Exp Brain Res (1998) 121:465–470

RESEARCH ARTICLE

J.J. Marotta · M.A. Goodale

The role of learned pictorial cues in the programming and control of grasping

Received: 10 December 1997 / Accepted: 11 March 1998

Abstract Binocular information has been shown to be important for the programming and control of reaching and grasping. Even without binocular vision, people are still able to reach out and pick up objects accurately – albeit less efficiently. It remains unclear, which of the many available monocular depth cues humans use to calibrate manual prehension when binocular information is not available. In the present experiment, we examined whether or not subjects could use a learned relationship between the elevation of a goal object in the visual scene and its distance to help program and control the required grasp. The elevation of the goal object was systematically varied with distance in some blocks of trials by presenting the object at different positions along a horizontal plane 35 cm below eye level. In other blocks of trials, elevation did not vary with distance because the objects were always presented along the subject's line of sight. When subjects viewed these two displays monocularly, they showed fewer on-line adjustments in the trajectory of the limb and the aperture of the fingers when the elevation of the target object in the visual scene could be used to help program the required movements. No such difference between performance on the two arrays was seen when subjects were allowed a full binocular view. This study confirms that subjects are indeed able to use a learned relationship between the elevation of an object and its distance as a cue for programming grasping movements when binocular information is not available. Together with evidence from work with neurological patients who have difficulty perceiving pictorial cues, these findings suggest that the visuomotor system might normally "prefer" to use binocular cues, but can fall back on learned pictorial information when binocular vision is denied.

J.J. Marotta · M.A. Goodale (✉)
Vision and Motor Control Laboratory,
Department of Psychology,
The University of Western Ontario, London,
Ontario, Canada, N6A-5C2
e-mail: goodale@uwo.ca,
Fax: 1-519-661-3961, Tel.: 1-519-661-2070

Key words Monocular · Binocular · Grasping · Pictorial cues

Introduction

Historically most studies of depth vision have concentrated on judgements about the relative distance of objects and the role of depth vision in figure-ground segregation. Recently, however, an increasing number of studies have examined the role of depth vision in the calibration and control of motor outputs, such as those involved in jumping (e.g. Ellard et al. 1984) or manual prehension (e.g. Jackson et al. 1997; Servos et al. 1992). For example, when we reach out to pick up an object such as a cup, our motor system must have access to information about the exact location of the cup in egocentric space and information about its actual size. In short, relative distance or size information is not sufficient. Instead, the absolute distance of the object (within a particular frame of reference) must be computed in order to program both the trajectory of the reach and the aperture of the grasp.

What are the critical sources of distance information used by the visuomotor system in controlling manual prehension? Binocular vision appears to be particularly important. When binocular vision is restricted by means of an eye-patch, people are much less "efficient" in reaching out and picking up objects. They reach more slowly, show longer periods of deceleration and execute more on-line adjustments of both their trajectory and their grip size during the closing phase of the grasp (Dijkerman et al. 1996; Jackson et al. 1997; Marotta et al. 1995a, b; Servos et al. 1992; Servos and Goodale 1994). One of the most striking differences between binocular and monocular reaches is the number of on-line adjustments made by the subject during the execution of the movement, particularly in the closing phases of the movement (Kruyer et al. 1997; Marotta et al. 1998). These adjustments appear to arise as a consequence of errors in the initial estimate of the target's distance. Although these findings demonstrate the importance of binocular vision in the programming and control of grasping, the particular roles of stereopsis, convergence and other binocular cues have yet to be disentangled.

Even though removing binocular vision has significant effects on performance, people are still able to reach out and pick up objects reasonably accurately. They must therefore be relying on monocular cues – but which ones? There are many that could be exploited. One source of monocular information about distance (and thus object size) that is potentially available is the motion of the object (and the scene) on the retina – particularly motion generated by head movements. We found that enucleated individuals, who have lost an eye due to an accident or disease, made larger and faster head movements during the execution of their reaching movements, suggesting that they were using this as a strategy to generate useful motion cues (Marotta et al. 1995a, b). Although subjects with normal vision do not adapt this strategy when faced with a monocular view, recent work in our lab indicates that restricting their head movements impairs reaching and grasping performance under monocular, – but not binocular – viewing conditions (Marotta et al., in press). However, head motion, and the cues it generates, cannot be the whole story. Even when the head is fixed, subjects can still generate reasonably competent grasping movements when only monocular information is available. So the question remains: what cues are they using?

