed from a higher vantage point (experiments 1 and 2). This viewing condition adds three important sources of information that could potentially correct for the compression of visual space. First, the vertical gaze angle (i.e., the angle formed between the straight ahead direction and the direction of gaze). Second, the pictorial information provided by the object contour to be combined with stereo disparity for a more accurate estimate of the object depth extent. Third, a better visibility of the grasp locations on the object, enabling a direct comparison between the grasping digits and the contact points when positions of both digits are visible (experiment 1). This was tested with (experiment 2) and without (experiment 1) a horizontal ground plane below the objects, providing a reference to a physical ground and adding information about the elevation. In addition to grip apertures, we then also measured perceptual estimates obtained from manual estimation (experiment 2) and an adjustment task (experiment 3), to test whether responses in both tasks would both show a bias depending on object distance.

If the negative relationship between egocentric distance and perceived depth reported in the previous studies, in which objects were presented at eye-height, is due to a limited source of depth and distance information, then we can predict an improvement of distance and front-to-back depth estimates for grasping actions directed at objects below eye-height. Alternatively, it is also possible that the planning of an action is not based on a veridical metric reconstruction of object properties, and that 3D estimates for action are based on the same mechanisms as in perception, therefore yielding the same biases (Domini and Caudek, 2013; Volcic et al., 2013).

Experiment 1: Grasping with multiple depth cues

In this experiment we tested whether viewing an object from above reduces or eliminates distance-dependent depth biases that are usually found in grasping when objects are seen at eye-height.

Participants

Nineteen undergraduate students from the University of Trento (mean age 23.5; 14 females) participated in the study and received a reimbursement of 8 Euros/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 1964 Declaration of Helsinki.

Apparatus and design

Participants were seated in a dark room in front of a high-quality, front-silvered 400 x 300 mm mirror slanted at 45° relative to the subjects' sagittal body mid-line. Participants maintained their head stable by placing it on a chin rest. The mirror reflected the image displayed on a monitor (19'' cathode-ray tube monitor running at 1024*768 px and 100 Hz) placed directly to the left of the mirror. For consistent vergence and accommodative information, the position of the monitor was adjusted on a trial-by-trial basis to equal the distance from the subjects' eyes to the virtual object. Visual 3D stimuli were presented with a frame interlacing technique in conjunction with liquid crystal FE-1 goggles (Cambridge Research Systems, Cambridge, UK) synchronized to the monitor frame rate. A disparity-defined high-contrast random-dot visual stimulus was rendered in stereo simulating a trapezoid (front-to-back depth: 60 mm) oriented with the smaller base (20 mm) towards the participants and the larger base (60 mm) away from the participants. Stimuli were simulated at four distances from the observers' eyes (420, 450, 480 and 510 mm along the depth axis) and two heights (eye-height and 130 mm below-eye-height), see Fig. 2a. In concomitance to the simulated virtual stimulus, a real wooden trapezoid of the same size as the simulated one was located at the designated distance and height in front of the participant to provide veridical haptic feedback.

The physical object was positioned by a mechanical arm. Each condition was presented in random order 10 times, for a total of 80 trials.

Before the experiment, subjects were tested for stereo-vision and their interocular distance was measured. All experimental sessions started with a calibration procedure. The position of the head, eyes, wrist, the thumb and index fingers pads was calculated with respect to infrared-emitting diodes. During calibration, the position of the center of each finger pad relative to the three markers fixed on the nail of each finger was determined through a fourth calibration marker. This information was used to render the visual feedback (virtual cylinders of 20 mm height and 10 mm diameter representing the fingers phalanxes, see Fig. 2b) and for calculating the grip aperture, defined as the Euclidean distance between the two finger pads (Nicolini et al., 2014). Head, wrist, index and thumb movements were acquired online at 100 Hz with sub-millimeter resolution by using an Optotrak Certus motion capture system (Northern Digital Inc., Waterloo, Ontario, Canada). Participant's interocular distance and the tracked head position and head orientation were used to deliver in real time the geometrically correct projection to the two eyes.

Trials started with the presentation of the stimulus at one of the different distances and heights. From the moment the target was visible, participants could move to grasp it front-to-back, placing the thumb on the frontal surface (smaller base) and the index finger on the posterior surface (larger base). One second after participants had grasped the object, the visual image of the object disappeared and the trial ended.

Data analysis, dependent variables

Data were processed and analyzed offline using custom software written in R (R Core Team, 2015). Each grasping movement was smoothed with a third-order Savitzky-Golay filter and spatially normalized using the total length of the movement, i.e., the length of the trajectory covered by the midpoint between the fingertip positions. Each thumb and index trajectory was resampled in 100

points evenly spaced along the three-dimensional trajectory in the range from 0 (movement onset) to 1 (movement end, defined as both fingers touching the target object) in 0.01 steps using cubic spline interpolation. For each trial and for each point of the space-normalized trajectory, we computed the Euclidean distance between the fingertips of the thumb and the index finger (grip aperture) and we used this information for the statistical analysis (see also Volcic and Domini, 2016), which consisted mainly of computing how grip aperture is related to the egocentric object distance across the trajectory. For this, we assumed a linear relationship *grip aperture~egocentric object distance* and focussed on its slope, which we calculated for the grip apertures measured at each point along the space-normalized trajectory. This analysis enabled us to investigate not just the effect of object distance on maximum grip aperture (MGA), but also potential corrections in the earlier and later portions of the grasping movement.

Note that this normalization method does not show the same spurious correlation with grip aperture that may be seen in time-normalization techniques (Whitwell & Goodale, 2013), see Fig. 9 of Volcic & Domini (2016). Small artefacts could occur in our design if the MGA was reached at substantially different points of the spatial trajectory for different distances – however, the differences in relative position of MGA between relatively near and far objects were extremely small (on the order of 1% of the movement) and more importantly, other space-normalization methods revealed very similar results (e.g., normalization relative to the MGA, or of fixed parts of the movement such as the last 20 cm). We also analyzed the MGA. Trials were excluded before any data analysis was performed if upon visual inspection of the trajectories there were signs of 'double grasping'.

Results

In order to compare this study with our earlier investigations, we first analyzed how the MGA changes with viewing distance. To this end, we preliminarily ran a repeated-measures ANOVA on the

MGA with distance on the depth axis (4 levels: 420, 450, 480, 510 mm) and height (2 levels: eye-height, 130 mm below eye-height) as within-participant factors. The analysis revealed a main effect of distance [$F(3, 54) = 3.61$, $p = .019$] and a significant height x distance interaction [$F(3, 54) = 3.44$, $p = .023$]. Additionally, for each participant and object height we fitted a linear regression model on the MGA as function of distance (centered on the distance mean of 465 and 483 mm for the eye-height and below-eye-height, respectively). When the object was grasped at eye-height we found a clear signature of visual space compression. MGA decreased with object distance (Fig. 4 *left*), as revealed by the significant negative slope of the linear function relating MGA to the egocentric object distance [$t(18) = -4.26$; $p < .001$, mean slope value -0.032; SEM = 0.007]. However, this effect was not found for grasps directed at objects below eye-height [$t(18) = -0.04$; $p = .96$, slope value -0.0004; SE = 0.01], in agreement with the interaction effect found in the ANOVA.

The apparent insensitivity of MGA to the distance manipulation seems to suggest perfect depth constancy in vision for action in the below-eye-height condition. However, examining how distance affects the grip aperture during the entire movement refutes this conclusion. We did so by calculating the slope relating grip aperture to egocentric object distance for each point along the space-normalized trajectory of the transport component of the movement (a typical grip-aperture trajectory and an example of how the slope was calculated can be seen in Fig. 3a). Fig. 3b, where this slope is plotted as function of the proportion of path traveled of the trajectory, clearly shows that in both experimental conditions the slope was negative for a large portion of the movement. This analysis shows a clear effect of distance on the grip aperture in agreement with a visual space compression.

