

The role of binocular vision in grasping: a small stimulus-set distorts results

Bruce D. Keefe · Simon J. Watt

Received: 10 October 2008 / Accepted: 15 January 2009 / Published online: 6 February 2009
© Springer-Verlag 2009

Abstract The role of binocular vision in grasping has frequently been assessed by measuring the effects on grasp kinematics of covering one eye. These studies have typically used three or fewer objects presented at three or fewer distances, raising the possibility that participants learn the properties of the stimulus set. If so, even relatively poor visual information may be sufficient to identify which object/distance configuration is presented on a given trial, in effect providing an additional source of depth information. Here we show that the availability of this uncontrolled cue leads to an underestimate of the effects of removing binocular information, and therefore to an overestimate of the effectiveness of the remaining cues. We measured the effects of removing binocular cues on visually open-loop grasps using (1) a conventional small stimulus-set, and (2) a large, pseudo-randomised stimulus set, which could not be learned. Removing binocular cues resulted in a significant change in grip aperture scaling in both conditions: peak grip apertures were larger (when reaching to small objects), and scaled less with increases in object size. However, this effect was significantly larger with the randomised stimulus set. These results confirm that binocular information makes a significant contribution to grasp planning. Moreover, they suggest that learned stimulus

information can contribute to grasping in typical experiments, and so the contribution of information from binocular vision (and from other depth cues) may not have been measured accurately.

Keywords Reaching and grasping · Depth perception · Cue integration · Visuo-motor control · Familiar size

Introduction

Over the past two decades there has been a steady growth in research aimed at understanding the roles of different depth cues in the control of reach-to-grasp movements. Much of this effort has focused on the importance of information from binocular vision (horizontal and vertical binocular disparities, and the eyes' angle of convergence) on the planning and online control of grasping (see Melmoth and Grant 2006, for a recent review). The majority of studies have employed similar experimental designs. Movement kinematics (maximum velocity, maximum grip aperture, etc.) are compared under normal binocular viewing and when binocular information is removed by covering one eye. The largest and most consistent effects of removing information from binocular vision have been found towards the end of the movement, which is thought to be predominantly under online control (Servos et al. 1992; Jackson et al. 1997; Watt and Bradshaw 2000; Loftus et al. 2004; Melmoth and Grant 2006). Based on these findings, a consensus view is emerging that binocular vision is particularly important for the online control of grasps, but less so for providing an initial estimate of object properties to plan movements (Melmoth and Grant 2006).

The above experimental design has been directly adopted from 'classical' perception experiments, which

Electronic supplementary material The online version of this article (doi:10.1007/s00221-009-1718-4) contains supplementary material, which is available to authorized users.

B. D. Keefe · S. J. Watt
School of Psychology, Bangor University, Wales, UK

S. J. Watt (✉)
School of Psychology, Bangor University, Adeilad Brigantia,
Penrallt Road, Gwynedd LL57 2AS, UK
e-mail: s.watt@bangor.ac.uk

have examined the effects on depth perception either of subtracting specific depth cues from the visual scene, or of introducing them into an impoverished scene (e.g. Künnapas 1968; see Sedgwick 1986). The logic is straightforward: the extent to which adding or removing a specific cue causes a change in bias or sensitivity is thought to index the role that cue plays in recovering depth. In order to interpret results unambiguously, however, great care must be taken to eliminate sources of depth information that are not under experimental control, and stimulus artefacts that can ‘artificially’ improve performance. In this paper we highlight ways in which the design of typical grasping experiments differs from that of typical perception experiments, with a resulting reduction in experimental control. We examine the consequences of this for a specific example: investigating the role of binocular information in providing the initial estimate of object properties to programme a grasp.