One possibility is that subjects can use *pictorial* information from the goal object itself, or the scene in which it is embedded, to help calibrate reaching movements. Depending upon the way in which the goal objects are presented, subjects can use learned pictorial cues like interposition, familiar size and perspective to help calibrate their grasp – the same cues that artists have been using for years to give an impression of three-dimensional structure on a two-dimensional canvas. For example, as the distance of an object changes, so does the angle from which it is viewed; as a consequence, the object's shape on the subject's retina also changes. The retinal image of a block will become more trapezoidal in shape as the viewing angle changes with increasing distance. If the size of the table on which the object is resting is known, then the position of the object with respect to the leading or far edge of the table can also be a reliable cue to distance and (with the help of retinal image size) the size of the object. Thus, local perspective and shape cues may not only enable us to construct the relations between objects in a scene, but with experience can be used to calculate the actual distance and size of objects we wish to pick up.

The idea that pictorial cues are used to compensate for the loss of binocular information is supported by work from Marotta et al. (1997). In that study, we found that individuals with visual-form agnosia, who are unable to perceive many of the pictorial cues present in visual scenes, are much more disadvantaged in the control of their grasp than are normal observers when binocular information is removed. In fact, in another study involving grasping in normal observers, we found that when the availability of both binocular and pictorial cues to distance and object size were restricted, performance became particularly poor (Kruyer et al. 1997). This work makes it clear that pictorial cues play an important role in programming and controlling reaching and grasping under monocular conditions. Nevertheless, which pictorial cues in particular are used in the programming and control of grasping remains to be determined.

The exploitation of pictorial cues undoubtedly depends both on hard-wired visual mechanisms and on experience. All of us have been reaching and grasping for a long time – both in terms of evolution and personal experience. Many pictorial cues, such as occlusion, may rely on years of experience with objects in natural settings. But even those cues that depend on long-term learning may be fined-tuned or 'calibrated' for the task at hand.

One important cue to object distance that we use every day and that could be fine-tuned for particular situations is the elevation of the object in the visual scene. For an observer positioned above a horizontal ground surface, the horizon of the ground is at eye-level and all locations on the ground are projected into the lower half of the visual field. As we look from near objects to ones further away, the further objects are typically higher in the visual scene. Thus, in a particular scene, objects situated higher in the visual array will tend to be perceived as being further away (Dunn et al. 1965; Epstein 1966; Sedgwick 1986; Smith 1958; Wallach and O'Leary 1982). This cue to depth occurs for successive as well as simultaneous presentations of objects; thus, even when objects are presented one at a time in total isolation, if elevation is allowed to vary across trials, it will have a powerful influence on perceived distance (Sedgwick 1986). Previous studies of this pictorial cue have utilized perceptual estimates of distance (Dunn et al. 1965; Epstein 1966; Smith 1958; Wallach et al. 1982). The role of elevation in the calibration of prehension movements has not been studied, even though most previous studies of grasping (e.g. Jakobson and Goodale 1991; Jeannerod 1986; Servos et al. 1992) have required subjects to reach out and pick up objects that were located at different distances below the subject's line of sight on a table top – a situation that provides subjects with an opportunity to learn that the higher an object is in the visual scene, the further away it is. If elevation information is to be of any use in these situations, subjects must calibrate their visuomotor system on the basis of their experiences with real metrics.

To test whether or not subjects could learn the relationship between the elevation of an object in the visual scene and its distance in a grasping task, we created two different arrays in which the goal objects could be presented. In one array or context, the goal objects were presented in a "traditional" fashion at different positions along a horizontal plane below the subject's line of sight. In the other array, the objects were placed at different distances along the subject's line of sight. These two arrays are illustrated in Fig. 1. Notice that we were able to compare movements made to objects placed at the same location, but in two different arrays. Notice too that only one of these arrays