Whereas the effect of distance on grip aperture persisted until movement completion in the eye-height condition, it disappeared in the below-eye-height condition around the moment of MGA. This change may indicate an online correction due to the visibility of the object contact points and the detection of an inappropriate safety margin between the object's grasp locations and the

fingertips (Fig. 3c). This sudden correction may also explain why at the end of the movement the slope's sign is even inverted.

Discussion

The present results replicate previous findings showing grip opening for grasps along the depth axis to be negatively related to an object's egocentric distance for grasps aimed at objects located at eye-height. Although participants repeatedly grasped the same object while seeing their hand and feeling the object at the end of the action, they show a clear modulation of grip aperture with object distance. As recently demonstrated (Bozzacchi and Domini, 2015; Volcic and Domini, 2016), the effect of distance on grip aperture is not only evident at the time of MGA, but characterizes a large portion of the movement. Note that this pattern of smaller grasps towards objects further away is in line with known perceptual effects (Campagnoli et al., 2017; Domini & Caudek, 2013; Foley, 1980; Volcic et al., 2013) but the opposite to what has repeatedly been shown in grasps towards flat, fronto-parallel objects, where grasps further away tend to show a larger MGA (Keefe, Hibbard, & Watt, 2011; Verheij, Brenner, & Smeets, 2012).

When the object is viewed at an angle, additional information *does not contribute* to the correction of front-to-back depth estimation biases, with the exception of the very last part of the grasping movement. During this phase, the relative disparity between contact points and fingertips can be easily monitored and corrected. We therefore interpret this finding, which agrees with previous research (Mon-Williams et al., 2001; Melmoth and Grant., 2006; Melmoth et al., 2007; Jakobson and Goodale, 1991; Servos et al., 1992; Marotta and Goodale 1998), as the result of an online control mechanism, not possible in the eye-height condition, in which the index finger tends to disappear at the time of MGA and beyond.

Experiment 2: Comparing grasping and perception with multiple depth cues

Having shown that a negative relationship between egocentric distance along the depth axis and grip aperture in grasping can also be found when elevation angle is added as distance cue, we wanted to test the generalizability of this finding by additionally including a planar surface to make perspective information more reliable (as used, e.g., by Greenwald, Knill, & Saunders, 2005). To be able to test that our stimuli indeed produced consistent perceptual biases, we also included a manual estimation (ME) task, in which participants estimated the depth of our stimuli.

Participants

Twenty volunteers (mean age 26.1 years, 12 female, all right handed by self-report) were invited to the lab for experiment 2. Recruitment, ethical approval and data protection were handled the same way as in experiment 1. Two participants could not complete the experiment since their arms were not long enough to comfortably grasp the object at the position that was further away, and one participant did not receive visual feedback for a number of trials due to the head markers being occluded by hair. A further participant had to be excluded due to a too high number of trials that had to be excluded in grasping, and two participants for ME, leaving us with N=16 and N=15 participants for the two tasks, respectively.

Apparatus and design

Like in experiment 1, subjects were seated in a dark room in front of a semi-transparent mirror that was slanted and adjusted to reflect the display of a CRT monitor located to the left of the mirror to a position in front of the subject. Subjects again wore liquid crystal goggles, as well as IREDs attached to the head, thumb, index finger, and wrist to track movements of the head and right hand. The head was resting in a chin rest. We presented similar disparity-defined random-dot images as in experiment 1, again displaying virtual 3D trapezoids. Unlike in experiment 1, a 380 mm * 480 mm

plane made of black cardboard, covered with blotches of phosphorescent green paint and encompassed by a band of lightly glowing green diodes, was located in front of the subjects and below the target objects, at a height of app. 155 mm below eye level. The mechanic arm moving the objects was protruding from below the plane through a 60 mm * 120-mm hole in the middle of the cardboard.

All trapezoid stimuli were 22.5 mm high, 26.5 mm wide at the base oriented towards the subject and 55 mm at the base further away, and either 50 mm or 60 mm deep. Stimuli were presented at one of four possible positions: at one of two heights (at eye level or 130 mm below eye level) and one of two distances along the depth axis (420 mm or 510 mm). Using three motors, one of two real metal trapezoids (of the same dimensions as the virtual stimuli) attached to a metal arm could be moved to the same location as the virtual stimulus to provide veridical haptic feedback. These objects were not illuminated, such that only the plane and the virtual object could be seen during the experiment.

Subjects completed two tasks: a grasping task, and a manual estimation task, with calibration performed before each task in the same way as in experiment 1. The grasping task was analogous to the grasping task in experiment 1, with subjects performing the same front-to-back precision grip on an invisible real object that was of the same size and in the same location as the 3D virtual object. The differences were (i) the added use of a plane below the objects, (ii) the fact that only the thumb position was visible during the grasp, and (iii) a different number of sizes and distances, as we used two different object sizes at two different distances rather than one size at four distances. Like in experiment 1, we used ten repetitions, resulting in 8 * 10 = 80 trials.

In the manual estimation task, participants were presented with the same virtual stimuli as well as the plane but did not reach out and grasp the objects. Instead, they were asked to indicate the perceived depth of the trapezoid with their right thumb and index finger (again measured using the Optotrak Certus). Subjects registered their responses by pressing the spacebar with the left hand, at

which point the virtual stimulus disappeared and subjects returned their right hand to the starting position. A new trial started following an interval of 1 second. Like in the grasping task, subjects completed 80 (8 cells * 10 repetitions) trials.

Results and discussion

We submitted the MGA values from the grasping task and the manual estimates each to a 2 (*distance*) * 2 (*height*) * 2 (*object size*) repeated-measures ANOVA. This was in order to establish whether grasping as well as ME were affected by stimulus distance. We then again used planned t-tests to test for the distance effect (i.e., the difference between responses to objects presented at different egocentric distances) specifically at both stimulus heights, here by simply comparing responses at the two distances. This was the equivalent to the t-tests on slopes conducted for experiment 1, except that the use of only two distances meant that we could run t-tests on means per distance directly rather than slopes.

The ANOVA on MGAs revealed that in the grasping task, participants scaled with object size [main effect of *size*: $F(1, 15) = 40.57$, $p < .001$] and showed a distance effect [i.e., a main effect of egocentric object *distance*: $F(1, 15) = 11.21$, $p = .004$]. However, we also saw a main effect of the factor *height*, $F(1, 15) = 6.86$, $p = .019$, as well as an interaction of *size*height*, $F(1, 15) = 9.18$, $p = .008$, which revealed that participants grasped larger, but MGAs scaled less when objects were presented at eye height. Planned t-tests also revealed that while the distance effect was clearly present at eye height [$t(15) = 3.58$, $p = .003$], this was not the case below eye height [$t(15) = 0.87$, $p = .399$]. However, these effects were not significantly different [$t(15) = -1.78$, $p = .095$]. In ME, there was also a main effect of object *size*, $F(1, 14) = 33.11$, $p < .001$, but no significant *distance* effect, $F(1, 14) = 0.50$, $p = .492$. Planned t-tests showed that the effect was present neither at nor below eye height (both *ps* > .13). Response scaling was rather weak in MGAs (mean scaling factor, indicating the unit of change in response per unit of change in target size: 0.33, SEM: 0.05), but stronger in manual estimates (mean

scaling factor: 0.98, SEM: 0.17). This scaling factor is smaller than the typical factors both for *MGA~size* (for objects grasped in the frontoparallel orientation typically around 0.8, Smeets & Brenner, 1999) as well as manual estimates (typically larger than 1 for relatively small ranges of objects, Franz, 2003; Kopiske & Domini, 2018). Fig. 5c shows how grip aperture scaled with object size across the entire trajectory.