Depth stimuli potentially contain many geometrical artefacts, which could allow an observer to complete a task using information other than the depth cue(s) under study. For example, consider a simple experiment in which an observer must indicate the perceived slant of a planar stimulus. The angular size of a plane of fixed dimensions varies directly as a function of its slant. Observers could therefore base their judgements of slant magnitude directly on the size at the retina (or size on the screen) of the stimulus, without recovering slant, since there is a one-to-one mapping between angular size and slant magnitude. To control for this, experimenters typically either match the angular size of the stimulus on each trial, or vary it randomly (within a certain range), so as to render this signal uninformative (Knill and Saunders 2003; Hillis et al. 2004).

Psychophysical depth perception studies routinely employ such ‘controls’ to eliminate stimulus artefacts. However, such measures are extremely uncommon in studies of visually guided grasping. Almost all grasping studies use a small set of widely spaced object distances and sizes—rarely more than three of each, and typically fewer—and the dimensions that are not manipulated experimentally (object height, for example) are of a fixed physical size. This means that low-level image properties vary directly with object properties in the manner described above. For example, since object width is normally held constant, an object’s width at the retina uniquely specifies its distance. Moreover, each object \times distance combination is likely to be unique in terms of low-level image properties such as retinal size so, in principle, these simple cues uniquely specify which stimulus configuration is being presented on a given trial.

These stimulus artefacts may be particularly problematic in grasping experiments, because of another difference

from classical perception experiments: the provision of feedback about performance. In depth perception experiments observers usually receive no feedback about the ‘correctness’ of their responses. In contrast, participants in reach-to-grasp experiments almost always successfully grasp the stimulus, thereby receiving kinaesthetic feedback about the true size and distance of the stimulus on every trial. Over repeated trials the consistent relationship between simple low-level image properties and stimulus parameters could therefore be learned.

Even if this lack of control over low-level stimulus properties is overcome (by matching retinal size, etc.) learning of the stimulus set via kinaesthetic signals could still present problems. If the visuo-motor system has learned the sizes and distances used in the study, the intended task of recovering an object’s properties from the available depth cues is reduced to *identifying* which object \times distance combination is presented on a given trial. Stimuli in grasping studies are normally widely spaced along the dimensions of interest (object distances typically differ by at least 100–150 mm, and object sizes by 15 mm or more). So even rather imprecise visual information, which on its own would result in poor estimates of object properties, could be sufficient to identify whether the stimulus is the ‘small’, ‘medium’ or ‘large’ object, for example, allowing its properties to be retrieved from memory.¹ Essentially, the stimuli could become familiar objects (and distances), through the course of the experiment. Familiar size is known to be an effective depth cue (Holway and Boring 1941; O’Leary and Wallach 1980), and information from familiar size has previously been shown to contribute to the control of grasping movements (Marotta and Goodale 2001; McIntosh and Lashley 2008).

The above discussion illustrates ways in which learning the stimulus set could provide an additional, unintended source of information about stimulus properties. What are the likely effects of this? Since the intended (binocular and monocular) depth cues remain available on each trial, it is informative to consider how multiple sources of information are combined to estimate depth. Recent work on sensory integration provides a framework to do this. Psychophysical studies have shown that the combined depth estimate from n independent information sources is well approximated by a weighted sum of depth estimates from each, where each signal is weighted according to its reliability (the reciprocal of its variance) (e.g. Ernst and Banks 2002). This model has been shown to give a good account of depth percepts from binocular disparity and (monocular)

¹ In psychophysical studies, using the same stimuli as this paper, we have found that just-noticeable differences in object depth under monocular viewing are typically 4–6 mm (Watt et al. 2008).

texture/perspective cues (Knill and Saunders 2003; Hillis et al. 2004). This general principle has also been shown to account well for visuo-motor performance towards similarly defined stimuli (Knill 2005; Greenwald and Knill 2008). A key feature of this cue combination ‘strategy’ is that the variance, and therefore the uncertainty, of the resulting depth estimate is always increased by removing signals. Assuming the noise in each signal is independent and Gaussian, the variance of the combined estimate ($\sigma_{\text{combined}}^2$) is given by Eq. 1 (see Oruç et al. 2003 for a derivation):

$$\sigma_{\text{combined}}^2 = \frac{\prod_n \sigma_n^2}{\sum_n \sigma_n^2} \tag{1}$$