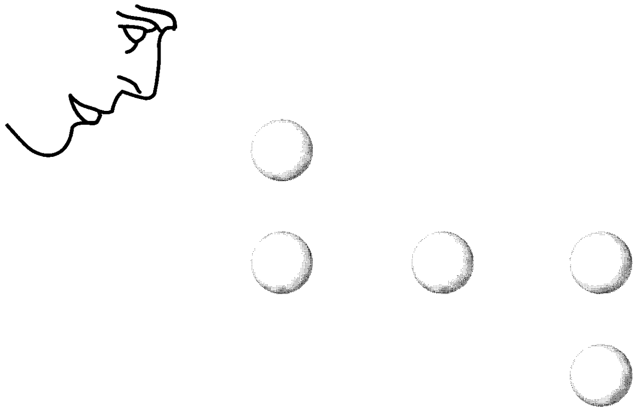

Fig. 1 Diagram of presentation arrays (flat and angled presentation arrays)

afforded the subject an opportunity to learn the relationship between elevation in the visual scene and location – just as is the case in many conventional studies of reaching and grasping where objects are placed at different positions on a horizontal surface. Subjects were tested with each of these arrays under both monocular and binocular viewing conditions. We anticipated that subjects would have to rely on elevation as a cue to location only when binocular information was denied. Thus, when subjects viewed the two arrays monocularly, we expected them to show more on-line corrections when reaching to objects in the array in which elevation could not be exploited.

Materials and methods

The experiment was carried out at the University of Western Ontario in compliance with the Social Sciences and Humanities Research Council (Canada) Guidelines (1981).

Subjects

Thirteen right-handed subjects (7 males, 6 females; age range 22–39 years old; mean age 26.69 years old) with normal or corrected-to-normal vision participated in the experiment, for which they were paid. Subjects were strongly right-handed, as determined by a modified version of the Edinburgh handedness inventory (Oldfield 1971). All subjects had stereoscopic vision in the normal range with assessed stereoacuity of 40″ of arc or better as determined by the Randot Stereotest (Stereo Optical, Chicago).

Apparatus

In this study, we utilised a cue-deprived test environment developed in our laboratory (Kruyer et al. 1997). In this test environment, a wide range of depth cues can be removed by presenting lit spheres in 3-D grasping space to subjects who are in the dark and viewing the scene with only one eye. By systematically reintroducing depth cues into this severely cue-deprived environment, we are able to examine the contribution of individual depth cues to the programming and control of manual prehension.

Three sizes of styrofoam spheres (6.25, 7.5 and 10 cm in diameter) were presented one at a time on a rod that could be positioned in one of a number of sockets in a vertical matte-black presentation board (183×120 cm). The center of each sphere contained 4 light-emitting diodes (LEDs) controlled by computer. The voltage sent to each sphere was controlled so that the surface luminance levels for each size of the sphere was equivalent (10 candelas per m^2 as measured by a light meter). (It should be noted that perfect spheres would offer no retinal disparity cues to depth or distance. Of course, the styrofoam spheres we used were not perfect and, in addition, had a textured surface. Moreover, even with perfect spheres, other binocular cues such as convergence would provide depth information.)

The spheres were presented in two different presentation arrays: a "flat" array was used in which a single sphere was presented approximately 35 cm below eye-level, at one of three different distances (11, 32.5 and 53.5 cm) from a start key located 69.5 cm in front of the board, and 57 cm below eye-level. Thus, because all three positions fell along a horizontal plane, elevation in the scene could be used as a cue to distance. In the other block of trials, an "angled" array was used in which the sphere was presented at one of three different positions (20, 35 and 50 cm below eye-level; 11, 32.5 and 53.5 cm from the start key, respectively), such that all three positions fell along the same line of sight when the subject gazed downwards. Thus, in this condition, elevation of the sphere in the visual scene did not vary with distance and could not be used as a distance cue. These two arrays are illustrated in Fig. 1. A common position in both arrays (35 cm below eye-level, 32.5 cm from the start key) was selected as the target position for data analysis, so that any differences in the kinematics of reaching and grasping movements would have to be due to differences in the information available in the two arrays.

Subjects sat in an adjustable chair with their hand on the start key. They wore PLATO spectacles (Translucent Technologies Inc., Toronto) throughout the testing sessions. These liquid-crystal shutter spectacles permitted monocular or binocular viewing and, when both shutters were closed, prevented subjects from viewing the spheres being put into position. Subjects also wore earphones that emitted white noise between trials to prevent subjects from using any audible cues from the spheres being put into position. The room was dark and subjects reached for the spheres which remained lit for 2.5 s under monocular viewing conditions.