Distance effects in mm can be seen in Fig. 4. Note that to compare these directly, it would be preferable to correct these effects for the response scaling of grasping measures and manual estimates, respectively. We chose not to do this to enable a clearer comparison to experiment 1, where no such scaling factor could be computed, reflecting our emphasis on comparing distance effects between stimulus setups, rather than between perception and action. Performance in the ME task, indicating how well participants estimated the different objects depths, can be seen in Table 1. Virtually all participants scaled their responses with object depth, but some showed rather poor discrimination performance. Discrimination performance was quantified using d' (the difference between the mean responses divided by the pooled standard deviation: $[M_{60mm} - M_{50mm}] / SD$), and was for some participants around 0, compared to the overall-mean d' of 0.97 ± an SEM of 0.15. We also conducted the same slope-trajectory analysis of grasping movements using a linear mixed-effect model, which is also plotted in Fig. 5 (*a*: showing slope-trajectories both at eye-height and below, *b*: showing the difference between slope-trajectories at eye-height and below).

Comparing results from experiments 1 and 2 allows several conclusions. First, introducing a planar surface as reference did not weaken the distance effect in grasping. Second, the distance effect in grasping persisted whether or not participants were given enough online visual information to correct their movements online. Third, we saw again the distance effect disappear at late portions of grasps below eye height, despite the fact that visual information could not be used to correct the grip aperture online. Fourth, the perceptual distance effect (Domini & Caudek, 2013; Foley, 1980) did not show up in our manual estimation task. This was in contrast to our predictions as well as previous

studies from our group (Campagnoli et al., 2017) and especially surprising given that we found the expected bias in grasping. While of course some have argued for a dissociation between perception and action (Goodale & Milner, 1992; Goodale, 2008), their account would predict unbiased action measures even where perception is biased, which is of course the opposite pattern to our results. Consequently, we conducted conducted another experiment to verify whether the specific setup used here would elicit the typical distance effect or whether, unexpectedly, perception would remain unaffected by the manipulation.

Experiment 3: Verifying the perceptual distance effect on estimated front-to-back depth

In experiment 2, we found the same distance effect in grasping as we did in experiment 1 – however, the manual estimation task did not show clearly showed such effect, which was quite unexpected. To investigate whether this was a problem with our method of investigation (as manual estimates can be noisy, especially when investigating small effects, see Kopiske & Domini, 2018) we conducted another experiment in which we tested for the known distance effect on perceived depth (Domini & Caudek, 2013) with the same stimuli, but a more traditional psychophysical size-adjustment task.

Participants

Twelve volunteers (mean age 29.4 years, 5 female, all but one participant right handed by self-report) participated in experiment 3. Recruitment, ethical approval and data protection were handled the same way as in experiment 1.

Apparatus and design

Like in experiments 1 and 2, subjects were seated in a dark room in front of a semi-transparent mirror that was slanted and adjusted to reflect the display of a CRT monitor located to the left of the mirror to a position in front of the subject. Subjects again wore liquid crystal goggles, as well as IREDs

attached to the head, which was resting in a chin rest, to ensure spatially accurate stimulus presentation.

Stimuli were the same disparity-defined random-dot 3D trapezoids as in experiment 2, and of the same dimensions. To the right (90 mm from the body midline) at a distance along the depth-axis of 465 mm was a flat vertical line of a random height between 1 mm and 100 mm and a width of 10 mm, which could be adjusted in height by pressing the ',' and '.' buttons of a standard Italian USB keyboard. Participants were instructed to adjust the height of the line to match the front-to-back depth of the centrally displayed trapezoid, and then press the spacebar to confirm their response.

Results and discussion

We analysed the data in the same way as the manual estimates in experiment 2, by conducting first a repeated-measures 2*2*2*2 ANOVA with factors *distance*, *height*, *plane*, and *object size*. This was followed up by planned t-tests comparing the two distances at eye-height and below eye-height to investigate the distance effect specifically (see Fig. 4).

We found main effects of object *size* ($F(1, 11) = 36.13$, $p < .001$) and of *distance* ($F(1, 11) = 9.71$, $p = .010$), as well as an interaction of *size*height* ($F(1, 11) = 22.28$, $p = .001$) indicating that estimated depth scaled more strongly with actual depth when the object was presented below eye-height. The factor *plane* showed no main effect ($F(1, 11) = 0.17$, $p = .691$) and no interaction with any other factor (all *ps* > .18). Critically, there was no interaction of factors *distance* and *height* ($F(1, 11) = 1.39$, $p = .263$) and planned t-tests confirmed that indeed the distance effect persisted both at eye-height ($t(11) = 2.65$, $p = .023$) and below eye-height [$t(11) = 3.07$, $p = .011$]. Responses scaled with a factor of only 0.63 (SEM: 0.1), which is consistent with both the notion of a negative relationship between egocentric distance and perceived depth and the literature on mapping of horizontal to vertical distances (see e.g. Higashiyama & Ueyama, 1988). Participants' overall depth discrimination

performance was slightly better than in experiment 2 (mean d': 1.18 ± 0.26) and can be seen in Table 1.

Overall, this confirms that despite mixed results in experiment 2, our setup elicited the standard perceptual depth underestimation at larger distances compared to objects that were closer to the observer (Domini & Caudek, 2013; Foley, 1980). We found no difference between the distance effect with and without a planar surface below the stimuli in a within-participant design. Descriptively, the distance-dependent underestimation was slightly larger when objects were presented below eye height. However, since this difference was small and there was no statistically significant interaction between object height and object distance ($p = .263$), it is unclear whether this is indicative of an inconsistency between the biases in grasping and perception or merely random noise.

General discussion

Our experiments show that a perceptual bias where the relative front-to-back depth of objects is estimated gradually smaller with increasing object distance can be found in grasping. This is true even when multiple perspective cues are given, including gaze angle, contour information, and a ground plane. While subtle differences persist between grasps at eye level and below eye level, overall the data clearly show (Fig. 3, Fig. 5) that in all conditions grasping kinematics differed by distance in accordance with this perceptual effect.

In order to confirm previous evidence of this distance effect with our experimental setting we measured it in two different ways: As manual estimates (experiment 2, Fig. 4 *second-to-right*) and with a classic adjustment task (experiment 3, Fig. 4 *right*). The former gave us no clear pattern of a distance effect. While we did see numerically smaller responses to objects further away, this difference was small and non-significant. This unexpected result (the negative relationship between egocentric distance and perceived front-to-back depth has been shown in ME tasks, see e.g.

Campagnoli et al., 2017) prompted us to conduct the latter experiment to test if our setup would elicit the usual distance effect in a simpler, less noisy perceptual task (mean SD for each cell: 9 mm, compared to 10.7 mm in ME. Compare also the discrimination performance given in Table 1, which reveals steeper scaling in ME [0.98 in ME compared to 0.63 in experiment 3], but a lower mean d' [0.97 to 1.18]). Here, we not only found the standard distance effect, but were also able to show in a within-participant design that it persisted regardless of whether we added a planar surface as a cue or not. Thus, perception and action behaved in a largely similar way.

