Perceiving and grasping in 3D share the same visual mechanisms

by

Carlo Campagnoli

B. S., Università degli Studi di Firenze, 2010

Sc. M., Brown University, 2013

A dissertation submitted in partial fulfillment of the

requirements for the degree of Doctor of Philosophy

in Cognitive, Linguistic, and Psychological Sciences at Brown University

Providence, Rhode Island

May 2017

This dissertation by Carlo Campagnoli is accepted in its present form by the Department of Cognitive, Linguistic and Psychological Sciences as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Fulvio Domini, Advisor

Recommended to the Graduate Council

Date _____ _____

William Warren, Reader

Date _____ _____

Volker Franz, Reader

Approved by the Graduate Council

Date _____ _____

Peter Weber, Dean of the Graduate School

Carlo Campagnoli

490 Angell Street Apt. 102A, Providence 02906 RI
Phone: (+1) 401-500-5745; carlo.campagnoli@gmail.com

Creative researcher, with main interest in how 3D vision and proprioception are integrated when performing accurate actions, such as pointing and grasping. Highly curious, both independent investigator and effective team member, with strong passion for virtual reality and 3D graphics, as well as for programming and data analysis.

Professional Experience

Aug 2011 – Present: **Graduate Student** - *Brown University, Department of Cognitive, Linguistic & Psychological Studies*
I focused my research on 3D information for grasping, and some of my activities included:

- helping building an enhanced virtual reality laboratory, that displays virtual objects and generates consistent tactile feedback through computer-controlled motorized devices
- programming and running psychophysics experiments with such apparatus
- writing an R package for data analysis of grasping trajectories
- designing and animating 3D graphics with software like Blender and Art of Illusion

Sep 2012 – Present: **Teaching Assistant** - *Brown University, CLPS Department*
Along with fulfilling the traditional duties (grading and lecturing), I actively participated to the planning of the classes and I prepared most of the materials for the following courses:

- Quantitative Methods for Psychology – Prof. Leslie Welch
- Visualizing Vision – Prof. Fulvio Domini
- Decision Making – Prof. Steven Sloman

Feb 2010 – Jul 2011: **Research Assistant** – *Istituto Italiano di Tecnologia, Center for Neuroscience and Cognitive Systems*
I managed all the research activities of the laboratory by recruiting participants, preparing and running experiments and handling the administrative paperwork. Thanks to what I learned during this period, I was able to put together my first experiment in all its aspects (designing, building, coding, running and analyzing).

Education

MSc Cognitive Science Brown University (2011-2013)
BSc & MSc Experimental Psychology Università degli Studi di Firenze (2002-2010), distinction

Languages

- Italian (Native)
- English (Professional working proficiency)

Conference Abstracts

Campagnoli, C., & Domini, F. (2016). Conscious perception and grasping rely on a shared depth encoding. *Journal of Vision*, *16*(12), 449-449.

Cesanek, E., Campagnoli, C., Walker, C., & Domini, F. (2015). Online vision of the hand supports accurate grasp performance in illusory contexts. *Journal of vision*, *15*(12), 185-185.

Walker, C., Campagnoli, C., & Domini, F. (2014). Visually guided grasping in depth is systematically inaccurate. *Journal of Vision*, *14*(10), 421-421.

Campagnoli, C., & Domini, F. (2014). Reach-to-grasp actions affect the perceptual scaling of disparity-defined depth. *Journal of Vision*, *14*(10), 407-407.

Lee, A., Campagnoli, C., & Domini, F. (2013). The taller you are, the smaller the world appears: perceptual distortion of binocular space depends on the size of personal body space. *Journal of Vision*, *13*(9), 215-215.

Campagnoli, C., Volcic, R., & Domini, F. (2012). The same object and at least three different grip apertures. *Journal of Vision*, *12*(9), 429-429.

Peer-reviewed Publications

Campagnoli, C., & Domini, F. (submitted). Stereovision for action reflects our perceptual experience of distance and depth. *Journal of Vision*.

Interests

Music has always been a big passion of mine. During my college years I collaborated with fellow musicians as well as with painters and writers in a number of contamination projects, recording and producing several works and live shows around Italy, as well as helping building a recording studio. I am a drummer by nature, and I am able to play other instruments as well.

Voluntary work

I was council member of two cultural associations of my home region (Circolo Culturale Aquaragia, 2002-2005; Associazione Culturale Leggermente, 2009-2011). I helped organizing events such as expositions, concerts, readings, local markets.

ACKNOWLEDGMENTS

This work is dedicated to all persons who shared their time, knowledge, wit and love with me during the years of graduate school.

Table of Contents

Table of Figures

Chapter 1

1.1 Absolute and relative distance estimation for perception and grasping

To pick up an object we must first estimate its shape and position relative to our body. Stereovision can unambiguously determine these properties, given that ocular vergence specifies the egocentric distance of an object while binocular disparities determine its 3D structure (Wheatstone, 1838; Howard and Rogers, 1995). Since this is the most reliable source of depth information within the personal space of an agent, it has been assumed that reach-to-grasp actions critically depend on accurate stereovision (Bradshaw et al., 2004; Loftus et al., 2004; Jackson et al., 1997; Melmoth & Grant, 2006; Mon-Williams and Dijkerman, 1999; Servos & Goodale, 1994; Servos et al., 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 report two phenomena: a) *visuospatial compression*, which makes objects at different distances look closer to each other than they really are; b) *depth underconstancy*, which makes objects close to the observer appear deeper than objects farther away (Foley, 1980; Johnston, 1991; Sousa et al., 2011; Tittle et al., 1995; Volcic et al., 2013).

Although recent investigations on reach-to-grasp actions have revealed a similar pattern of visual distortions (Bozzacchi et al., 2015), it is still unclear whether stereovision for action yields the same 3D information as stereovision for perception. The aim of the first study is to show that the same mechanisms underlying perceptual biases affect the planning of goal-directed actions.

1.2 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. In everyday viewing conditions, many signals can inform the observer about distance. Within the peripersonal space, where things are reached and manipulated, there are two important binocular sources of information for distance. The first, which is extraretinal, is the vergence angle formed by the eyes when fixating at a specific location (Foley, 1980; Cormack, 1984). 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 (Mayhew & Longuet-Higgins, 1982; Bishop, 1989; Rogers & Bradshaw, 1995; Brenner et al., 2001; Bradshaw et al., 1996). In this investigation, we focused on the role of vergence and on the effects that distance-from-vergence has on the scaling of binocular disparities. When vergence is the primary 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. This phenomenon has been attributed to errors in the translation of the oculomotor signals into angles, particularly at great distance (*specific distance tendency*; Gogel, 1969) and in absence of other relevant cues (Baird & Biersdorf, 1967; Gilinsky, 1951; Brenner & van Damme, 1998). In the seminal work of Foley (1980) it is hypothesized that perceived distance is systematically biased. According to Foley, “Perceived distance of near targets exceeds physical distance; perceived distance of far targets is less than physical distance”. In other words, objects closer than a specific distance z_{fA} are seen more distant than they are and those farther than z_{fA} are seen closer than they are. The perceptual space of egocentric distances is therefore compressed, since, as indicated in the diagram of Figure 1.1, the spacing between egocentric distances in the physical domain maps onto a smaller spacing in the perceptual domain.

If we assume that for a small range of distances (Foley 1977) this mapping is linear, then a set of evenly spaced egocentric distances is still evenly spaced in its perceptual representation. In this case, the relation between physical (z_f) and perceptual (z'_f) 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_{fA}, $z'_{fA} = z_{fA}$) (Fig. 1.1). The slope of the egocentric function, which we will refer to as *compression factor*, decreases as the magnitude of visuospatial compression increases.

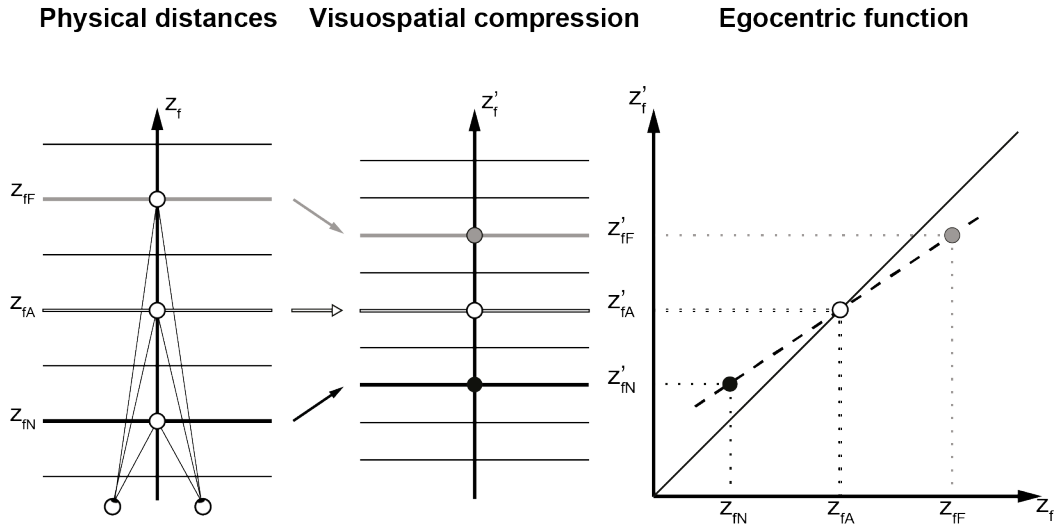

Figure 1.1. An observer fixates at three distances. The *egocentric function* defines the relationship between physical (z_f) and estimated (z'_f) distances. If the range of estimated distances is smaller than the physical range then the slope of this function is smaller than 1, resulting in *visuospatial compression*.

The compression of visual space could affect depth judgments in turn, causing *depth underconstancy*. Consider an object of depth Δz viewed at a distance z_f from the observer (Fig. 1.2A). Two points belonging to the near and far surfaces of the object subtend two slightly different angular extents at each eye's nodal point (α_L and α_R in the figure). Relative disparity, obtained by subtracting these angles, is the projection of an infinite family of objects, whose depth increases with distance. Therefore, depth reconstruction requires an estimate of z_f, commonly referred to as *scaling distance* (z_s), in order to scale binocular disparities (Rogers & Bradshaw, 1995; Glennerster et al., 1996; Brenner & van Damme, 1999).

If the viewing distance is accurately estimated via sensed ocular vergence ($z_s = z_f$), then the true depth of the object can be estimated as well (Fig. 1.2A, left). Otherwise, an underestimation of distance ($z_s < z_f$) leads to an underestimation of the object depth (Fig. 1.2A, center) and vice versa ($z_s > z_f$; Fig. 1.2A, right) (Gregory, 1963; Howard and Rogers, 1995).

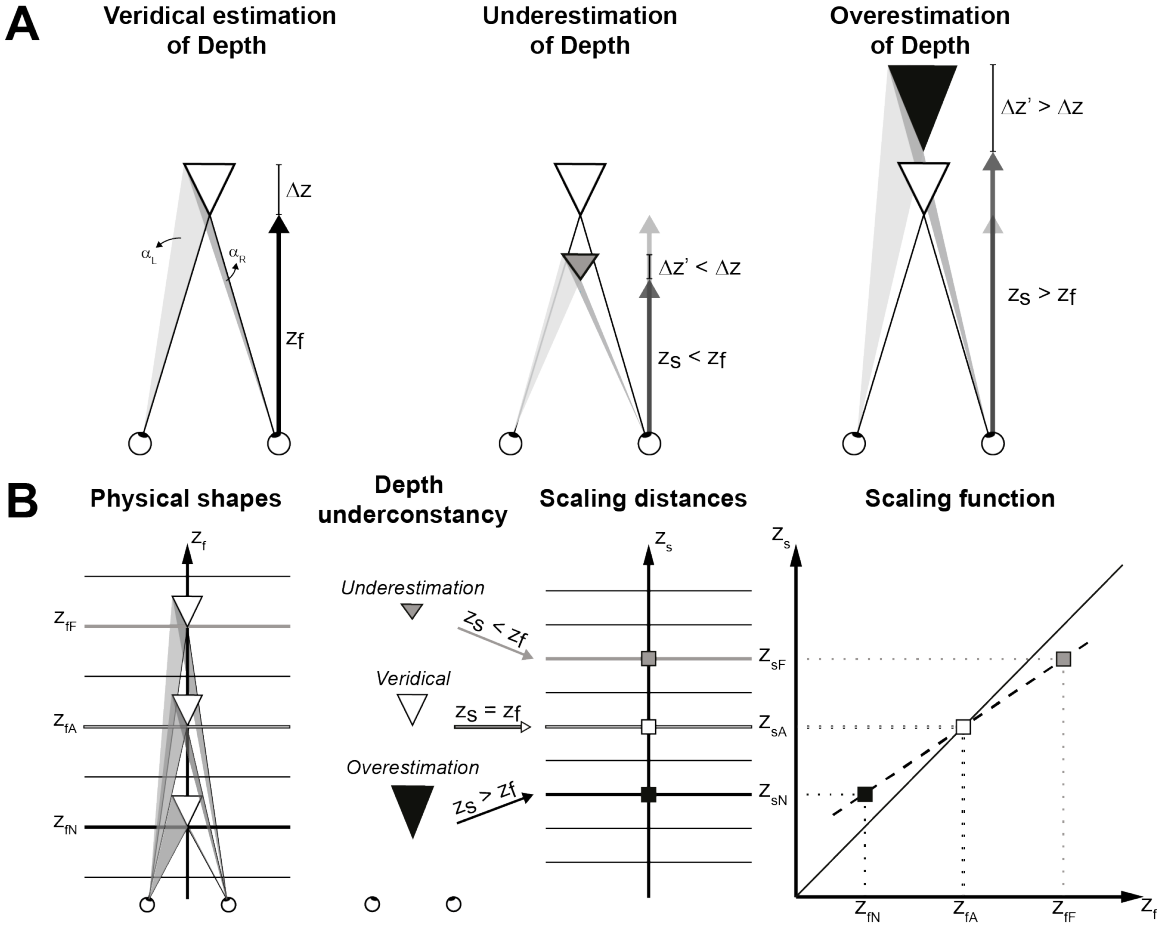

Figure 1.2. (A) The relative disparity between two points (the difference between the angles α_L and α_R) uniquely specifies their depth separation Δz once the viewing distance z_f is known (left). If retinal disparities are scaled by a distance z_s (dark arrow) smaller than z_f (gray arrow) the same disparity gives rise to depth underestimation ($\Delta z' < \Delta z$, center) and vice versa (right). (B) The scaling function defines the relationship between physical (z_f) and scaling (z_s) distances. If the scaling distance of near objects is overestimated ($z_{sN} > z_{fN}$) and that of far objects is underestimated ($z_{sF} < z_{fF}$) the scaling function has a slope smaller than 1. Therefore, *visuospatial compression* causes near objects (black) to appear deeper than far objects (gray). Distance and depth estimation should be accurate at a theoretical distance z_{fA} (see also Appendix).

As a consequence, objects closer than the point of veridicality z_{fA} should appear deeper than they are, since visuospatial compression leads to the overestimation of their egocentric distance. The

opposite should happen for objects farther than z_{fA} (Fig. 1.2B). The relationship between the scaling distance, which can be derived directly from depth estimates, and physical distance will be termed *scaling function*.

Finally, it is important to note that depth underconstancy and visuospatial compression are, in principle, geometrically incompatible, because the former implies distortions of shape that cannot be predicted by a homogeneous stretching of space produced by the latter. In fact, if the perceived egocentric space is linearly compressed with respect to the physical world, then an object presented at any distance within such space should always look flatter than it is, and by a constant amount. An explanation of this paradox is that the allocentric properties of an object (i.e., its depth) can be analyzed independently from the egocentric properties (i.e., its absolute distance), effectively giving rise to different representations of the same 3D shape.

This mechanism is especially suitable for action, since different movements often require focusing on specific properties of an object at the expense of others. In this regard, several studies have shown that observers can indeed change the weighting of allocentric and egocentric cues selectively, depending on the task that they are performing (Bingham et al., 2004; Bingham et al., 2000; Bingham, 2005), or even whenever the goal of a certain action changes (Neely et al., 2008; Smeets et al., 2002). These results suggest that egocentric and allocentric tasks (i.e., judgment of distance and size) performed on the same shape can be based on different spatial mappings. Reaching-to-grasp is an interesting case, because it depends upon both types of information at the same time. An open question is whether multiple representations of an object can be accessed simultaneously, even though they are inconsistent with each other.

CHAPTER 2

2.1 Depth cues combination for perception and grasping

A critical goal of the visuomotor system is to recover the best estimate of an object's shape given the available depth cues to ensure that, during a grasp, the grip aperture can be accurately adjusted to fit the object's physical structure. The most important source of information to this end is represented by stereovision, and in particular by the relative orientation of the eyes at fixation, or vergence angle, and by the pattern of horizontal disparities. These two signals combined can uniquely specify the depth map of a 3D structure (Wheatstone, 1838; Howard and Rogers, 1995).

Consistent with this theoretical framework, several studies have shown that grasping relies primarily on stereovision. The reaching component is mostly affected by changes in the vergence angle (Mon-Williams & Dijkerman, 1999; Loftus et al., 2004), while the scaling of the grip aperture is mainly dependent on binocular disparities (Watt & Bradshaw, 2000; Melmoth & Grant, 2006; Melmoth et al., 2007). Since grasping movements are normally highly smooth and efficient, it is usually assumed that stereovision produces unbiased estimates of distance and shape during the planning and execution of an action (Loomis et al., 1992; Goodale, 2008; Bradshaw et al., 2004; Loftus et al., 2004; Jackson et al., 1997; Servos & Goodale, 1994; Servos et al., 1992).

The latter assumption is in apparent contradiction with evidence from the literature on depth perception, which shows that the visual space is subject to systematic distortions. In previous studies, we found evidence that these biases actually do impact the execution of a grasp as well, although they are rapidly corrected through online visual and tactile feedback from the hand (Campagnoli et al., 2012; study 1 of this thesis).

Aside from online control, grasping movements might also be highly efficient because, under normal viewing conditions, the visuomotor system can take advantage of other depth cues than just binocular disparities, like texture and motion gradients. These monocular signals provide additional 3D information that can be combined with stereo signals to potentially improve the accuracy of shape reconstruction. If this hypothesis is true, then consequently a grasp directed at an object specified by multiple depth cues should be characterized by smaller or no biases, and consequently by less online corrections, than a grasp directed towards an object defined only by stereo cues.

2.2 Shape constancy

If binocular information is the only depth source available, judgments of depth and width indirectly involve the estimation of egocentric distance. Although binocular disparities and retinal size covary with depth and width respectively, the signals themselves are inherently ambiguous. The same retinal size can be projected by a small object near the observer or by a large object far from the observer (Fig. 2.1C, left panel). Similarly, although horizontal disparities directly specify the depth order between the points on a surface, the same disparity can be generated by a flat object near the observer, or by a very deep object far away from them (Fig. 2.1C, right panel). Viewers face the problem of somehow constraining a set of infinite possible shapes to only one plausible candidate.

In this study, we consider the case in which stereovision is the only source to 3D shape, and we restrict the discussion to just the peripersonal space, where things are reached and manipulated. In this hypothetical situation, the main signal that directly specifies absolute distance is represented by the vergence angle (assuming a negligible role of vertical disparities, which is the case when an object subtending a relatively small visual angle is viewed in isolation).

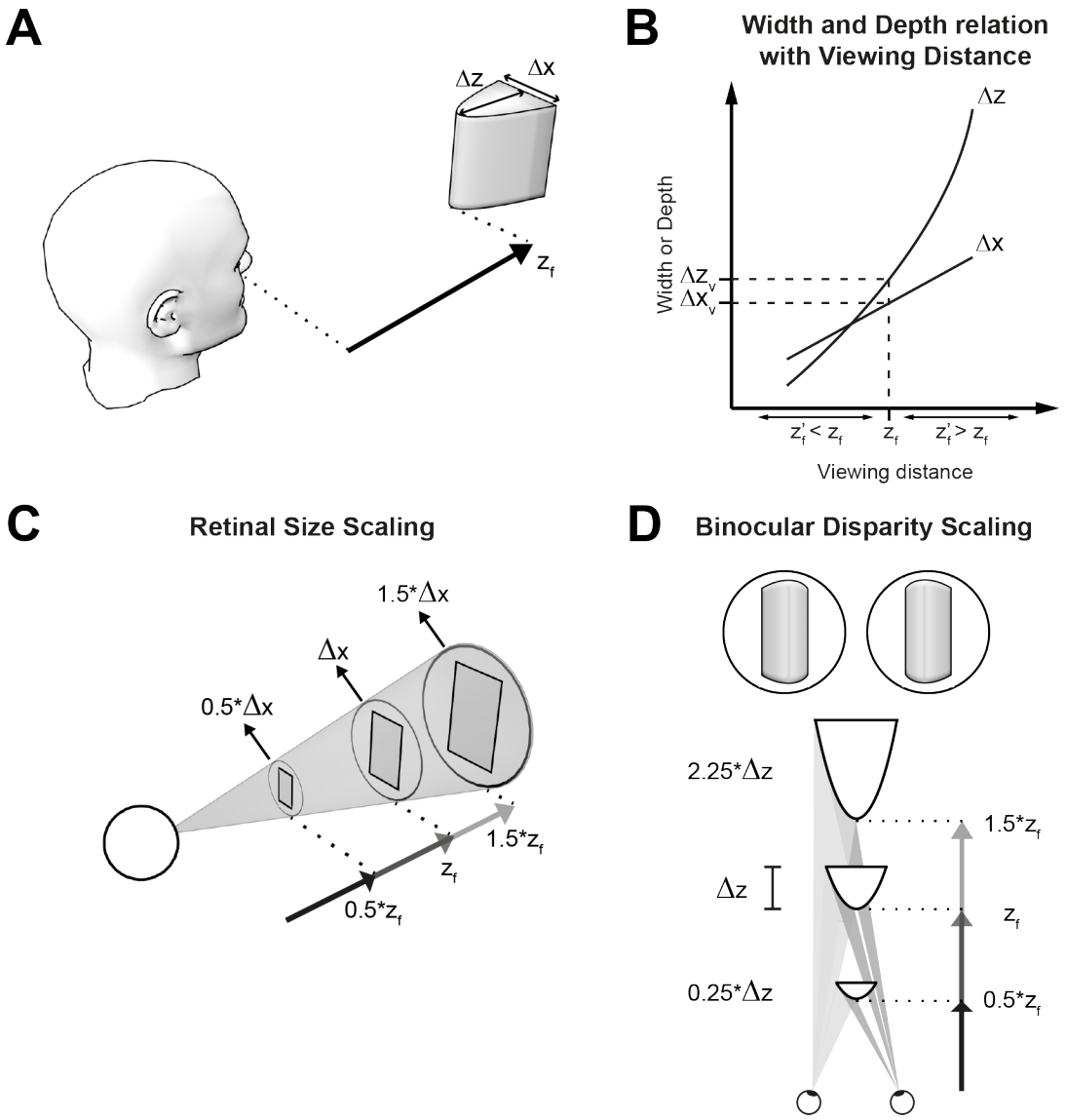

Figure 2.1. Shape estimation from binocular disparity and retinal size information. (A) The stimulus used in the study was a parabolic cylinder whose shape was defined by the ratio between its depth (Δz) and its width (Δx). (B) Depth-from-disparity and width-from-retinal-size (in ordinate – their veridical values are represented by Δz_v and Δx_v respectively) have a different geometric relationship with the viewing distance (z_f in abscissa): Δx scales linearly with z_f, Δz scales quadratically with z_f. Overestimation ($z'_f > z_f$) and underestimation ($z'_f < z_f$) of the true viewing distance have different effects on the estimation of width and depth. (C) The same retinal projection can be generated by a distant and large object or by a close and small object, following a linear mapping between distance and size (width and height dimensions). (D) The same disparity can be generated by a distant and deep object or by a close and flat object, following a quadratic mapping between z_f and Δz.

It is well known that egocentric distance is poorly estimated via vergence signal (Baird & Biersdorf, 1967; Gilinsky, 1951; Brenner & van Damme, 1998). In fact, several psychophysical experiments have documented systematic biases, and large interindividual variability in the

responses (Foley, 1980; Mon-Williams & Tresilian, 1999; Servos, 2000; Wallach & Zuckerman, 1963). Most commonly, the distance between a far and a near object tends to be underestimated, a phenomenon that we refer to as *visuospatial compression*: objects near the body appear more distant than they are, while objects far from the body appear closer than they are.

Interestingly, judgments of shape often show systematic biases and great intersubject variability too. One phenomenon in particular is *depth underconstancy*: near objects usually appear deeper than far objects (Johnston, 1991; Brenner & van Damme, 1999). This bias is consistent with the hypothesis that visuospatial compression affects the scaling of binocular disparities. As a result, near the body both distance and depth are overestimated, whereas far from the body they are both underestimated (Fig. 5.1, "Shape from Stereo" panel). While the link between visuospatial compression and depth underconstancy has been theorized before (Johnston, 1991), we recently found remarkable evidence in its favor, indicating that the interindividual variability of both distance and depth judgments is also linked to visuospatial compression too: subjects experiencing a large compression of the egocentric space show pronounced depth underconstancy and vice versa, during both perceptual and grasping tasks.

The fact that participants report changes in the estimated depth along the line of sight raises the question as to whether shape is overall constant in return, since size estimation depends on distance as well (van Damme & Brenner, 1997). Results from studies using, for example, the apparently circular task suggest the opposite (in the task, participants adjust the depth of a cylinder until it appears perfectly circular; i.e., Johnston et al., 1994). One explanation is that retinal size decreases linearly with distance, while binocular disparities decrease with the square of the distance. Consequently, visuospatial compression has a smaller impact on size estimation than on depth estimation, particularly within a small spatial region. In agreement with this hypothesis, we also found that subjects showed a negligible bias in the predicted direction in a width-matching task performed at two distances that were 18 cm apart.

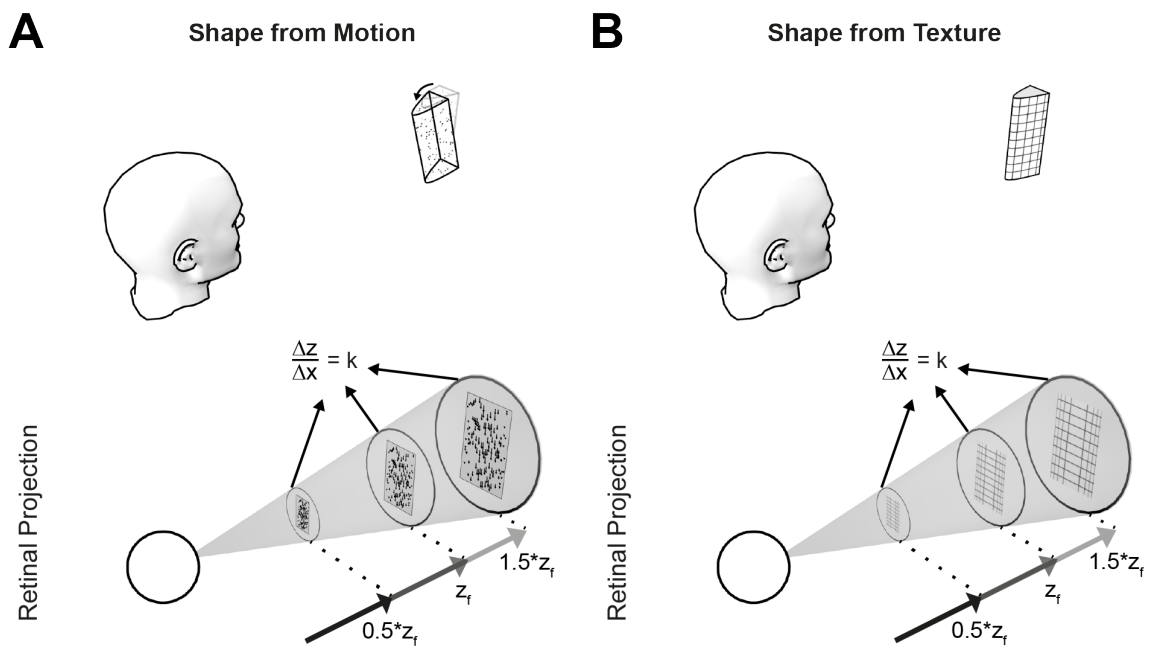

Figure 2.2. Shape constancy from motion and texture gradients. (A) The same gradient of retinal velocities can be generated by a small object rotating at a near distance, or by a large object rotating at the same angular speed at a farther distance. In all cases, the aspect ratio specified by motion is constant ($\Delta z/\Delta x = k$). (B) The same principle applies to the texture gradient.

One way in which shape recovery could improve is through the integration of additional monocular sources. Texture and motion gradients seem especially appropriate for this goal, since they possess two distinctive features. First and foremost, like binocular disparities, they directly specify the entire 3D layout of a shape. Secondly, as can be seen in figure 2.2, both gradients do not change with distance, unlike retinal size and disparities (Vishwanath, 2014).

All these characteristics make motion and texture signals relevant for studying both perception and grasping in this context, since in principle they both offer additional 3D shape information that is consistent with stereovision. At the same time, both have the potential to counteract or at the very least reduce the biases caused by visuospatial compression on the scaling of binocular disparities.

In conclusion, the structure of an object defined by stereo and motion cues or by stereo and texture cues should be perceived as more constant across distances and grasped accordingly.

CHAPTER 3

3.1 Are perception and grasping affected by the use of virtual reality?

Interacting with the world requires that we process a rich visual environment. Even mundane tasks demand complex computations; simply picking up a pencil involves that we extend our arms and close our fingers with enough precision to ensure a successful grip. In order to succeed at this, the visuomotor system fundamentally requires three-dimensional information about the scene.

However, it is known that depth estimates are affected by underconstancy: objects are usually reported to appear deeper when seen near the body than far from the body (Johnston, 1991; Sousa et al., 2011; Glennerster et al., 1998; Volcic et al., 2013). In the first study of this thesis, we explained this behavior as the result of the systematic misinterpretation of the viewing distance, a phenomenon we called *visuospatial compression* (egocentric distance is overestimated near the body and underestimated far from the body), which in turn affects the scaling of depth cues, in particular binocular disparity (Foley, 1980; Rogers & Bradshaw, 1995; Glennerster et al., 1996; Brenner & van Damme, 1999).

Critically, visuospatial compression seems to persist even when multiple cues specify the same 3D shape. In the second study of this thesis, we used stimuli that were defined by a combination of disparity plus a rich monocular gradient to depth (motion or texture). Despite the more abundant quantity of 3D information, participants kept showing underconstancy.

What was most interesting in both studies was that the same bias affected the reaching trajectories of grasping movements: the in-flight grip aperture (*GA*) during the last phase of the grasp was systematically bigger when aiming towards a close object than towards a more distant one,

regardless of whether the stimulus was defined by just stereovision, stereo and motion or stereo and texture.

We concluded that these three cues do not yield accurate 3D vision for neither perception nor action, and for this reason, the goal of the visuomotor system is not to build perfect Euclidean representations of objects from those signals. Instead, we proposed that stereo and monocular gradients are combined in order to maximize the discriminability of depth intervals, at the expense of accuracy. In general terms, the rationale held that the probability of detecting that two points, A and B, are separated in depth increases as their relative distance is specified by more and more cues. As a consequence, a stereomotion (or stereotexture) display should produce a greater impression of depth than a stereo-only display, because the observer correctly detects that the two points are at different depths more reliably in the former case than in the latter, even if the actual separation between A and B is constant. A critical point of this theory is that it does not constrain depth estimation in terms of accuracy, allowing for overestimation (Domini et al., 2006; Di Luca et al., 2010; Domini et al., 2011).

Interestingly, the data showed that stimuli were in fact seen as deeper (not more veridical) when specified by two cues-to-depth than when specified by just stereovision. Moreover, these changes did not improve shape constancy: even when multiple cues were available, objects still appeared flatter when seen far from the body than close to the body. Hence, we argued that the success of grasping movements is strictly dependent on an efficient online control system.

There is yet another possible explanation to these results, which assumes that during our experiments the accuracy of shape estimates bias may have been compromised by the use of virtual reality. VR inevitably entails a certain amount of flatness (or frontoparallel) cues, which cause a conflict of information: even when disparity, motion and texture consistently specify the veridical depth of an object, residual cues such as blur and accommodation indicate a depth of zero (that of the monitor), leading to underestimation (Buckley & Frisby, 1993; Frisby et al., 1995; Willemsen et al., 2008; Hoffman et al., 2008; Watt et al., 2005). Consequently, the more

depth signals, the weaker the influence of flatness information, and thus the deeper the resulting percept. Frontoparallel cues may also explain underconstancy: since the reliability of binocular disparity decays with distance, the flattening effect is greater far from the body than close to the body (Landy et al., 1995).

This alternative interpretation has a fundamental impact for the understanding of grasping actions: if VR is truly the reason why shapes are seen as more frontoparallel, then prehensile movements are usually successful because they typically take place in natural contexts, where there is no conflict of depth information at all. From this hypothesis it follows that the visuomotor system normally supports accurate 3D vision, in contrast to our previous conclusions. In partial agreement with this view, studies have shown that the kinematics of actions performed in VR are different compared to when the same movements are executed in a natural setting (Cuijpers et al., 2008; Magdalon et al., 2011; Mon-Williams & Bingham, 2008; Hibbard & Bradshaw, 2003).