Three 8-mm-diameter infrared light-emitting diodes (IREDs) were attached with small pieces of cloth adhesive tape to the radius at the wrist, the ulnar border of the thumbnail and the distal portion of the index fingertip of the subject's right hand. The tape allowed complete freedom of movement of the hand and fingers.

The IREDs were monitored by an infrared sensitive camera system (Optotrak) positioned approximately 2 m from the subject. The 3-dimensional coordinates of the IREDs were stored by the Optotrak's data acquisition unit and later filtered off-line (with a low-pass second-order Butterworth filter with a 7-Hz cut-off).

Procedure

At the beginning of the test session, subjects were given the handedness questionnaire and tested for eye dominance (viewing preference). Subjects were then seated in front of the presentation board with the tips of their index finger and thumb of their right hand on the start button. Their chair was adjusted so that the spheres in the angled array would all fall along the same line of sight. This was the subjects' "start" position, which they returned to after they had completed a reach. Subjects were informed that, in one block of trials, a sphere would be presented at one of three positions along a flat horizontal plane, while in the other block of trials, a sphere would be presented at one of three different heights along an angled plane. Subjects were instructed that as soon as they saw the lit sphere, they were to reach out quickly, accurately and as "naturally" as possible and grab hold of it with their whole hand, but not to pull it off of the rod and continue to hold onto the sphere until they heard a tone signalling the end of the trial. The experimenter initiated the start of a trial by signalling the computer to simultaneously activate the goggles, which allowed the preferred eye to view the scene, and illuminate the spheres for a period of 2.5 s.

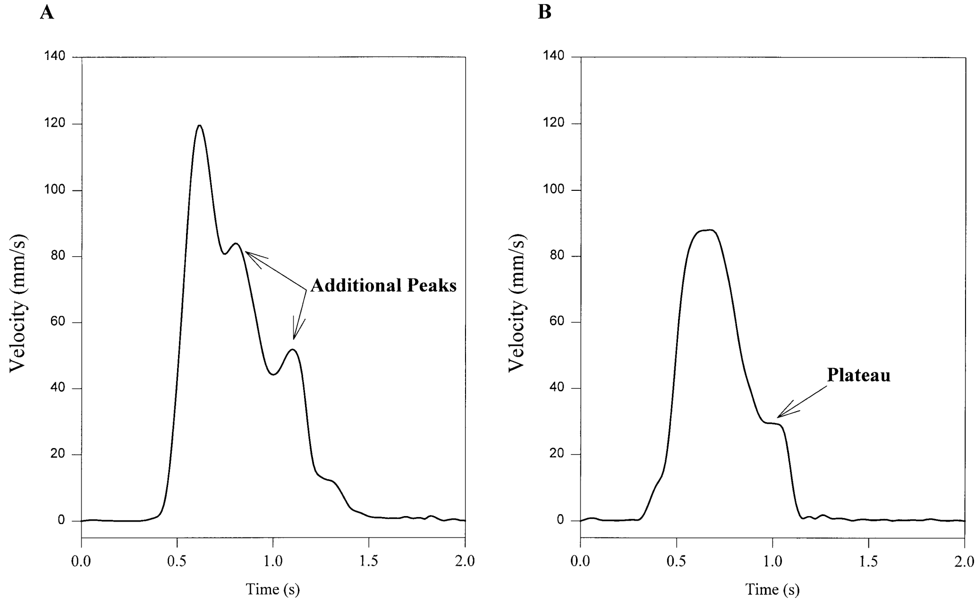

Fig. 2 Velocity profiles from two individual trials which exhibit additional peaks (**A**) and a plateau (**B**). Similar patterns are seen in aperture profiles

Subjects were administered four testing blocks of 27 experimental trials, each consisting of five instances of each of the three sphere sizes at the target position (35 cm below eye-level, 32.5 cm from the start key) that were used for the data analysis, along with two instances of each of the remaining six distance × sphere size combinations. Trial presentation was random and each testing block was preceded by five practice trials. The testing session lasted for approximately 90 min.