Why we found no distance effect in ME we can only speculate. In fact, the third experiment was run exclusively to investigate the possibility (unlikely, given the results in grasping) that our setup might not have produced the usual effect. It did produce the effect, leaving us with, in essence, two options which we cannot distinguish based on our data: either some of our pattern of results (i.e., the distance effects in the two grasping tasks as well as the judgment task on the one hand, or the lack thereof in ME on the other hand) was a product of chance, or something about the ME task made the difference. Whether ME is indeed useful as a standard measure of perception has been discussed (see e.g. Franz, 2003) but not definitively answered, and may be a target for future studies. We would point out that a distance effect has been found in ME tasks in similar experiments (see, e.g., Campagnoli, et al., 2017).

To the degree that the subtle differences between grasps at eye height and below eye height are robust, this raises the question of what may have caused them. Two candidates were tested in our study: perspective information about the stimulus (i.e., a better depth estimate due to gaze angle, contour, available in the below-eye-height conditions in experiments 1 and 2, and the reference given by the planar surface, available in experiment 2), and online visual information about the hand (i.e., online comparison of aperture vs. object size, possible in experiment 1). The fact that perceptual tasks showed better discrimination performance below eye height (mean d' in experiment 2: 0.83 at eye height, 1.1 below eye height; experiment 3: mean d' of 0.86 at eye height to 1.51 below eye height)

but revealed no effect of stimulus height or the presence of a surface on the distance effect in depth estimation speaks against the notion of the differences being driven by distinct depth estimates. On the other hand, online information about the fingers or the lack thereof is known to affect the trajectory of precision grasps (Bozzacchi et al., 2018), although it should be noted that a strong effect of online control on the distance effect would predict pronounced differences between the results from experiment 1 (where both digits were visible during grasping) and experiment 2 (where only the thumb was visible), which we did not see. Thus, we are not in a position to make strong claims on this question based on the data from this study.

From a methodological perspective, MGA is often considered a diagnostic feature of a grasping movement, since it is empirically observed that it reliably correlates with the size of the grasped object (Jeannerod, 1981; 1984; Jakobson and Goodale, 1991; Goodale and Milner, 1992; Aglioti et al., 1995; Westwood and Goodale, 2003). However, the results of the present study show that the lack of modulation of MGA with distance is only partly indicative of a correct estimate of front-to-back object depth. In fact, to truly understand how distal information is encoded at planning, grip aperture must be observed at movement instants before online control come into play (Glover & Dixon, 2002; Heath, Rival, & Binsted, 2004). In order to do so, we analyzed how object distance modulated the grip aperture at finely chosen locations along the spatially normalized transport trajectory. The slope of the function relating grip aperture to egocentric object distance was negative throughout most of the movement trajectory and did not significantly differ from that observed when the object was grasped at eye-height. This is consistent with the measured effect of distance on MGA overall and at eye height, as well as other work showing an effect of distance on grasping in depth also at MGA (Campagnoli et al, 2017; Campagnoli & Domini, submitted), and clearly indicates that vision for action is subject to systematic biases related to depth estimates, showing, like perception, a negative relationship between egocentric target distance and the dependent measure (Johnston, 1991; Hibbard and Bradshaw, 2003; Volcic et al., 2013).

Another methodological point is our choice of stimuli, as we used random-dot stereograms presented on a computer screen. Our motivation was wanting to isolate certain cues, such as stereo information, which is readily done in such a setup, but much more difficult with real objects as targets. However, it might raise the question of whether the processing of such virtual stimuli is really comparable to that of real objects. To our knowledge, there have been studies suggesting that there may be a difference in processing for 2-dimensional pictures (e.g. Snow, et al., 2011), but as far as we are aware such studies do not exist for 3D stimuli comparable to the ones used in our study. Thus, we consider the advantages of a virtual setup (being able to directly manipulate our target cues in isolation) to outweigh this potential concern. We would also point out that the effect of underestimated depth at larger distances has been found not just with virtual targets, but under certain conditions also with real objects in perception (Norman, Todd, Perotti, & Tittle, 1996) as well as grasping (Campagnoli & Domini, 2016). Thus, we believe that the effects we study here are a matter of cues, not realism.

In summary, our data further support the hypothesis that 3D processing is biased both in perception and action, and are in contrast with current visuomotor theories postulating that the visual information used to program grasping movements is metrically accurate and immune to visual distortions (Milner and Goodale, 1993; Goodale and Milner, 1992; Goodale 2011). How can we reconcile this fact with the observation that humans achieve successful motor actions effortlessly, without missing an object or knocking it off? We propose that the visual system does not pursue a veridical metric reconstruction of the scene and that online control (visual feedback of the hand and the final contact with the object) can refine the movement to compensate for visual inaccuracy (Domini and Caudek, 2013).

Compliance with Ethical Standards

References

Bozzacchi, C., Brenner, E., Smeets, J. B., Volcic, R., & Domini, F. (2018). How removing visual information affects grasping movements. *Experimental Brain Research, 236,* 985-995. http://doi.org/10.1007/s00221-018-5186-6

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. http://doi.org/http://dx.doi.org/10.1152/jn.00350.2015

Bozzacchi, C., Volcic, R., & Domini, F. (2014). Effect of visual and haptic feedback on grasping movements. *Journal of Neurophysiology*, *112*(12), 3189–3196. http://doi.org/10.1152/jn.00439.2014

Bozzacchi, C., Volcic, R., & Domini, F. (2016). Grasping in absence of feedback: systematic biases endure extensive training. *Experimental Brain Research*, *234*, 255–265. http://doi.org/10.1007/s00221-015-4456-9

Brenner, E., & Smeets, J. B. J. (1997). Fast responses of the human hand to changes in target position. *Journal of Motor Behavior, 29*(4), 297–310.

Brenner, E., & Van Damme, W. J. M. (1999). Perceived distance, shape and size. *Vision Research*, *39*(5), 975–986. http://doi.org/10.1016/S0042-6989(98)00162-X

Campagnoli, C., Croom, S., & Domini, F. (2017). Stereovision for action reflects our perceptual experience of distance and depth. *Journal of Vision*, *17*(9), 1–26. http://doi.org/10.1167/17.9.21.doi

Campagnoli, C., & Domini, F. (2018). *Depth-cue combination yields identical biases in perception and grasping*. Manuscript submitted for publication.

Campagnoli, C., & Domini, F. (2016). Conscious perception and grasping rely on a shared depth encoding. *Journal of Vision, 16*, 449. doi:doi:10.1167/16.12.449

Chen, J., Sperandio, I., & Goodale, M. A. (2018). Proprioceptive distance cues restore perfect size constancy in grasping, but not perception, when vision is limited. *Current Biology*. doi:10.1016/j.cub.2018.01.076

Domini, F., & Caudek, C. (2013). Perception and action without veridical metric reconstruction: An affine approach. In *Shape perception in human and computer vision* (pp. 285–298). London, UK: Springer.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*(5), 411–434. http://doi.org/10.1037/h0021465

Fukui, T., & Inui, T. (2006). The effect of viewing the moving limb and target object during the early phase of movement on the online control of grasping. *Human Movement Science, 25,* 349–371. doi:10.1016/j.humov.2006.02.002

Franz, V. H. (2003). Manual size estimation: a neuropsychological measure of perception? *Experimental Brain Research*, *151*, 471–477.

Gardner, P. L., & Mon-Williams, M. (2001). Vertical gaze angle: Absolute height-in-scene information for the programming of prehension. *Experimental Brain Research*, *136*(3), 379–385. http://doi.org/10.1007/s002210000590

Glover, S., & Dixon, P. (2002). Dynamic effects of the Ebbinghaus illusion in grasping: Support for a planning/control model of action. *Perception & Psychophysics, 64*(2), 266–278.