For these reasons, we decided to test our previous findings in a completely natural environment, using real objects that were seen at one fixed distance. We chose a viewing distance of 42 cm because we previously found that judgments of depth are usually accurate around that position, even in VR (Volcic et al., 2013; first study of this thesis). This apparatus eliminated flatness cues altogether and avoided the interference of depth underconstancy. Through a within-subjects design, we measured both the perceived front-to-back extent of cardboard-made prisms, and the in-flight *GA* of grasping movements aimed at the same objects. We allowed participants to see their own hand and limb during the entire execution of the grasp, as well as to touch the objects at movement completion, effectively reproducing all circumstances typical of natural actions. We manipulated the amount of stereo and motion information specifying the shape of the stimuli, and studied how perceptual judgments and motor actions changed as 3D information varied from trial to trial.

CHAPTER 4

Study 1: Depth and distance estimation in perception and action

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

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

The main goal of this study was to test two hypotheses about the visual space: (1) It is characterized by visuospatial compression of egocentric distances, which is in turn related to depth underconstancy; (2) it varies considerably across individuals, but it is the same for perception and reach-to-grasp actions. If the first assumption is true then the egocentric and scaling functions must be the same. If the second assumption is correct then this function characterizes a unique visual representation for perception and action, which is specific for each individual.

These hypotheses were tested by asking participants to judge the depth of a three-rod configuration and to grasp it (see figure 4.2). In order to avoid effects of learning, which may reduce or eliminate the very biases we are interested in measuring, subjects grasped virtual objects without feeling their physical presence at the end of the grasp. 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 direct 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 movement completion, this quantity provides a direct assessment 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.

According to the first hypothesis, the perceived distance of an object should be affected by visuospatial compression, while the perceived depth of the same object should be affected by depth underconstancy. As a result, objects that are near to the body should appear farther away and deeper than they are, while objects that are far from the body should be perceived closer and shallower than they are. Hence, a direct test of the first hypothesis is only possible in the grasping task, where *FHP* and *FGA* directly measure estimated distance and depth respectively. Let's say for instance, that two subjects, A and B, were looking at the same visual scene. Subject A experiences a largely compressed visual space, whilst subject B shows almost veridical mapping between perceived and physical space. When subject A is asked to grasp an object at two distances z_{fN} and z_{fF}, 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 ($z_{fF} - z_{fN}$). 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. The shaping of the grip aperture should adhere to the same principles: subject A should grasp the near object as much deeper than the far object, while subject B grasps them as almost identical. The *FGA* is a measure of the depth estimate and therefore it determines, through inverse projection, the scaling function (see Fig 1.2 and Appendix). As a result, the first hypothesis predicts that A should show a compression factor of .5 for the scaling function as well as for the egocentric function, while B should show a compression factor close to 1 for both functions.

According to the second hypothesis, the same distortions of the visual space should affect perception as much as action. In a Manual Size Estimation we can only measure the estimated depth of an object, not its perceived position, meaning that with this task we can only parameterize the scaling function, but not the egocentric function. Nevertheless, the second hypothesis predicts that grasp and perception are affected by the same biases in the recovery of distance. As a consequence, depth estimation during an *MSE* task should lead to similar results to those previously described for the *FGA* of the grasp: subject A should show a much larger *MSE* for the near object than for the far one (scaling function with compression factor of .5), while B should perceive the two objects as almost identical (scaling function with compression factor of almost 1). The egocentric function as predicted by the first hypothesis (that is, derived from the *FHP* of the grasp) should therefore match the scaling function derived through inverse geometry from the *MSE*.

In summary, the biases in distance estimation as revealed by the hand position in a grasp should be, for each individual, predictive of the biases in depth estimation during both grasping (first hypothesis) and perception (second hypothesis).

4.2 Overview of the experiments

We conducted four experiments to investigate whether stereovision for action reflects our perceptual experience of distance and depth. In the first experiment we compared a perceptual task (manually estimate the 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, to ensure that the movement was based purely on planning information, without in-flight corrections due to online feedback. We found three important results: i) the final position of the hand in the grasp was consistent with visuospatial compression (near and far objects were grasped as if they were closer to each other than in the real world); ii) the final grip aperture in

both tasks was consistent with depth underconstancy (the near object elicited a larger grip than the far object); iii) although the magnitude of the biases varied from person to person, everyone was remarkably consistent across conditions: if one showed a mild visuospatial compression when grasping, they also showed small depth underconstancy in both tasks, and vice versa. We hypothesized that visuospatial compression directly influenced the object's perceived egocentric position, and that in turn this distortion affected the estimation of the object's depth. We studied a model that supported this hypothesis and found that the errors of the final hand position of the grasp indeed correlated with the biases of the manual depth estimation, on a subject-by-subject level.

If, as indicated by the results of the first experiment, perception and action stem from the same visual representation of space, then two physically different objects, which appear perceptually identical when located at different distances, should be grasped with the same grip aperture. In the second experiment, each participant adjusted the physical depth of two objects presented at two different distances until they looked identical. As predicted, when every person reached-to-grasp their own pair of perceptually matched objects, they showed identical grip apertures.

In the third experiment, we tested whether the biases found in the previous grasping tasks were due to the lack of visual and haptic feedback from the hand. However, we found clear evidence of depth underconstancy even when feedback was present.

Visuospatial compression and depth underconstancy suggest distortions of both egocentric and allocentric maps. Interestingly, it can be shown that these maps are geometrically incompatible, even though they both originate from the same systematic biases in the estimates of objects' distances. 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. Indeed this is what we found in the fourth experiment, where participants either reached separately to an object's front and back surfaces or grasped the object front-to-back. Remarkably, although the object's depths resulting from the two tasks were different from one another, we

successfully applied the model used in the first experiment to show that they are both predicted by the same distortion of the egocentric space.

4.3 General Methods

Participants

Sixty-four students (twenty-six males, thirty-eight females) 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/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/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 (Fig. 4.1). Subjects sat with their head on a chin rest installed along one of the short sides of a rectangular table, facing a semi-transparent mirror which was slanted 45 degrees 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 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 the y axis. This latter system was used for different purposes depending on the experiments (see below).

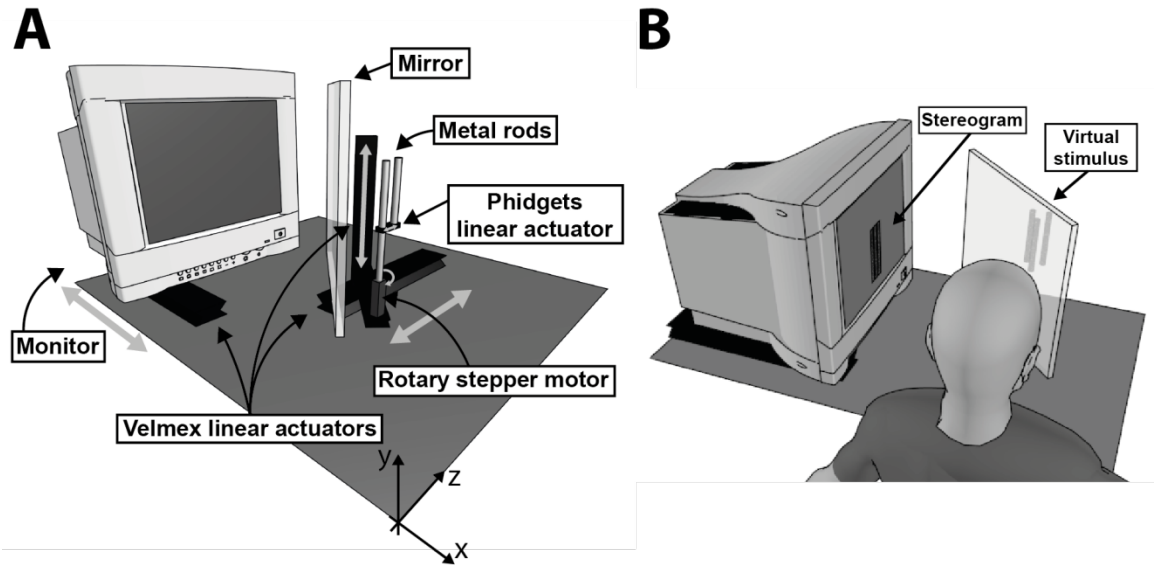

Figure 4.1. (A) Schematic of the setup used in all four experiments. Monitor and physical objects (when used) were moved by a series of linear actuators, gray arrows indicate the direction of motion provided by each motor. (B) In all experiments, the stereogram of the stimulus rendered on the monitor 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.

In the third experiment, a combination of small linear actuators attached to a stepper motor (Phidgets Inc.) was also mounted on the platform. Using shutter glasses (liquid-crystal FE-1 goggles by Cambridge Research Systems at *CNCS*; NVIDIA 3D Vision® 2 wireless glasses 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.). We used an Optotrak 3020 Certus motion capture system and small infrared emitters (Northern Digital) to record index and thumb trajectories. A calibration procedure allowed us to track the 3D position

of the index and thumb's finger pad, 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 the other cues (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 experiment 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 was adopted to make sure that the vergence-accommodation conflict was negligible (Watt et al., 2005; Frisby et al., 1995; Buckley and Frisby, 1993).

Data Analysis

Raw movement kinematics of the fingertips of index and thumb were processed and analyzed offline using R (R Core Team, 2016) with custom functions. From the raw 3D positional data of index and thumb we calculated the raw profiles of grip aperture (*GA* — the Euclidean distance between the fingertips) and hand position (*HP* — the middle point between the fingertips). The raw data were then smoothed using a cubic spline with a small smoothing parameter (lambda = .0005). Missing data due to sudden invisibility of the markers were interpolated up to 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, *GA* and *HP*.

The movement was first segmented into two main sections: i) from the movement onset (defined as the moment when the hand was more than 5 cm away from the home position along the z-axis)

to the time of the Maximum Grip Aperture (*MGA*), and ii) from the time of the *MGA* till the end of the movement. In trials without visual and haptic feedback (experiment 1, 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 where the *HP* 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 extent-wise, see for example Messier & Kalaska, 1997); finally, from the resulting window, we extracted the *FGA*, which corresponded to the value of the *GA* that minimized both net velocity and net acceleration. The *FHP* was extracted next, as the value of the *HP* at the time of the *FGA*. In trials with feedback (when the hand touched the object and the fingertips were visible – experiment 3 and the other half 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 *GA* 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 *GA* 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. Except for few cases where the trajectory was noisy and characterized by many local minima and maxima (despite the fact that we trained subjects prior to starting the actual experiment, few still felt unsure in the first trials — overall 1% of the 5240 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 et al., 2015). Significance test on the models' coefficients was performed using the Kenward-Roger approximation for degrees of freedom (Luke, 2016).

4.4 Experiment 1: A shared mechanism of for perception and grasping

This experiment 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? Both perceptual and grasping tasks required the depth estimate of an object, consisting of a virtual three-rod configuration placed in front of the observer in an otherwise dark environment. In the perceptual task, observers matched the perceived depth extent of the object with the sensed distance between their index finger and thumb. In the action task, they carried out a movement to reach and grasp the virtual object, without feeling its actual presence or seeing their hand.

4.4.1 Methods

Stimuli

The target consisted of patterns of random dots representing the surface of three cylinders, each being 0.5 cm wide in diameter and 6 cm high (Fig. 4.2B). 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 others were 2 cm to its right or left (total width of 4 cm). The middle rod was simulated in front of the two flanking rods at one of four depth separations (30, 40, 50, 60 mm) and never occluded them. Subjects were instructed to focus on the overall shape that the three rods suggested: most participants self-defined 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. Since we only varied the depth but never the width, all stimuli represented isosceles triangles except the 40 mm deep object in experiment 1, which was equilateral.

Procedure

In two separate blocks participants either judged the depth of the three-rod configuration or grasped it. In each block they viewed 5 times a total of 8 stimuli presented in random order: 4 depths (30, 40, 50, 60 mm) X 2 distances (450 and 550 mm) (Figure 4.2A). The width of the stimulus (the horizontal distance between the rear rods) was always 4 cm (Figure 4.2B). In each trial, subject saw only one stimulus at a time, either the near or the far one.

In the Manual Size Estimation (*MSE*) task, participants rested their hand on a wooden platform which was located close to their body (approximately 30 cm down, 26 cm right and 26 cm ahead with respect to the cyclopean eye) and each trial began with the fingers closed together. At the signal of a beep, the target appeared automatically and the participants were instructed to open their index and thumb along the line of sight without moving the hand from home, as to indicate the perceived depth of the object. We made sure that the fingers were lifted from the platform during the *MSE*, instead of being slid along its side, since subjects could feel the irregularities of the wooden surface and develop alternative strategies not directly related to the estimated depth. Each trial lasted 1500 ms from the stimulus onset.

In the Reach to Grasp (*RG*) task, participants began each trial by having their index and thumb closed together at a start position, which was the same as in the *MSE* task. At a sound of a beep, the target appeared. At this point 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. Although

participants performed the *RG* task without seeing their hand and without receiving haptic feedback, they underwent a short training before starting the experimental session, in order to be able to act in the most smooth and natural way, as if they reached-to-grasp a real object (Figure 4.2C).

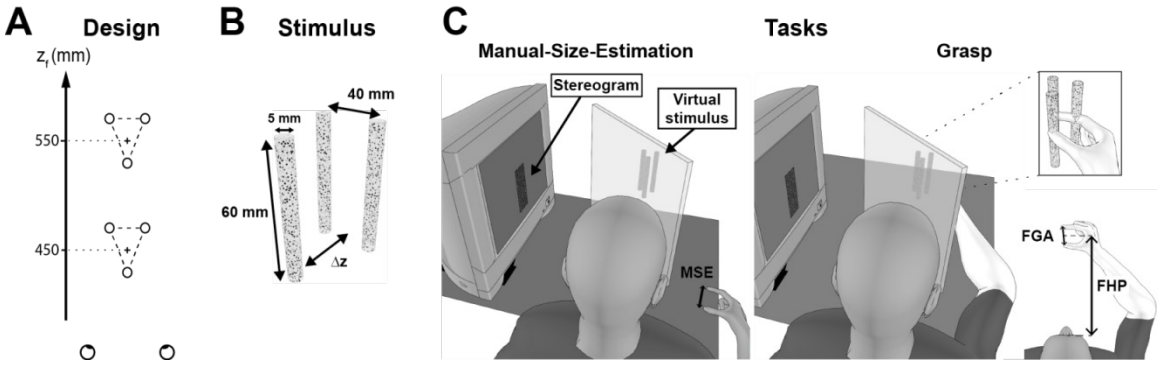

Figure 4.2. (A) Bird's eye view of the experimental design: in both tasks of experiment 1, subjects saw the stimulus at one of two egocentric distances (450 and 550 mm). (B) The stimulus consisted of three clouds of random dots each representing the surface of a cylinder of 5 mm of diameter and 60 mm tall. The three cylinders were positioned as the three vertexes of a triangle having a constant base of 40 mm and a variable height (the stimulus depth *Δz*, which could take one out of 4 values: 30, 40, 50, 60 mm). (C) Tasks and variables that were measured. During the grasp, participants did not see their hand and limb, nor touched any physical object (as depicted by the inset).

4.4.2 Results and Discussion

Figure 4.3 shows the average *FHP* as function of viewing distance for each object depth. In Figure 4.4 we plotted the average *MSE* (left) and *FGA* (right) as function of object depth for each viewing distance.

The egocentric function for the grasping task, represented by the relation between *FHP* and physical distance, clearly indicates a compression of visual space, since its slope is smaller than 1 ($slope_{FHP\text{-}Zf}$ = .69, 95%*CI* = [.43, .95]). A repeated measures ANOVA on the *FHP* also found a significant effect of the object depth ($F_{(3,36)}$ = 7.86, p < .001), as shallower objects were grasped farther than deeper objects. This suggests that subjects tended to bias their responses based on the distance of the thumb's contact point: a deeper object entailed a closer front surface. In agreement

with our hypotheses, the compression of space affected depth estimates for both the perceptual and grasping tasks, since *FGA* and *MSE* were larger for the near stimulus than for the far stimulus. A repeated-measures ANOVA on the depth estimates (*FGA* or *MSE*) with distance (450, 550 mm), object's depth (30, 40, 50, 60 mm) and task (grasping vs. perceptual) as within-participants variables revealed significant main effects of object's depth ($F_{(3,36)}$ = 190.32, p < .001), task ($F_{(1,12)}$ = 7.66, p < .02), and distance ($F_{(1,12)}$ = 16.9, p < .01). Interestingly, no significant interaction between distance and task was found ($F_{(1,12)}$ = 0.05, p = .82), indicating the same magnitude of compression across tasks.

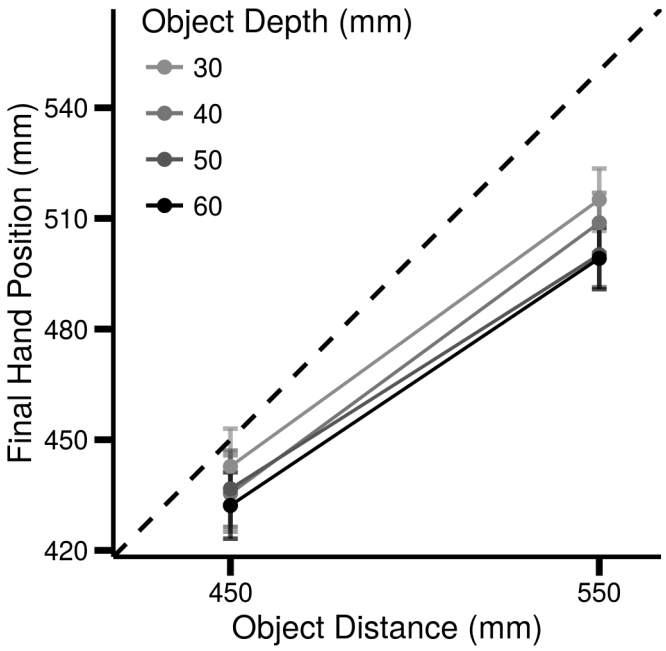

Figure 4.3. Mean Final Hand Position as function of object distance for each object depth. Error bars represent the standard error of the mean. The dashed line represents accurate performance.

In order to directly test the two main hypotheses, we estimated the egocentric function from the *FHP*, and the scaling functions from the *MSE* and *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). Figure 4.5 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 one individual.

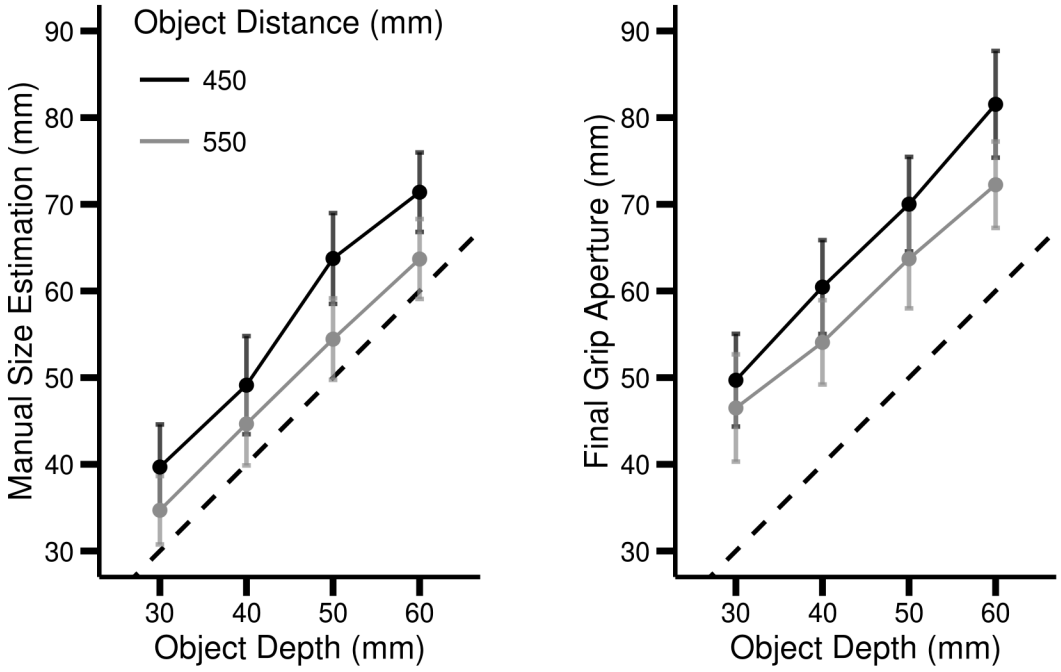

Figure 4.4. Mean Manual Size Estimation (left) and mean Final Grip Aperture (right) as function of object depth for each viewing distance.

As can be clearly seen, these results support the two hypotheses stated above. In agreement with the first hypothesis, the results of the grasping task directly confirm that distance estimates and depth estimates come from the same spatial distortions. Indeed, the compression of visual space has a predictable effect on both *FHP* and *FGA* (figure 4.5A): for each 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 (data point enclosed in a square in figure 4.5A) reached the two objects as if they were very close to

each other and also grasped the near one as much deeper than the far one, producing very small compression factors for both egocentric and scaling functions respectively. Instead, another subject (encircled data point in figure 4.5A) exhibited a weak visuospatial compression, thus she reached to the stimuli at almost their true distance and grasped them with nearly the same *FGA*, yielding compression factors close to 1 on both axes.

The second hypothesis, which predicts a unique 3D representation for perception and action is confirmed as well, since the egocentric function derived from the grasping task predicted the scaling function derived from the perceptual task as well. In figure 4.5B, the values on the x axis are the same of figure 4.5A (the egocentric function obtained from the *FHP* in grasp), while the values on the y axis represent the scaling function calculated from the *MSE*. The two data points highlighted in figure 4.5B are the same two persons considered in the previous example. The figure shows that the visuospatial compression affecting the grasp also affected the *MSE* in the same way: if a subject exhibited a small compression factor in the egocentric function via *FHP* (dot enclosed in a square), they showed the same small compression factor in the *MSE*-derived scaling function (they perceived the near object to be much deeper than the far object); similarly, if a subject showed an almost veridical estimation of distance in the grasp (dot surrounded by a circle), she also perceived the stimuli's depth fairly accurately.

The compression factors derived from the *FGA* and the *MSE* were each fitted by a linear model that had the *FHP*-derived compression factor as the only predictor. Consistent with the two hypotheses, the intercepts of both fits were not significantly different from zero ($intercept_{FHP\text{-}FGA}$ = -.25, 95%*CI* = [-.61, .10]; $intercept_{FHP\text{-}MSE}$ = -.07, 95%*CI* = [-.49, .34]), while the slopes of both fits were not significantly different from 1 ($slope_{FHP\text{-}FGA}$ = 1.34, 95%*CI* = [.80, 1.87]; $slope_{FHP\text{-}MSE}$ = 1.08, 95%*CI* = [.45, 1.71]), providing no evidence that X and Y were different in neither model. This result is particularly remarkable, 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.

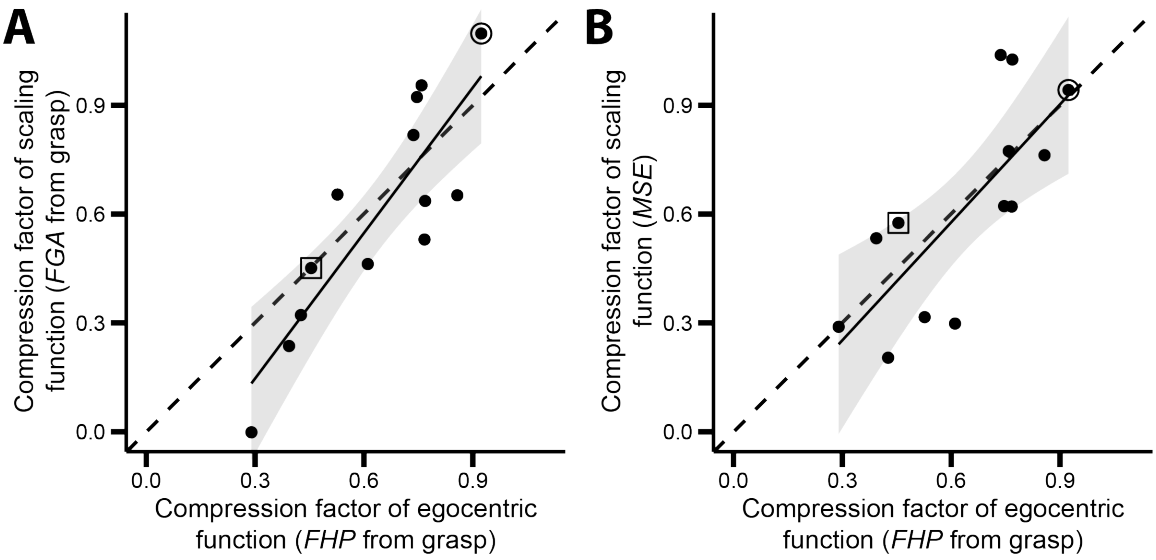

Figure 4.5. Individual compression factor of the scaling function estimated from the *FGA* data (A) and *MSE* data (B) in ordinate, as function of the compression factor of the egocentric function estimated from the *FHP* data in abscissa. Each data point is a subject. The dashed line identifies the prediction of the first (A) and the second (B) hypotheses. The black lines are the result of a linear fit, with the gray regions specifying the 95% confidence interval of the fits.

4.5 Experiment 2a: Correcting perceptual biases corrects grasping

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. 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 the perceptual task by setting the simulated depth and size of a three-rod-configuration at two egocentric distances, so to match its structure to that of a standard full-cue stimulus. Because of visuospatial compression, we predicted that a far object must be set deeper than a near object for achieving a perceptual depth match.

In a successive task, participants were asked to grasp the two objects they produced in the matching procedure. According to our hypothesis, the Final Grip Aperture should be the same for the two objects, in spite of their different physical structure.

4.5.1 Methods

Procedure

Each observer performed the perceptual matching task followed by a reach-to-grasp task. In the Visual Depth Adjustment task observers 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, comprised of three wooden cylinders covered with white paper, was a physical reproduction of the virtual stimulus. The physical front rod was installed on the movable platform behind the mirror, whereas the two rear rods were attached to a stand that was fixed on the table. Since the whole experiment was conducted in a completely dark room, the physical target was made visible by a small black light positioned next to it. To guarantee uniform illumination on the object, a small mirror was positioned at the opposite side of the object from the neon. The reference stimulus was located roughly 25 mm to the left of the participant's cyclopean viewpoint (see figure 4.6). While the two physical rear rods always stood at 350 mm from the observer, 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, whereas its size (i.e. the distance between the back rods) was constantly 40 mm.

On each trial, observers varied the size and depth of the test stimulus via keyboard key press until it appeared to match the dimensions of the reference stimulus. As in the previous experiment, the

test stimulus was presented straight ahead in front of the observer at one of two distances (270 and 450 mm). The depth of the reference stimulus and the distance of the test stimulus were chosen randomly at each trial. Subjects performed a total of 36 perceptual trials (3 depths x 2 distances x 6 repetitions). The average estimated depths and widths set by each observer were then used to simulate six virtual stimuli that participants were required to grasp during the second part of the experiment. The Perceptually-Equalized Stimuli (*PES*), appearing to have the same depth and width at both distances, were therefore unique for each participant.

In the reach-to-grasp task only the *PES* were presented, one in every trial, in random order. In all respects the procedure of this task was identical to that of Experiment 1. The grasping task comprised a total of 30 trials: 6 stimuli (3 per distance) x 5 repetitions.

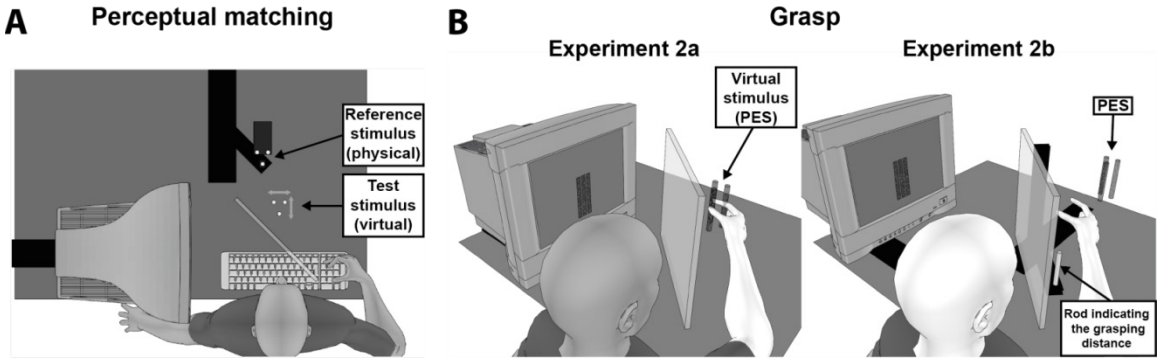

Figure 4.6. Schematic of the tasks in experiment 2a and 2b. (A) In the perceptual task, subjects viewed a physical (*reference*) stimulus always at the same distance, and adjusted the width and the depth of a virtual (*test*) stimulus with the keyboard until the *test* looked identical to the *reference*. 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 (Fig. 4.2C). In the grasp of experiment 2b (right), a small rod positioned in the lower visual field indicated the distance at which the subject had to perform the grasp, that was different from that of the virtual stimulus. When the *PES* was presented at 450 mm the small rod appeared at 270 mm (as in the example), and vice versa.

4.5.2 Results and Discussion

Figure 4.7 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. Like in Experiment 1,

the results confirm the phenomenon of visuospatial compression. In order to achieve a perceptual match, observers adjusted the depth of the far object to be larger than that of the near object, showing that distant objects are perceived shallower than close objects. We conducted a repeated-measures ANOVA on the adjusted depth with distance (270 and 450 mm) and reference object's depth (30, 40, 50 mm) as within-participants factors and we found significant main effects of both variables (distance: $F_{(1,9)} = 15.82$, $p < .01$; reference object's depth: $F_{(2,18)} = 43.58$, $p < .001$), as well as a significant interaction between them ($F_{(2,18)} = 3.89$, $p < .05$). The same analysis on the adjusted width revealed instead no significant effects (distance: $F_{(1,9)} = 0.73$, $p = .41$; object's depth: $F_{(2,18)} = 1.98$, $p = .17$; distance vs. object's depth interaction: $F_{(2,18)} = .08$, $p = .92$). Hence, the compression of visual space did not have any significant influence on the adjusted width. This latter result can be expected, since retinal size only varies linearly with distance: given the small range of distances that was tested, any bias on the scaling of width due to visuospatial compression may have not been statistically detectable in our design (but see van Damme & Brenner, 1997).

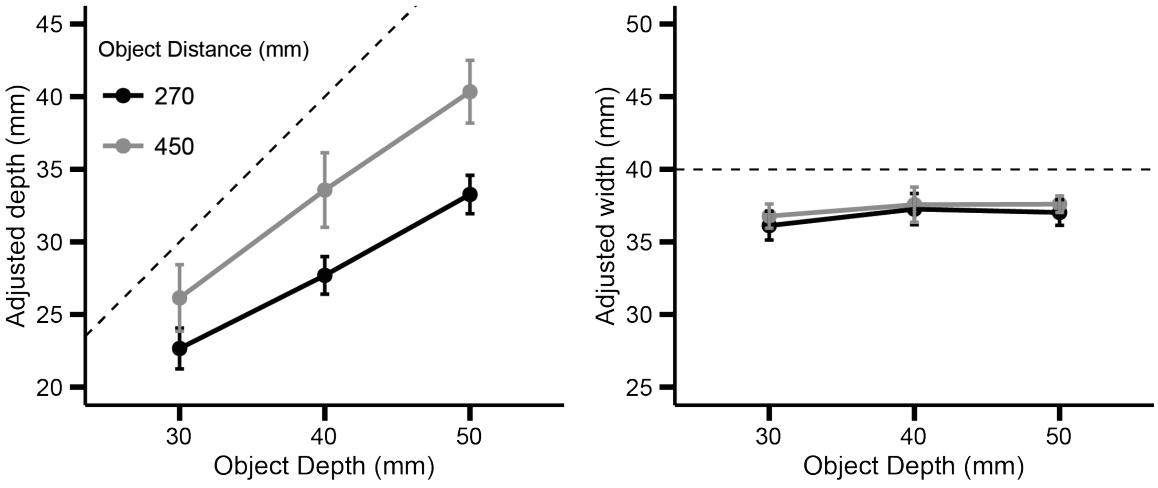

Figure 4.7. Mean adjusted depth (left) and width (right) of the test stimulus as function of object depth for each distance.

As a consequence of the depth settings, in the following task observers grasped objects that were on average 5.48 mm deeper at the far distance than at the close distance. Figure 4.8 (left panel) shows, for each distance, the average *FGA* as function of reference stimulus depth to which the test stimuli were perceptually matched. In spite of the difference in simulated depth between the near and far objects, 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 ($F_{(2,18)} = 9.64$, $p < .01$) and a near-to-significant effect of distance ($F_{(1,9)} = 2.9$, $p = .12$). In figure 4.8 (left panel), it should be noticed in fact that the FGA showed a tendency to be larger at the far distance. Even though the average difference between FGA (about 2.5 mm) was smaller than the difference between physical depths, we tested whether subjects responded to the physical depth in a control experiment.