Dependent measures

If subjects program their grasp on the basis of an incorrect estimate of target distance, then they will have to make an on-line correction in order to acquire the target. If they overestimate the distance, then they will sometimes collide with the target. If they underestimate the distance, then they will have to adjust the trajectory (and grasp) during the closing phase in order to make successful contact. These latter movements in particular have been observed in a number of different experiments in our laboratory where cues to distance (and thus object size) were either ambiguous or absent (Kruyer et al. 1997; Marotta and Goodale 1996; Marotta and Goodale 1997). The methods we have developed for quantifying these adjustments are outlined below.

Velocity

In a typical reach, subjects accelerate smoothly to a peak (or maximum) velocity and then decelerate as their hand approaches the object to be grasped. Occasionally, however, subjects show on-line adjustments in the reach, which are evident as "additional peaks and plateaus" in the velocity profile. A peak is defined as a sharp increase in velocity followed by a decrease (see Fig. 2a); a plateau is a flat portion of 30 ms or more on the velocity profile, which is preceded and followed by a slope of the same sign (see Fig. 2b). The number of these additional velocity peaks and plateaus were recorded for each trial.

Aperture

In a typical grasp, subjects open their hand smoothly to a peak (or maximum) aperture and close it as their hand approaches the object. As with their reach, subjects occasionally adjust their grasp on-line. Again these adjustments are reflected as additional peaks and plateaus in the aperture profile. The number of these additional aperture peaks and plateaus were recorded for each trial.

These measures provide a more accurate representation of the "efficiency" of a manual prehension movement than do more "traditional" kinematic measures (e.g. maximum velocity, maximum grip aperture). We have recently shown that these on-line corrections can reveal differences in performance that are not evident in the pattern of traditional kinematic measures (Marotta et al. 1997).

Results

For each of the subjects, mean values of each of the dependent measures in each viewing condition were calculated. (Equipment failure resulted in some loss of data, but this constituted less than 3% of the trials). The values were entered into separate 2×2×3 (presentation array × viewing condition × sphere size) repeated-measures analyses of variance. All tests of significance were based on an alpha level of 0.05. Post hoc Neuman-Keuls analysis were performed where necessary.

As was seen in previous studies (Kruyer et al. 1997; Marotta et al. 1997), under monocular viewing conditions, subjects produced more on-line corrections in their reaching and grasping movements than they did when binocular vision was available. As can be seen in Fig. 3, monocular reaches exhibited significantly more peaks ($F_{(1,12)}$=71.12, P<0.0001) and plateaus ($F_{(1,12)}$=5.33, P<0.05) in their velocity profiles than did binocular reaches. Similarly, as can be seen in Fig. 4, monocular grasping movements showed significantly more peaks in their aperture profiles than binocular grasping movements ($F_{(1,12)}$=37.34, P<0.0005).

When the spheres were presented in the angled array, subjects showed significantly more peaks (P<0.01) and plateaus (P<0.01) in their velocity profiles than when the spheres were presented in the flat array – but only when they viewed the displays with one eye covered. Un-