Goodale, M. A. (2011). Transforming vision into action. *Vision Research*, *51*(13), 1567–1587.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Greenwald, H. S., Knill, D. C., & Saunders, J. A. (2005). Integrating visual cues for motor control: A matter of time. *Vision Research*, *45*, 1975–1989. http://doi.org/10.1016/j.visres.2005.01.025

Heath, M., Rival, C., & Binsted, G. (2004). Can the motor system resolve a premovement bias in grip aperture? Online analysis of grasping the Müller-Lyer illusion. *Experimental Brain Research, 158,* 378–384. doi:10.1007/s00221-004-1988-9

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. http://doi.org/10.1007/s00221-002-1295-2

Higashiyama, A., & Ueyama, E. (1988). The perception of vertical and horizontal distances in outdoor settings. *Perception & Psychophysics*, *44*(2), 151–156. http://doi.org/10.3758/BF03208707

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. http://doi.org/10.1007/BF00231054

Jeannerod, M. (1984). The timing of natural prehension movements. *Journal of Motor Behavior*, *16*(3), 235–254.

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

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*, 1351–1360. http://doi.org/10.1016/0042-6989(91)90056-B

Keefe, B. D., Hibbard, P. B., & Watt, S. J. (2011). Depth-cue integration in grasp programming: No evidence for a binocular specialism. *Neuropsychologia*, *49*, 1246–1257. http://doi.org/10.1016/j.neuropsychologia.2011.02.047

Keefe, B. D., & Watt, S. J. (2017). Viewing geometry determines the contribution of binocular vision to the online control of grasping. *Experimental Brain Research*. http://doi.org/10.1007/s00221-017-5087-0

Kopiske, K. K., & Domini, F. (2018). On the response function and range dependence of manual estimation. *Experimental Brain Research, 236,* 1309-1320. http://doi.org/10.1007/s00221-018-5223-5

Marotta, J. J., & Goodale, M. A. (1998). The role of learned pictorial cues in the programming and control of grasping. *Experimental Brain Research*, *121*, 465–470. http://doi.org/10.1007/s002210050482

Melmoth, D. R., & Grant, S. (2006). Advantages of binocular vision for the control of reaching and grasping. *Experimental Brain Research*, *171*(3), 371–388. http://doi.org/10.1007/s00221-005-0273-x

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. http://doi.org/10.1007/s00221-007-1041-x

Mon-Williams, M., McIntosh, R. D., & Milner, A. D. (2001). Vertical gaze angle as a distance cue for programming reaching: Insights from visual form agnosia II (of III). *Experimental Brain Research*, *139*(2), 137–142. http://doi.org/10.1007/s002210000658

Nicolini, C., Fantoni, C., Mancuso, G., Volcic, R., & Domini, F. (2014). A framework for the study of vision in active observers. In B. Rogowitz, T. Pappas, & H. de Ridder (Eds.), *Proceedings of the SPIE* (Vol. 9014, p. 901414). San Francisco, CA. http://doi.org/10.1117/12.2045459

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. doi:10.1037/0096-1523.22.1.173

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Servos, P. (2000). Distance estimation in the visual and visuomotor systems. *Experimental Brain Research*, *130*, 35–47. http://doi.org/10.1007/s002210050004

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. (1999). A new view on grasping. *Motor Control*, *3*(3), 237–271.

Snow, J. C., Pettypiece, C. E., McAdam, T. D., McLean, A. D., Stroman, P. W., Goodale, M. A., & Culham, J. C. (2011). Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Scientific Reports, 1*, 1–10. doi:10.1038/srep00130

Verheij, R., Brenner, E., & Smeets, J. B. J. (2012). Grasping kinematics from the perspective of the individual digits: A modelling study. *PLoS ONE*, *7*(3), e33150. http://doi.org/10.1371/journal.pone.0033150

Volcic, R., & Domini, F. (2016). On-line visual control of grasping movements. *Experimental Brain Research*, *234*, 2165–2177. http://doi.org/10.1007/s00221-016-4620-x

Volcic, R., Fantoni, C., Caudek, C., Assad, J. A., & Domini, F. (2013). Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination. *Journal of Neuroscience*, *33*(43), 17081–17088. http://doi.org/10.1523/JNEUROSCI.2936-13.2013

Wallach, H., & O'Leary, A. (1982). Slope of regard as a distance cue. *Perception & Psychophysics*, *31*(2), 145–148. http://doi.org/10.3758/BF03206214

Whitwell, R. L., & Goodale, M. A. (2013). Grasping without vision: Time normalizing grip aperture profiles yields spurious grip scaling to target size. *Neuropsychologia, 51*(10), 1878–1887. doi:10.1016/j.neuropsychologia.2013.06.015

Tables

Table 1: Depth estimates and discrimination performance in ME (experiment 2) and visual adjustment (experiment 3).

			Z_f = 420 mm			Z_f = 510 mm		
			Mean (COV), 50 mm	Mean (COV), 60 mm	d'	Mean (COV), 50 mm	Mean (COV), 60 mm	d'
Manual Estimation		y = 0 mm	72.4mm (0.14)	83.15mm (0.13)	0.96	73.36mm (0.14)	81.13mm (0.14)	0.71
		y = -130 mm	69.42mm (0.13)	80.54mm (0.11)	1.29	71.35mm (0.14)	81.23mm (0.14)	0.91
Adjustment	plane	y = 0 mm	59.81mm (0.13)	64.42mm (0.12)	1.0	59.02mm (0.13)	62.52mm (0.11)	0.71
		y = -130 mm	61.37mm (0.10)	70.03mm (0.08)	1.73	59.29mm (0.10)	67.01mm (0.09)	1.5
	no plane	y = 0 mm	61.49mm (0.11)	66.19mm (0.09)	0.93	59.02mm (0.11)	64.08mm (0.11)	0.78
		y = -130 mm	62.03mm (0.09)	69.7mm (0.07)	1.38	59.21mm (0.09)	67.55mm (0.09)	1.42