Understanding changes in the uncertainty of estimates of object properties is of central importance because the increases in peak grip apertures observed when binocular information is removed are thought to reflect an adaptive compensation for increased uncertainty in these estimates, as well as of hand position (Watt and Bradshaw 2000; Melmoth and Grant 2006; see also Wing et al. 1986). That

is, in response to increased probability of an error, a ‘safety margin’ is added to the grasp, reducing the chance of it being unsuccessful. Consistent with this, it has been shown that perceptual uncertainty, introduced by presenting objects in the visual periphery, causes systematic increases in peak grip aperture (Schlicht and Schrater 2007). Using the reliability-based cue-weighting model, above, we can therefore make general predictions about the effects on grip apertures of removing binocular cues in the presence and absence of learned stimulus information, by considering the changes in variance (uncertainty) of depth estimates in each case.

Figure 1 plots changes in the variance of hypothetical depth estimates when binocular information is removed, both when learned information is available (left-most column) and when it is not (middle column, see caption for details). In the top row, we assume the simplest case in which all three sources of information have equal variance ($\sigma_{\text{monoc}}^2 = \sigma_{\text{binoc}}^2 = \sigma_{\text{learned}}^2$). Removing binocular information always increases the variance of the depth estimate (plot c; Eq. 1), but this effect is smaller when learned

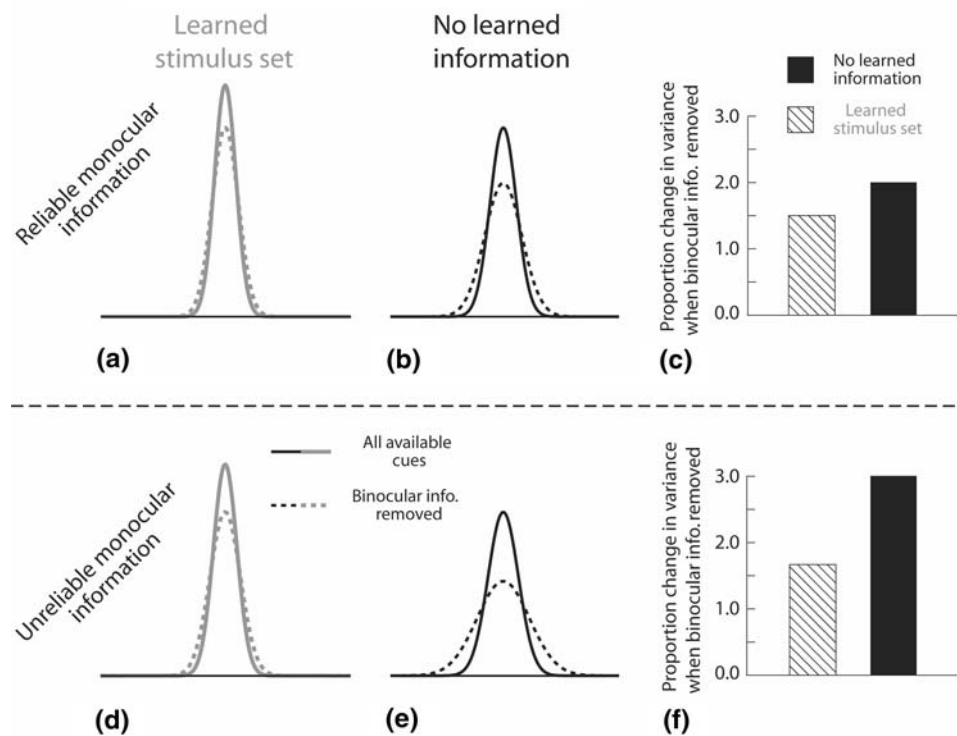


Fig. 1 Changes in the variance of estimates of object properties when binocular information is removed. **a** Hypothetical probability density functions for an estimate of object size when learned information is available. The solid line shows the estimate when all cues are available, and the dashed line shows the estimate when binocular information is removed. **b** Similar probability density functions for an estimate of size from binocular and monocular information only, when learned stimulus information is unavailable. **c** The proportion