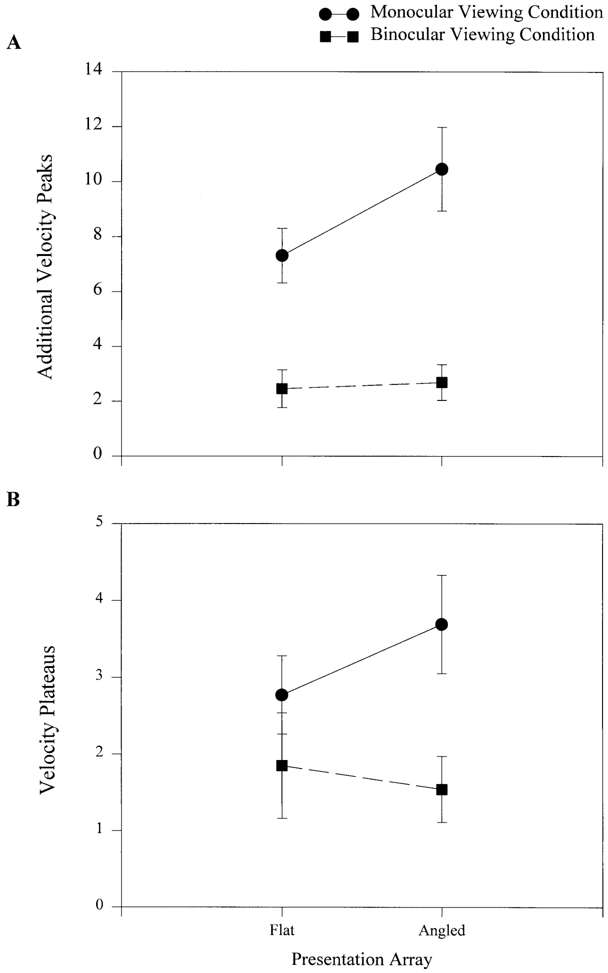

Fig. 3 The effects of presentation array and viewing condition on additional velocity peaks (**A**) and velocity plateaus (**B**). *Error bars* SEMs, *closed circles* monocular viewing condition, *filled squares* binocular viewing condition

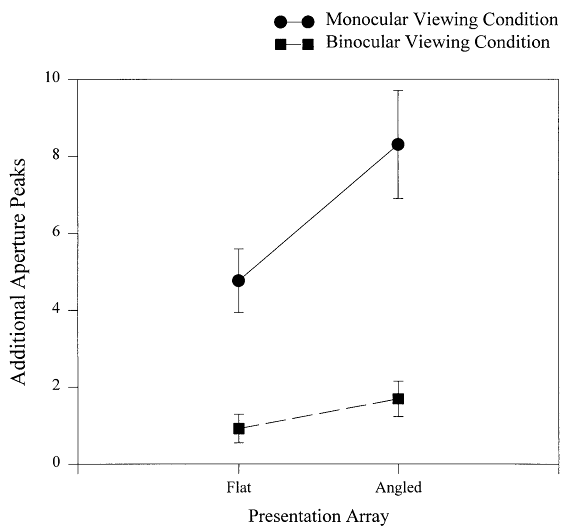

Fig. 4 The effects of presentation array and viewing condition on additional aperture peaks. *Error bars* SEMs, *closed circles* monocular viewing condition, *filled squares* binocular viewing condition

der binocular conditions, the number of additional velocity peaks ($P>0.05$) and plateaus ($P>0.05$) did not vary as a function of the presentation arrays. This interaction between the presentation arrays and the viewing conditions is evident in Fig. 3 for both peaks ($F_{(1,12)}=8.39$, $P<0.05$) and plateaus ($F_{(1,12)}=8.35$, $P<0.05$). As can be seen in Fig. 4, this pattern of results also holds for the number of additional aperture peaks ($P<0.01$) produced during these reaches ($F_{(1,12)}=4.98$, $P<0.05$). There were too few plateaus in the aperture profiles (on average, one or less per subject) for any meaningful differences to emerge between conditions.

Discussion

When the elevation of an object in the visual scene was not a useful cue to distance in the task we used, subjects produced more on-line corrections in their reaching and grasping movements – corrections that were evident as additional peaks and plateaus in their velocity and aperture profiles – but only under monocular viewing conditions. This result suggests that subjects can learn to exploit the relationship between the elevation of an object and its distance in a particular context to help program and control reaching and grasping movements, particularly when binocular cues are not available.

These results thus suggest that subjects were applying a learned algorithm relating elevation and distance, which was calibrated for the particular plane below eye level on which the objects were positioned in this experiment. One must be cautious, however, before making this inference. It is conceivable that subjects were not using this strategy at all. Instead, as subjects became familiar with the three different places at which objects could appear along the horizontal plane, they would simply learn to use a different trajectory for each of the different positions. If this were the case, one would expect to see the performance of subjects improve over the course of the testing block, as they became more familiar with the three different positions. There was no evidence for such learning. The fact that subjects knew in what arrangement the spheres would be presented (coupled with the brief amount of practice that they had) was enough to allow them to exploit the relationship between elevation and distance in these blocks of trials. Indeed, the vital piece of information the subjects needed to perform this task was given to them by the experimenter – the spheres would be positioned along a horizontal plane. Once the height of the plane was known, the visuomotor system was able to adjust the function relating elevation to distance in this new scene.

The on-line corrections presumably occurred on trials in which subjects underestimated rather than overestimated the distance of the sphere. Thus, if subjects decelerated and began to close their grasp too early, they would be forced to correct their trajectory to acquire the target. Of course, when they overestimated the distance of the target, they would encounter the target sooner than anticipated, would be unable to decelerate properly and would collide with it with some force. On occasion, such collisions did occur, and when they did, were obvious to both the subject and the experimenter. Unfortunately, such collisions (or near collisions) were difficult to measure unambiguously.