4.6 Experiment 2b: Control for a possible bio-mechanic confound

It has been long shown that grip apertures are also planned based on the expected accuracy of the transport component (Bootsma et al., 1994; Wing et al., 1986; Jakobson and Goodale, 1991): 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 the previous experiment (Fig. 4.8 left): the position where one plans to do a grasp should affect the grip aperture irrespectively of the estimated depth of the grasped object. We therefore reasoned that the opposite effect ought be expected if the subjects were induced to grasp the closer object by reaching at the far distance and the farther object by reaching at the close distance.

4.6.1 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 (Fig. 4.6B, right panel).

4.6.2 Results and Discussion

In agreement with the predictions, the results showed the opposite effect of grasping distance found in the previous experiment (Fig. 4.8, 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 both object's depth ($F_{(2,18)} = 7.07$, $p < .01$) and object's distance ($F_{(1,9)} = 8.71$, $p = .02$) on the *FGA.* Notably, the effect of distance we found was opposite to the effect found in the previous experiment: this time, near objects were grasped as deeper than far objects. This cannot be explained by the fact that subjects recovered the veridical depth of the stimuli, since, as in the experiment 2a, the near objects were actually shallower.

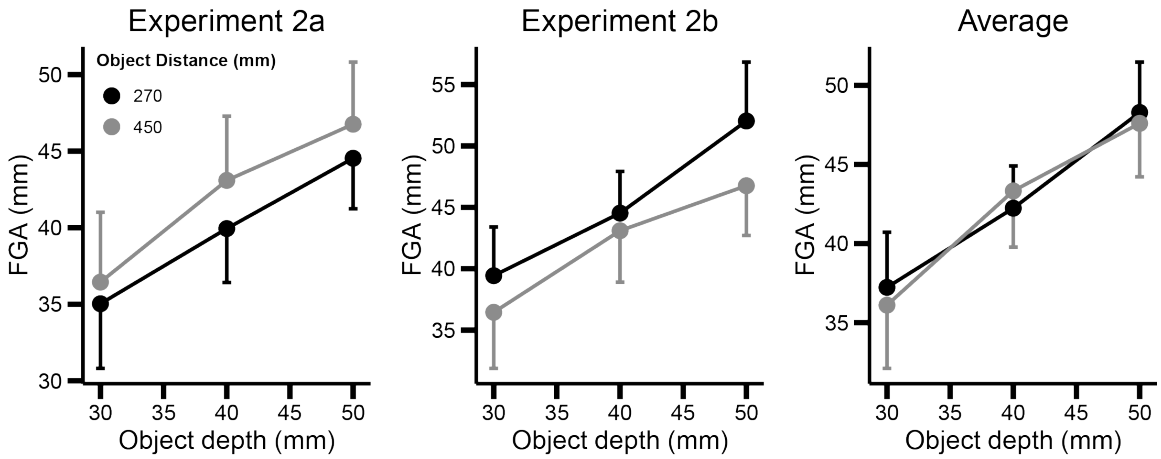

Figure 4.8. *FGA* as function of the object depth for each viewing distance (270 mm in black; 450 mm in gray). Left, results from experiment 2a (reaching distance identical to viewing distance). Center, results from experiment 2b (reaching distance opposite to viewing distance). Right, average *FGA* across experiments 2a and 2b.

The grasp tasks of experiments 2a and 2b were identical with the exception of where the hand was brought in order to grasp the object. Recall that the near and far virtual stimuli appeared identical because of the 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 ($F_{(2,18)} = 10.24$, $p = .001$), object distance and hand distance as within-subjects factors: as expected, the hand distance was the only variable affecting the *FGA* ($F_{(1,9)} = 9.30$, $p = .01$), while the object distance had no influence ($F_{(1,9)} = 0.09$, $p = .77$). 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. This result represents converging evidence that perception and action share the same representation of the visual space.

4.7 Experiment 3: Online control does not eliminate biases at planning

The previously described grasping tasks, 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 et al., 1994; Westwood et al., 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 as 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, interpretation which we will define as the *pantomime grasp* hypothesis. A pantomime grasp is defined as a grasp towards a remembered object, that is, that usually disappears at movement onset. As a consequence, subjects can see their hand but not the stimulus throughout the action, so they have to mimic the grip aperture that they assume would be appropriate to hold that object between their fingers. According to this hypothesis, convergent visual and haptic feedback are required for the veridical estimation of an object size. Therefore, only when both are present, vision-for-action is posited to be veridical (Jazi et al., 2015; Holmes et al., 2013; Hosang et al., 2016).

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 not due to the absence of haptic feedback then they should be detected also when the visible grasping digits make contact with the physical object. 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.

4.7.1 Methods

Procedure

The procedure was the same as in the grasp task of Experiment 1, with the following exceptions. First, while reaching, participants could see two luminous dots which coincided with the centers of their index and thumb fingertips (Fig. 4.9A). Since the index is normally occluded by an object at the end of a front-to-back grasp, the index's virtual tip also disappeared when it was behind the plane defined by the two rear rods (the object's "back surface"). Second, we provided haptic feedback by positioning two metal rods at specific locations behind the mirror (Fig. 4.9B). The metal rods were mounted on a moving platform and the linear actuators set their position so to perfectly superimpose them to the virtual image. In two conditions, subjects either grasped the configuration along the depth dimension, as in the previous experiments (*front-to-back* grasp), or they grasped the front and right-side rods (*along-the-side* grasp) (Fig. 4.9C). In the *front-to-back* grasp, one metal rod matched the position of the virtual front cylinder, while the other metal rod was brought at the same depth of the two virtual cylinders, and positioned in the middle of them. In the *along-the-side* grasp, one metal rod was again aligned with the virtual rod in the front, while the other metal rod matched the position of the right rear virtual rod. The metal rods were only felt by the subjects but never seen. While grasping, participants only saw the virtual stimulus and the rendering of their fingertips. In the *front-to-back* grasp, the index virtual fingertip

disappeared as soon as it went behind the object's back surface delimited by the two back rods. In the *along-the-side* grasp, the index remained visible until object contact. A 2 (depth: 30 and 50mm) X 2 (distance: 270 and 450mm) X 2 (grasp type: front-to-back, along-the–side) design with 5 repetitions for each combination yielded a total of 40 trials. 18 subjects executed the front-to-back grasps, as in Experiments 1 and 2, and 17 subjects grasped the objects along the side.

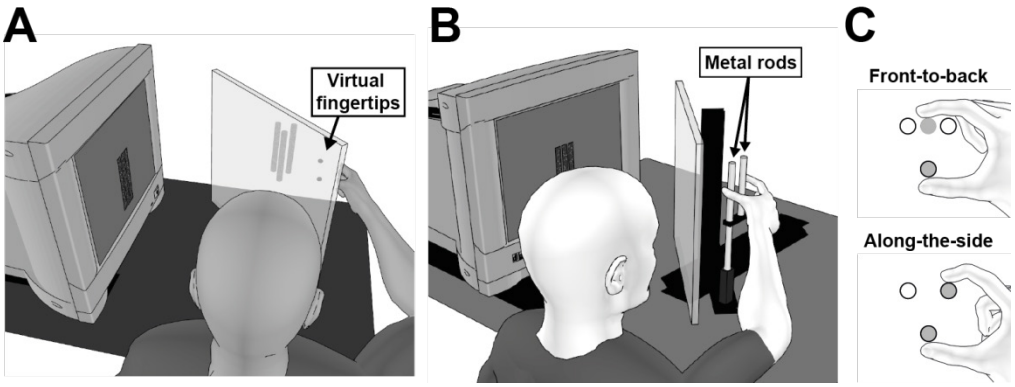

Figure 4.9. (A) Subjects reached-to-grasp the virtual stimulus while seeing two additional small 3D discs aligned with their index and thumb fingertips. (B) Behind the mirror, not seen, were two metal rods that were aligned with the virtual stimulus in two different ways, to allow for two different grasps. (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. Consequently, the index disappeared in the last moments of the *front-to-back* grasp, whereas it remained visible until object contact in the *along-the-side* grasp.

4.7.2 Results and Discussion

Since the fingers inevitably made contact with the physical object at the end of each grasping movement, we could not use the *FGA* as independent variable. Instead, we looked at how the Grip Aperture (*GA*) evolved over the course of the reaching trajectory. Figure 4.10A shows the average trajectories of index and thumb in the final stages of the grasp in the two conditions (left, *front-to-back*; right, *along-the-side*). Our hypothesis states that the grasp is planned based on distorted estimates of the object's depth, even when visual and tactile feedback are available. In

particular, because of depth underconstancy, we expected the grip aperture to be on average smaller when reaching to the farther object than when reaching to the near object.

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 *GA* was averaged and analyzed as a linear function of the object's distance and depth. The slope of that function ($slope_{GA\text{-}Z}$) was compared across four critical instants along the trajectory: at *MGA* time and when the thumb was 20, 10 and 5 mm away from its final position (fig. 4.10B, bottom row). In the *front-to-back* grasp, we found a negative $slope_{GA\text{-}Z}$, meaning depth underconstancy, already at *MGA* time, although it was not yet significant ($F_{(1,49)} = 2.08$, $p = .15$). The effect then progressively increased in magnitude and remained significant until 5 mm before the thumb contacted the object (20 mm: $F_{(1,49)} = 35.74$, $p < .001$; 10 mm: $F_{(1,49)} = 21.28$, $p < .001$; 5 mm: $F_{(1,49)} = 14.98$, $p < .001$). A similar pattern was found for the *along-the-side* grasp (MGA: $F_{(1,52)} = 1.44$, $p = .24$; 20 mm: $F_{(1,52)} = 38.56$, $p < .001$; 10 mm: $F_{(1,49)} = 26.35$, $p < .001$; 5 mm: $F_{(1,49)} = 11.55$, $p = .001$). We also looked at the overall evolution of the $slope_{GA\text{-}Z}$ during the full thumb's reaching trajectory. As can be seen in figure 4.10B (top row), we found that it was continuously negative from the time of the *MGA* until the movement end, that is, when the grip was finely adjusted in preparation to grasp.

The fact that we found depth underconstancy in both grasps suggests that the grip aperture was always planned based on a wrong estimate of the object's depth (or side), even when, as in the *along-the-side* grasp, subjects could theoretically use online visual feedback to navigate their fingertips on each finger's contact surface, without the need to recover the depth of the stimulus at all.

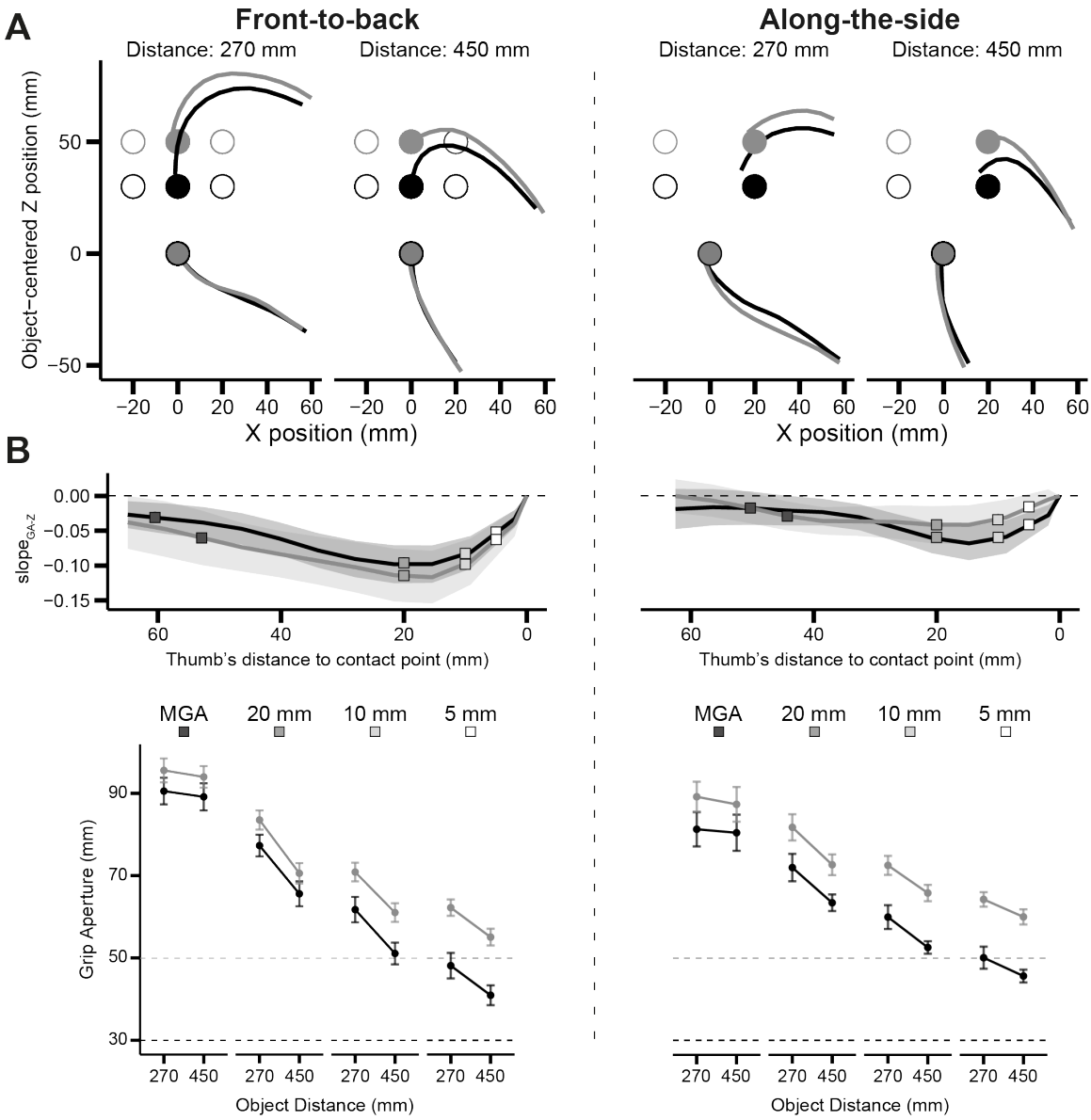

Figure 4.10. Trajectory analysis from the data of experiment 3. Left, front-to-back grasp; right, along-the-side grasp. Throughout the figure, results and trajectories relative to the 30 mm object are in black, whereas the data relative to the 50 mm object is in gray. (A) Average index and thumb trajectories for each object distance (left, 270 mm; right, 450 mm). The positions of the virtual and physical objects are marked by large circles (open circles represent the virtual rods, filled circles represent the physical rods, as in figure 4.9C). (B) Evolution of the slope of the function relating the *GA* to the object distance ($slope_{GA\text{-}Z}$) as the thumb approaches its contact location. Negative values mean depth underconstancy: the *GA* is smaller for the far object than for the near object. Snapshots of the *GA* were taken at four moments along the trajectory (at the time of the *MGA* and when the thumb was 20, 20, and 5 mm away from its contact point), which are marked by little squares gradually turning from dark gray to white as the thumb's distance to the object decreases. The four snapshots are visualized in the graph below: *GA* as function of the object distance in abscissa for each object's depth.

Even though these findings show that the pantomime grasp hypothesis cannot explain the results of the previous experiments, we found further evidence of the systematic biases affecting grasping actions by analyzing the effect of previous trial. Here we present the analysis of the grip aperture when the thumb was 2 cm away from its contact point, but the results are the same in the rest of the trajectory. Along with the significant main effects of object's depth and current trial's distance discussed previously, we also found a significant main effect of the previous trial's distance ($F_{(1,33)} = 13.5$, $p < .001$), as the *GA* for trials preceded by grasps towards far objects was larger than when the previous grasp was directed at near objects (Figure 4.11). 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.

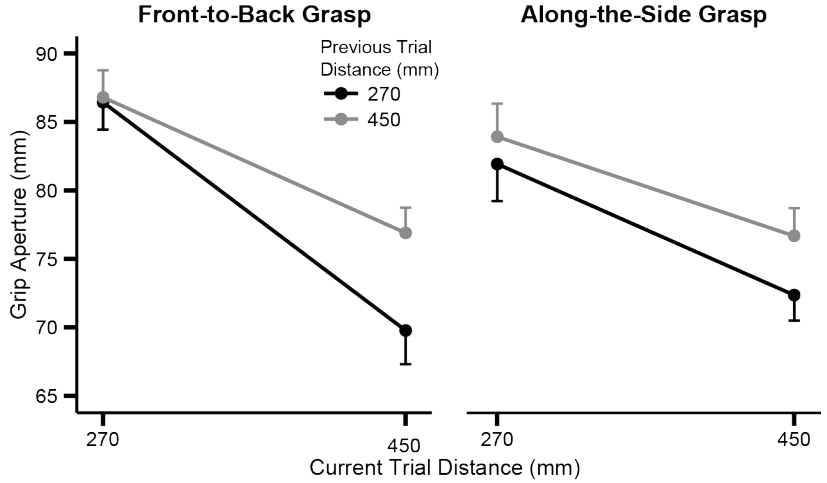

Figure 4.11. *Grip Aperture* profile 20 mm prior to the thumb's contact with the object's surface, as function of the current trial's distance (in abscissa) for each possible distance of the previous trial (270 mm in black and 450 mm in gray).

4.8 Experiment 4: Grasping is planned in both allocentric and egocentric coordinates

The co-occurrence of both visuospatial compression and depth underconstancy in experiment 1 represents an interesting paradox, since in principle the two phenomena are geometrically inconsistent with one another. According to the former, if a region of egocentric space, and presumably all the objects within it, underwent a homogenous compression, then the same object should always appear compressed by a constant amount in return, no matter which distance one viewed it from. However, as we have previously seen, depth underconstancy causes an object to appear either deeper or flatter depending on how far it is from the body.

Many 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). Different 3D properties of a shape are in fact encoded within different frames of reference: 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). There is strong evidence of at least a partial dissociation between egocentric and allocentric maps of visual space during perception as well as during reaching and grasping (Loomis et al., 2002; Loomis et al., 1996; Bingham et al., 2004; Bingham et al., 2000; Neely et al., 2008; Binsted & Heath, 2004; Thaler & Goodale, 2010; Chen et al., 2011; Gentilucci et al., 1997). If different spatial maps affect the planning of the grasp in the same way as perception in our previous experiments, then we should expect subjects to behave differently depending on which 3D property is the focus of their actions. Let's consider an experiment in which an observer is asked to either grasp the front-to-back extent of an object or to reach to the location of its front and back surfaces independently. As shown in figure 4.12, two possible hypotheses can be formulated.

According to what we will define as the *single map hypothesis* (Fig. 4.12A), both absolute distance and relative depth are encoded within the same frame of reference. As a consequence: i) when asked to reach to the front and back surfaces of two objects (z_{front} and z_{back} – physical world), the observer will aim at four distances (z'_f and z'_b – reaching task) that are uniformly closer to each other than their physical counterparts, because of visuospatial compression; ii) since relative depth is encoded as the separation between the same set of distorted distances, the Final Index and Thumb Positions of the grasp (*FIP* and *FTP*) will match the endpoints of the reaching, thus falling on the bisector of the final plot of figure 4.12A.

On the contrary, the *dual map hypothesis* (Fig. 4.12B) states that absolute distance is encoded in an egocentric frame of reference, while relative depth is retrieved in allocentric coordinates. According to this hypothesis, the endpoints of the reaching task are subject to the same visuospatial compression predicted by the single map hypothesis (z'_f and z'_b – reaching task in Fig. 4.12B). However, during the grasping task, a global estimate of the object's distance is used to scale binocular disparities to obtain an estimate of the object's depth. As a consequence, visuospatial compression leads to depth underconstancy: if the absolute position of the object is overestimated due to compression, the object's depth will be overestimated in turn and vice versa (black and gray shapes respectively – grasping task in Fig. 4.12B). If we assume that the *FTP* will match the front pinch (final plot of figure 4.12B), then the *FIP* will not coincide with the location of the back pinch. Instead, it will deviate from the bisector by an amount that varies with distance.

In this experiment, we compared a front-to-back grasp with a reaching task where subjects pinched either the front or the back edges of the grasped object. The pinch was adopted to maximize the similarity between tasks (i.e., both were grasp movements). Furthermore, we used a modified version of the stimulus: subjects saw only one rod at the center of the visual field and a rectangular plane behind it, at one of two depths (40 and 50 mm). During the grasp, the thumb touched the rod while the index's contact location corresponded to the right edge of the plane in the back. This ensured that the index contact point was always visible (in the first experiment, the

index's contact location was an imaginary point between the two back rods). We tested the two hypotheses shown in figure 4.12 by aligning the *FIP* and *FTP* of the grasp with the endpoints of the pinch and by comparing the resulting spatial maps.

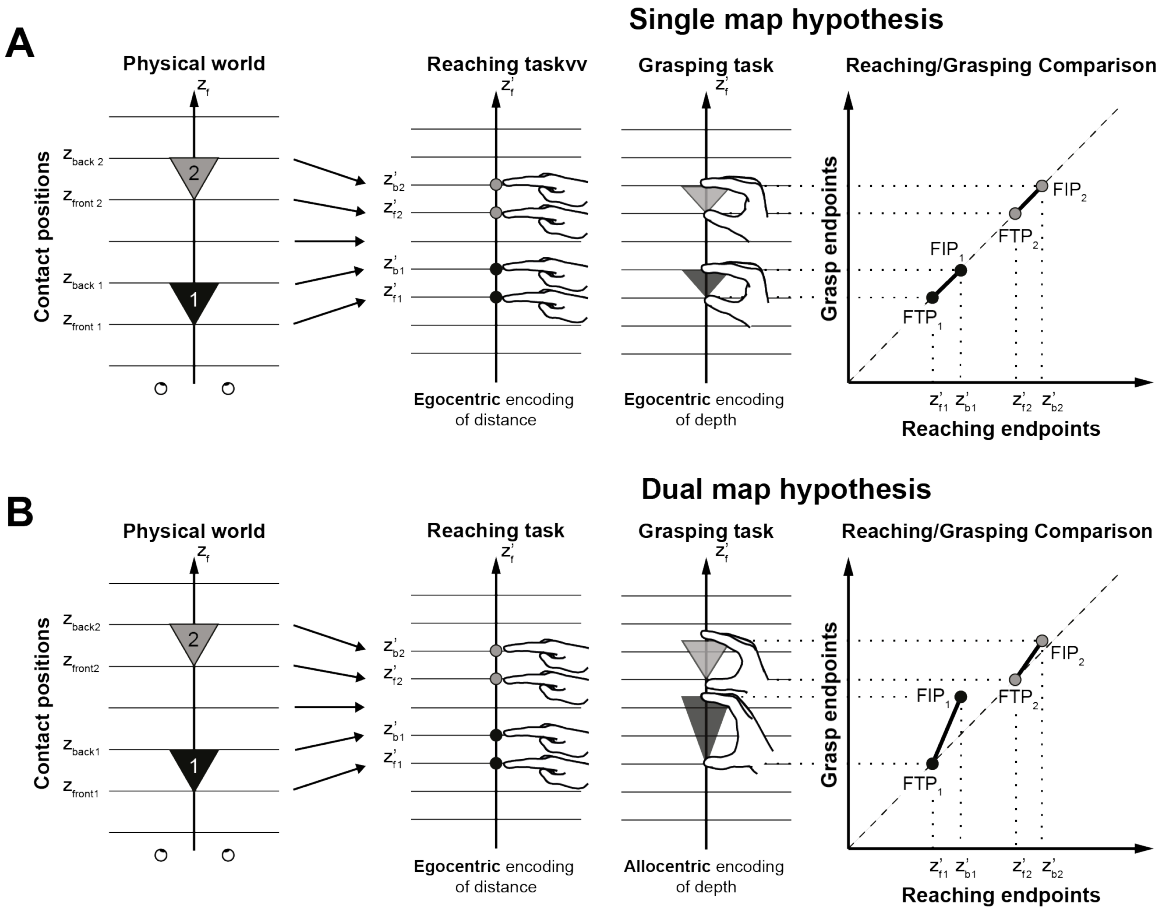

Figure 4.12. Different frames of reference for planning a front-to-back grasp of two identical objects (1 and 2) positioned at two successive distances (the front and back edges the objects are marked with z_{front1}, z_{back1} and z_{front2}, z_{back2}). (A) If all the 3D properties of the visual scene are encoded in egocentric coordinates, then both index and thumb final positions of the grasp (*FIP* and *FTP* respectively) should match the endpoints of a reaching task (z'_f and z'_b) where subjects aimed at the front and back surfaces. In the final plot, the oblique dashed line represents perfect correspondence between pinch and grasp (*single map hypothesis*). (B) If distance is retrieved in egocentric coordinates while depth is encoded in allocentric coordinates, then the endpoints of reaching and grasping tasks should not match. In the example, the position of the front surface is recovered in egocentric coordinates during both tasks, therefore the *FTP* is identical to z'_f. Instead, the position of the back surface is estimated in relation to z'_f (that is, allocentrically, as the distance relative to a biased estimate of the front surface), therefore the *FIP* deviates from the bisector by an amount that depends on distance.

4.8.1 Methods

Procedure

Each participant did four blocks in which they performed one of two movements on every trial: i) grasp the front-to-back separation (40 or 50mm) between one rod and a rectangular plane (touching the rod with the thumb and the plane with the index) or ii) pinch only the rod with the index and thumb (Fig. 4.13 A&B). The rod and the plane thus represented the front and the back edge of the stimulus respectively.

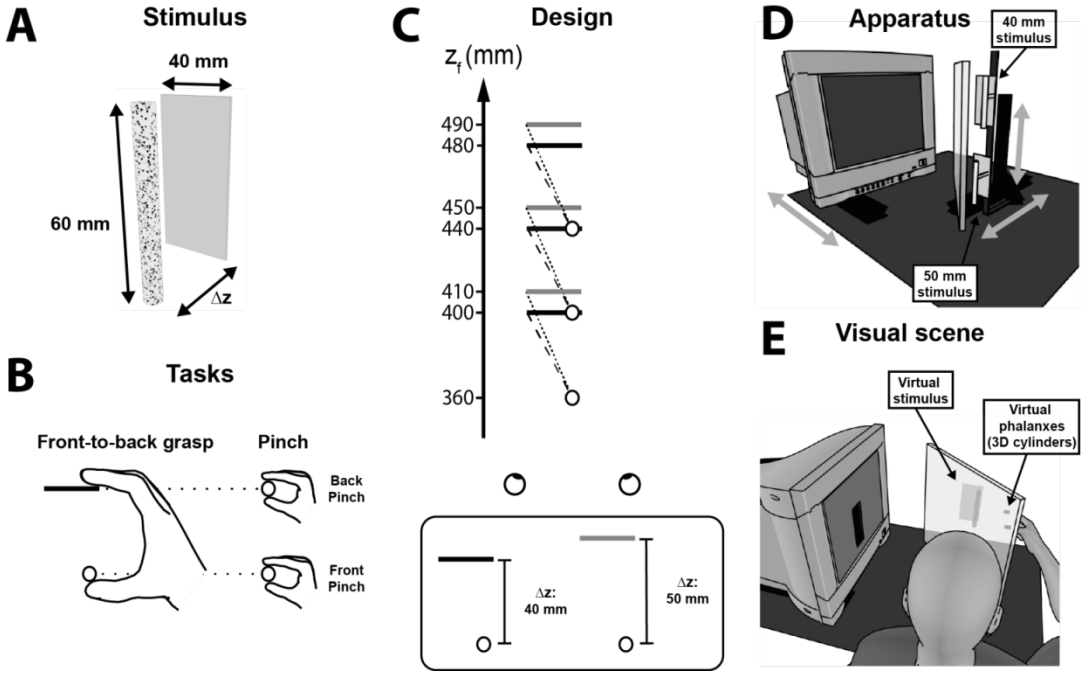

Figure 4.13. Methods used in experiment 4. (A) The stimulus comprised of two 3D virtual objects: a rod (front) and a plane (back). (B) Subjects either grasped the stimulus front-to-back or pinched the rod located either in the front or in the back position. (C) Bird's eye view of the experimental design. Subjects were presented with 40 or 50 mm deep stimuli (shown in the box at the bottom) that were each seen at one of three successive distances. The oblique dashed lines are only there to facilitate the reader in discriminating the different shapes. The front and the back surfaces of all six objects defined a set of 7 egocentric distances (360, 400, 410, 440, 450, 480, 490 on the distance vector z_f). The stimuli appeared with the rod at the center of the visual field. (D) and (E) Schematic representation of the setup and of the visual imagery that were used.

The stimulus was presented at one of 6 distances (3 distances per object size) in each trial. These distances were chosen such that the front rods of the 40mm and of the 50mm stimuli always coincided. While the grasp was performed towards one of the six distances mentioned above, the pinch required reaching to just the egocentric points corresponding to the front or the back of each grasped object (front and back pinches, fig. 4.13B), thus resulting in 7 distances total (Fig. 4.13C).

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 (Fig. 4.13D). 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 automatically, and participants had 2 seconds 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 pinch. Participants were never able to see the physical objects as the stand and the Velmex system were blocked by an occluder behind the mirror.

Each trial was performed in either a full-feedback condition (FF), or in a no-feedback condition (NF), and at any given trial, participants did not know in advance whether they would have feedback or not. During FF trials, the Velmex system first aligned the correct physical stimulus with the location of the virtual images, and then the participants performed the task. While reaching, their index and thumb's final phalanxes were represented by two additional 3D cylinders, aligned and updated in real time using the coordinates of the Optotrak markers attached to the fingernails, so the subjects could accurately navigate their fingers onto the objects to touch them (Fig. 4.13E). During NF trials, the motors first pushed the stand out of reach, and then participants performed the task without seeing the rendering of their fingers nor touching a physical object at the end of the movement. FF and NF trials were intermixed within the same

session, resulting in four equally possible actions: a pinch with or without feedback or a grasp with or without feedback. To summarize, over the course of 4 blocks, participants performed 13 different actions (6 grasps + 7 pinches) x 2 feedback conditions x 20 repetitions (5 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).

4.8.2 Results and discussion

The data from the *NF* grasps supported the hypothesis that subjects exploited both egocentric and allocentric information when grasping. As in experiment 1, the *FHP* was related to the stimulus distance by a slope less than 1 (slope: .70, 95%*CI* = [.64 .77]; $F_{(2,10)}$ = 77.58, p < .001), in agreement with visuospatial compression, and was also modulated by the stimulus depth ($F_{(1,5)}$ = 23.47, p < .001), such that the flatter object (40 mm) was grasped as slightly closer than the deeper object (Fig. 4.14A), suggesting that the object's estimated distance was biased towards the position of the front surface.

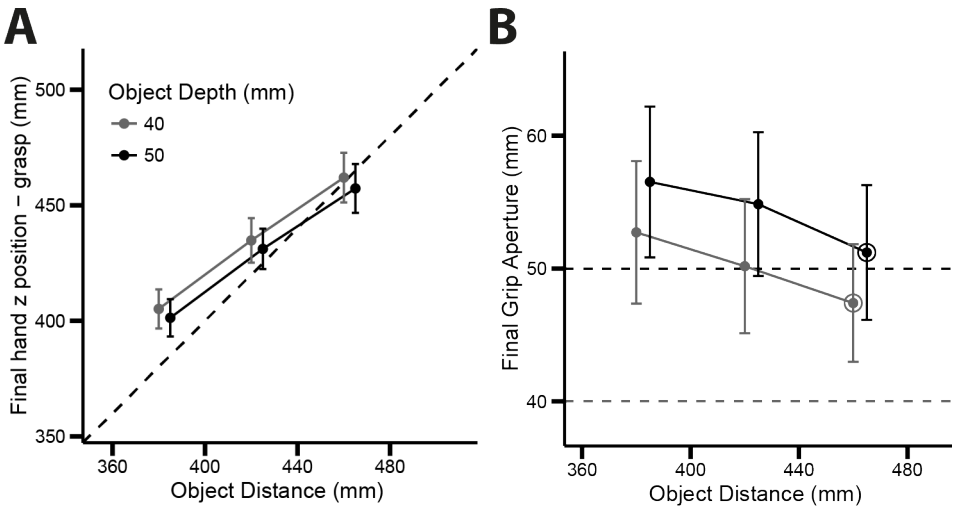

Figure 4.14. Results from the *No-Feedback* grasps in experiment 4. (A) Visuospatial compression: *FHP* as function of the stimulus distance for each object's depth (gray, 40 mm; black, 50 mm). (B) Depth underconstancy: *FGA* as function of distance for each depth. Horizontal dashed lines represent veridicality. The two circled dots represent the closest-to-veridical grasps (see in reference to Fig. 4.13 and relative explanation in the text).