Note: Mean responses for each cell are given with the corresponding coefficients of variation (COV). Mean d' (i.e., the mean difference between responses to large vs. small objects, divided by the standard deviation) is reported as an index of mean discrimination performance across participants between each pair of 50-mm and 60-mm objects.
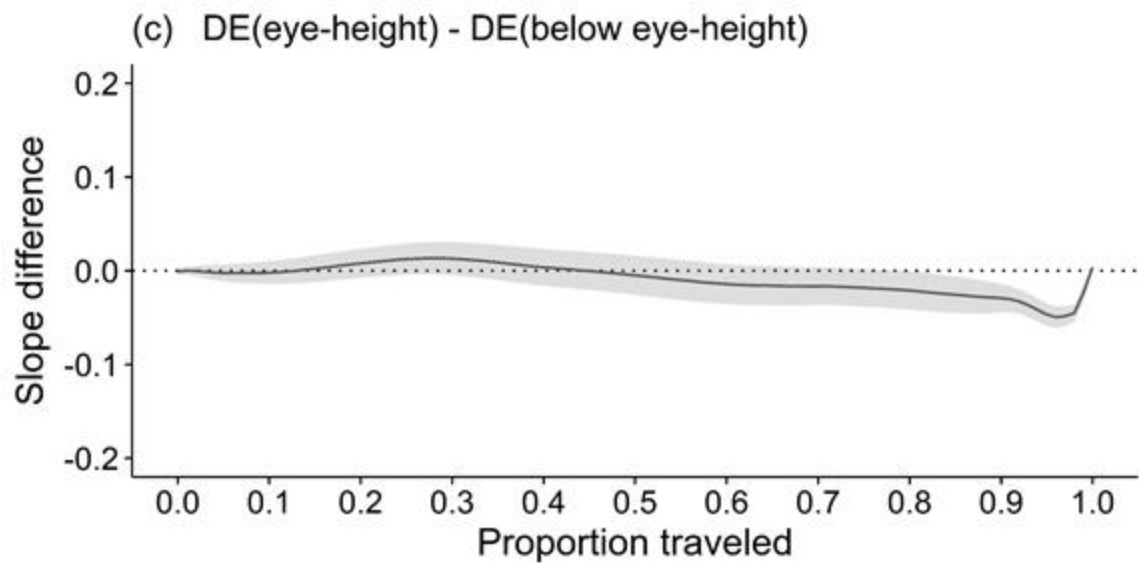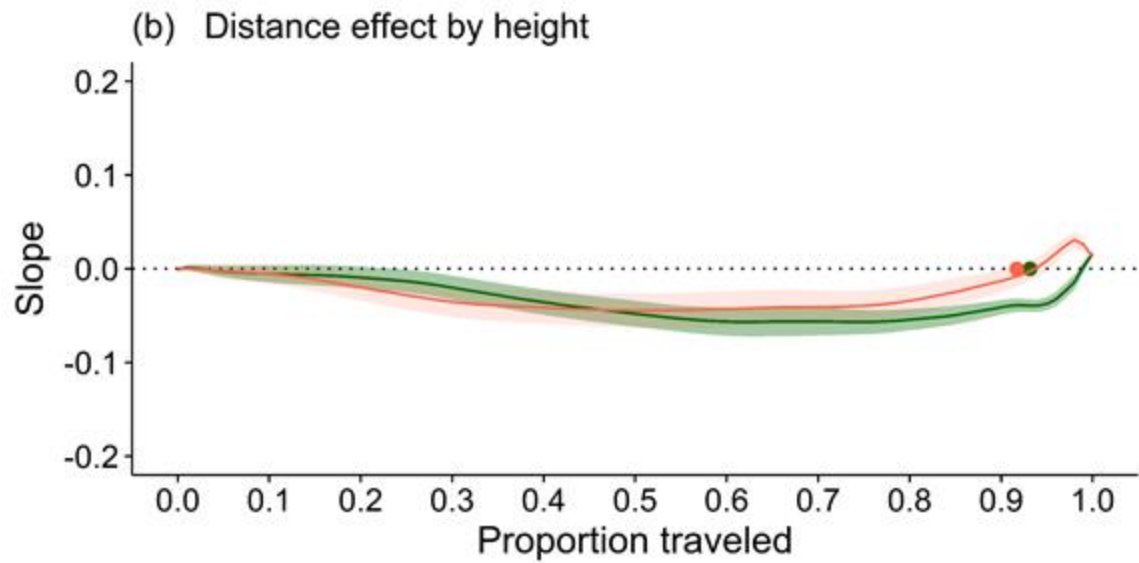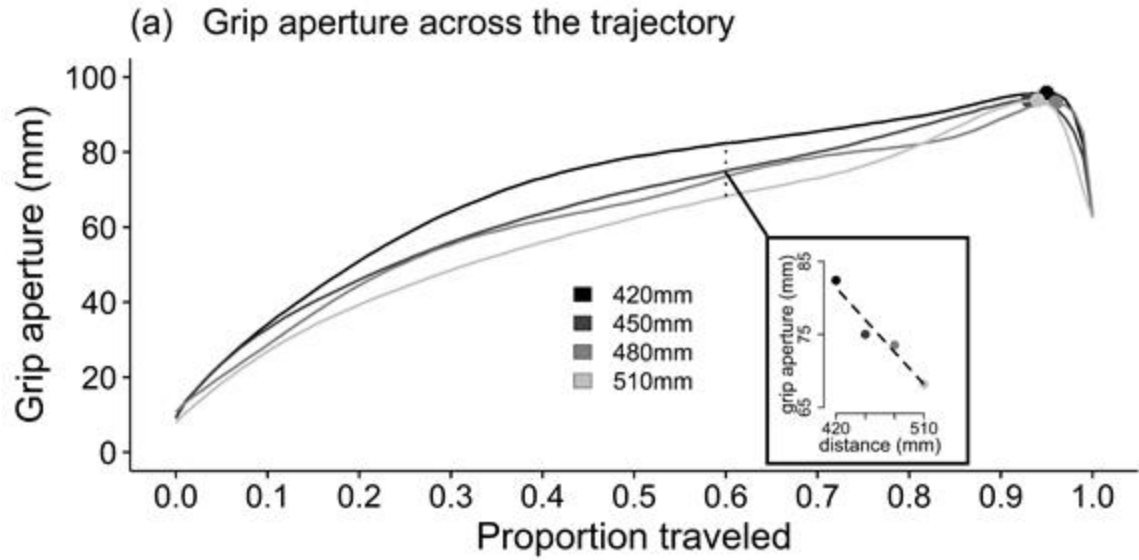
Figures and figure captions
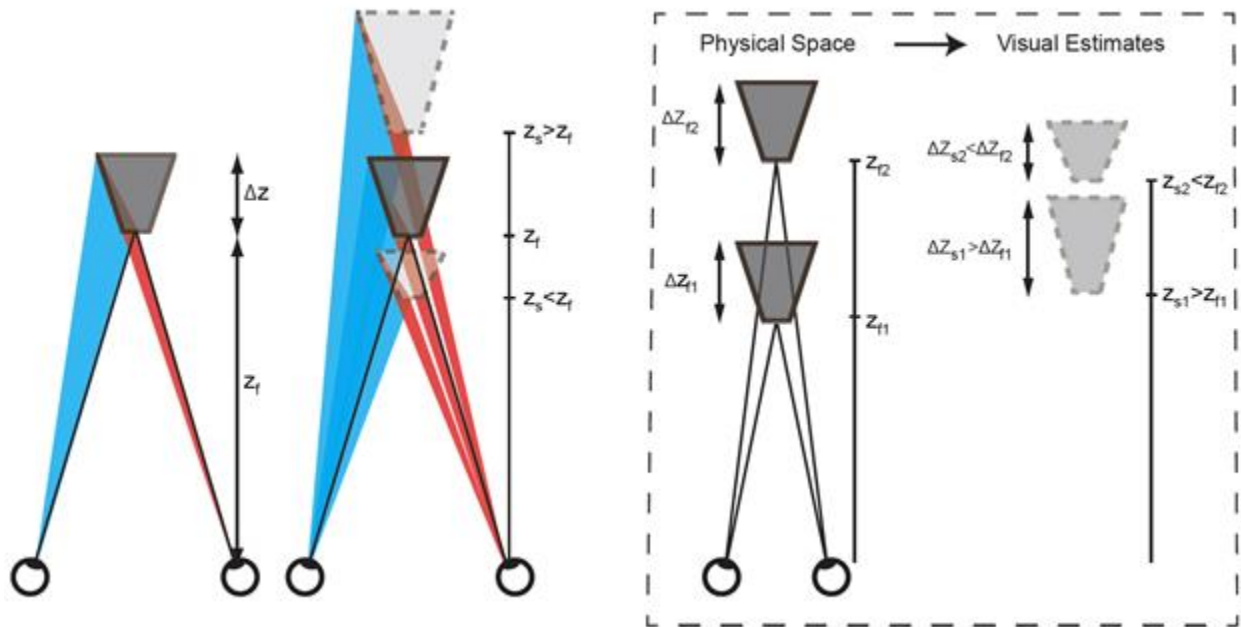
Figure 1. *Left.* The relative disparity between two points, defined as the difference between the angles the two points subtend at each eye location (blue and red), is directly related to the relative depth (Δz) between the points. In order to correctly estimate Δz, the egocentric distance (Z_f) of the object must be accurately gauged to scale relative disparity. *Center*. If the scaling distance (Z_s) is an overestimation of the actual egocentric distance ($Z_s > Z_f$) then the relative depth of the object is overestimated as well ($\Delta Z_s > \Delta Z_f$) and vice versa. *Inset.* Previous research showed that near distances are overestimated ($Z_{s1} > Z_{f1}$), giving rise to an overestimation of the object depth ($\Delta Z_{s1} > \Delta Z_{f1}$). Far distances, on the other hand, are underestimated ($Z_{s2} < Z_{f2}$), causing objects to be perceived as having a smaller depth ($\Delta Z_{s2} < \Delta Z_{f2}$).
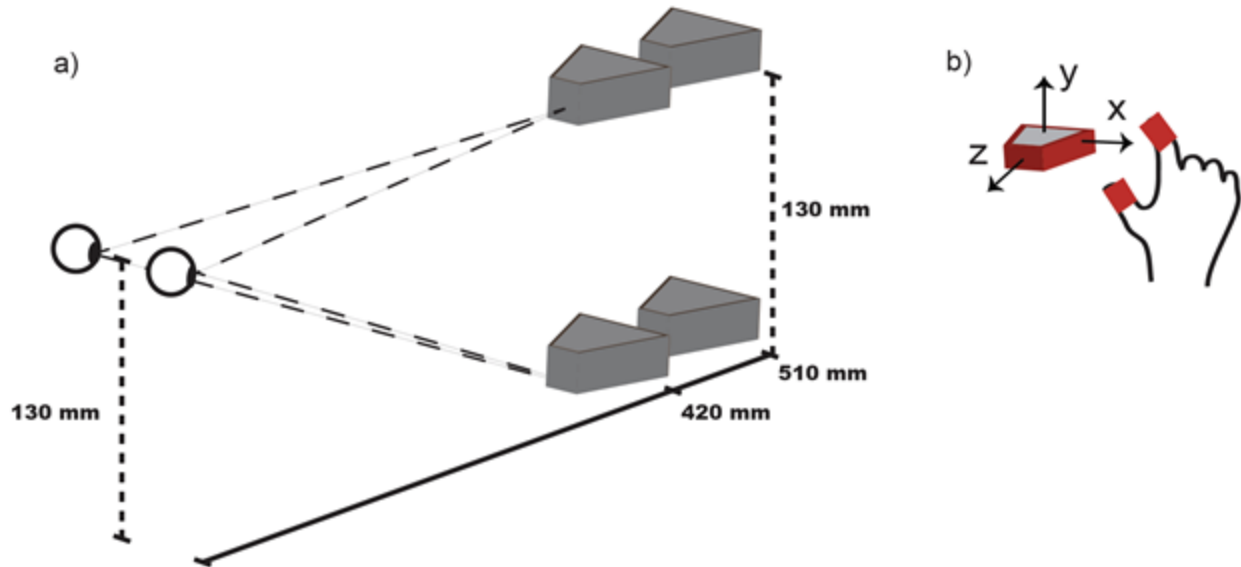
Figure 2. *a)* Schematic representation of the experimental setup. The object was randomly presented at two different heights: eye-height and 130 mm below eye-height. For each condition, four different distances were used within a range of 90 mm (from 420 mm to 510 mm from the eye position). *b)* The object was a trapezoid having the smaller base pointing towards the participant. The virtual object was paired with a real one having the same dimension and located at the same distance. The object was grasped front-to-back with the thumb and index finger virtually rendered by means of virtual cylinders (height: 20 mm, ø 10 mm).