This study provides clear evidence that systematic variations in the elevation of the goal object in the visual field can be used to predict the position and distance of that object for the programming and control of a grasping movement when binocular information is not available. The results of this study imply that the visuomotor system can use learned pictorial information when binocular information is not available. Of course, even when binocular information was available, subjects could have been using elevation information in addition. The availability of such pictorial information, however, did not improve their performance over what they could do without it, provided they still could use binocular vision. Furthermore, the availability of learned elevation cues in the monocular viewing condition, even though improving performance, did not completely substitute for binocular vision. Subjects were never as good monocularly as they were binocularly. Taken together with the neuropsychological work by Marotta et al. (1997), discussed earlier, these results suggest that, under normal viewing conditions, the visuomotor system "prefers" to use binocular cues, but is able to fall back on pictorial information when binocular vision is denied.

Acknowledgements This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to M.A. Goodale and a postgraduate scholarship from the Ontario Graduate Scholarship Program to J.J. Marotta. We would like to thank J. Orphan, D. Pulham and Y. Hu for their technical support and T. James for his assistance with illustration preparation. Last but not least, we would like to thank the subjects for the cooperation they showed throughout the testing.

References

Dijkerman HC, Milner AD, Carey DP (1996) The perception and prehension of objects oriented in the depth plane. 1. Effects of visual form agnosia. Exp Brain Res 112:442–451

Dunn BE, Gray GC, Thompson D (1965) Relative height on the picture plane and depth perception. Percept Mot Skills 21:227–236

Ellard CG, Goodale MA, Timney B (1984) Distance estimation in the Mongolian gerbil: the role of dynamic depth cue. Behav Brain Res 14:29–39

Epstein W (1966) Perceived depth as a function of relative height under three background conditions. J Exp Psychol 72:335–338

Jackson SR, Jones CA, Newport R, Pritchard C (1997) A kinematic analysis of goal-directed prehension movements executed under binocular, monocular, and memory-guided viewing conditions. Vis Cognit 4:113–142

Jakobson LS, Goodale MA (1991) Factors affecting higher-order movement planning: a kinematic analysis of human prehension. Exp Brain Res 86:199–208

Jeannerod M (1986) The formation of finger grip during prehension: a cortically mediated visuomotor pattern. Behav Brain Res 19:99–116

Kruyer A, Marotta JJ, Goodale MA (1997) Scene-based binocular cues improve grip trajectory accuracy. Invest Ophthalmol Vis Sci 38: 988

Marotta JJ, Goodale MA (1996) Height and distance in the visual field: calibrating a monocularly guided reach. Soc Neurosci 22:885

Marotta JJ, Goodale MA (1997) Elevation in the visual scene: calibrating a monocularly guided reach. Invest Ophthal Vis Sci 38:988

Marotta JJ, Goodale MA, Kruyer A (1998) The role of head movements in the control of manual prehension. Exp Brain Res 120:134–138

Marotta JJ, Perrot TS, Nicolle D, Goodale MA (1995a) The development of adaptive head movements following enucleation. Eye 9:333–336

Marotta JJ, Perrot TS, Nicolle D, Servos P, Goodale MA (1995b) Adapting to monocular vision: grasping with one eye. Exp Brain Res 104:107–114

Marotta JJ, Behrmann M, Goodale MA (1997) The removal of binocular cues disrupts the calibration of grasping in patients with visual form agnosia. Exp Brain Res 116:113–121

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–112

Sedgwick HA (1986) Space perception. In: Boff KR, Kaufman L, Thomas JP (eds) Handbook or perception and human performance, vol 1, Sensory processes and perception. Wiley-Interscience, New York, pp 1–57

Servos P, Goodale MA (1994) Binocular vision and the on-line control of human prehension. Exp Brain Res 98:119–127

Servos P, Goodale MA, Jakobson LS (1992) The role of binocular vision in prehension: a kinematic analysis. Vis Res 32:1513–1521

Smith OW (1958) Judgments of size and distance in photographs. Am J Psychol 71:529–538

Wallach H, O'Leary A (1982) Slope of regard as a distance cue. Percept Psychophys 31:145–148