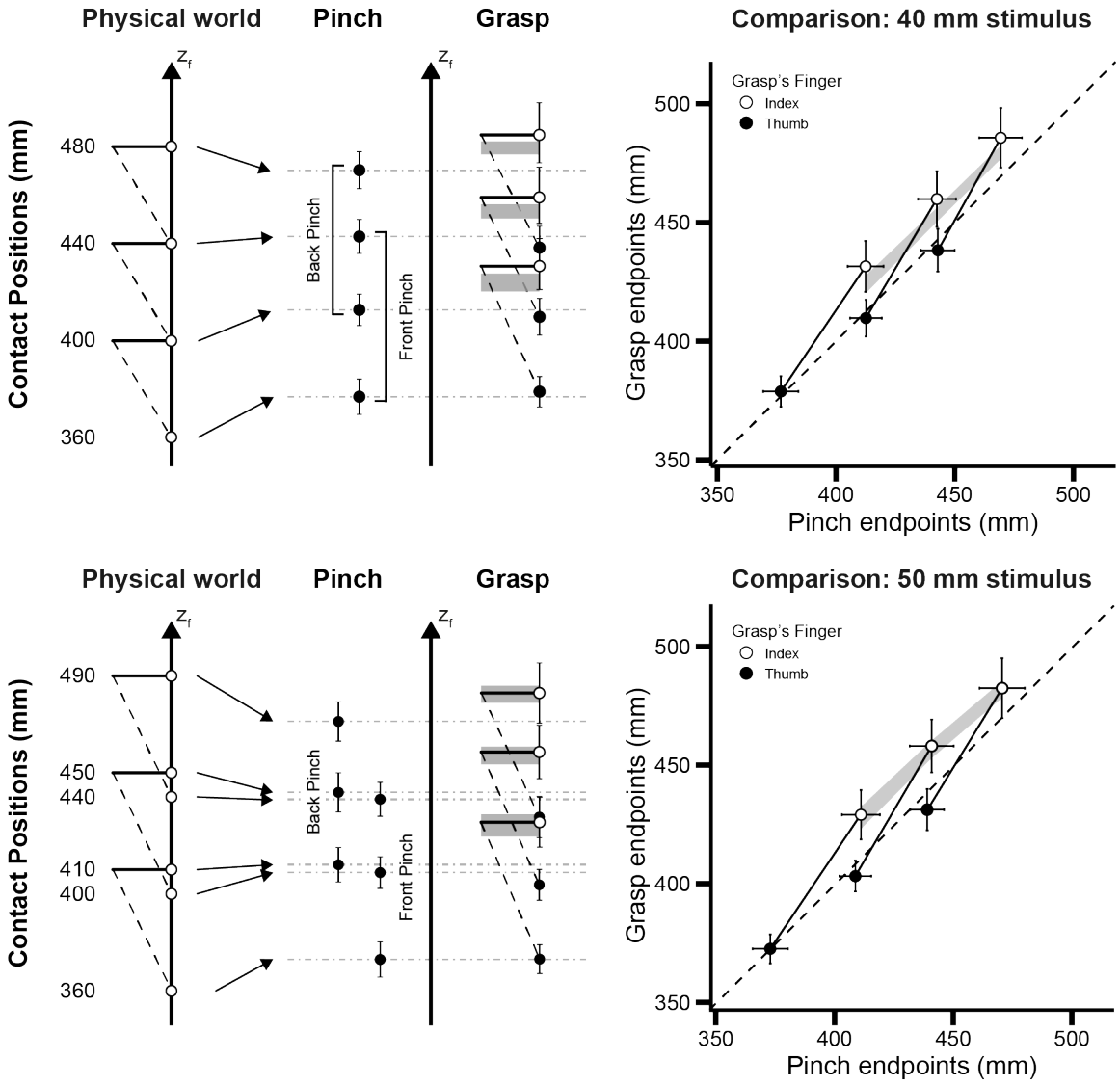

Figure 4.15. Comparison between grasp and pinch tasks, following the predictions in figure 12. Upper row: 40 mm stimulus; lower row: 50 mm stimulus. In each row, from left to right: 1) "Physical world": diagram of the experimental desing in bird's eye view; 2) "Pinch": arrows connect the physical distances to the corresponding average endpoints of the pinching task, showing visuospatial compression – the reached distances corresponding to the front and back surfaces of the objects are labeled "Front Pinch" and "Back Pinch" respectively; 3) "Grasp": horizontal dashed lines allow to compare the average endpoints of the grasp with their corresponding pinch (*FTP*, filled circles; *FIP*, open circles); 4) "Comparison": the same endpoints of the previous two diagrams are directly compared with each other in a Cartesian plane (grasp endpoints on the y axis, pinch endpoint on the x axis). The bisector (dashed line) illustrates the prediction of the single map hypothesis (distance and depth estimation share the same distorted space). Errorbars represent 1 *SE*. The gray areas depict the 95% confidence intervals of the prediction for the *FIP* according to the dual map hypothesis.

At the same time, depth estimation followed depth underconstancy: a repeated-measures ANOVA showed in fact that the *FGA* significantly decreased with distance (object's depth: $F_{(1,5)} = 33.18$, $p = .002$; object's distance: $F_{(2,10)} = 12.43$, $p = .002$) (Figure 4.14B). Overall, we successfully replicated the effects found in experiment 1. Note that these effects emerged notwithstanding the presence of trials with full visual and tactile feedback within the same session.

The comparison between pinch and grasp, done by aligning the spatial maps resulting from the two tasks (Fig. 4.15), revealed three main results in support of the dual map hypothesis.

First, we found compelling evidence of visuospatial compression from the data of the pinch. In figure 4.15, the bird's eye view of the experimental design is presented in the leftmost column ("Physical world"). Arrows connect the contact positions in that diagram to the endpoints of the pinch in the next column ("Pinch", filled circles). The slope of the function mapping physical distances to endpoints was .76 (95%*CI* = [.71, .81]), in line with the previous experiments.

Second, the final thumb position of all grasps (filled circles in the "Grasp" column of figure 4.15) was almost coincident with the front pinch, suggesting that the front surface was encoded in egocentric coordinates during the planning of both tasks.

Third, if the single map hypothesis was true, then the index final position should be encoded in egocentric coordinates as well, meaning that the *FIP* should match the back pinch. However, we found that the *FIP* (open circles in the "Grasp" column) was systematically farther than the egocentric predictions (gray dashed lines in both "Pinch" and "Grasp" columns), which is actually consistent with the dual map hypothesis. According to this hypothesis, in fact, the estimated depth of an object is not the interval between the egocentric estimates of the index and thumb's contact points, but it is obtained via scaling binocular disparity by an estimate of the viewing distance (Fig. 1.2). If we assume that subjects fixated the front surface of the stimulus (the thumb's contact point), as the data suggests, then the *FIP* is equivalent to the *FTP* plus an estimate of the object's depth derived by scaling binocular disparities by the *FTP* (see Appendix for the equations).

The "Comparison" column of figure 4.15 shows the prediction of the single map hypothesis (the bisector) and of the dual map hypothesis (gray areas): the observed *FIP* were fairly in agreement with the dual map hypothesis, while at the same time they systematically overshot the bisector. Two linear models were fitted, where the raw *FIP* was a linear function of the predictions of either hypothesis. A goodness-of-fit test confirmed that the dual map hypothesis fitted the data significantly better ($\chi^2_{(0)} = 14.11$, $p < 0.001$).

It is also worth commenting on the role of visual and haptic feedback in this experiment. Participants' performance during the *NF* trials did not improve, despite receiving consistent visual and haptic feedback 50% of the time and many repetitions across four experimental blocks. Indeed, we found no effect of the block on neither the *FGA* of the *NF* grasping task ($F_{(3,15)} = 1.85$, $p = .18$) nor the *FHP* of the *NF* pinching task ($F_{(3,15)} = 0.41$, $p = .74$). This underscores two important findings. One, that the visuomotor system depends critically on online feedback, especially when very few cues to depth are available, in agreement with the results of experiment 3. Two, that both grasps with and without feedback were planned based on wrong estimates of the 3D scene: the trajectories of the *NF* grasps and the biases we found therein inevitably reflected the same planning behind the *FF* grasps, because participants did not know whether they were going to receive feedback prior to the movement onset.

Figure 4.16 shows the bird's eye view of the trajectories of *FF* and *NF* trials (black and gray lines respectively) superimposed at each distance of each task (top row, pinch; bottom row, grasp). The reader can observe the gradual drift from overestimation to underestimation during the *NF* trials, due to visuospatial compression, as movements were aimed at more and more distant targets. It is also interesting to note how the trajectories of the trials with and without feedback were very similar to each other up to the last moments of the reach, that is, when the visual feedback of the fingers was eventually exploited to correct the errors at planning. When comparing the average trajectories of the trials with and without feedback we also noticed one more piece of evidence in support of the dual map hypothesis. At one particular distance, 440 mm, the *FF* and the *NF*

trajectories of the pinch were remarkably similar, implying that little or no online correction was produced when reaching at that location (highlighted panel in Fig. 4.16). This suggests that around 440 mm the estimation of egocentric distance was almost veridical at planning, and therefore, according to the dual map hypothesis, subjects should estimate the object's depth in an almost accurate fashion when grasping around that distance. In agreement with this hypothesis, the most accurate *FGA* was found for the two grasps were the thumb was aimed at 440 mm (see figure 4.14).

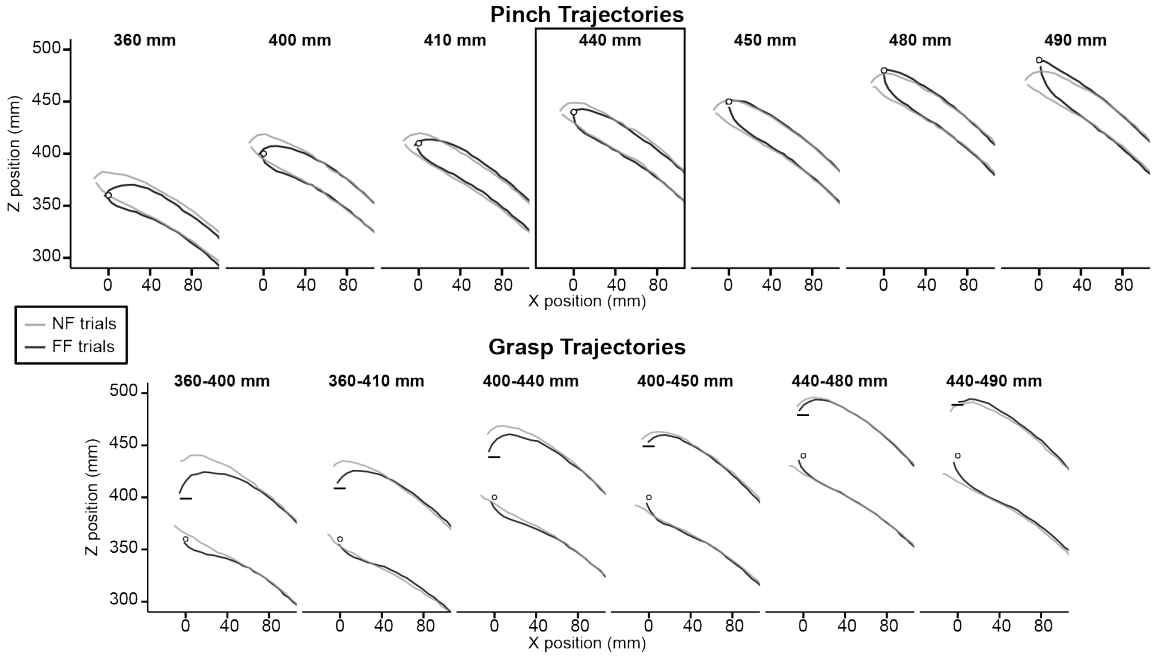

Figure 4.16. Bird's eye view of the trajectories of the pinch (upper row) and of the grasp task (bottom row). Each panel shows the reaching task at one particular distance, labeled at the top. The x and z positions of index and thumb are in abscissa and ordinate respectively. The trajectories of the trials with feedback are in black, those of the trials without feedback are in gray. The highlighted panel in the upper row (440 mm – pinch) corresponds to the case where *FF* and *NF* trajectories overlapped the most, indicating veridical estimation of egocentric distance: in agreement to visuospatial compression, the figure shows that subjects overshot distances that were smaller than 440 mm, and undershot distances greater than 440 mm.

Finally, we analyzed the evolution of the grip aperture during the *FF* grasps with the same procedure utilized in experiment 3, and we found an identical pattern of results: depth underconstancy affected the grasp until the final stages of the reaching (Fig. 4.17) (MGA: $F_{(1,28)}$ =

11.64, p = .002; 20 mm: $F_{(1,28)}$ = 5.05, p = .03; 10 mm: $F_{(1,28)}$ = 19.09, p < .001; 5 mm: $F_{(1,28)}$ = 4.98, p = .03).

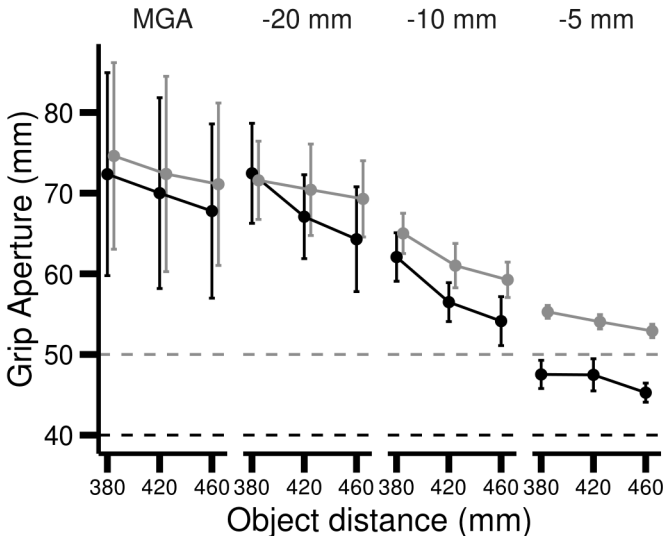

Figure 4.17. Grip Aperture profile taken at four consecutive points along the trajectory of the *Full-Feedback* trials in experiment 4 (at the time of the MGA, and when the thumb was 20, 10 and 5 mm away from its final position), as function of the object's distance for each depth (40 mm in gray, 50 mm in black).

In conclusion, the results of the fourth experiment highlighted three important aspects regarding the planning of a grasping movement. First, the planning can take advantage of both egocentric and allocentric maps of the visual space, which are selectively used to determine the endpoint of each grasping digit: the thumb position is encoded in egocentric coordinates, whereas the index position results from a combination of both egocentric and allocentric processing of visual information. Second, both egocentric and allocentric frames of reference are affected by the same distortion of the visual space: the thumb's contact point is consistent with visuospatial compression, the index's contact point is consistent with depth underconstancy (which stems from visuospatial compression). Third, visual and tactile feedback from the hand is actively involved in the execution of the movement, since it helps to correct the errors at planning in real time during the reaching.

CHAPTER 5

Study 2: Depth cue combination in perception and action

5.1 Rationale of the study

Suppose that an observer views an object at one of two different distances over the course of successive presentations, and each time she is either asked to judge its shape or to grasp it. One way to quantify the perceived shape is by asking to report the apparent depth and width and calculate their ratio. If the 3D structure is always correctly estimated, then the depth-to-width aspect ratio should be constant irrespective of the viewing distance. The same method can be applied to measure shape constancy during grasping. The participant grasps an object along either the depth or the width dimension, and then the in-flight grip apertures of the two movements are compared. If the same estimates deployed for perception are also used to guide action, the aspect ratio should be constant during grasping too.

Consider now the object in figure 2.1A: a cylinder with parabolic cross-section. If the cylinder's shape is specified only by binocular information, judgments of depth are expected to suffer from underconstancy: near the body, disparities are scaled with respect to an overestimated viewing distance, resulting in an apparently deeper object; far from the body, disparities are scaled by an underestimated viewing distance, thus the same object looks flatter (Fig. 5.1, “Shape from Stereo” panel). As a result, if the estimated depth of a stereo-defined stimulus is plotted against the viewing distance, within a limited range of space the best fitting curve should have a negative slope (“Stereo” curve in fig. 5.1B).

Suppose then, that the same cylinder either rocks around the x axis, or is covered with a grid wrapped around its surface. In both these cases, the introduction of motion or texture cues allows the observer to gather richer and more reliable information about the layout of the 3D structure

and its curvature ("Motion/Texture" curve in fig. 5.1B). Critically, these new signals are independent of the viewing distance: in fact, as seen in figure 2.2, although the retinal projection of the stimulus scales with distance, neither the motion gradient nor the texture gradient change along the line of sight. As a consequence, adding either monocular cue should increase the accuracy of depth estimation both near and far from the body. In other words, judgments of the depth of a multiple-cues object should exhibit less underconstancy. If the estimated depth of the stereo-motion or the stereo-texture object is plotted against the viewing distance, the slope of the best fitting function should be closer to zero ("Combined" curve in fig. 5.1B).

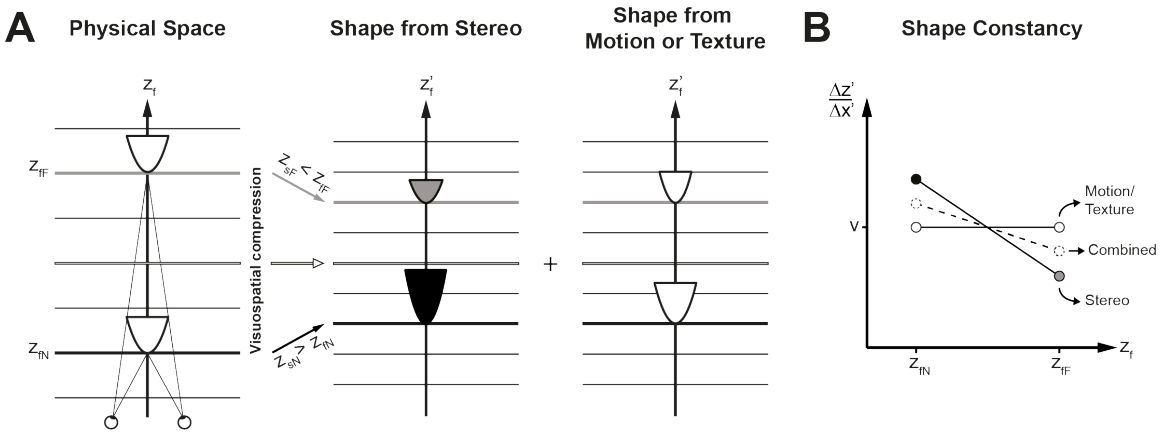

Figure 5.1. Predictions of the study. (A) "Physical Space": bird's eye view of the experimental design (a parabolic cylinder was viewed at two distances, $z_{fN}$and z_{fF}). "Shape from Stereo": visuospatial compression (arrows connecting the previous panel with the current) is predicted to make the objects appear at different distances than their true ones. As a consequence, the aspect ratio of the near stimulus should be greater than veridical, whereas the aspect ratio of the far stimulus should be smaller than veridical. "Shape from Motion of Texture": visuospatial compression is predicted to impact the scaling of motion and texture gradients too, but the resulting aspect ratio of the stimuli should be constant nonetheless. (B) The predicted aspect ratio is plotted as function of the viewing distance: the predictions for the stereo cue are represented by the filled circles, those of the motion/texture cues by the open circles. The point v on the y axis represents the veridical aspect ratio. Negative slope represents depth underconstancy. The combination of stereo and motion/texture should attenuate depth underconstancy.

In Experiment 1, we asked participants to manually estimate the depth of the parabolic cylinder presented at one of two distances (300 and 450 mm), or to reach-to-grasp it front-to-back. The stimulus was defined by dots randomly distributed on its surface and it was either static (stereo condition) or rocking around the x axis (stereomotion condition). Consistent with visuospatial

compression, we predicted to find depth underconstancy in the stereo condition ("Shape from Stereo" panel in figure 5.1A, and negative slope of the "Stereo" curve in fig. 5.1B). We also expected a significant reduction of the bias, consistent with increased shape constancy, in the stereomotion condition ("Combined" curve in fig. 5.1B). Finally, we hypothesized that the same pattern of results would affect perception and grasping tasks equally.

Experiment 1 was conducted assuming that the stimuli's width was constant at the two viewing distances, based on the results from our previous experiments. We verified this assumption in a control experiment (1a), where participants were asked to judge the width of the same parabolic cylinders used in Experiment 1, and to grasp the objects along the width as well. Consistently with our previous data, depth cues and distance did not affect the estimation of width, neither during perception nor grasping.

In Experiment 2, we repeated the same exact design of the first experiment, only this time we used texture information instead of motion parallax.

5.2 General Methods

Participants

105 students (36 males, 71 females) participated to the experiments [experiment 1, n = 39; experiment 1a, n = 20; experiment 2, n = 41; experiment 2a, n = 13], of whom 7 did both experiments 1 and 1a, while another person did all four experiments. All participants self-identified as right-handed with normal or corrected-to-normal vision. The experiments were conducted at Brown University, where the subjects were paid $8/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

Subjects sat with their head on a chin rest installed along one of the short sides of a rectangular table, facing a semi-transparent mirror which was slanted 45 degrees with respect to the frontoparallel 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 was installed on the table: one actuator carried the monitor in 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 located behind the mirror and carried a platform along the z and the y axis. A combination of small linear actuators attached to a stepper motor (Phidgets Inc.) was installed on the movable platform. The primary function of this arrangement of motors was to position physical objects in the workspace and align them with the virtual imagery, so that observers could "touch" and interact with the virtual stimuli with their hand.

Using shutter glasses (NVIDIA 3D Vision® 2 wireless glasses) synchronized with the refresh rate of the screen (85 Hz), participants saw 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.). We used an Optotrak 3020 Certus motion capture system and small infrared emitters (Northern Digital) to record index and thumb trajectories at a sampling frequency of 85 Hz. A calibration procedure allowed us to track the 3D position of the index and thumb's finger pad, 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. In experiment 2 and 2a, additional material and texture features were first generated using Blender (www.blender.org) and then rendered with OpenGL.

In all experiments, the stimulus was a parabolic cylinder, presented at eye height and oriented vertically such that the convex side faced the observer. The top and bottom edges of the cylinder were occluded by two rectangular flat surfaces, so that the overall visible portion was 6 cm in height. The cylinder was colored in red, while the occluders were dark red. The depth and width of the stimulus were varied according to the equation:

$$\Delta z = 0.2 \cdot \Delta x^2 \qquad [1]$$

The exact values and graphic specs of the stimuli will be described in detail in each experiment's methods section. The width of all stimuli subtended a small visual angle, to maximize the role of vergence for the estimation of distance, at the expense of the other cues (Bradshaw et al., 1996). The accommodative distance, determined by the reflected monitor's surface, coincided with the front surface of all stimuli. The maximum distance between the surface of the monitor and the simulated back edge of the stimuli was 5 cm. This made sure that the vergence-accommodation conflict was negligible (Watt et al., 2005; Frisby et al., 1995; Buckley and Frisby, 1993).

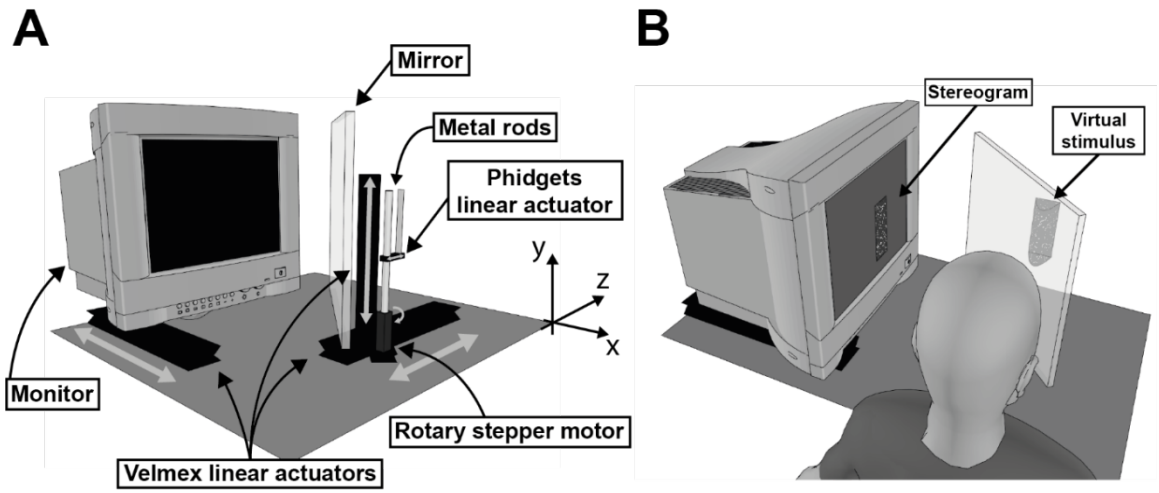

Figure 5.2. (A) Schematic of the setup used in all four experiments. Monitor and physical objects (when used) were moved by a series of linear actuators, gray arrows indicate the direction of motion provided by each motor. (B) In all experiments, the stereogram of the stimulus rendered on the monitor 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.

All experiments comprised of a perceptual task and a grasping task (with the exception of experiment 2a, which consisted only of a grasping task). In the grasping task, participants saw the virtual rendering of their index and thumb's finger pads, represented as luminous red dots of about 5 mm in diameter. This visual feedback about the grasping hand, in combination with the haptic feedback received at the end of the grasp, made the experience of interacting with the stimuli highly compelling. Subjects adapted rapidly and naturally to the virtual environment after a quick training session that was performed prior to starting the actual experiment.

5.3 Experiment 1: Adding motion to stereovision

In the first experiment we tested whether the addition of motion to stereo information improves shape estimation during perception and grasping. Subjects saw a disparity-defined object that was either static or rocking around the x axis, appearing at one of two distances. We tested three hypotheses: 1) depth underconstancy should occur when estimating the shape of the static (stereo) object, because of visuospatial compression (*depth underconstancy hypothesis*); 2) depth underconstancy should affect judgments of the rocking object to a lesser extent, because the additional 3D information provided by the motion gradient should improve shape constancy (*shape-from-multiple-cues hypothesis*); 3) grasping should be characterized by both previous effects, because our previous results indicate that 3D information is shared between perception and action (*perception-action identity hypothesis*).

5.3.1 Methods

Stimuli

All stimuli were parabolic cylinders represented by a cloud of red dots randomly distributed on its surface. We tested two viewing conditions: stereo versus stereomotion. In the stereo condition, the target object appeared vertically oriented and static; in the stereomotion condition, the same

object also rocked around the horizontal axis, describing a 15 degrees angle of travel (± 7.5 deg around the vertical position) at a constant angular velocity of 60 deg/sec**.** The depth of the stimulus was varied across four discreet values (20, 30, 40 and 50 mm), which corresponded to four widths calculated through equation [1] (20, 24.5, 28.3 and 31.6 mm). This method ensured that the curvature of the object was constant for all four stimuli.

The stimuli were presented at two viewing distances: 300 and 450 mm. The viewing distance corresponded to the egocentric distance of the front edge of the stimulus.

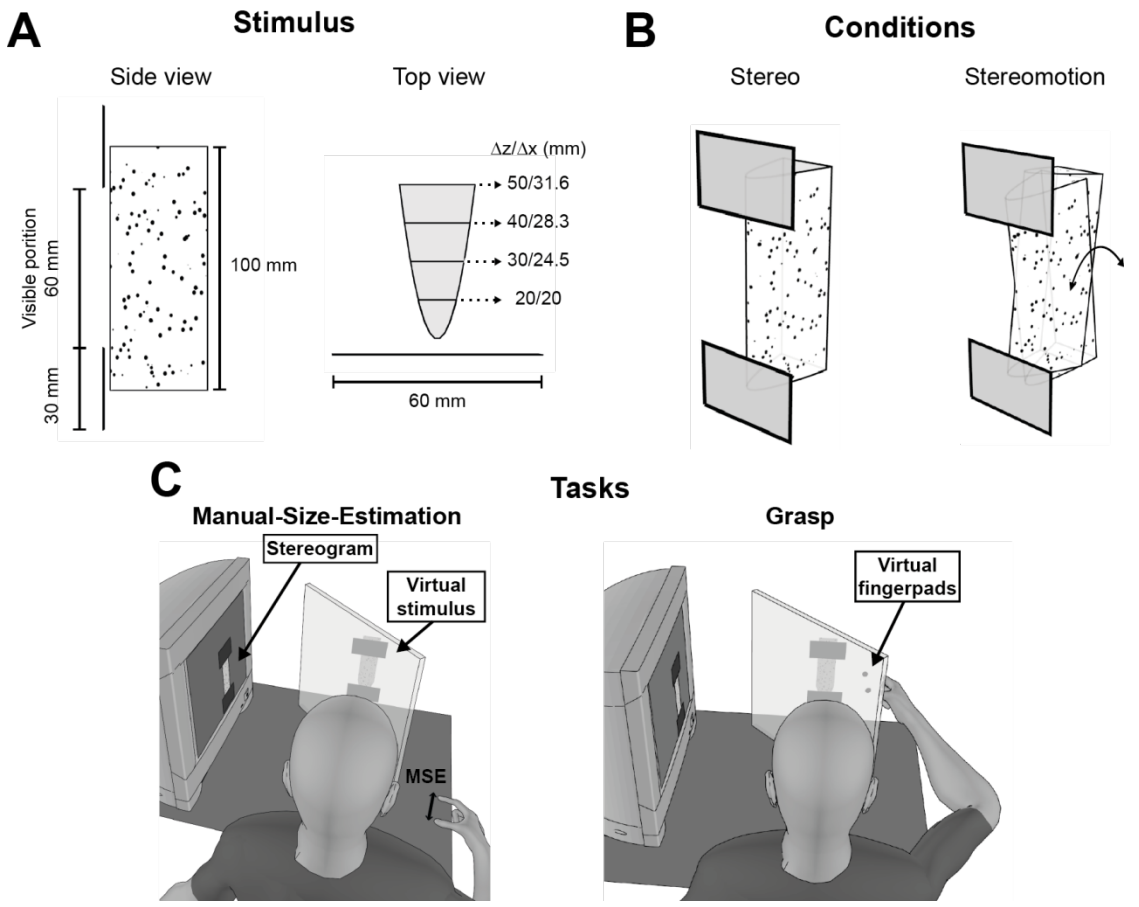

Figure 5.3. Methods of Experiment 1. (A) The stimulus was a parabolic cylinder positioned with the tip facing the observer at one of two distances (300 and 450 mm) from the cyclopean eye. Two occluders allowed the view of only a portion of the stimulus, which was 60 mm wide in height (Side view panel). The Top view panel shows the four shapes used in the experiment, shown by their depth-to-width ratio. All four shapes had the same curvature. (B) Schematic representation of the viewing conditions: left, Stereo (the object was static); Stereomotion (the object rocked around the x axis describing an angle of 15 degrees total (±7.5 from vertical position) at a constant angular speed of 60 deg/s. (C) Tasks: left, Manual-Size-Estimation; right, grasp.

Procedure

Participants were asked to perform two tasks: a perceptual task and a grasping task. In the perceptual task, subjects opened their index and thumb apart until they felt that the separation between the digits matched the perceived front-to-back extent of the stimuli (Manual-Size-Estimation, *MSE*). The stimulus appeared automatically at the beginning of each trial, after which the subject moved their right hand away from a "start" position and performed the *MSE*. When they felt their response to be appropriate, they hit a key on a numeric pad to record their judgment and move on to the next trial. Subjects could never see their hand while doing the task. There was no time limit to perform this task.

In the grasping task, subjects reached-to-grasp the target object, by landing with their thumb and index on the front and back edges of the object respectively. While moving the hand towards the stimulus, participants could see the virtual 3D rendering of their index and thumb's finger pads. The simulated index disappeared once it passed behind the object, mimicking the occlusion occurring when grasping a real object. Prior to the beginning of each trial, the system of linear actuators positioned a tiny metal cylinder and a metal plate such that they would align exactly with the rounded front tip and the flat back edge of the stimulus. Thus, at the end of the grasp participants actually touched two physical objects corresponding to the contact locations of the two grasping digits. The resulting experience was described as compelling and natural as a grasp in the real world by all subjects. The stimulus appeared automatically at the beginning of each trial. Subjects could observe the object for as long as they wanted, and then they performed the grasp. After the onset of the movement, they were given 1800 msec to complete the task. Once the object was contacted, typically within 1200 msec, subjects were instructed to keep holding it until it disappeared, and then return to the start position, which triggered the onset of the next trial. If the object was rocking (stereomotion condition), it stopped moving as soon as the fingers touched it.

The experiment consisted of two sessions, one for each task, which were ran in two separate days. Each experimental session comprised of four brief blocks. In each block, participants performed either the *MSE* or the grasp task while seeing the stimuli at one distance, 300 or 450 mm. The distances were alternated across the four blocks following an ABAB design and the order of presentation (300/450/300/450 or 450/300/450/300) was counterbalanced across subjects. Depending of which sequence was picked for the perceptual task, the other sequence was used for the grasping task.

Overall, the experiment was a full within-subjects 2x2x4x2 factorial design, with the following independent variables as factors: 1) the depth information specifying the shape of the stimulus (stereo vs stereomotion); 2) the viewing distance (300 and 450 mm); 3) the depth of the stimulus (20,30,40,50 mm) and 4) the task (*MSE* or grasp). Over the course of eight blocks distributed across two experimental sessions, participants performed a total of 320 trials, since each individual block included 40 trials (2 depth cues by 1 distance by 4 depths by 5 repetitions).The trials were pseudo-randomized within each block: participants performed five adjacent strings of 8 unique trials (2 depth cues by 4 physical depths). In every string, the order of the 8 trials was randomized. This method prevented that a particular combination of unique trials would be presented more often than the others due to chance.

5.3.2 Results and Discussion

The results from the stereo condition of the perceptual task replicated the depth underconstancy effect that we had previously found, thus confirming the *depth underconstancy hypothesis*. If its shape was specified only by binocular disparity, the stimulus was reported to look deeper when viewed at 300 mm than at 450 mm, as shown in the “Stereo” panel of figure 5.4A. To highlight this bias, we plotted the average *MSE*, collapsed across the four object’s depths, as function of the viewing distance (gray data points and line in fig. 5.4B). As expected, the best fitting curve had a significant negative slope (-0.018; $F_{(1, 38)} = 8.89$, $p = 0.005$).