(a) Grip aperture across the trajectory

(b) Distance effect by height


# (c) DE(eye-height) − DE(below eye-height)

Figure 3: **a**: Four trajectories of grip-aperture (i.e., euclidean distance between thumb and index finger) across the space-normalized grasp. Data are presented from four typical (but otherwise arbitrarily chosen) grasps by one participant towards a 60-mm object at eye-height and at each of the four different distances. A third-order Savitzky-Golay filter was used to smooth the kinematic data. The single dot indicates the MGA. The inset shows graphically how the slope displayed in panel b is computed with (again arbitrarily chosen) points at 60% of path travelled: Dots indicate data points, the dashed line indicates the linear function whose slope is plotted in the curves below in panel b. This calculation is done in the same way for all points along the trajectory. **b**: Effect of distance on the grip aperture as function of the hand position (as proportion of path traveled) within the space-normalized reaching trajectory. The effect of distance is represented by the slope of the linear fit of the grip aperture (in mm) as function of egocentric object distance (in mm), calculated at 100 instantaneous positions. The light red line shows the effect of distance in the below-eye-height condition and the dark green in the eye-height condition. Since the slope shows the change in grip aperture in mm per change in egocentric object distance in mm, it is unitless. Dark green and light red bands represent the 95% confidence interval. Dots in corresponding colours represent the moments along the trajectory where the MGA was reached. **c**: The difference between distance-effect slopes for grasps at eye-height and below eye-height along the trajectory. Thus, where the two curves in panel b intersect, this curve is at 0. The shaded area represents 95% confidence interval for the within-participant difference.
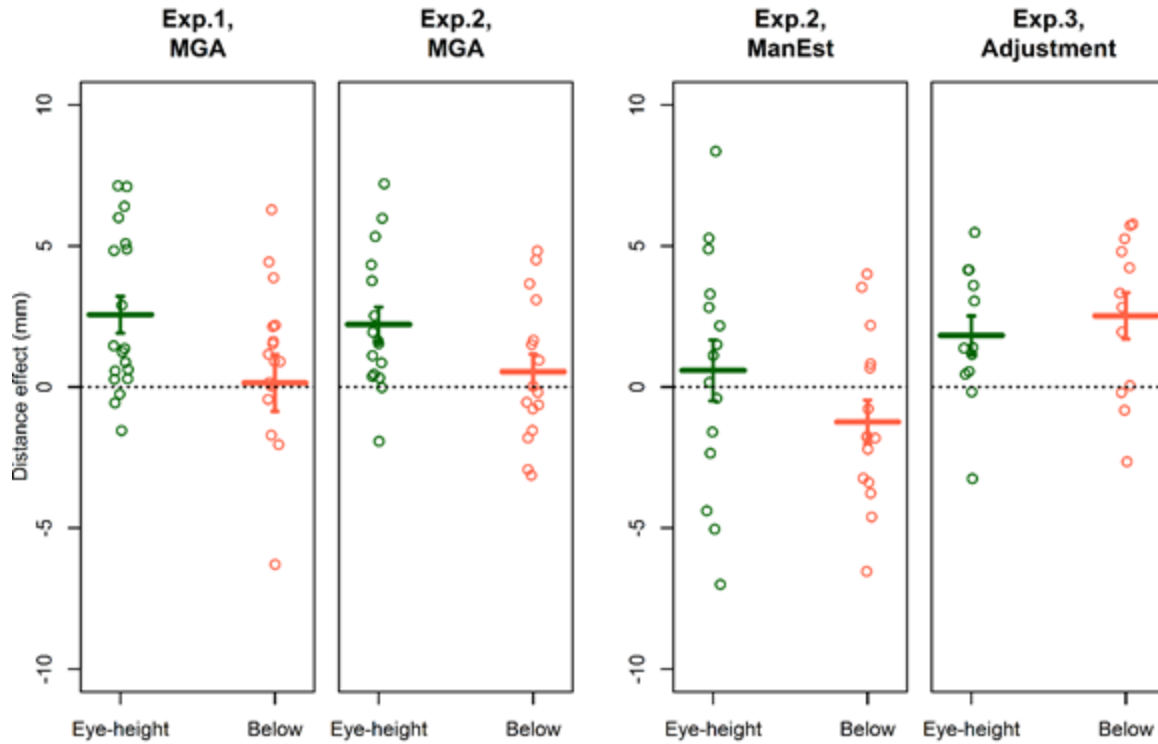
Figure 4: Distance effects in mm, for each task and with each data point indicating one participant. These indicate the mean difference between responses to objects at the furthest distance subtracted from responses at the closest distance. Dark green plots indicate responses to stimuli at eye level, light red shows responses below eye level. Thick horizontal lines and error bars indicate mean and within-participant SEM of distance effects, respectively.