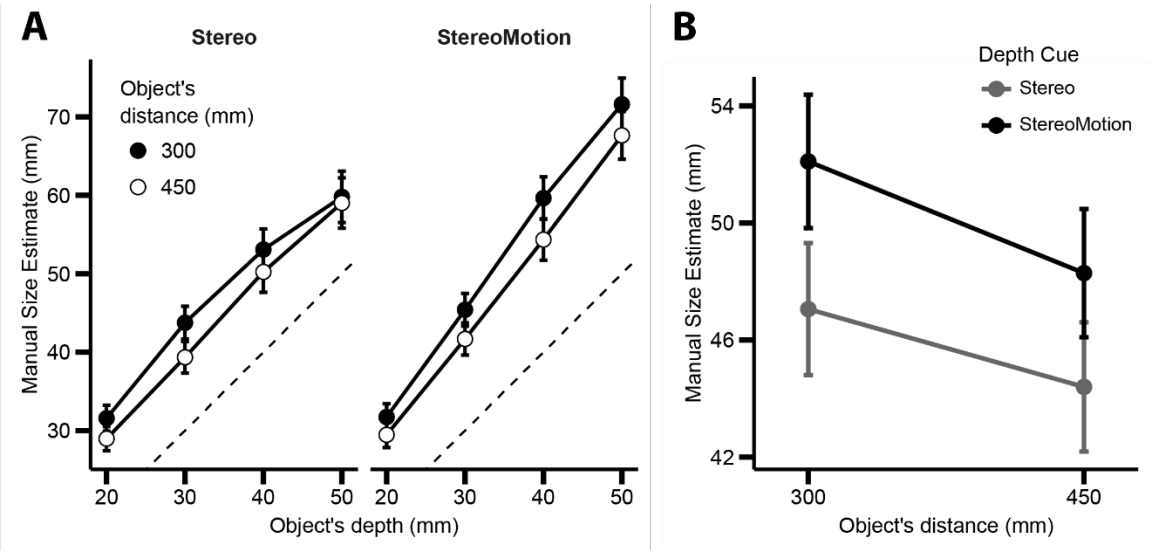

Figure 5.4. Perceived depth measured in the *MSE* task of Experiment 1. (A) *MSE* as function of the object's depth showed for each distance (300 mm, filled circles; 450 mm, open circles). Filled circles above the open circles represent depth underconstancy. The dashed line indicates veridicality. Left panel, results from the Stereo condition. Right panel, results from the Stereomotion condition. (B) Same results but collapsed across object's depth and presented as function of the object's distance. The stereo condition is in gray, the stereomotion condition is in black. Negative slopes represent depth underconstancy.

The *shape-from-multiple-cues hypothesis* was instead dismissed by the data, as depth underconstancy did not diminish as a result of adding a motion gradient to the stimulus. This is visible in the "StereoMotion" panel of figure 5.4A: in the stereomotion condition, participants still reported seeing the stimulus as deeper at 300 mm than at 450 mm, despite the availability of motion information. Even more so, the magnitude of the bias actually increased. The slope of the curve fitting the average *MSE* as function of the object's distance was -0.025 (Fig. 5.4B, black data points and curve).

Overall, a repeated-measure ANOVA on the *MSE* found a significant main effect of object's depth ($F_{(3, 114)} = 193.29$, $p < 0.01$), object's distance ($F_{(1, 38)} = 18.43$, $p < 0.01$). In addition, the analysis of the ANOVA table showed two more interesting findings.

First, we found a significant main effect of depth cue ($F_{(1, 38)} = 36.64$, $p < 0.01$) and a significant interaction between depth cue and object's depth ($F_{(3, 114)} = 35.77$, $p < 0.01$). The estimated depth increased with the physical depth more rapidly in the stereomotion condition than in the stereo

condition, causing the multiple-cues stimulus to appear on average deeper than the disparity-defined stimulus (Fig. 5.6A). Assuming that the object's width was constant along the z axis, we hypothesized that flatness cues biased the judgment of shape-from-stereo and caused a severe underestimation of the object's depth (Fig. 5.8). If that were truly the case, then it is possible that the addition of motion information helped the observers recover a closer-to-veridical shape at both viewing distances, although not enough to compensate for the underconstancy bias. To test this hypothesis directly, one should calculate the exact aspect ratio between the estimated depth and width of the stimulus: as a result, this ratio should be closer to the veridical value in the stereomotion condition than in the stereo condition. We could not perform this test based only on the data from this experiment, as we did not have a direct measure of the estimated width. We ran a control experiment (1a) to investigate this issue.

Second, we found a nearly significant interaction between object's depth and distance ($F_{(3, 114)}$ = 2.52, p = 0.06) and a significant three-way interaction between object's depth, distance and depth cue ($F_{(3, 114)}$ = 2.91, p = 0.04). Both these effects are related to the fact that depth estimation was poorer at the near distance for the deepest object (50 mm) in the stereo condition (this is visible in the "Stereo" panel of figure 5.4A). Subjects found it difficult to fuse the disparities when that object was close to the body, therefore in some cases they said that it looked flat. This bias vanished at the farther distance, since the same object projected smaller disparities which were easier to fuse. The bias disappeared in the stereomotion condition as well, since the motion gradient allowed to recover depth information whenever depth-from-disparity was poor. For this reason, we decided not to use the 50 mm object in the next experiments.

To analyze the performance in the grasp task, we looked at the evolution of the Grip Aperture (*GA*) during the entire reaching trajectory. In particular, we analyzed the final stage of the movement, from the time of the Maximum Grip Aperture (*MGA*) to the moment when the fingers finally contacted the target object. It is during this time window, in fact, that participants typically refine the scaling of the *GA* according to the estimated size of the target object, making it possible

to measure the latent parameters of vision-for-action. Furthermore, it is in the final stages of a grasp that subjects can actively exploit full online control to guide their hand, and implement corrections if necessary.

In agreement with the *perception-action identity hypothesis*, the results from the grasping task mirrored those of the perceptual task quite remarkably.

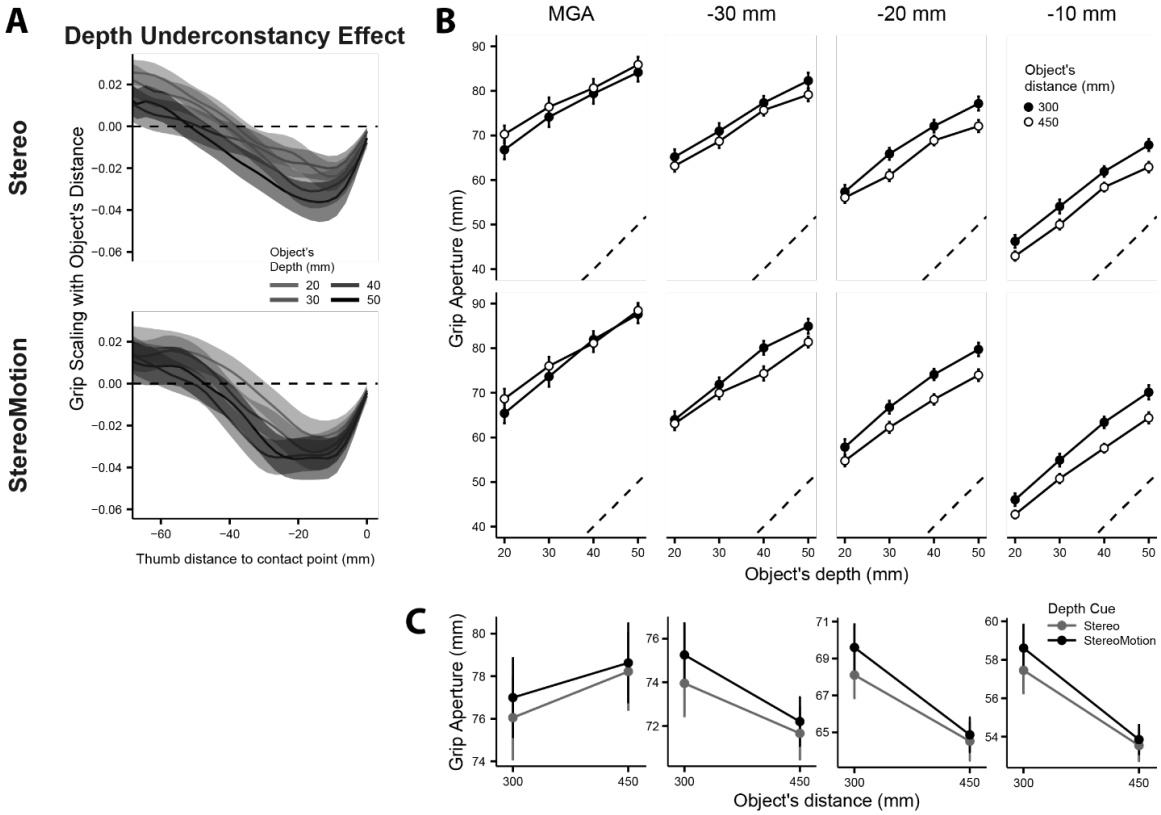

Figure 5.5. Results from the grasping task of Experiment 1. (A) In the ordinate, we plotted the slope of the function relating the *GA* to the object's distance, as function of the distance of the thumb to its contact point. The data is presented for each object's depth, coded with a grayscale. Areas around the curves represent ±1 *SEM*. Upper panel, stereo condition; lower panel, stereomotion condition. Negative values indicate depth underconstancy. The horizontal dashed line corresponds to absence of bias (slope of zero). (B) Four snapshots of the *GA* were taken along the trajectory, and in each of them the *GA* was plotted as in figure 5.4A. From left to right, grip aperture taken at the time of the *MGA*, and when the thumb was 30, 20 and 10 mm far from its contact point. (C) Same data of panel (B) collapsed across object's depth and plotted as in figure 5.4B. Notably, the *MGA* shows an unexpected overconstancy bias, in sharp contrast with the rest of the trajectory.

Consistent with depth underconstancy, we found that, during the last phases of the movements in the stereo condition, the *GA* was bigger when subjects aimed at the object near to the body than

when reaching-to-grasp it at the far distance. To determine this result, we calculated the slope of the function relating the *GA* to the object's distance and observed its evolution along the trajectory. In the final stages of the grasp, this slope sharply became negative, and the bias was corrected basically at movement completion (Fig. 5.5A, upper panel). To better compare grasp and perception with one another, we further took four snapshots of the *GA* at different instants along the reaching trajectory: at the time of the *MGA*, and when the thumb was 30, 20 and 10 mm away from its contact point. At each of these instants, we plotted the *GA* as function of the object's depth for each distance of the stereo condition (Fig. 5.5B, upper row panels). The progression is clear: as the subjects got closer to the object with their hand, and shaped the *GA* to prepare for grasping, the depth underconstancy bias arose steadily, culminating when the thumb was roughly 20 mm far from the front surface of the stimulus. The negative slope of figure 5.5A is also visualized for each of the four snapshots in fig. 5.5C: the results shown by the gray data points and curves are identical to those found in the *MSE* task.

The results in the stereomotion condition disconfirmed the *shape-from-multiple-cues hypothesis* for grasping as well, as shape constancy did not improve despite more depth information was available. Instead, as in the perceptual task, the depth underconstancy bias became even stronger than in the stereo condition. We took the data of the 20 mm snapshot and performed a repeated-measure ANOVA on the *GA*. The results table revealed a significant main effect of object's depth ($F_{(3, 114)} = 198.46$, $p < 0.01$) and object's distance ($F_{(1, 38)} = 23.68$, $p < 0.01$).

The ANOVA table also displayed a significant main effect of depth cue ($F_{(1, 38)} = 5.46$, $p = 0.02$), as the *GA* was greatest in the stereomotion condition. Furthermore, the trajectory analysis showed that this effect was modulated by the object's depth: as can be seen in figure 5.6B, the effect of the cue increased for the deeper objects. Similarly, figure 5.6C shows that the slope of the grip scaling function in the stereomotion condition was greater than in the stereo condition. In other words, the grip aperture increased with the object's depth more rapidly when the stimulus's shape was specified by two cues. These two results matched the interaction found in the perceptual task

between depth cue and object's depth (Fig. 5.6A), remarking the identity between 3D vision for perception and 3D vision for action.

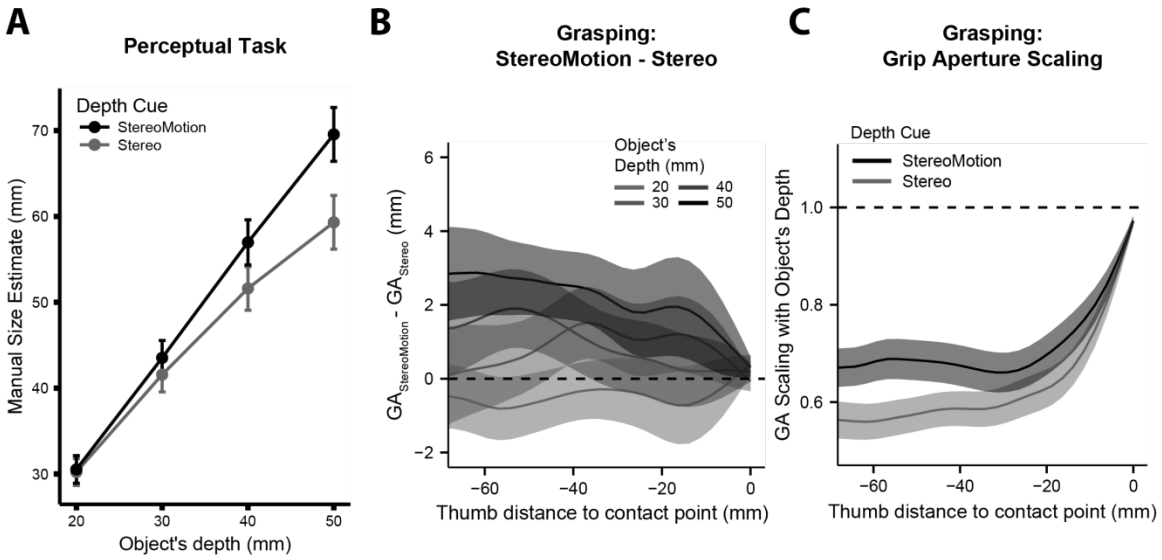

Figure 5.6. Comparison between perception and action. (A) Same results of figure 5.4A (*MSE*), collapsed across distance and presented for each combination of depth cues (black: stereomotion; gray: stereo): an interaction between depth cue and object's depth is clearly visible. (B) Effect of the depth cue on the grip aperture in the grasping task. In ordinate we plotted the difference between the *GA* in the stereomotion condition and the *GA* in the stereo condition, as function of the distance of the thumb to its contact point. Color codes for object's depth (the deeper the object, the darker the curve). Positive values mean that the stereomotion object was grasped as deeper than the stereo object. Areas around the curves represent ±1 *SEM*. Similar to the interaction shown in 7A for the perceptual task, the effect of the depth cue in the grasp was greater for deeper objects. The dashed line corresponds to a difference of zero. (C) Grip scaling function (slope of the function relating the *GA* with the object's depth) as function of the distance of the thumb to its contact point (black, stereomotion condition; gray, stereo condition). The dashed line corresponds to a slope of 1 (veridical scaling). Similar to the interaction shown in 7A, the *GA* was more responsive to the object's depth in the stereomotion than in the stereo condition.

5.3.2.1 Effects of visual control on the scaling of the grip aperture

In figure 5.6C it is also possible to notice that the *GA* scaled reasonably well with the object's depth (slope greater than .8) only very late in the reaching, when the thumb's distance to the stimulus was much smaller than 20 mm. We reasoned that this late scaling might had been caused by the sudden disappearance of the index towards the end of the grasping movement. During the final approaching phase of the action, participants could not use vision to guide the index because the stimulus occluded it. We hypothesized that this lack of visual control during the most

important phase of the grasp had two important consequences. First, it was the reason why the biases were stronger at the end of the movement: when the forefinger suddenly disappeared, the participants positioned it where the back surface would be if the object had the depth estimated at planning, thus exhibiting depth underconstancy. Second, after guessing the final index position, participants simply kept pulling the index until it touched the object, which made the slope of the scaling function suddenly increase. If this hypothesis is true, then the depth underconstancy bias should disappear when both thumb and index can be guided visually until movement completion, because the observer no longer has to guess the position of the index. We tested this prediction in Experiment 2.

5.3.2.2 Sensorimotor correction from haptic feedback

The visuomotor system receives shape information also through the tactile sensation generated at the end of the grasp when contacting an object. Since the results showed that the stimulus's 3D structure was systematically biased at planning, we hypothesized that the haptic feedback could trigger an error signal, whenever the physical object's depth felt with the hand was either greater or smaller than expected.

We assumed that the grip aperture in a given trial depended on both an estimate of the current object's depth ($\Delta z'$) and on an internal state (M) that mapped $\Delta z'$ to a specific configuration of muscles and joints. As explained previously (Fig. 5.1), $\Delta z'$ required information about the object's physical distance and depth. Our design was such that the distance was fixed within each of the four blocks of the grasping task, hence in every trial $\Delta z'$ changed only as a function of the physical Δz and the amount of depth cues available. When the object was touched at the end of the trial, the difference between the expected and the physical depth was registered as an error signal and used to update the mapping between the visual estimate and its corresponding motor plan. Therefore, according to our hypothesis, the *GA* depended on both the current set of depth

cues (feeding the generation of Δz') and the depth cues in the previous trial (that affected the internal state M).

In agreement with our rationale, the *GA* was systematically greater than average during the trials that immediately followed the presentation of a stereo stimulus, and likewise, it was smaller than average after touching a stereomotion stimulus (Fig. 5.7). We ran the same repeated-measures ANOVA on the 20mm-far-*GA* of before, but this time we also included a factor coding for the depth cue of the previous trial and we found a main effect ($F_{(1, 38)} = 6.13$, $p = 0.02$). This "previous trial" effect suggests that the visuomotor system was indeed able to integrate the tactile sensation, compare it with the expected depth from the planning, and produce an error signal: since the stereo object was seen as flatter in comparison with its stereomotion homologous, the motor plan of the subsequent action was updated to produce a larger grip aperture to compensate for the error. The opposite happened after touching the stereomotion object.

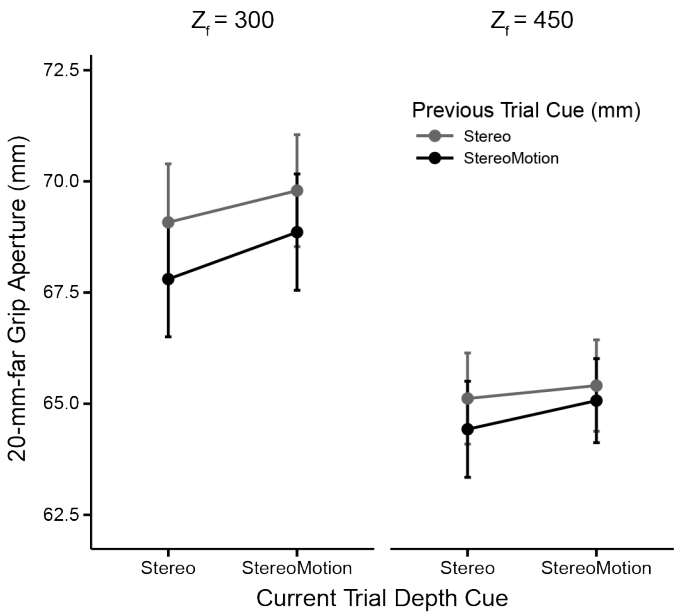

Figure 5.7. Previous trial effect during grasping. In ordinate is the *GA* taken when the thumb was 20 mm far from its contact point, plotted as function of the depth cues during trial N, and presented for each cue of trial N-1 (gray, previous trial = stereo; black, previous trial = stereomotion). Left panel, trials with object's distance of 300 mm. Right panel, trials with object's distance of 450 mm. The *GA* of the current trial was systematically larger after touching the stereo object and smaller after touching the stereomotion object.

5.3.5 Conclusions

In summary, Experiment 1 confirmed two fundamental hypotheses about the way in which depth information is processed during perception and grasping: i) when only binocular disparities and vergence signals are available, objects close to the body look deeper than objects away from the body; ii) this lack of depth constancy affects the execution of grasping movements as well as perception. In addition, results also suggested that the availability of additional shape information, such as the motion gradient generated by the rhythmic oscillation of an object along its x axis, did not improve shape constancy. On the contrary, the depth underconstancy bias seemed to become even stronger, and the stereomotion stimulus was reported to look deeper than its disparity-defined analogous.

In a subsequent control experiment, we investigated whether the increase of the estimated depth corresponded to a more accurate recovery of 3D shape.

5.4 Experiment 1a: Shape constancy in perception and action

In this control experiment, we wanted to verify empirically that size estimation is reasonably constant along the line of sight, at least within the distances used in this study. This test was a necessary requisite for the correct interpretation of the data from the previous experiment, in particular for two reasons.

On the one hand, we wanted to make sure that the variable we used in Experiment 1 actually measured what we were interested in. In Experiment 1, subjects were asked to estimate the front-to-back extent of stimuli presented at two distances. Our rationale was the following: under the assumption that 2D size (width or height) would be invariant within the space delimited by the two viewing distances, the estimated depth can be taken as a proxy for measuring shape constancy. Although the results of the second study of this thesis had previously supported this hypothesis, we tested it again using the stimulus of the current study. Furthermore, we wanted to

compare size estimation during grasping as well as perception, so we asked participants to also grasp the object along the vertical dimension.

On the other hand, the findings of Experiment 1 suggested that the main theory behind the study (Fig. 5.1) may needed to be adjusted from its original formulation. In the previous experiment we found that i) subjects experienced depth underconstancy regardless of the depth cues specifying the stimulus's shape; ii) at the same time subjects reported seeing the stimulus as overall deeper when defined by stereo and motion cues then when defined only by stereo information.

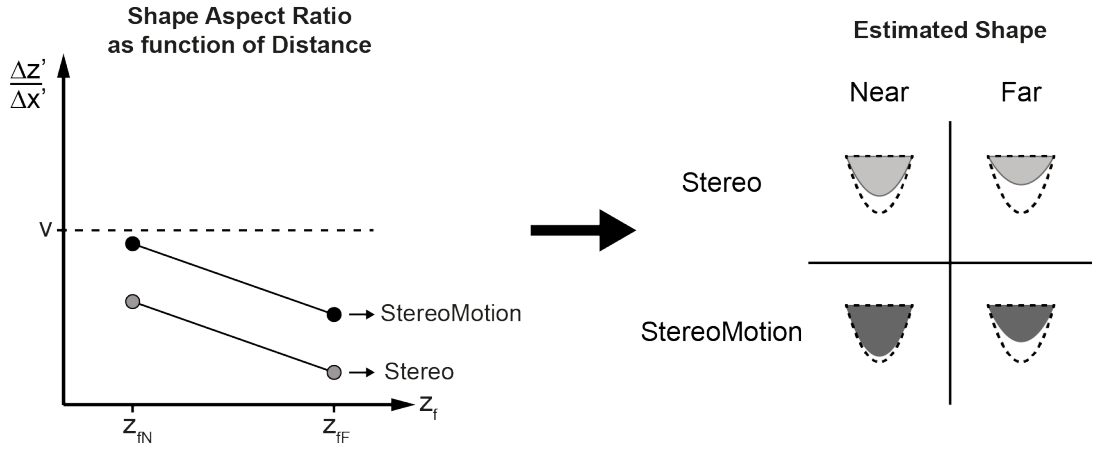

Figure 5.8. A possible interpretation of the results of Experiment 1. Left graph: if we hypothesize that the estimated aspect ratio of the stimulus ($\Delta z'/\Delta x'$) in the stereo condition was underestimated and affected by underconstancy (gray points in the graph are below the veridical value v, and the best fitting line has a negative slope), then the increase of the object's estimated depth in the stereomotion condition corresponded to an increase of the overall aspect ratio (black points), assuming that $\Delta x'$ was constant across distance and independent of the depth cues. In other words, the stereomotion object was seen as closer to veridical: in the right diagram, the dashed lines represent the true shape of the object and the filled areas the estimated shape.

Under the assumption that the primary goal of the visuomotor system is to recover the most accurate estimate of an object's 3D shape (Landy et al., 1995), and thus that the more depth sources are available, the closer to veridical the recovered shape will be, the above results can be reconciled as follows: when only binocular disparities are available, depth is underestimated and also undergoes underconstancy, likely because of interfering flatness cues (Fig. 5.8A). If size estimation is instead accurate and constant, then the depth-to-size aspect ratio of the stimulus's shape is smaller than its veridical value (represented in the figure by the point *v* on the ordinate),

and its relation with the viewing distance is best described by a curve having a negative slope (gray data points). The addition of motion information counteracts the flatness cues, yielding an improvement of the shape aspect ratio at both distances. However, at the same time it is possible that motion may not be strong enough to also compensate for the underconstancy bias (black data points in figure 5.8A; see also figure 5.8B).

5.4.1 Methods

Apparatus

The apparatus was identical to that used in Experiment 1, except that in the grasping task the Phidgets actuators were substituted by a carousel controlled by the stepper motor. Mounted on the carousel were four wooden blocks all roughly 1 cm by 1 cm in width and depth, but each of different height. The heights of the wooden blocks matched the widths of the four target objects, as calculated through equation [1]: 20, 24.5, 28.3 and 31.6 mm. In each trial, the carousel was positioned at a distance and rotated such to align the appropriate wooden block to the current virtual stimulus.

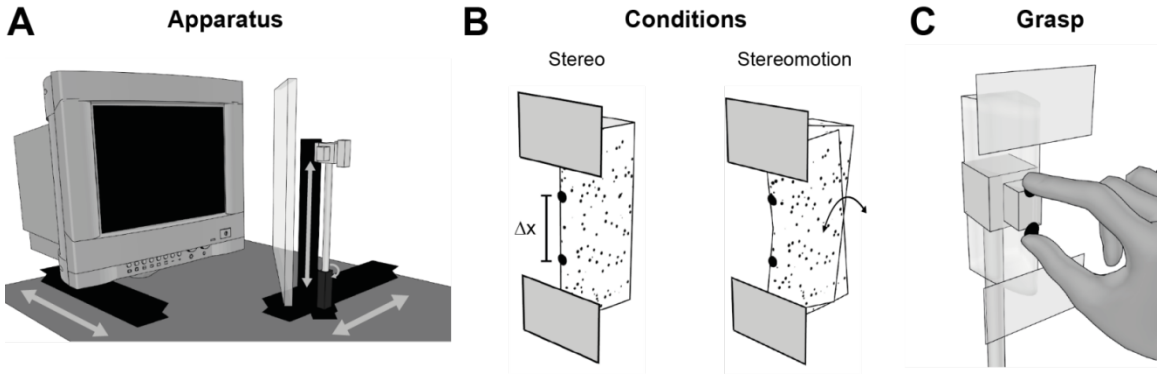

Figure 5.9. Methods of Experiment 1a. (A) The apparatus was identical to that used in Experiment 1, except for an arrangement of four wooden blocks that was mounted on the stepper motor. (B) Schematic representation of the viewing conditions: left, Stereo; right, Stereomotion (same as in Experiment 1). In both conditions, two large dots visible on the front tip of the stimulus defined a vertical separation equal to the width of the cylinder (Δx). (C) During the grasp, one wooden block was superimposed on the virtual image and its height corresponded to one of the four vertical separations (one for each stimulus's aspect ratio) defined by the large dots. When the subjects grasped that separation, they in fact contacted the upper and the lower edges of the matching wooden block.

Stimuli

The stimuli were identical to that of Experiment 1, with the following exception: an additional couple of large dots (5 mm in diameter) were vertically aligned on the stimulus's front tip. The separation between the two target dots matched the object's horizontal extent (the width). The depth and distances of the stimuli were the same used in Experiment 1.

Procedure

In this control experiment, subjects judged and grasped the stimulus's width instead of its depth. The procedure for the *MSE* task was identical in every respect to that of the first experiment, except that subjects had to manually estimate the distance between the two large dots visible on the object's front. The grasping task was also performed in the same fashion as before, except that this time, participants grasped the object vertically, by positioning the thumb and index on the lower and upper dot respectively. At the end of the movement, participants actually contacted the upper and lower edges of a wooden block that were aligned with the visual stimulus. To increase the realism of the task, when the Euclidean distance between either digit and the front surface was less than 2 mm, the portions of the parabolic cylinder above and below the target dots disappeared, so that subjects could complete the grasp by enclosing the target area with their fingers. If the object was rocking, it stopped when the fingers touched it.

The experimental design was identical to that of Experiment 1.

Data analysis

The analysis of the results comprised of two parts. First, we considered the data of the experiment on its own, and we looked at the effects of the manipulations on the dependent variables (*MSE* and *GA*) to test whether or not there was size constancy during perception and grasping. Second, we compared judgments of size with judgments of depth from Experiment 1, to study shape

constancy. In order to do so, we modeled the *MSE* as a simple linear function of the 3D property of interest:

$$MSE_p = a + b \cdot \hat{p} \quad [2]$$

where *a* is a parameter absorbing the offset related to the thickness of the finger and to variability in the calibration of the markers, *b* is a scaling parameter accounting for the mechanics of the effector, that is, of the hand (i.e., muscle resistance and joint constraints) and $\hat{p}$ is the pure estimate of size ($\widehat{\Delta x}$) or depth ($\widehat{\Delta z}$) that the observer intends to express. For any given shape, the theoretical estimated aspect ratio corresponds to the fraction $\widehat{\Delta z}/\widehat{\Delta x}$, which, after solving equation [2] by $\hat{p}$ for size and depth, becomes:

$$\frac{\widehat{\Delta z}}{\widehat{\Delta x}} = \hat{R} = \frac{MSE_{\Delta z} - a}{MSE_{\Delta x} - a} \quad [3]$$

We fitted the data from this control experiment with a linear mixed-effects model defined by equation [2], with intercept and slope as random components, and estimated the parameter *a*, under the assumption that size estimation is independent of distance and depth information (which was confirmed empirically, see the Results section). The analysis of the coefficients using the Kenward-Rogers approximation for degrees of freedom (Luke, 2016) revealed that *a* was in fact not significantly different from zero (-3.82, $F_{(1, 36)} = 1.63$, $p = 0.21$), thus equation [3] reduced to $MSE_{\Delta z}/MSE_{\Delta x}$.

Next, we estimated $MSE_{\Delta z}$ and $MSE_{\Delta x}$ from Experiment 1 and 1a respectively, for each distance (300 and 450 mm) and visual condition (stereo and stereomotion). Finally, we calculated $\hat{R}$ for all four combinations of distance and depth cue (corresponding to the four data points shown in the graph of figure 5.8).

5.4.2 Results and Discussion

As expected, the results from the perceptual task proved that size estimation was indeed constant between 300 and 450 mm (Fig. 5.10A). A repeated-measure ANOVA on the *MSE* found in fact

only a significant effect of the object's size ($F_{(3, 57)}$ = 313.32, p< 0.01). Logically, since retinal size is also scaled by an estimate of the viewing distance, participants did show a tendency to report the near object as larger than the far object, consistent with visuospatial compression, but the effect was negligible (.24 mm) and not significant ($F_{(1, 19)}$ = .39, p= 0.54). Contrary to the predictions, the ANOVA table revealed a significant effect of the depth cue ($F_{(1, 19)}$ = 7.09, p= 0.02). However, a closer look at the data showed that this result was entirely driven by the largest (and deepest) object, which for some reason elicited a greater response in the stereo condition with respect to the stereovision condition. As a matter of fact, this effect was actually negligible, because the average difference between the two conditions amounted to only .69 mm.

The trajectory analysis of the grasp movements showed an identical pattern of results: the *GA* scaled only in response to changes of the simulated size of the stimulus, and not with changes in depth cue information or distance. Figure 5.10B shows the results during two key instants along the trajectory: at the time of the MGA and when the thumb was 20 mm far from its contact point[1]. As expected by the *perception-action identity hypothesis*, the grip aperture in the final stages of the grasp was only modulated by the object's size ("-20 mm" panel of figure 5.10B). A repeated-measures ANOVA on the *GA* 20 mm before movement completion showed in fact only a significant effect of size ($F_{(3, 57)}$ = 59.63, p< 0.01), but neither of the depth cue ($F_{(1, 19)}$ = 0.001, p= 0.97) nor of distance ($F_{(1, 19)}$ = 0.02, p= 0.87). Interestingly, the same analysis on the *MGA* revealed a significant effect of distance ($F_{(1, 19)}$ = 5.96, p= 0.02), although in the direction opposite to the underconstancy bias: the *MGA* was larger when reaching to the far object compared to the near reaches (open circles higher than the filled circles in fig. 5.10B, *"MGA"* panel). This overconstancy bias was unexpected but identical in the previous experiment (see the *"MGA"* panel of figure 5.4B, and the positive slopes in the leftmost graph of fig. 5.4C), but we had no

[1]We chose to present only these two instants for synthesis, but also because they are most representative of the data: the MGA is widely recognized as a reliable variable for studying the scaling of the *GA*, whereas the 20mm-far-*GA* corresponds to the instant at which the biases affecting the scaling, if any, had usually the greatest magnitude, as emerged from the previous experiment; in any case, the *GA* in the rest of the trajectory is consistent with the findings reported below.

explanation for it within the context of shape estimation. Instead, we hypothesized that it could reflect a collision avoidance strategy, which is what we tested in Experiment 2a.

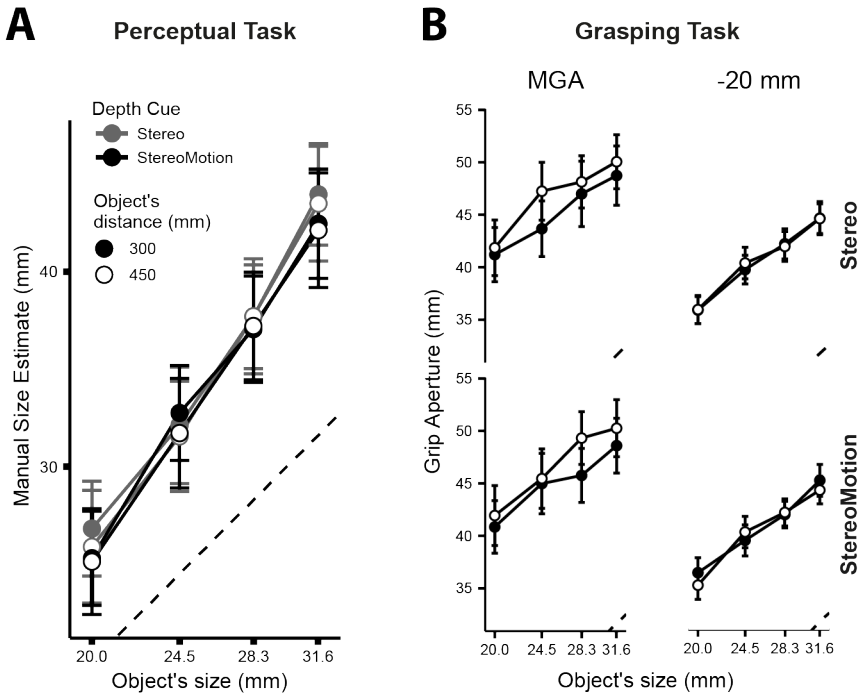

Figure 5.10. Results of Experiment 1a. (A) Perceptual task: the *MSE* is plotted as function of the object's size. Color and shape code for depth cue and object's distance respectively. However, there was a great overlap in the data from the various conditions, signaling that size was unaffected by the experimental manipulations. (B) Grasping task: two snapshots are presented, where the grip aperture is plotted against the object's size as in figure 5.4B. When the thumb was 20 mm far from its contact point (right graph), there was no effect of either depth cue or distance. An apparent overconstancy bias affected the scaling of the *MGA*, as in Experiment 1 (figure 5.4B&C).

Finally, we compared the estimated sizes obtained in this experiment with the depth estimates measured in Experiment 1, and calculated an index of estimated shape, the depth-to-width aspect ratio, for each stimulus (except the 50 mm deep object, which was discarded from the analysis because it was found that its projected disparities were difficult to fuse). The results are shown in figure 5.11: contrary to our predictions, in the stereomotion condition the estimated aspect ratio was greater than veridical on average, and especially at the near distance. Furthermore, a repeated measures ANOVA on the $\hat{R}$ showed an almost identical pattern of effects to that observed in Experiment 1. In particular, we found significant main effects of object's aspect ratio ($F_{(2,\ 76)}$ =

41.51, $p < 0.01$), depth cue ($F_{(1, 38)} = 20.91$, $p < 0.01$) and object's distance ($F_{(1, 38)} = 26.4$, $p < 0.01$), and a significant interaction between depth cue and object's ratio ($F_{(2, 76)} = 21.47$, $p < 0.01$).

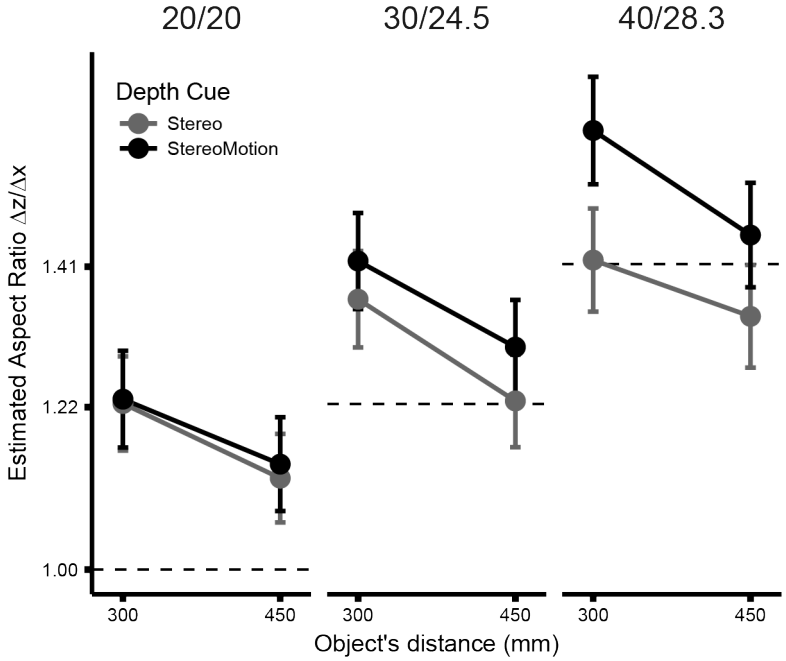

Figure 5.11. Estimated depth-to-width aspect ratio of the stimuli, as function of the viewing distance and divided by cue combination (gray, stereo condition; black, stereotexture condition). The data is presented for each object, whose specific depth/width ratios are marked at the top of each sub-panel. Dashed horizontal lines represent the veridical values of the aspect ratios. Three out of four combinations of distance and depth cue were overestimated.

Experiment 1a allowed us to draw two important conclusions about shape information for perception and grasping. First, the present experiment confirmed that within the peripersonal space, the 2D properties of an object are essentially invariant across distance, since the range of distances that we used to test size constancy in the laboratory is essentially the same space where objects are usually grasped, and independent of depth information. Second, and most remarkably, the data suggests that observers were not able to achieve an accurate Euclidean representation of the stimuli's 3D structure, at least under the conditions that were studied, but that instead they overestimated its depth systematically. In the next experiment, we extended the investigation to consider the effects of texture gradient.

5.5 Experiment 2: Adding texture to stereovision

In the second experiment we asked participants to compare and grasp a stereo-defined object and an object defined by stereo and texture information using the same design of Experiment 1, and thus testing the same hypotheses with a different combination of depth cues. In addition to the existing tasks, we also tested another type of action task, where participants grasped the stimuli along the side, that is, by contacting the front of the object with the thumb and the right rear edge with the index. This change ensured full visibility of both digits until movement completion, which was different from the front-to-back grasp performed before, where the index was occluded by the stimulus few instants before contact. We hypothesized that the disappearance of the forefinger during the final stages of the grasp was one of the main reasons why depth underconstancy had affected the action task in the first place: participants could not visually track the index when preparing to grasp, hence they could only direct it based on the depth estimate produced at planning (*visual control hypothesis*). Therefore, we predicted that the bias would get highly reduced or even disappear entirely, in the case observers were allowed to track both fingers until object contact.

5.5.1 Methods

Stimuli

We tested two viewing conditions: stereo versus stereotexture. In the stereo condition, the stimulus was identical to that of the same condition of Experiment 1. In the stereotexture condition, the stimulus was a parabolic cylinder with a regular grid mapped onto its surface. The elements of the grid were 5 mm by 5 mm squares, and the grid was colored in red. All other aspects of the visual imagery were the same as in the previous experiments.

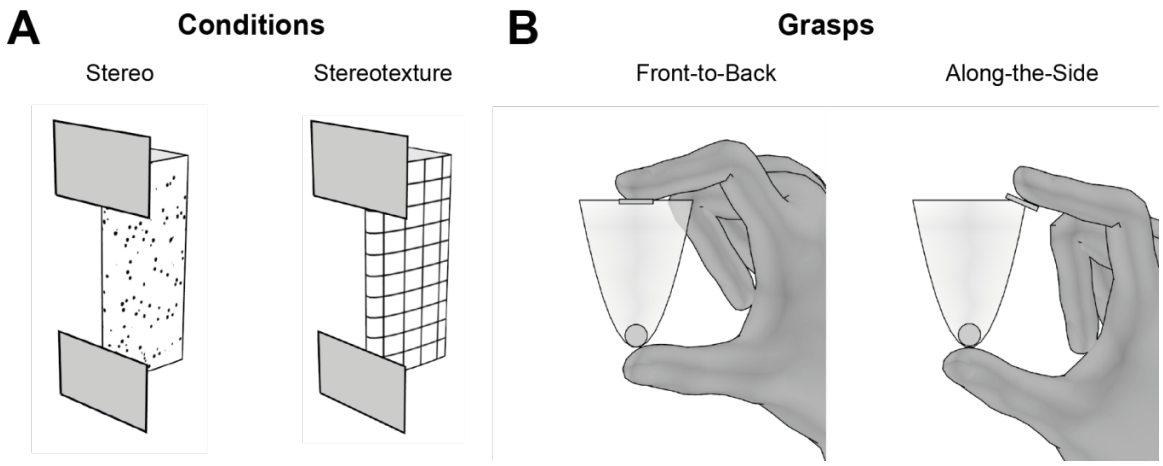

Figure 5.12. Methods of Experiment 2. (A) Schematic representation of the viewing conditions: left, Stereo; right, Stereotexture. (B) Types of grasps tested: left, Front-to-Back (the index's fingerpad was occluded by the stimulus at the end of the grasp); right, Along-the-Side (the index's fingerpad was visible until object contact).

Procedure

The procedure was identical to that of the Experiment 1, with few exceptions. Participants performed the same perceptual task (*MSE*), whereas they performed two grasp tasks. One grasp was identical to before (Front-to-Back), while the second was a grasp Along-the-Side. This second type of grasp differed from the other with respect to the contact location of the index: instead of aiming to the back surface of the cylinder, in the along-the-side case the forefinger contacted the right rear edge of the stimulus. This small change allowed observers to track the virtual index's fingertip until full movement completion, as opposed to the front-to-back grasp, in which the fingertip was occluded by the object during the final critical stages of the grasp.

The experiment was a within-subjects 2x2x3x2 factorial design, with four independent variables as factors: depth cue combination (stereo or stereotexture), object's depth (20, 30, 40 mm) and distance (350 and 450 mm), task (*MSE* or grasp). The experiment was divided into two sessions, one for each task, which were ran in two separate days. Each experimental session consisted of four brief blocks, in which all factors were randomly intermixed with the exception of the object's distance, which was alternated across blocks following an ABAB sequence. During the session

involving the grasp task, subjects also performed one of the two grasps in each block following a ABBA sequence.

Each individual block consisted of 42 trials (2 depths cues by 1 distance by 3 depths by 7 repetitions), therefore, over the course of eight blocks, each subject performed a total of 336 trials.

5.5.2 Results and Discussion

Participants in this experiment exhibited essentially the same pattern of results found in the first experiment. In the perceptual task, depth judgments in both stereo and stereotexture conditions were affected by depth underconstancy. At the same time, the stereotexture stimulus appeared on average deeper than the stereo stimulus (Fig. 5.13 A&B). A repeated-measure ANOVA found three main effects: object's depth ($F_{(2, 80)} = 280.89$, $p < 0.01$), object's distance ($F_{(1, 40)} = 15.28$, $p < 0.01$) and depth cue ($F_{(1, 40)} = 37.23$, $p < 0.01$), plus a nearly significant interaction between depth cue and object's depth ($F_{(2, 80)} = 3.08$, $p = 0.051$).

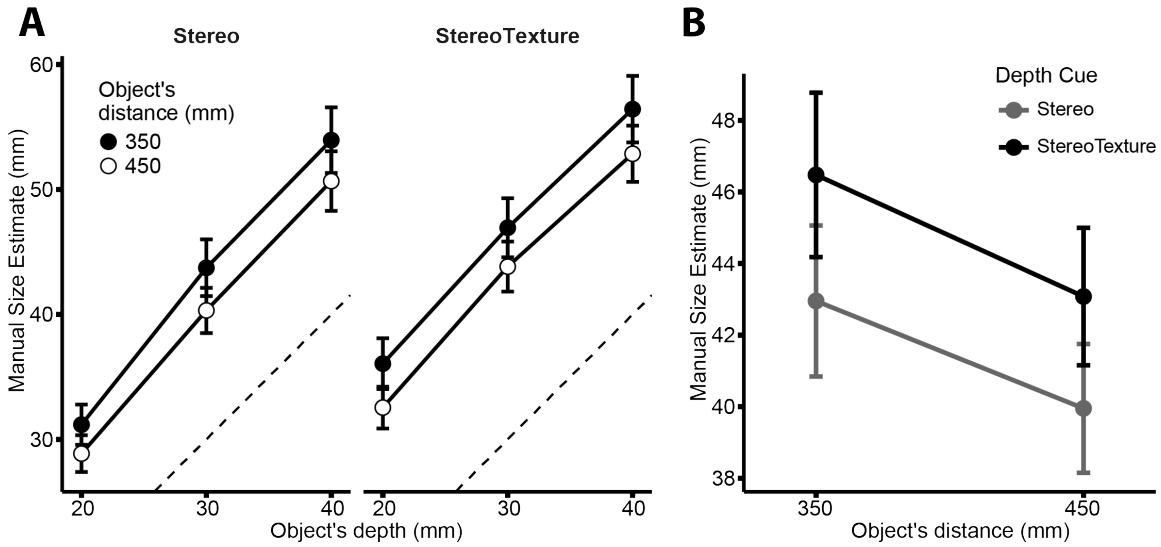

Figure 5.13. Results from the perceptual task of Experiment 2. (A) *MSE* as function of the object's depth for each viewing distance (filled circles, 350 mm; open circles, 450 mm). Left panel, results from the stereo condition; right panel, results from the stereotexture condition. (B) Same results collapsed across object's depths and presented as function of the viewing distance: the negative slope in the data denotes depth underconstancy.

The interaction was caused by a reduction of the effect of the texture cue as the object's depth increased, as can be seen in figure 5.17A. This was opposite to what we found in Experiment 1, where stereomotion stimuli were more and more overestimated with respect to stereo stimuli as the depth increased. We have no clear explanation for this effect.

Depth judgments were then compared to size judgments from Experiment 1a to obtain an estimate of the shape's aspect ratio ($\hat{R}$) and evaluate its constancy in this task (Fig. 5.14). Similar to what found with stereo and motion (Fig. 5.10), the aspect ratio was largely overestimated, especially for the shallower objects, suggesting that the most of the stimuli looked more convex than they actually were, particularly when seen close to the body and in presence of texture. Not surprisingly, the repeated-measure ANOVA on the $\hat{R}$found three main effects, the object's aspect ratio ($F_{(2, 80)} = 32.16$, $p < 0.01$), the object's distance ($F_{(1, 40)} = 15.41$, $p < 0.01$) and depth cue ($F_{(1, 40)} = 40.3$, $p < 0.01$), and a significant interaction between depth cue and aspect ratio ($F_{(2, 80)} = 8.06$, $p < 0.01$).

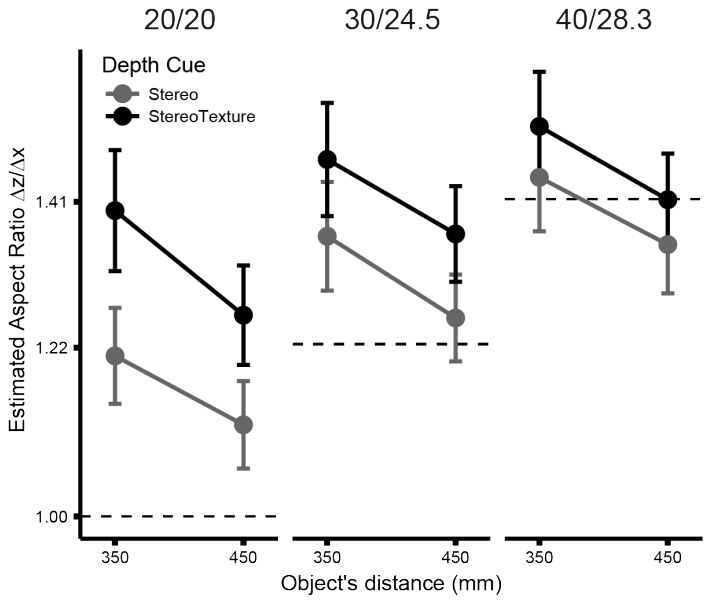

Figure 5.14. Estimated depth-to-width aspect ratio of the stimuli in Experiment 2, as function of the viewing distance and divided by cue combination (gray, stereo condition; black, stereotexture condition). The data is presented for each individual object, whose specific depth and width are displayed at the top of each sub-panel. Dashed horizontal lines represent veridicality.

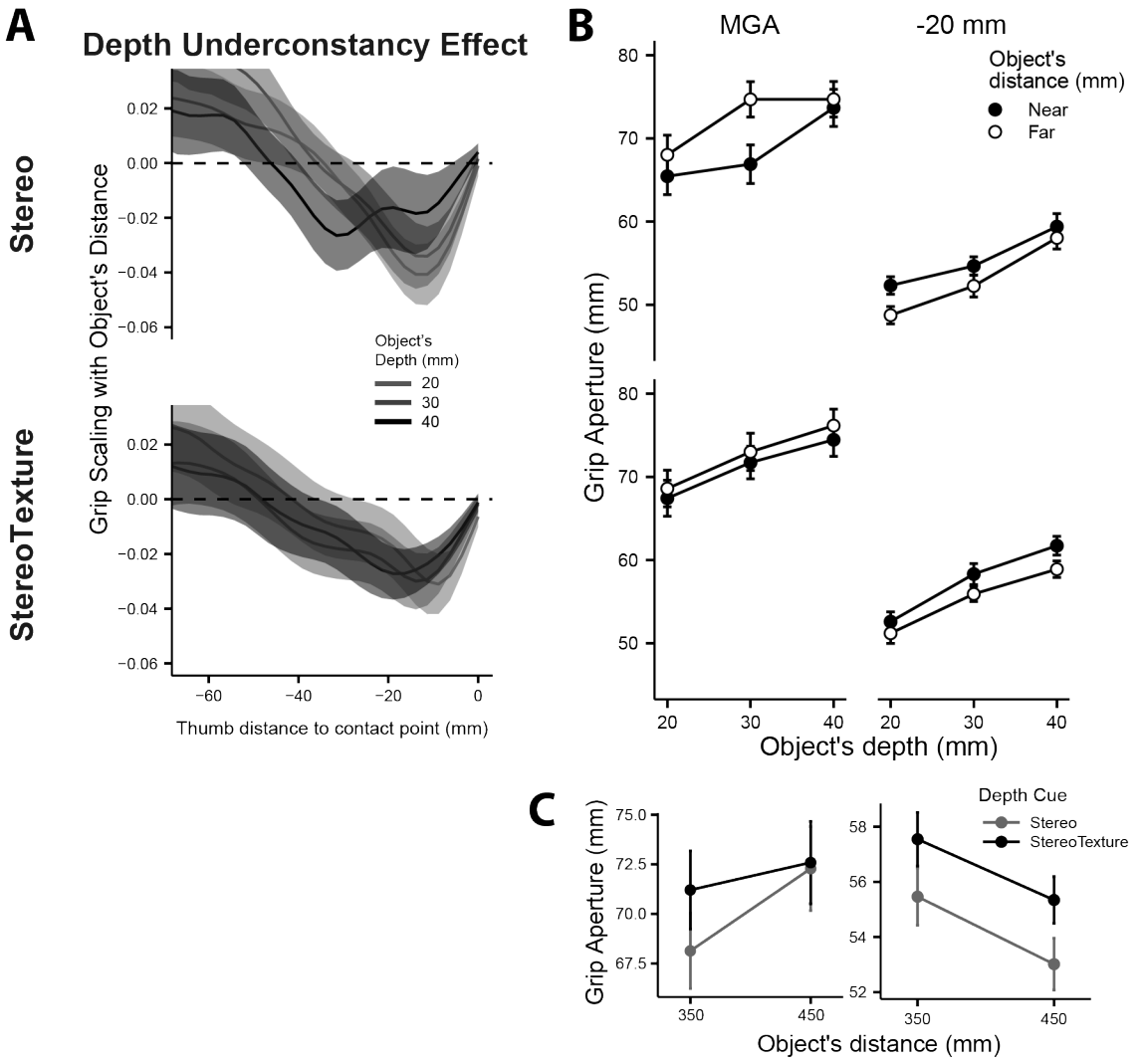

Figure 5.15. Results from the Front-to-Back grasp task of Experiment 2. (A) Slope of the scaling function relating the GA to the object's distance as function of the thumb's contact point. Color codes for the object's depth. Upper panel, stereo condition; lower panel, stereotexture condition. Areas around the curves represent ±1 *SEM*. Negative values indicate depth underconstancy. The horizontal dashed line corresponds to absence of bias (slope of zero). (B) *GA* as function of the object's depth presented for each viewing distance (filled circles, 350 mm; open circles, 450 mm) at two instants along the trajectory (left graph, at the time of the *MGA*; right graph, when the thumb was 20 mm far from its contact point). (C) Same data of point (B) collapsed across object's depths and plotted against the viewing distance. Color codes for depth cue condition (gray, stereo; black, stereotexture).

The results from the grasping task further confirmed the *perception-action identity hypothesis*, as they resembled the results of the perceptual task quite remarkably.

Similar to the *MSE*, the grip aperture of the grasping phase of the Front-to-Back movement (from *MGA* to object contact) exhibited strong depth underconstancy. As can be seen in figure 5.15A,

the slope relating the *GA* with the object's distance became increasingly negative as the grip was shaped to prepare to grasp. This effect reached its peak when the thumb was about 20 mm far from its contact point, and was independent of the depth cues (top and bottom panels of fig. 5.15A respectively). To help visualize this bias, in figure 5.15B&C we present snapshots of the *GA* at two critical instants along the reaching (the *MGA* and around the bias's peak). The ANOVA table found significant effects of object's depth ($F_{(2, 80)} = 83.68$, $p < 0.01$), object's distance ($F_{(1, 40)} = 7.70$, $p < 0.01$) and depth cue ($F_{(1, 40)} = 25.69$, $p < 0.01$). As in experiments 1 and 1a, the grip aperture at the time of the *MGA* was seemingly affected by overconstancy (positive slopes in the left panel of fig. 5.15C), which we had not predicted (and which was the subject of Experiment 2a).

On the contrary, the grip aperture of the Along-the-Side grasp showed a marked reduction of the underconstancy bias, consistent with the *visual control hypothesis*, as visible in both panels of figure 5.16A. The bias did seem to affect the grasping of the smaller objects, but its magnitude was negligible and the overall effect of object's distance was not significant when the thumb was about 20 mm far from object contact ($F_{(1, 40)} = 25.69$, $p < 0.01$) (Fig. 5.16B&C). Interestingly, at the same time, the grip aperture was still significantly affected by the depth cue ($F_{(1, 40)} = 25.69$, $p < 0.01$), in addition to the not surprising effect of object's side ($F_{(2, 80)} = 47.41$, $p < 0.01$). In other words, even though the continuous vision of both fingers allowed to eliminate the depth underconstancy bias, subjects still approached the stereotexture object with a larger *GA* than when aiming at the stereo object.

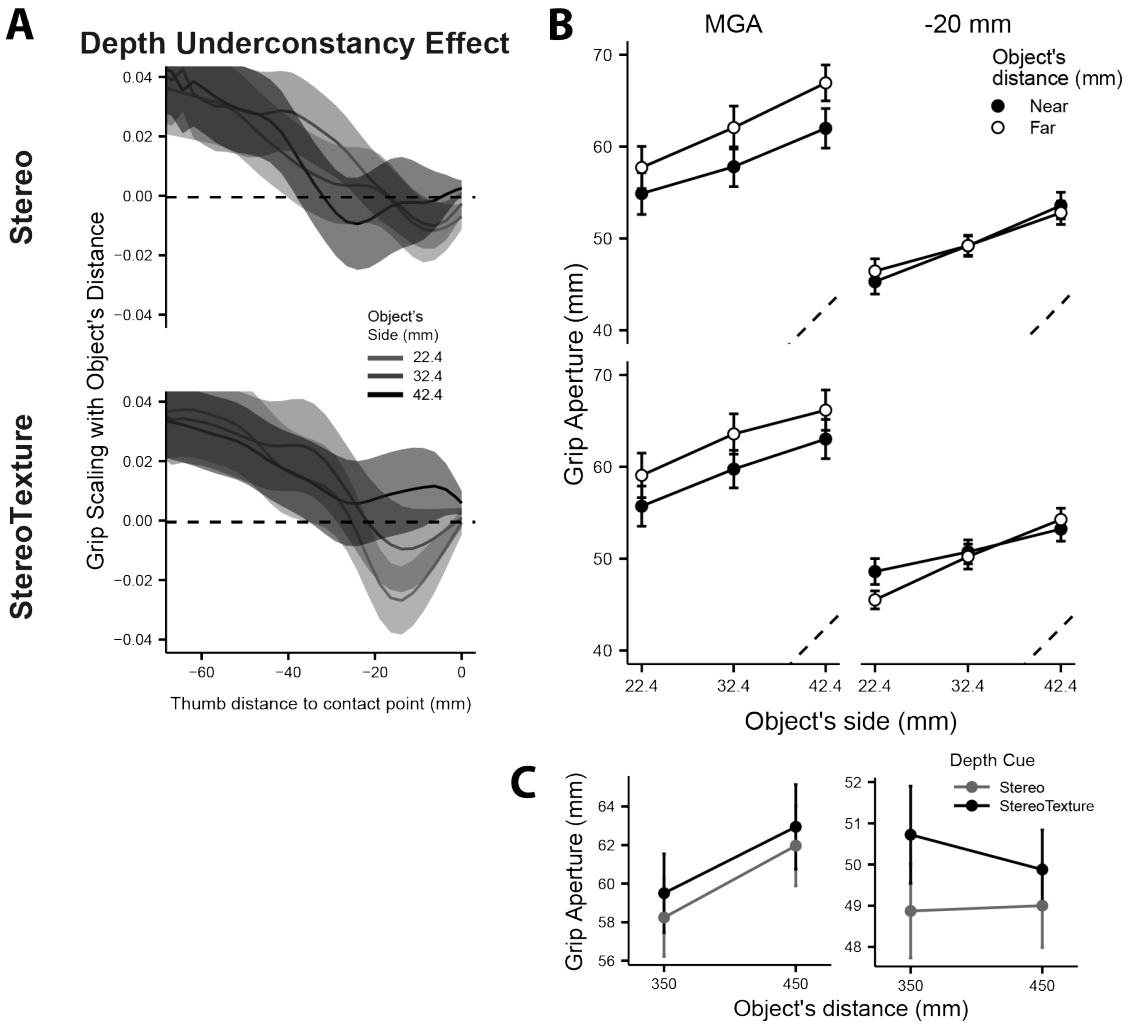

Figure 5.16. Results from the Along-the-Side grasp task of Experiment 2. (A&B&C) Data is presented as in figure 5.15. Unlike in the Front-to-Back grasp, the *GA* was not affected by underconstancy in the terminal phase of the action (slope of zero in the right graph of section C).

Overall, the effect of the depth cue was robust throughout all tasks: it affected perception (Fig. 5.17A) and both action tasks (Fig. 5.17B). Furthermore, the results of the perceptual task showed not only that depth judgments increased overall in the stereotexture condition, but also that depth information interacted moderately with the physical depth. As a result, the slope relating the *MSE* to the object's depth was similar in both visual conditions, although slightly smaller in the stereotexture condition. In agreement with the *perception-action identity hypothesis*, the results of both grasp tasks showed that the slope relating the *GA* to the object's depth was similar in both

visual conditions (Fig. 5.17C). In fact, in the Along-the-Side grasp, the slope tended to be smaller in the stereotexture condition (lower panel of figure 5.17C).

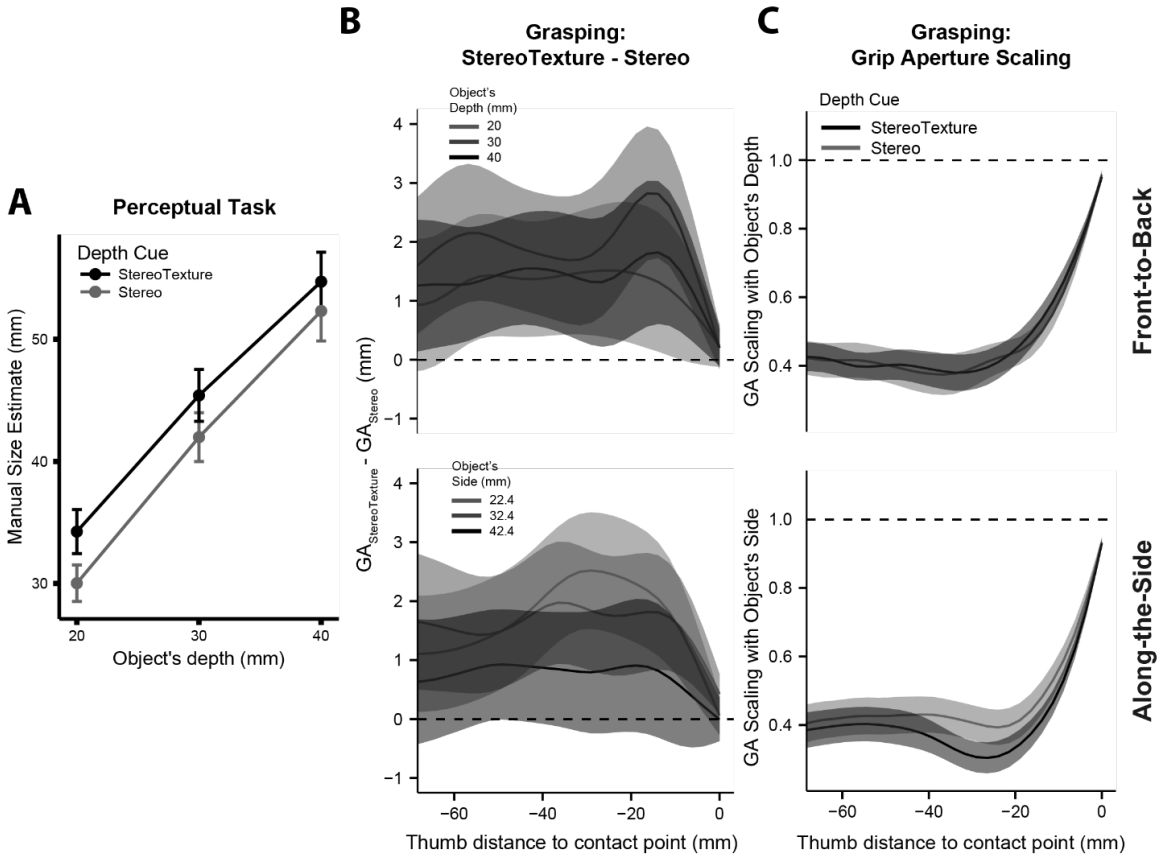

Figure 5.17. Comparison between perception and action tasks in Experiment 2. (A) Same results of figure 5.13A (*MSE*), collapsed across distance and presented for each combination of depth cues (black: stereotexture; gray: stereo). (B) Effect of the depth cue on the grip aperture in the grasping task (upper panel, Front-to-Back; lower panel, Along-the-Side), expressed as the difference between the *GA* in the stereotexture condition and the *GA* in the stereo condition. Color codes for object's depth (the deeper the object, the darker the curve). Positive values mean that the stereotexture object was grasped as deeper than the stereo object. Areas around the curves represent ±1 *SEM*. (C) Grip scaling function as function of the distance of the thumb to its contact point (black, stereotexture condition; gray, stereo condition). Upper panel, Front-to-Back; lower panel, Along-the-Side. The dashed line corresponds to a slope of 1 (veridical scaling).

Finally, we found an effect of the previous trial as well, in line with the results of the grasp task in Experiment 1: 20 mm before object contact, the grip aperture was larger than average if subjects had touched the stereo object in the previous trial, and smaller than average if the previous trial's stimulus was the stereotexture cylinder ($F_{(1, 40)} = 5.28$, $p = 0.03$).

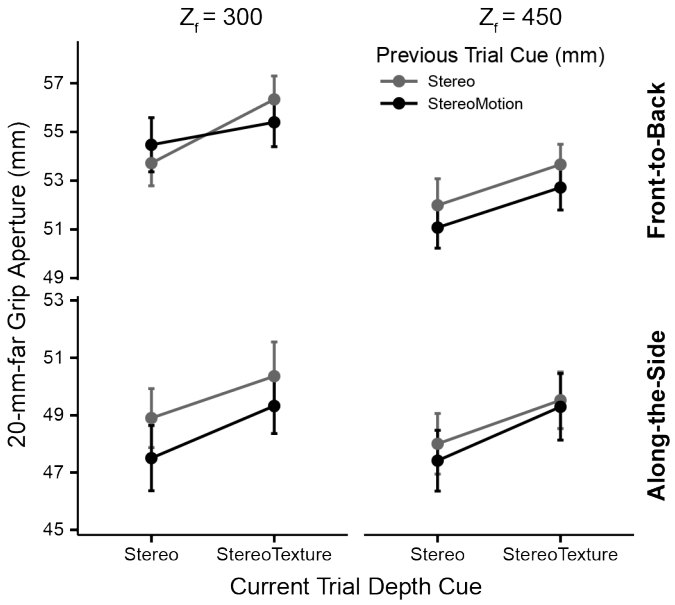

Figure 5.18. Previous trial effect in the grasp tasks of Experiment 2. The 20mm-far *GA* is plotted as function of the depth cues during trial N, and color-coded for each cue of trial N-1 (gray, previous trial = stereo; black, previous trial = stereotexture). Left column, trials with object's distance of 350 mm; right column, trials with object's distance of 450 mm. Upper row, data from the Front-to-Back grasp; bottom row, data from the Along-the-Side grasp. Results are identical to those found in the grasp task of Experiment 1.