(a) Distance effect by height

(b) DE(eye-height) - DE(below eye-height)

(c) Scaling of grip aperture by object size
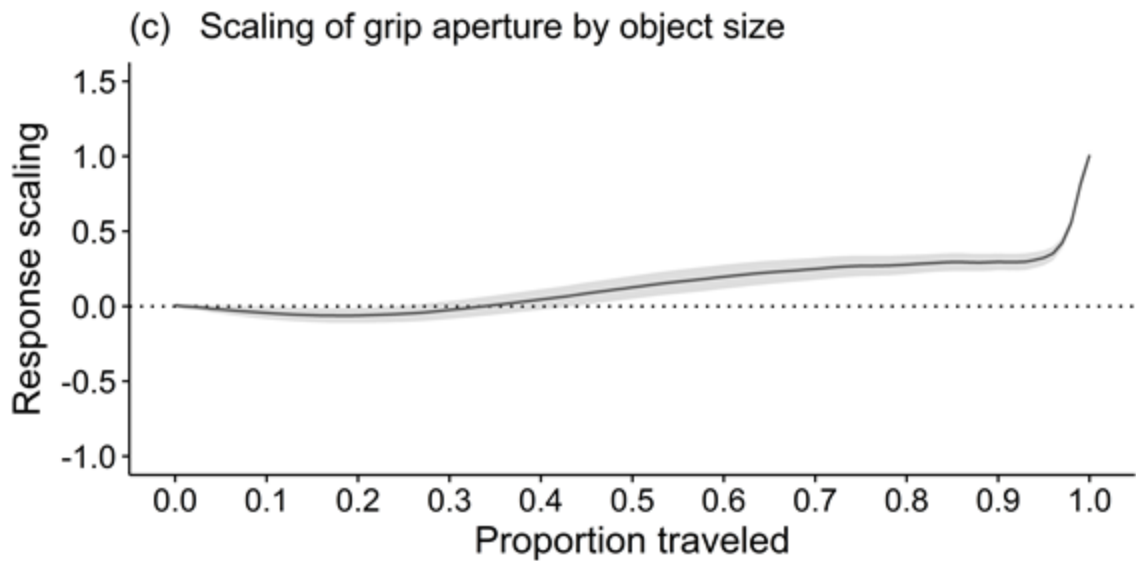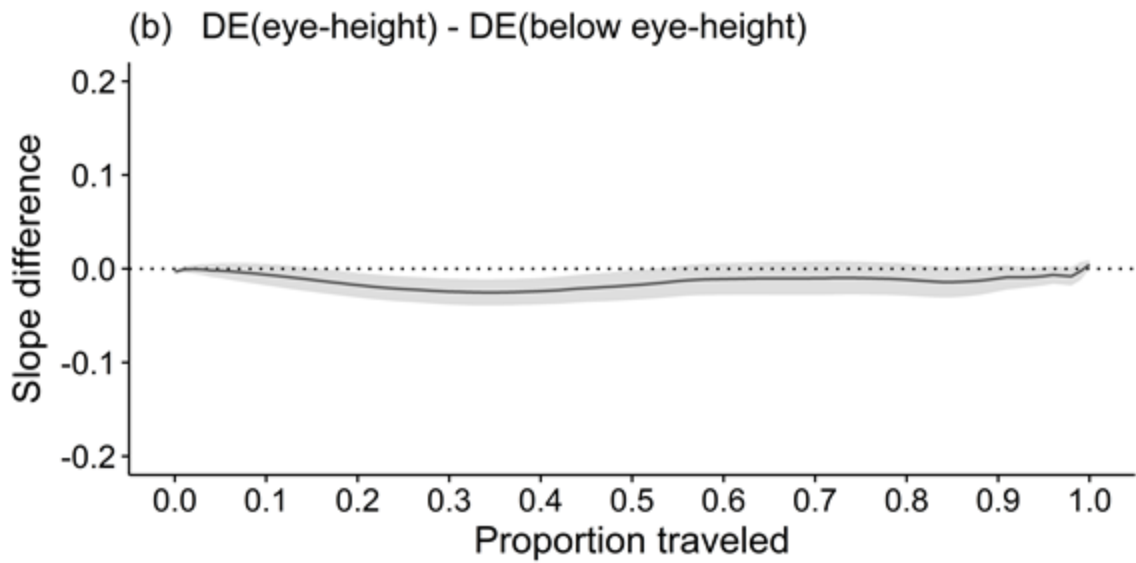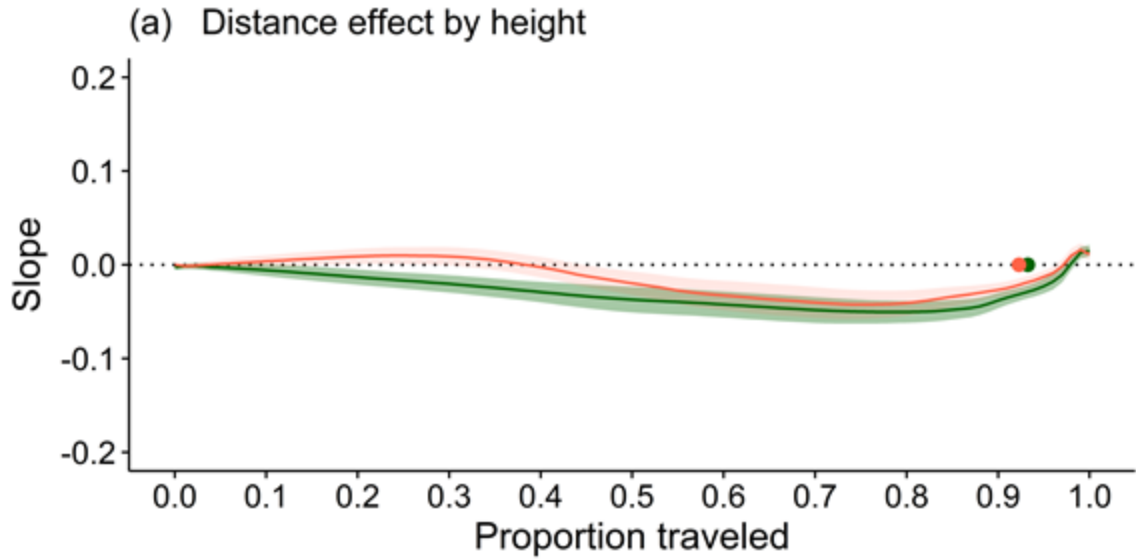
Figure 5: **a**: Distance effect in grasping obtained from experiment 2, expressed like in Figure 3 as the unitless slope of *grip aperture~egocentric object distance*, for each point across space-normalized trajectories. These grasps were performed open-loop and with a planar surface below the objects. Light red indicates targets below eye height, dark green at eye height. Data were collapsed over the two object sizes (50-mm depth and 60-mm depth). Shaded areas depict 95% CIs, dots depict time of MGA. **b**: The difference between distance-effect slopes for grasps at the two heights across the trajectory, plus 95% CI. **c**: Response scaling of grip aperture depending on target size (*Grip aperture ~ object depth*) across the trajectory, collapsed across both heights. Again, the x-axis shows 100 equidistant points between movement onset and movement end, while the y-axis shows a unitless slope. Shaded areas indicate 95% CIs.