Overall, Experiment 2 successfully replicated all the findings of the Experiment 1. First, the perceptual task showed once again that binocular disparity alone leads to underconstancy of shape. Second, this bias was not reduced when texture was added to the scene (as we saw with motion as well): the only effect of texture was that stimuli looked more convex than when specified only by stereo information. Third, the results from the grasping task reflected those found in the MSE task in basically every respect. Fourth, a comparison between two different types of grasp showed that online visual control of the digits helps eliminating the biases at planning. Finally, an effect of the previous trial indicated that the haptic feedback was used to register an error signal at the end of the movement, which was used to update the motor plan of the subsequent grasp.

5.6 Experiment 2a: Control for a possible collision avoidance strategy

This second control experiment was designed to investigate one specific finding that recurrently emerged from the trajectory analysis of the grasps of experiments 1, 1a and 2: at the time of its maximum, the grip aperture was unexpectedly affected by overconstancy (see the "*MGA*" panels of figures 5.5C, 5.10B, 5.15C and 5.16C), in discontinuity with the rest of the trajectory, and in contrast with the results of the perceptual task. One possible explanation was that the subjects took into account the presence of the physical apparatus, which stood in the middle of the reachable space, while planning the action. It is well known that collision avoidance strategies impact the motor execution of grasping movements: the grip aperture tends to be larger when aiming at distant objects, net of other perceptual biases, because farther reaches are usually faster and thus predisposed to an increased level of uncertainty (*collision avoidance hypothesis*). To test this hypothesis, we replicated Experiment 2, only this time we took away the motors and supports behind the mirror, effectively removing the physical object from the scene. To avoid that subjects would pantomime the movement as a result of this modification, we carefully trained them to fully rely on the vision of their fingertips to complete the movement accurately. Two grasp tasks were tested (Front-to-Back and Along-the-Side), and all aspects of them, except for the final touch, were identical to Experiment 2.

5.6.1 Methods

Stimuli and Procedure

Stimuli, procedure and design were identical to those of the grasping task of Experiment 2, with the only exception that in this experiment participants did not contact any physical object at the end of the movement. This was achieved by removing all motors and physical objects from the workspace. Thus, participants completed the task by relying solely on the visual feedback from the hand. A training session was ran prior to the beginning of the actual experiment. We carefully

instructed the subject to accurately place the virtual fingertips on the virtual object's surface, as if they actually touched it, by nulling the relative disparity between the stimulus and the virtual fingertips. This strategy was only partially practicable during a Front-to-Back grasp, due to the occlusion of the index by the object's shape at the end of the movement, hence with compared it again with the Along-the-Side grasp. A post-hoc visual inspection of the grasping trajectories was performed to check the accuracy of the movements.

Each individual block consisted of 42 trials (2 depths cues by 1 distance by 3 depths by 7 repetitions), therefore, over the course of four blocks (only the grasping task was tested), each subject performed a total of 168 trials.

5.6.2 Results and Discussion

In agreement with the *collision avoidance hypothesis*, the removal of potential obstacles from the workspace had its most visible impact on the portion of the reaching around the time of the *MGA*. As a result, the grip aperture was consistently affected by underconstancy throughout the whole trajectory of the Front-to-Back grasp (Fig. 5.19A). A repeated-measures ANOVA found a significant effect of the object's distance on the *MGA* ($F_{(1, 12)} = 9.40$, $p < 0.01$) in the predicted direction (the *MGA* was smaller when aiming at the far distance and larger in the other case). The same effect was also found at later instants of the reaching. On the contrary, the grip aperture of the Along-the-Side grasp never exhibited underconstancy (Fig. 5.19B), as there was no difference between near and far reaches neither at the time of the *MGA* ($F_{(1, 12)} = 1.08$, $p = 0.32$), nor in the rest of the trajectory.

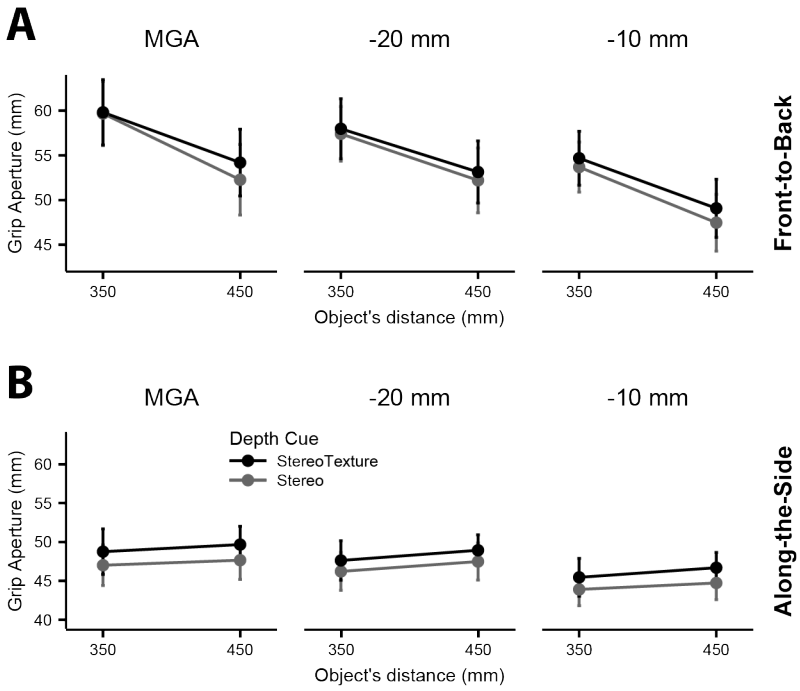

Figure 5.19. Results from the grasp tasks of Experiment 2a. The *GA* is collapsed across object's depth, plotted against the object's distance and color-coded for depth cue condition (gray, stereo; black, stereotexture). (A) Front-to-Back grasp: the underconstancy bias affected the *MGA* as well as the rest of the trajectory, as indicated by the negative slope in all three plots. (B) Along-the-Side grasp: no effect of distance was found at any point in the trajectory.

We also noticed that, under these conditions, the *MGA* was smaller than in the previous experiments, and happened much later in the trajectory. In all previous experiments, the distance from the thumb to its contact point at the time of the MGA was consistently around 50 mm. In this experiment, the *MGA* was measured around 40 mm in the Front-to-Back grasp, and around 30 mm in the Along-the-Side grasp. This is shown in figure 5.19: the MGA and the 20mm-far grip aperture are quite similar to each other, and the *GA* becomes visibly smaller only when the thumb is about 10 mm far from its final position. These are interesting indications of how top-down information about the presence of physical objects interacts with the online visual control of the hand. The removal of the physical object made the participant adopt a straighter path while reaching to the target. Since the in-flight grip aperture is usually greater during curved trajectories (Brenner &Smeets, 1998), this change in reaching path caused a reduction of the *GA*. Evidently,

since no risk of collision was to be taken into account any more, no safety margin was necessary during the transport phase. Furthermore, in this experiment the only way to complete the task successfully was to rely on a disparity nulling strategy, because final touch was not provided. Thus, it is presumable that participants postponed the terminal phase of the grasp as much as possible, hence the delayed *MGA*, because this strategy allowed them to compare the position of the fingers relative to the surface while terminating the movement. Interestingly, this strategy did not seem to transfer to the Front-to-Back grasp: when the index disappeared at the end of the grasp, the underconstancy bias re-appeared systematically.

CHAPTER 6

Study 3: Grasping in depth in natural environments

6.1 Predictions

This study aimed to test whether stereo and motion information are combined to produce veridical estimates of shape for perception and grasping, when flatness cues are removed from the visual scene.

Our stimuli consisted of 60 mm wide, 40 mm tall and 20 or 40 mm deep prisms with triangular cross-section, constructed from cardstock. The visible portion of their surface was matt and painted with white stripes that ran parallel to the long side. We actively modulated the amount of binocular disparity and of motion specifying the shape of the stimulus, which added to the intrinsically present cues of accommodation and blur (or baseline depth cues).The prisms were presented in four different configurations: static horizontal (baseline condition), horizontal with motion (motion condition), static tilted (stereo condition), and tilted with motion (stereomotion condition).

Under baseline condition, the prism was oriented horizontally along its long side, tip facing the observer, and was viewed as static. When in this position, the horizontal disparity gradient caused by the object's physical depth was aligned with the white stripes on its surface: disparities essentially disappeared into the stripes and subjects were left to rely on only accommodation and blur cues. Participants verbally described the total percept as a substantially flat surface crossed by white lines.

When the same horizontal object rocked around the horizontal axis (motion condition), the pattern of relative velocities of the white stripes helped the observer disambiguate the prism's 3D layout very effectively. In the stereo condition, the object was rotated by either 10 or 20 degrees around an axis perpendicular to the subject’s frontoparallel plane and centered on the cyclopean

eye, and viewed as static. Tilting the prism provided the observer with reliable stereo information about shape: the binocular disparities did not blend into the oblique stripes, becoming better and better detectable as the prism's tilt increased.

Finally, in the stereomotion condition, the prism was tilted as in the stereo condition, and it also rocked as in the motion condition. During this presentation, the binocular disparities generated from tilting the object were reinforced by a motion gradient that specified the same geometry (see figure 6.1).

We measured the perceptual appearance of the prism using a Manual-Size-Estimation (*MSE*) procedure: participants opened their index and thumb apart by an amount equal to the perceived front-to-back extent of the upper side of the prism. In this task, we expected one of two possible outcomes.

Following our previous results obtained in VR, we maintained that stereo and motion are not attuned to the stimulus's veridical 3D structure, even when flatness cues are absent. Instead, we expected that the probability of correctly detecting the presence of depth when there physically is would increase as more 3D information is available. By extension, the observer should perceive a deeper object as more cues are provided.

The prediction that follows is that when only blur and accommodation cues are present, the prism should look flatter than in any other viewing condition. Hence, we expected the baseline condition to elicit the smallest *MSE*. For the same reason, we expected that subjects would exhibit a greater *MSE* in both the stereo and the motion conditions. Finally, we predicted to find the largest *MSE* in the stereomotion condition. These three predictions constituted the *non-veridical shape hypothesis*.

Alternatively, it can be assumed that all depth signals convey veridical estimates of 3D shape, each with a different reliability, and their combination simply maximizes the precision of these true estimates. If this is the case, then participants should produce theoretically the same *MSE* in

all four viewing conditions, because there are no flatness cues in the visual scene (*veridical shape hypothesis*).

In a separate session, participants also performed grasping movements directed towards the same prism, under the same four visual conditions, and we analyzed the grip aperture over the course of their reaches. Since we assumed that 3D vision for action is the same as 3D vision for perception (*perception-action identity hypothesis*), we hypothesized that the *GA* should show the same pattern of results of the *MSE*, regardless of which of the two previously stated hypotheses was eventually supported by the data of the perceptual task.

6.2 Methods

Participants

Forty-three students participated in the experiment (14 males, 29 females). All participants self-reported normal or corrected-to-normal vision. The participants received either a reimbursement of $8/hour or coursework credit for their effort. The experiment was approved by the Institutional Review Board at Brown University. Each participant gave informed consent prior to the experiment.

Apparatus

Subjects sat with their head on an optometric chin and forehead rest installed along one of the short sides of a rectangular table, directly facing the physical stimulus.

The stimulus consisted of an isosceles triangular prism made from cardstock, oriented with the tip facing the participant. The horizontal and vertical dimensions of the prism were 97 by 40 mm. We built two of these prisms, one with a depth (tip-to-base extent) of 20 mm and the other 40

mm. The prisms were attached base-to-base such that only one was visible per trial. Printed onto the prism surface were pure white pinstripes on a matte black ink base, spaced 6.5 mm apart.

Trials were conducted in darkness. During the trials, two black lights guaranteed uniform illumination of the presented stimulus; the net effect was a series of floating white stripes. The two prisms were mounted on a custom, laser-cut wooden base, which was then installed on a vertical metal stand. The position and height of the stand was adjusted such that the tips of the prisms were 42 cm above the tabletop, matching the average eye height of the subjects, and at a viewing distance of 42 cm. Rectangular wooden occluders, also painted matte black, covered the far left and right sides of the stimulus to prevent participants from detecting the disparities at the edges of the prisms. The visible portion of the stimulus thus was 60 mm wide by 40 mm tall.

Two stepper motors (Phidgets Inc.) were used to switch between stimuli as well as manipulate their orientation and motion. The first motor rotated the prisms along the horizontal axis centered on their plane of contact (the prisms' bases), producing a rocking motion at 60 deg/sec that swept a total angle of 15 degrees (±7.5 degrees, see Figure 6.1).The second motor rotated the stimulus and occluders around an axis perpendicular to the subject's frontoparallel plane and centered on the cyclopean eye. For our viewing conditions, we selected the angles of 0, ±10 and ±20 degrees relative to the positive x-axis**.**

Prior to each trial, the black lights turned off. The first motor then rotated the prisms into an intermediate position, and rotated them again to present the appropriate stimulus. Thus, subjects could not use audio cues to anticipate the next stimulus. When the apparatus was ready, the black lights turned on. This was achieved with a custom *USB* relay that synced the lights' power source to the motion of the motors.

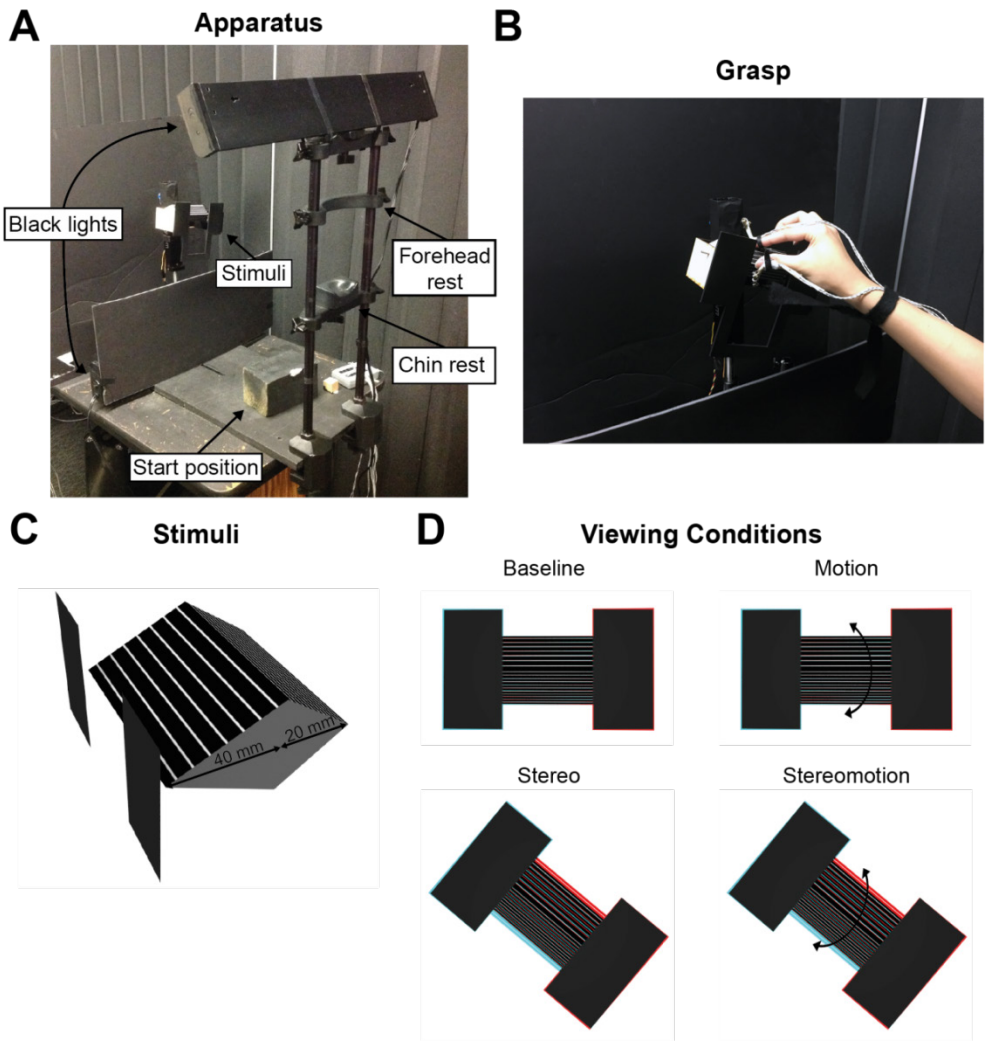

Figure 6.1. Methods used in the study. (A) Picture of the apparatus. Subject rested their head on a chinrest with the forehead firm against a forehead rest, to prevent self-induced motion parallax. The stimuli appeared at eye height and were seen at a distance of 42 cm from the cyclopean eye. Black lights guaranteed uniform illumination of the stimuli and of the hand in the dark. (B) In the grasp task, subjects contacted the tip of the prism with their thumb and the upper edge with the index. If the prism was rocking, it stopped immediately prior to object contact. (C) The stimuli were attached together along the base of the prisms and rotated by a motor such that only one would face the observer in every trial. (D) Schematic of the viewing conditions. From the upper left quadrant going clockwise: Baseline, Motion, Stereomotion and Stereo. The anaglyphs show that binocular disparities became available only when the object was tilted.

To record the location of the subject's fingers, we used an *NDI* Optotrak Certus motion capture system, with the position sensor located above and to the left of the workspace at a distance of 224 cm from the subject. Subjects wore index and thumb markers with a protruding flag that held three *NDI* infrared *LEDs* each, oriented such that the infrared light would directly hit the Optotrak

cameras. Optotrak recordings, the motion of both Phidgets motors and the *USB* relay were all controlled by custom C++ programs using specific *API* routines provided by each vendor.

Procedure

In this study, we utilized a mixed factorial within-subject design, composed of two main factors, cue combination (four levels) and physical depth (two levels), plus a third factor (tilt angle, two levels) nested inside two levels of cue combination.

The first factor concerned the amount of 3D information specifying the prism's depth. This was varied on four levels: Baseline, Stereo, Motion and Stereomotion. The Baseline level corresponded to the prism oriented horizontally along its long side and still, in which the stimulus's depth was mainly specified by accommodation and blurring gradient cues. In the Motion condition the prism rocked ±7.5 degrees around the x-axis at a constant angular speed of 60 deg/sec. The Stereo condition was obtained by tilting the prism ±10 or ±20 degrees from the x-axis (the sign of the tilt was randomized from trial to trial to prevent habituation to a specific direction), thus adding binocular disparity information. Finally, the Stereomotion condition involved rocking the tilted prism. With the second factor, we varied the physical depth of the stimulus across two discrete values: 20 mm and 40 mm.

Each subject underwent thorough calibration upon putting on the infrared *LEDs*. First, the experimenter calibrated the Optotrak position sensor with respect to a custom coordinate system. Then, the Optotrak recorded the position of the subjects' fingers on a designated home position. Participants performed a grasping task and a perceptual task in 4 separate blocks, following an ABBA order. All trials were fully randomized within blocks.

In the perception task, the stimulus was illuminated at the beginning of each trial, and subjects performed a *MSE* by raising the hand a few centimeters above home position and matching the perceived tip-to-upper-edge distance with their thumb and index fingers. Subjects were not allowed to look at the hand while performing the task. When they felt confident with their

estimate, they pressed a button with their other hand to confirm. The lights turned off with the button press, and the next trial would load as the subjects returned to home position.

In the grasping task, once the lights turned on, subjects reached out and physically contacted the tip and upperedge of the stimulus with their thumb and index fingers, respectively. The grasp was performed naturally; subjects could see their hand and feel the prism. If the prism was moving, it stopped once the index markers passed 42 cm in the z direction, to ensure that subjects could comfortably contact the object. When subjects returned their fingers to the home position, the lights went off, and the motors prepared the next stimulus.

Each task comprised of144 trials, divided in two blocks: in each block, Baseline and Motion conditions were tested for each depth 6times (total of 24 trials),while Stereo and Stereomotion conditions were presented 6 times for each depth and tilt angle (total of 48 trials). The four conditions were pseudo-randomized within each block: participants performed six adjacent strings of 12 unique trials (2 trials in Baseline and Motion condition, 4 trials in Stereo and Stereomotion condition). In every string, the order of the 12 trials was randomized. This method prevented that a particular combination of unique trials would be presented more often than the others due to chance. Over the course of the entire experiment, participants thus performed a total of 288 trials.

6.3 Results

In the perceptual task, participants adjusted their *MSE* based on the visual condition. The same prism was judged as flatter under Baseline view, intermediate estimates were given under both Motion and Stereo conditions, while the greatest *MSE* was measured under Stereomotion view (left panel of figure 6.2 and figure 6.3). We first ran a repeated-measure ANOVA on the *MSE* using the object's depth and cue combination as predictors and we found a significant effect of

stimulus depth ($F_{(1, 42)}$ = 252.46, p < 0.01), cue combination ($F_{(3, 126)}$ = 15.3, p < 0.01) and a significant interaction between the two predictors ($F_{(3, 126)}$ = 7.02, p < 0.01).

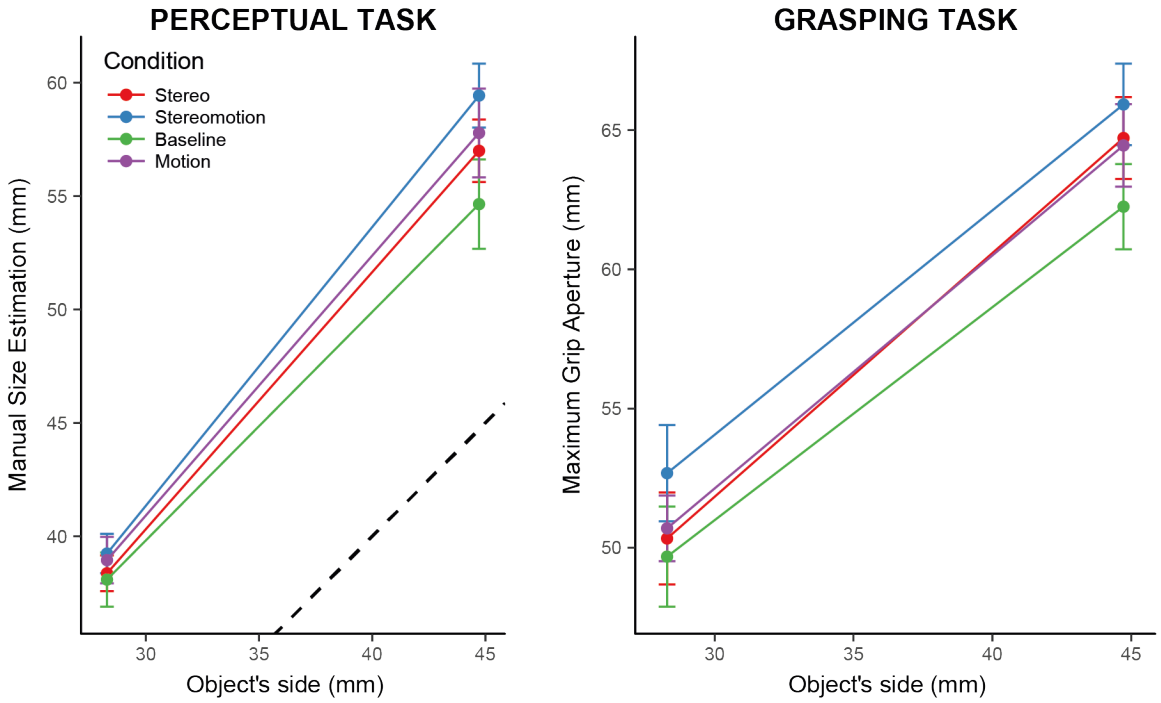

Figure 6.2. Results from the perceptual task (left panel, *MSE* as function of the object's side) and from the grasping task (right panel, *MGA* as function of the object's side). In both tasks, the smallest response was recorded in the Baseline condition whilst the largest response was recorded in the Stereomotion condition. The results of the Stereo and of the Stereomotion condition are collapsed across tilt angle.

The interaction was explained by a simple effect of the visual condition for the deeper object ($F_{(3, 126)}$ = 31.06, p < 0.01), but not for the smaller object ($F_{(3, 126)}$ = 1.16, p = 0.32). The side of the 20 mm deep prism was presumably too small to yield detectable changes in the *MSE* across visual conditions.

We also found a significant main effect of the tilt angle on both Stereo and Stereomotion conditions ($F_{(1, 42)}$ = 5.21, p = 0.03), since the *MSE* was bigger when the prism was oriented 20 degrees away from the x-axis, as predicted by the increased reliability of binocular disparity (filled bars in the left panel of figure 6.3).The data of the grasping task mirrored the perceptual results. Participants reached-to-grasp the prism with a grip aperture that increased with the amount of depth information available in the visual scene: the Maximum Grip Aperture (*MGA*)

was smallest in the Baseline condition, intermediate in both Motion and Stereo conditions, and largest in the Stereomotion condition (right panel of figure 6.2 and figure 6.3). An omnibus repeated-measure ANOVA on the *MGA* found a main effect of the stimulus's depth ($F_{(1, 42)}$ = 315.8, $p < 0.01$) and of cue combination ($F_{(3, 126)} = 11.24$, $p < 0.01$).

Simple effects of cue combination were found for both object's depth (20 mm: $F_{(3, 126)} = 3.09$, $p = 0.03$; 40 mm: $F_{(3, 126)} = 20.04$, $p < 0.01$). Unlike the *MSE*, the *MGA* was not affected by the tilt angle in the Stereo and Stereomotion conditions ($F_{(1, 42)} = 0.44$, $p = 0.51$).

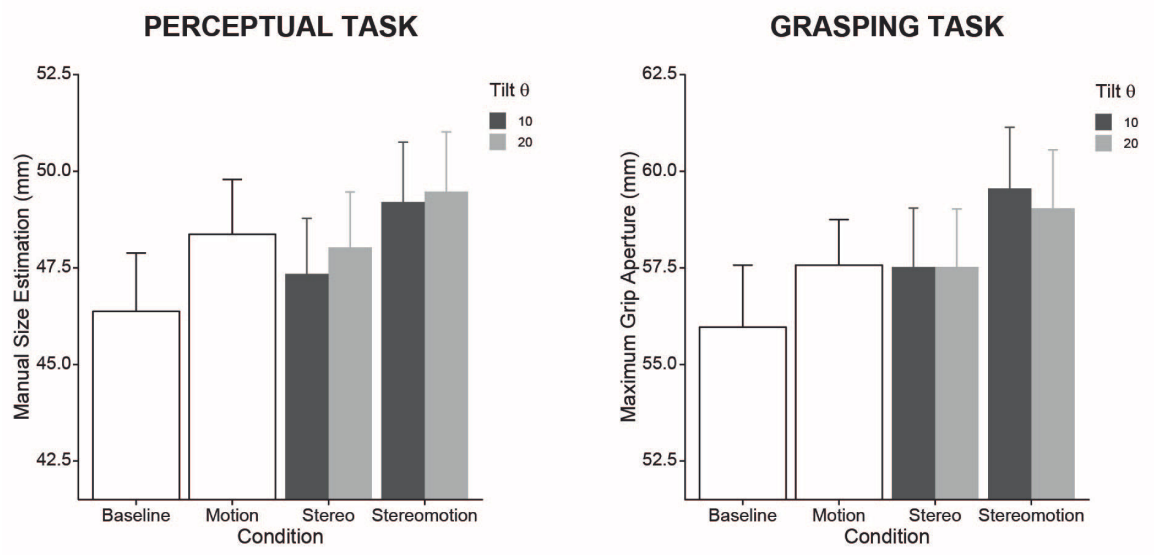

Figure 6.3. Same results of figure 6.2 collapsed across object's side to show the effect of the tilt angle on the Stereo and the Stereomotion condition (dark gray bar: tilt = 10 deg; light gray bar: tilt = 20 deg). Left panel, perceptual task (*MSE* for each viewing condition); right panel, grasping task (*MGA* for each viewing condition).

Taken together, these results reveal that the biases we found in our previous study were not related to *VR*, but instead they seem to be built in the visuomotor machinery.

Secondly, these data further confirm that perception and grasping share the same processing of 3D information, in agreement with our previous findings. The estimated prism's shape was affected by the amount of available 3D information, such that the stimulus looked "pointiest" in the Stereomotion condition compared to the other conditions. This latter result is particularly significant because it cannot be explained by the *veridical shape hypothesis*, in that it contradicts

the argument that motion and stereo produce the same estimates of depth: if that were the case, in fact, the responses in the Motion, the Stereo and the Stereomotion conditions should have had the same magnitude. Furthermore, although it is true that we reduced the reliability of the stereo cue with respect to motion by tilting the stimulus by small angles, it is not possible, under the same hypothesis, that their combination would cause changes in the appearance of the prism itself, but at most in the precision of the responses.

An additional analysis strengthened this conclusion. We compared the *MSEs* with the *MGAs* measured across the four conditions for each individual subject and both object's depths. We centered both variables with respect to each participant's average, to avoid spurious correlations due to differences in the individual offsets. We found that perceptual reports and grip apertures were significantly positively correlated (20 mm object: $r = .27$, $p < .01$; 40 mm object: $r = .49$, $p < .01$) (Fig. 6.4), suggesting that if a person underestimated the prism's shape with respect to the sample's average, they also grasped it with a smaller than average *MGA* and vice versa. In other words, the amount of depth information affected the recovery of shape to a different extent across subjects, but all participants were consistent across perception and grasping.

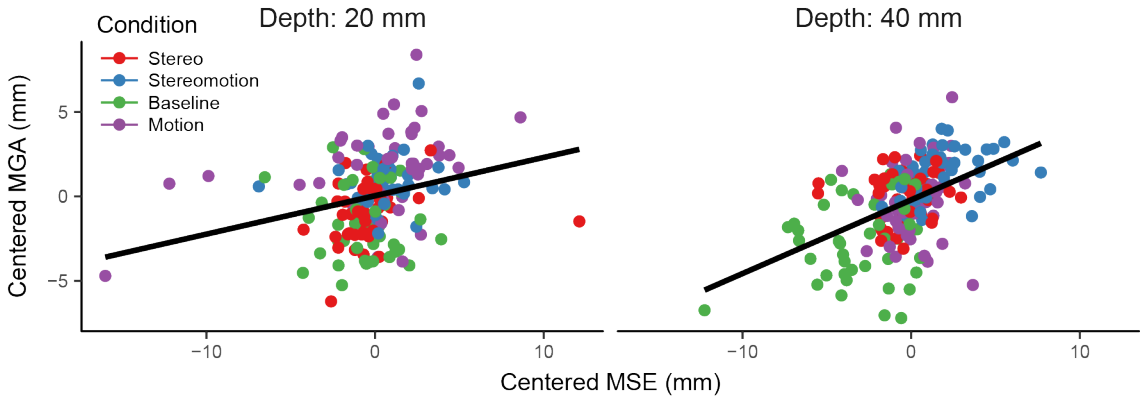

Figure 6.4. Scatterplot showing the correlation between the individual *MGA*s and the individual *MSE*s. Four data points are presented for each subject, one for each viewing condition. Both variables are centered with respect to each subject's average. Left panel, results for the smaller object (depth of 20 mm); right panel, results for the bigger object (depth of 40 mm). Black lines indicate the direction of the correlation between the two variables.

CHAPTER 7

General Discussion

7.1 How are absolute and relative distance estimated for perception and grasping?

The fundamental role of binocular information for guiding action is largely undisputed. Indeed, several investigations have shown that stereovision is one of the most powerful cues mediating goal directed actions (Servos et al., 1992; Harris, 2004) and that removing it from visual input has detrimental effects on both the programming and control of reaching and grasping in humans (Marotta et al., 1995; Keefe & Watt, 2009; Bradshaw & Elliott, 2003). It is also clear that 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 et al., 2007; Mon-Williams & Dijkerman, 1999; Watt & Bradshaw, 2000). What remains unclear is whether stereovision provides the action system with an accurate representation of an object 3D shape and distance.

The results of four experiments yield clear evidence of an inaccurate mapping between the physical space and visual estimates for reach-to-grasp movements. This mapping is characterized by two interrelated distortions: (1) The range of reaching distances is smaller than the range of physical distances, resulting in visuospatial compression and (2) the grip aperture of two grasps directed at the same object at two different distances was greater when reaching near the body than when reaching far from the body, indicating depth underconstancy. The results of the first experiment show that both phenomena stem from distorted estimates of egocentric distances. The data also reveals that perceptual depth judgments follow the same pattern of biases. The fundamental relationship between the *FHP* of the grasp and the *MSE* is the most relevant finding

of the experiment: it shows that the estimation of an egocentric property of an object (its distance to the observer) during an action task predicts the estimation of an allocentric property (the object's front-to-back extent) during a perceptual task. This finding is particularly remarkable, considering that the two estimates were measured through two motor responses (*FHP* and *MSE*) that are utterly distinct from a biomechanical standpoint. Even more surprisingly, figure 4.5 shows that this relation can be predicted at a subject-by-subject level.

Evidence of a common distorted representation for perception and action was confirmed in the second experiment, where subjects grasped objects at different distances after setting their physical structure in order to match their perceptual appearance. Even though perceptual matching produced larger depth settings for the farther objects than for the closer objects, to compensate for depth underconstancy, they were grasped with identical grip apertures. These findings strongly suggest that visual estimates at action planning are systematically biased as a result of visuospatial compression.

However, a methodological caveat characterizing these experiments requires a word of caution in the interpretation of the results. Since participants were executing grasping movements without vision of their hand and without feeling the physical presence of the object at action completion, it may be argued that their actions were “pantomimed”, as if participants pretended to grasp, which can be different then performing a movement with full vision. In fact, the action task in Experiment 1 was somewhere in between an “open-loop” grasp, where neither the hand nor the stimulus can be seen (but the stimulus is touched) and a “pantomimed grasp”, where the stimulus is removed prior to the movement onset but the hand can be seen during the movement. It is commonly assumed that the open-loop grasp is a true action (as in, not guided by perceptual biases), whereas the pantomimed grasp is instead more akin to a perceptual judgment, essentially because the absence of final touch in the latter case forces participants to shape their grip aperture based on a memory of the object's shape (Goodale et al., 1994; Westwood et al., 2000; Schenk et al., 2011; Schenk, 2012).

To control for these possible sources of additional biases (lack of online control and memory) we adopted two solutions. First, in experiment 1 we designed a grasping task where subjects received continuous vision of the stimulus but no final touch nor visual feedback of the hand. This satisfied two criteria: i) memory was not required for the execution of the movement; ii) the object's estimated distance and depth were explicitly measurable via the *FHP* and *FGA* respectively, and could be directly compared with the *MSE*. Second, we ran two control experiments where we replicated the same grasp task, while also introducing consistent visual and haptic feedback from the fingertips.

In the first control experiment (experiment 3), two distinct analyses of the results support the previously formulated hypotheses and deliver further insights on the processes regulating grasp control. First, in agreement with the *FGA* data of the first two experiments, the grip aperture, measured at different moments before the thumb reached its contact point, was systematically smaller when reaching to farther objects than when aiming to close objects. Ruling out the aforementioned alternative explanation, these findings constitute converging evidence of biased estimates of 3D properties from stereo information at action planning that persist until the very last instants of a grasping movement. Second, a significant effect of the previous trial suggests that these biases are detected when the fingers make contact with the object, inducing a correction on the subsequent grasp: after grasping an object whose depth had been overestimated (the near stimulus), participants produced a greater grip aperture in the following grasp, and vice versa.

The findings of the second control experiment (experiment 4) confirm that the planning of an action performed with online control is no different from the planning of an action performed without feedback, as clearly indicated by the trajectories shown in figure 4.16. Furthermore, Experiment 4 unveils another interesting characteristic of the visual processes underlying the execution of a grasp in depth. While estimating depth is indispensable in performing a manual estimation task, subjects could in principle avoid doing it altogether during a grasp, and simply guide their fingers towards their respective contact points. If this was true then the final positions

of the index and thumb in a front-to-back grasp (*FIP* and *FTP* respectively) should coincide with the endpoints of two pointing movements aimed at the same contact locations. In Experiment 4, subjects were asked to either pinch the egocentric position of the front and the back surfaces of a 3D structure, or to grasp the entire structure front-to-back. The endpoints of the front and back pinches were compared with the *FTP* and *FIP* afterwards. We found that the estimate of the front surface's position was identical in the two tasks (the *FTP* almost overlapped with the front pinch), suggesting that the reaching distance was encoded in egocentric coordinates (the thumb normally leads the reaching in a front-to-back grasp – Volcic & Domini, 2014). However, the *FIP* and the back pinch did not match, suggesting that the back surface was encoded according to different frames of reference depending on the task. In agreement with this hypothesis, we found that the *FIP* was correctly predicted by assuming that the index's contact location was encoded in allocentric coordinates (that is, relative to a biased estimate of the front surface's distance).

This result does not necessarily imply that a division between egocentric and allocentric processing of visual information is the standard in all kinds of grasp. Many studies have shown clear evidence that prehensile actions can still be successfully executed by guiding the fingers to specific contact locations while disregarding any information about an object's size or other allocentric properties (Smeets & Brenner, 1995, 1999, 2008a, 2008b; Smeets et al., 2002). Interestingly though, Sousa and collaborators (Sousa et al., 2010, 2011) found that when observers are asked to judge the position of an object, the presence of a second object visible in the back affects their responses, even if in principle the far object could be completely ignored. This suggests that perceived distance is indeed influenced by both egocentric and allocentric information. In agreement with these findings, the present study suggests that individual digits can be guided within different frames of reference, at least in the context of a front-to-back grasp: the thumb is positioned according to an egocentric coordinate system, while the index is positioned according to an interaction between egocentric and allocentric coordinate systems.

It may also be argued that the biases we found were caused by the use of virtual reality (*VR*) instead of real objects in a natural environment, primarily due to the absence of online visual control of the hand (Cuijpers et al., 2004, 2006; Magdalon et al., 2011; Levin et al., 2015). We do acknowledge that certain biases in 3D perception may be peculiar to *VR*, because it inevitably entails some degree of cues-to-flatness, but most critically because not all participants may have enough familiarity with virtual environments. However, recent investigations actually disprove this latter point: for example, Bozzacchi et al., 2014 found that subjects interact in the very same way with virtual and real objects, as long as they are provided with both consistent haptic feedback and online vision of the hand. In agreement with this, our subjects reported feeling equally confident when grasping both in absence and in presence of online control, and the great overlap between the trajectories of the trials with and without visual feedback in the fourth experiment demonstrates this unequivocally. Thus, we can safely conclude that *VR* was absolutely apt for the study. Besides, the ecological validity of *VR* was not at the center of this investigation. Rather, the goal was to test whether grasp and perception lead to similar or different biases *given the same visual conditions*. Whether depth estimation per se could be different in *VR* than in the real world, and we do agree that under certain circumstances it probably is, is absolutely important considering the increasing interest for this technique in many fields, and demands for further investigation.

The data from the present study, together with those of recent investigations (Bozzacchi et al., 2014, 2015, 2016; Volcic et al., 2013; Volcic & Domini, 2016), start delineating a clearer picture of the role of stereovision for the control of reaching and grasping. Contrary to common assumptions (Servos et al., 1992; Knill, 2005), binocular information is not sufficient for an accurate estimate of an object egocentric and allocentric properties. Instead, the egocentric map, resulting from systematic errors in distance estimates from ocular vergence, is compressed. In turn, this visuospatial compression leads to erroneous estimates of object depth (Brenner & van Damme, 1999). On that note, our results offer a novel insight on the apparent inconsistencies

between egocentric and allocentric maps, which have been extensively documented as affecting reaching and grasping movements (Bingham et al., 2004; Bingham et al., 2000; Bingham, 2005; Smeets et al., 2002; Thaler and Goodale, 2010; Coats et al., 2008; Mon-Williams and Bingham, 2007; Lee et al., 2008). In agreement with these findings, our data indicate two coexisting but inconsistent mappings between the physical space and the perception/action space. The compression of visual space, which affects the estimate of egocentric distances, is geometrically incompatible with the pattern of distortions in depth estimates. The incongruity is glaring: a homogenously compressed world should contain homogeneously compressed objects and distances. Instead, as experiment 4 shows, participants experienced expanded and contracted shapes embedded in a linearly compressed set of distances. These biases reinforce the idea that specific aspects of grasping imply specific 3D information processing (Jeannerod, 1986; Brenner & Cornelissen, 2000) and that reaching distance and grip aperture can be calibrated independently (Coats et al., 2008).

Given this state of affairs, planning is in general based on wrong estimates of distal properties, which naturally leads to the conclusion that the success of goal directed actions is critically dependent on the availability of feedback signals for online corrections. This was confirmed by the results of experiment 4, where subjects could never reach to the veridical distances unless provided with online vision of the fingertips and consistent tactile feedback.

Furthermore, the effect of previous trial highlights the dynamic aspect of the visuomotor mapping, since these corrections are clearly registered as sensorimotor prediction errors. These findings support the idea that online control plays an active role during grasping: continuous vision of the hand and final haptic feedback do not simply guarantee the successful execution of the movement, they also help the visuomotor system register the errors that might have occurred in the course of previous actions. Naturally, these errors are either negligible or completely absent in normal grasping conditions, because multiple visual cues contribute to the fine estimation of the 3D properties of objects, both at planning and during the reaching. Nevertheless, the presence

of predictable systematic motor corrections when only vergence and binocular disparity are available remarks the fundamental interplay between proprioception and vision.

The overall important aspect of these results is that visual estimates at action planning are consistent with the appearance of an object's allocentric and egocentric properties. Remarkably, the compression of visual space, quantified at the level of individual participants as the ratio of the span of the actual reached distances to the span of physical distances, can predict depth judgments in a perceptual task. In other words, the geometry of a subjectively unique visual space, determines each individual's phenomenological experience of 3D structure and kinematics of motor actions.

7.2 Mechanisms of depth cues combination

There are many open questions regarding the mechanisms underlying the execution of visually guided actions. In four experiments, we addressed two of them, related to each other: first, it is assumed that 3D vision for action is different from 3D vision for perception, because prehensile movements appear to resist to virtually all the biases that characterize depth judgments (Carey, 2001; Haffenden et al., 2001); second, following the last point, it is consequently hypothesized that 3D vision for action yields accurate estimates of shape and distance (Goodale, 2011). A central point in this theory is the notion that stereo information is especially important, since removing binocular vision significantly impairs the ability to carry out a grasp smoothly and successfully. Hence, the conclusion that follows is that stereovision is accurate during grasping (Bradshaw et al., 2004; Loftus et al., 2004; Jackson et al., 1997; Melmoth & Grant, 2006; Mon-Williams and Dijkerman, 1999; Servos & Goodale, 1994; Servos et al., 1992; Watt & Bradshaw, 2000; Knill, 2005).

In contrast with these hypotheses, recent experiments have shown that, when binocular information is the only source of depth information, grasping and perception are affected by the

same biases (Foster et al., 2011; Bozzacchi et al., 2014, 2016; Bozzacchi et al., 2015, Volcic et al., 2013). The question then is how do these findings fit with the existing evidence in the literature. One alternative explanation that has been proposed starts from the assumption that motor plans and perceptual judgments actually originate from the same distorted representation of space, which is affected by visuospatial compression (see the first study of this thesis): near the body, distances are overestimated, whereas far from the body distances are underestimated. During grasping though, online feedback (vision of the hand and touch) is exploited to rapidly correct the biases of the planning, and to generate an error signal. Through this signal, the mapping between the visual estimates and the motor command is constantly updated. Under everyday viewing conditions, errors at planning are negligible, and therefore so are the compensatory in-flight corrections. If, however, the visual input is impoverished, such as when only stereovision is available, the visuomotor system needs to make a heavier and more extensive use of online control signals to terminate an action. This would explain why systematic biases can be measured during grasping when stereovision is the only source of depth information available. Moreover, grasping normally happens under visual conditions that include plenty of depth sources, such as motion parallax and texture gradients. For all these reasons, it is possible that successful grasps are not necessarily dependent on an accurate analysis of 3D information, but that they critically depend both on rich visual scenes and on the availability of online control feedback.

In the present study, we tested this alternative explanation. In the first experiment, we presented the participants with a target object, a parabolic cylinder, whose shape was specified either by binocular disparity alone (stereo condition), or by disparity and motion (stereomotion condition). The stimulus was seen at two distances (300 and 450 mm), and subjects either manually judged its depth (*MSE* task) or grasped it front-to-back. In the grasp task, participants had full online control over their movements: they could see the virtual rendering of their fingertips while reaching, and they also touched the object at the end of the movement.

In the stereo condition, the *MSE* showed an underconstancy bias that was compatible with visuospatial compression: objects near the body appeared deeper then objects far from the body, because binocular disparities were scaled by an overestimated distance in the former case, and by an underestimated distance in the latter. Similarly, the grasp kinematics indicated that depth underconstancy affected action as well. At the end of the movement, the grip aperture (*GA*) was larger when reaching for the stimulus at the near distance than at the far distance. Overall, these results supported the hypothesis that perceptual judgments and grasping motor plans stem from the same distortion of space.

In the stereomotion condition, we expected to see an improvement in both tasks. The rationale was that, since the motion gradient does not change with distance (unlike binocular disparity), depth-from-motion should be less affected by visuospatial compression than depth-from-stereo. As a result, the stereomotion-defined stimulus should appear more constant along the line of sight. On the contrary, participants reported that the moving object looked even more distorted, as if the depth underconstancy bias was magnified, instead of reduced: the near object looked even deeper than the far object. Interestingly, the same exact pattern of results was found in the grasping task.

Overall, the results of the first experiment confirmed the initial hypothesis only partially. On the one hand, the data clearly indicated that perception and action share the same distorted space, not different ones. On the other hand, and paradoxically, when the stimulus was specified by multiple depth cues, shape recovery was seemingly less accurate than when the same object was specified by only stereovision. This second conclusion was further supported by the results of a control experiment (1a), in which we asked participants to estimate or grasp the width of the same stimulus. This control paradigm served two goals: i) to demonstrate that 2D size estimation is constant within small ranges of distances and is also not affected by modulations of the depth information; ii) to estimate the depth-to-width aspect ratio of the stimuli, by dividing the estimates of depth by those of size obtained in experiments 1 and 1a respectively. The analysis of

the aspect ratio revealed that three-fourth of the objects of the stimulus set looked more convex than they actually were. That is, their depth was overestimated. Within them, stereomotion stimuli looked overall more convex than their stereo-defined homologues.

An explanation of these results assumes that the goal of the visuomotor system is to maximize the discriminability of depth differences. In other words, depth signals directly specify depth order information. Let's suppose that two points are separated by a tiny distance along the line of sight, and that A is in front of B. Individual cues (binocular disparity, motion, texture et cetera) can provide information about the depth order between A and B. Although each one does so with a different reliability, the probability that an observer correctly detects that A is in front of B increases constantly as more depth cues are available. Let's imagine that the observer was asked to report the distance between the points in two conditions: using only disparity information or through a combination of stereo and motion information. Even if the physical position of the A and B did not change in the two cases, the observer might end up reporting that A and B are more distant during the stereomotion condition, simply because in that case they appear to be separated in depth more reliably than in the stereo condition. In addition, the metric estimate of the exact separation between point A and point B may be subject to other biases, because depth interpretation often depends on other scene parameters that act as scaling factors (for example, the viewing distance – see figure 5.1).

A model that predicts this behavior is the Intrinsic Constraints model (Domini et al., 2006). According to IC, the visual system first combines the depth cues available at a given moment (disparity, motion, texture and so forth) to obtain a composite signal. The combination rule is optimal, in that the composite signal thus obtained has a signal-to-noise ratio (*SNR*) that is greater than that of each individual cue in isolation. The increase in *SNR* corresponds to a greater probability of detecting even small separations along the z axis. Therefore, the prediction of the IC model is that the apparent depth should increase monotonically with the *SNR* of the composite signal. In a second stage, the combined signal is then scaled by a scene parameter (i.e., the

viewing distance) to obtain a metric estimate of depth. On this account, the results of the first experiment can be interpreted as follows: in the stereo condition, the disparity signal constituted a composite signal on its own, yielding a certain estimate of depth; however, in the stereomotion condition, the integration of disparity and motion increased the *SNR* of the composite signal, leading to a greater estimate of depth in return (Di Luca et al., 2010; Domini et al., 2011).

This hypothesis was supported by the results of a second experiment, in which we substituted motion with texture. Since the design and tasks were otherwise identical, at the light of the previous findings, if the IC model is correct, then the stereotexture stimulus should be judged and grasped as deeper than the stereo stimulus, and that this increment should not necessarily compensate for depth underconstancy. Indeed, this is what we found. Interestingly, we also tested a second type of grasp, where subjects touched the object along the side, which ensured full visibility of both fingers till object contact. This was different from the previous front-to-back grasps, where the index disappeared during the final critical moments of the grasp, because it was occluded by the object itself. In this latter case, we reasoned that the subjects' strategy was to first position the index roughly where the back surface was expected to be, according to the object's depth as estimated at planning. Then, the forefinger was pulled mechanically until it would actually touch the physical surface. By contrast, the thumb was guided visually for the entire duration of the movement. This could explain why the depth underconstancy bias peaked typically at the end of the trajectory. In agreement with this hypothesis, we found a dramatic reduction of the underconstancy bias in the along-the-side grasp, compared to the front-to-back grasp.

Another fundamental source of information is represented by the final haptic feedback. Subjects could feel the physical object at movement completion, and thus potentially compare the ground truth obtained via touch with the initial estimate produced at planning. The result was an error signal which significantly affected the execution of the subsequent movement on a trial-by-trial basis. After touching the multiple-cue stimulus (stereomotion or stereotexture), which was

normally overestimated, the *GA* of the next grasp was smaller than average. Conversely, the *GA* was larger following a grasp where the stereo stimulus had been touched, because subjects realized to have underestimated its depth.

These findings remark that the feedback coming from the vision of the hand and tactile sensations is not simply exploited to guide the reaching in real time, but it is actively incorporated by the visuomotor system as an error detection signal, which is then used to update the motor plans and refine the movement kinematics of the subsequent action. The data from this study indicates that the outcome of grasping actions cannot be predicted and interpreted based exclusively on the visual information available at planning, and converging evidence has been increasingly collected in this direction in the recent years. The high accuracy in the motor execution is the product of the dynamic integration of visual analysis and proprioceptive experience.

Finally, this investigation leaves few questions open for future research, particularly with respect to the exact mechanisms behind the processing of 3D information for grasping. In this and other previous studies, we observed that stereo information was not sufficient to achieve shape constancy. This is a remarkable finding already on its own, because binocular disparity is normally the most reliable depth cue within the peripersonal space (Howard & Rogers, 1995), and it has repeatedly proved to be the most important signal for grasping (Watt & Bradshaw, 2003; Jackson et al., 1997; Melmoth & Grant, 2006). It is possible that this behavior was in part due to the use of Virtual Reality (*VR*), which inherently entails some degree of flatness information or prior for frontoparallel (Adams & Mamassian, 2004) that can potentially contrast or limit the influence of the depth cues. Although the results from the control experiment 1a suggest otherwise, we cannot rule out this possibility entirely.

Even more surprisingly, experiments 1 and 2 showed that the presence of additional monocular cues did not seem to make shape estimation or grasping become any more accurate. In fact, the data suggest quite the opposite: with two cues available, shape constancy was always worse than with only stereo information. Parenthetically, it should be noticed that the grid we used in the

stereotexture condition was intentionally richer than a pure texture patch, as it also contained some linear perspective information. However, participants' responses did not seem to benefit from this additional feature either, at least in terms of accuracy.

All these observations raise the question as to whether accurate 3D shape reconstruction for grasping requires even more depth information than what we used, or maybe a different set of cues. We think that the latter case is the least likely, since we picked the three most reliable signals for this study, so any other cue would be weaker. Instead, it may be that the visuomotor system needs to combine multiple sources of depth information at the same time to be able to compensate for biases such as visuospatial compression. Another possibility is that accuracy is simply not the goal of 3D vision for grasping to begin with. If this is true, then one potential heuristic is to plan the movement based on a first approximation of few key 3D parameters (for example, an object's distance and depth), and then guide the fingers at the intended location using online control as the movement unfolds (akin to the planning/execution model proposed in Glover, 2004). The comparison between the front-to-back and the along-the-side grasps seems to suggest that this could be the case.

7.3 Virtual versus natural environments

Grasping normally occurs in environments where there is no conflict of 3D information. When preparing a grasp in a natural scene, the observer can rely on a rich set of depth cues, ranging from blur gradients to binocular disparity, all specifying the same geometry of an object. The result is typically an efficient and accurate movement, which is apparently unsusceptible to the many systematic biases affecting 3D shape estimation during perceptual tasks. In contrast with this conclusion, in the second study of this thesis, we found that the Grip Aperture (*GA*) during the final phase of a grasp does not reflect the true shape of an object, but rather it increases with the amount of depth information specifying that shape. Critically, the same pattern of results was

found for perceptual judgments of depth. Based on this evidence, we concluded that 3D shape estimation is not veridical for both perception and grasping.

Although similarities between perception and action have been previously reported in literature (Franz et al., 2000; Franz et al., 2001; Franz, 2001; Christiansen et al., 2014; Vishton et al., 1999; Smeets & Brenner, 1995; Smeets & Brenner, 1999), in our specific case we could not rule out the possibility that the errors in both tasks were caused by the use of virtual reality. In *VR*, blur and accommodation are usually in conflict with the rest of the visual information because the monitor is flat, thus causing depth underestimation. As a consequence, it can be hypothesized that this "flattening" effect is reduced as more depth cues become available: the more layers of depth information are added to the visual scene (stereo, motion, texture and so forth), the weaker is the weight of the flatness cues. Consequently, a virtual object should appear deeper when specified by stereo and motion than when specified by stereovision alone. Furthermore, *VR* could affect the execution of grasping movements in a similar way, since it has been shown that participants act less smoothly and spontaneously in virtual environments than in natural settings (Cuijpers et al., 2008). An increased caution during the execution of the grasp, or even the simple cognitive awareness of the presence of a flat monitor, may allow flatness cues to affect the movement kinematics.

In order to test these two alternative accounts, we designed an experiment where participants judged and grasped a physical cardboard-made prism, whose shape was defined by different combinations of depth cues. Despite the whole study was performed in a natural environment, hence there was no conflict of depth information, participants showed the exact same biases found using *VR* in both tasks: the prism was judged and grasped as deeper when blur, accommodation, stereo and motion cues were all available at the same time, compared to when only blur and accommodation (baseline cues) were present. Furthermore, intermediate results were found for partial combinations of the above signals (baseline plus stereo or baseline plus motion). This data provides further converging evidence that 3D vision is intrinsically inaccurate

for both perception and action, in line with a number of previous studies (Foster et al., 2011; Bozzacchi et al., 2015; Bozzacchi et al., 2014, 2016), and that *VR* cannot be the main responsible for the biases we found in our previous experiments.

These evidence reinforce the hypothesis that the same processing of depth information feeds the perceptual estimation of shape as well as the planning of grasping movements. Importantly, the outcome of this processing is not compatible with an optimal combination of unbiased estimates of depth, based on the reliability of their underlying cues: if all depth signals specify the same veridical shape, only with different precision, then we should expect the same *MSE*/*MGA* in all four conditions, with decreasing noise as more cues become available. However, we found otherwise. It could also be hypothesized that the smaller *MSE* and *MGA* in the Baseline condition were caused by the interference of a prior to flatness, that is, a known tendency to underestimate depth in absence of reliable depth information (Adams & Mamassian, 2004). However, the fact that the Stereomotion condition elicited an even greater response than the other conditions indicates that stereo and motion signals were combined in a superadditive fashion: the estimate of depth resulting from their integration was bigger than the sum of the estimates produced by the individual cues. This is especially significant, considering that binocular disparities are extremely reliable within the reachable space (the viewing distance in our experiment was only 42 cm): if depth estimates were combined in an optimal fashion, then adding motion to stereo should at most increase the precision of the perceptual (or motor) response, and not of its magnitude.

Consistently with our previous conclusions, we argue that the goal of cue integration for 3D vision is to maximize the ability to discriminate depth intervals (as opposed to maximizing the accuracy of depth estimates): the visuomotor system aims to detect the presence of depth when it is physically present, responding stronger to richer array of signals. Remarkably, although each observer does so to a different extent, every individual is consistent across perception and action: the comparison between the responses of the manual-estimation and of the grasp tasks on a

subject-by-subject level showed that participants who did not modulate their *MSE* significantly across the viewing conditions, also exhibited a weak modulation of the *MGA*, and vice versa.

Appendix

Egocentric function

Consider the scenario depicted in figure 1.2A: an observer is asked to estimate the egocentric distance of a set of probe locations. Results from previous researches (Foley, 1980; Bozzacchi et al., 2015) have shown that distances near to the body are overshot whereas distances far from the body are undershot (phenomenon that we refer to as *visuospatial compression*). This kind of relationship between estimated and physical distance can be described by a linear function (which we name *egocentric function*):

$$z_f' = z_A + c \cdot (z_f - z_A) \qquad [1]$$

where z_f' is the estimated distance, z_A is a hypothetical distance at which the estimation is accurate, c is a compression parameter that is less than 1, and z_f is the actual viewing distance. Note that the parameter c is also the slope of the egocentric function (compression factor), and it is inversely related to the magnitude of the visuospatial compression. For example, a mild compression (near distances slightly overestimated, far distances slightly underestimated) corresponds to values of c that are just below 1. On the contrary, as c becomes progressively smaller than 1, the egocentric function gets flatter and the visual space collapses over a given z_A. Finally, suppose that the observer were instead always veridical. This situation corresponds to the case of $c = 1$, when equation [1] is simply the equality $z_f' = z_f$.

In summary, the parameter c is the key term that allows to quantify the presence and magnitude of a visuospatial compression, given a set of estimated distances and of corresponding physical distances.

Scaling function

When the eyes fixate one of two points that are at two different distances from the observer, the point that is not being fixated projects on two different locations in each retina, generating a horizontal disparity that is approximately equal to

$$d \cong \frac{\Delta z \cdot IOD}{z_f^2} \qquad [2]$$

whereΔz is the relative depth between the two points, *IOD* is the Inter-Ocular Distance and z_f is the viewing distance. We can solve equation [2] by Δz to calculate the relative depth that would generate a given disparity *d* when fixating at a distance z_f given a certain *IOD*:

$$\Delta z = \frac{d \cdot {z_f}^2}{IOD} \qquad [3]$$

Consider now the scenario depicted in figure 1.2B: a participant is asked to estimate the relative depth of an object that is presented at some distance from them. Just like Δz, the estimated depth $\Delta z'$ can be defined using equation [3] as

$$\Delta z' = \frac{d \cdot {z_s}^2}{IOD} \qquad [4]$$

wherez_s is the scaling distance, that is, the distance at which an object of depth $\Delta z'$ needs to be brought in order to project the disparity *d* (for a given *IOD*). In the last equation we can substitute *d* with equation [2], simplify, and update the definition of estimated depth as follows:

$$\Delta z' = \Delta z \cdot \left(\frac{z_s}{z_f}\right)^2 \qquad [5]$$

Equation [5] conveniently summarizes the relationship between estimated depth, physical depth, scaling distance and physical distance. In fact, according to equation [5], depth estimation is veridical ($\Delta z' = \Delta z$) only when the ratio $\frac{z_s}{z_f}$ is equal to 1, that is, when the scaling distance matches the viewing distance. If, however, the participant scales the binocular disparity by a distance that is greater than the viewing distance ($\frac{z_s}{z_f} > 1$) , they end up overestimating the object's depth as well. The opposite happens when $\frac{z_s}{z_f} < 1$.

Equation [5] 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 [1]). Therefore, we define the *scaling function* as:

$$z_s = z_A + c \cdot (z_f - z_A) \quad [6]$$

Comparison of visuospatial compressions

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 in fact two alternative definitions of the same mechanism); b) to determine whether the same visuospatial compression also underlies both perception and grasping.

In order to answer to point a) in experiment 1, we used a grasping task to take direct measures of z_f' and $\Delta z'$ at the same time (the FHP and the FGA respectively) and we compared them. To answer to point b), we took the MSE as a direct measure of $\Delta z'$ during the perceptual task and we compared it with the FHP from the grasping task.

The next step was to quantify the magnitude of the visuospatial compression from these measures. We fitted the FHP with equation [1] and we extracted the compression parameter c_{FHP} for each subject (values on the abscissa of both graphs in figure 4.5). Notably, the mean z_A of the sample was 449.9 ± 17.7 mm, which is in line with what we found in Volcic et al., 2013.

To estimate the same parameter from the FGA and the MSE, we substituted z_s with the scaling function in equation [5], and we fitted both variables with the following model:

$$\Delta z' = m + \Delta z \cdot \left(\frac{z_A + c \cdot (z_f - z_A)}{z_f}\right)^2 \quad [7]$$

where the parameter *m* accounted for the average individual differences in the opening of the grip aperture (individual margin). Using equation [7] we extracted c_{FGA} and c_{MSE} for each subject (values on the ordinate of figure 4.5-left and 4.5-right respectively). The average z_A extracted from the FGA and the MSE were 404.4 ± 7.9 and 402.5 ± 10.21 respectively, again remarking the close similarity between during the tasks in terms of visual space.

Predicted Final Index Position according to the dual map hypothesis (experiment 4)

As described by equation [5], the apparent depth of an object is a function of the physical depth and of the squared ratio between the apparent distance and the physical distance of that object. We used the thumb's final position from the pinch as an estimate of the apparent distance of the front rod during the planning of the grasp (since they coincided), and plugged it in equation [5] to obtain an estimate the object's depth:

$$\Delta z' = \Delta z \cdot \left(\frac{FTP_{pinch}}{z_f}\right)^2 \qquad [8]$$

where Δz is the physical depth (40 or 50 mm) and z_f is the front rod's distance. The predicted *FIP* was simply the sum of *FTP* plus $\Delta z'$.

References

Adams, W. J., & Mamassian, P. (2004). Bayesian combination of ambiguous shape cues. *Journal of vision*, *4*(10), 7-7.

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.

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.

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.

Binsted, G., & Heath, M. (2004). Can the motor system utilize a stored representation to control movement?. *Behavioral and Brain Sciences*, *27*(01), 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., & 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.

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

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, F. (2012). The same object and at least three different grip apertures. *Journal of Vision*, *12*(9), 429-429.

Carey, D. P. (2001). Do action systems resist visual illusions?. *Trends in cognitive sciences*, *5*(3), 109-113.

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.

Christiansen, J. H., Christensen, J., Grünbaum, T., & Kyllingsbæk, S. (2014). A common representation of spatial features drives action and perception: grasping and judging object features within trials. *PloS one*, *9*(5), e94744.

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), 409-409.

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.

Di Luca, M., Domini, F., & Caudek, C. (2010). Inconsistency of perceived 3D shape. *Vision research*, *50*(16), 1519-1531.

Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision research*, *46*(11), 1707-1723.

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

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.

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.

Franz, V. H. (2001). Action does not resist visual illusions. *Trends in cognitive sciences*, *5*(11), 457-459.

Franz, V. H., Fahle, M., Bülthoff, H. H., & Gegenfurtner, K. R. (2001). Effects of visual illusions on grasping. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(5), 1124.

Franz, V. H., Gegenfurtner, K. R., Bülthoff, H. H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychological Science*, *11*(1), 20-25.

Frisby, J. P. Buckley, D. Wishart, K. A. Porrill, J. Garding, J. Mayhew, J. E. (1995). Interaction of stereo and texture cues in the perception of 3-dimensional steps. *Vision Research*, 35, 1463–1472.

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

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.

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

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception*, *27*(11), 1357-1365.

Glover, S. (2004). Separate visual representations in the planning and control of action. *Behavioral and brain sciences*, *27*(01), 3-24.

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

Gogel, W. C. (1977). The metric of visual space. InW. 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. (2008). Action without perception in human vision. *Cognitive Neuropsychology*, *25*(7-8), 891-919.

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

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*(678-91), 1.

Haffenden, A. M., Schiff, K. C., & Goodale, M. A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp. *Current Biology*, *11*(3), 177-181.

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.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, *8*(3), 33-33.

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*. Oxford University Press, USA.

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 visuomotor pattern. *Behavioural brain research*, *19*(2), 99-116.

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

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision research*, *34*(17), 2259-2275.

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.

Knill, D. C. (2005).Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception. Journal of Vision, 5(16), 103–115.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision research*, *35*(3), 389-412.

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. (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., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(4), 906.

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.

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. (2016). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 1-9.

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.

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.

Mon-Williams, M., & Bingham, G. P. (2008). Ontological issues in distance perception: Cue use under full cue conditions cannot be inferred from use under controlled conditions. *Perception & psychophysics*, *70*(3), 551-561.

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

Mon-Williams, M., & Tresilian, J. R. (1999). Some recent studies on the extraretinal contribution to distance perception. *Perception*, *28*(2), 167-181.

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.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.URL 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. (2000). Distance estimation in the visual and visuomotor systems. *Experimental Brain Research, 130*(1), 35-47.

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. (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.

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

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, 18, 11-16.

Thaler, L., & Goodale, M. A. (2010). Beyond distance and direction: The brain represents target locations non-metrically. *Journal of vision*, *10*(3), 3-3.

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.

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.

Vishton, P. M., Rea, J. G., Cutting, J. E., & Nuñez, L. N. (1999). Comparing effects of the horizontal-vertical illusion on grip scaling and judgment: relative versus absolute, not perception versus action. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1659.

Vishwanath, D. (2014). Toward a new theory of stereopsis. *Psychological review*, *121*(2), 151.

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*, 1-13.

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.

Wallach, H., & Zuckerman, C. (1963). The constancy of stereoscopic depth. *The American journal of psychology*, *76*(3), 404-412.

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., & Bradshaw, M. F. (2003). The visual control of reaching and grasping: binocular disparity and motion parallax. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 404.

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10), 7-7.

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.

Willemsen, P., Gooch, A. A., Thompson, W. B., & Creem-Regehr, S. H. (2008). Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, *17*(1), 91-101.

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