increase in the variance of size estimates that results from removal of binocular information in each case. The top row models a case in which the reliability (inverse variance) of all three information sources is equal ($\sigma_{\text{monoc}}^2 = \sigma_{\text{binoc}}^2 = \sigma_{\text{learned}}^2$). The bottom row (**d, e, f**) plots the same information for a case in which monocular cues are relatively unreliable ($\sigma_{\text{monoc}}^2 = 2\sigma_{\text{binoc}}^2 = 2\sigma_{\text{learned}}^2$). Probability density functions were computed using Eq. 1

information is also available. Therefore, we would expect the removal of binocular information to result in a smaller increase in grip apertures when learned information is available than when it is not (Schlicht and Schrater 2007).² The bottom row in Fig. 1 models the (arguably more realistic) case in which monocular cues are relatively unreliable ($\sigma_{\text{monoc}}^2 = 2\sigma_{\text{binoc}}^2 = 2\sigma_{\text{learned}}^2$). Here the confounding effect of learned stimulus information is greater. Removing binocular information results in a much larger increase in variance when learned stimulus information is unavailable, and so would be expected to have a much larger effect on grip apertures under such conditions. The availability of learned stimulus information would therefore be expected to cause misestimation of the effects of removing binocular information.

We examined this prediction directly by comparing the effects of removing binocular information on grasp kinematics in two conditions: (1) a conventional ‘small stimulus-set’ condition, and (2) a ‘randomised stimulus-set’ condition, designed to prevent learning. In the small stimulus-set condition, participants reached to three fixed object sizes at three fixed distances, so learning could occur in the manner described above. In the randomised stimulus-set condition, each trial was different. We used a near-continuous range of stimulus properties, and there was no consistent relationship between low-level image features and object properties. Therefore, participants could do the task only by using information available from the intended binocular and/or monocular depth cues presented on a given trial. We examined visually open-loop grasping to explore the contribution of binocular vision to the initial estimate of object properties, used to programme the movement. The role of binocular information in this phase remains unclear, since closing an eye in open-loop grasping often results in rather small effects (Melmoth and Grant 2006). If this is caused, at least in part, by learning of the stimuli, we should see a clear effect of removing binocular information when stimulus learning is prevented (randomised stimulus set), allowing us to determine unambiguously the contribution of binocular information in this initial phase.

Methods

Participants

Separate groups of 14 right-handed participants completed the small (7 female, 7 male; aged 19–35) and randomised (10 female, 4 male; aged 19–43) stimulus-set conditions, in

² In principle this analysis applies to the effects of removing any depth cue, but we concentrate on binocular cues here.

return for payment. All had normal or corrected to normal vision and stereoacuity < 40 arcsec. Participants gave informed consent prior to taking part, and all procedures were in accordance with the Declaration of Helsinki.

Apparatus and stimuli

The experimental setup and an example of the stimuli are shown in Fig. 2. A chin rest was used to stabilise head position and to locate the participants’ eyes 400 mm above a table surface, directly above a small ‘start button’ fixed to the table on the body midline. Each participant was carefully positioned, and his or her inter-pupillary distance taken into account, to ensure that the visual stimuli were geometrically correct. The visual stimuli were virtual objects, presented on a TFT monitor, and drawn stereoscopically using red–green anaglyph glasses. The monitor was placed face down, 500 mm above the table, and participants viewed a reflection of the stimuli in a horizontal mirror, placed so that the monitor surface was optically coincident with the table surface. The mirror occluded the participants’ hands.

The visual stimuli were rectangular blocks, lying on a textured ground plane coincident with the table surface. The ground plane and object surfaces were defined by the perspective projection of Voronoi diagrams (de Berg et al. 2000; see Fig. 2b), which have been shown to provide effective binocular and monocular depth cues (Knill and Saunders 2003; Hillis et al. 2004). We used anti-aliasing to

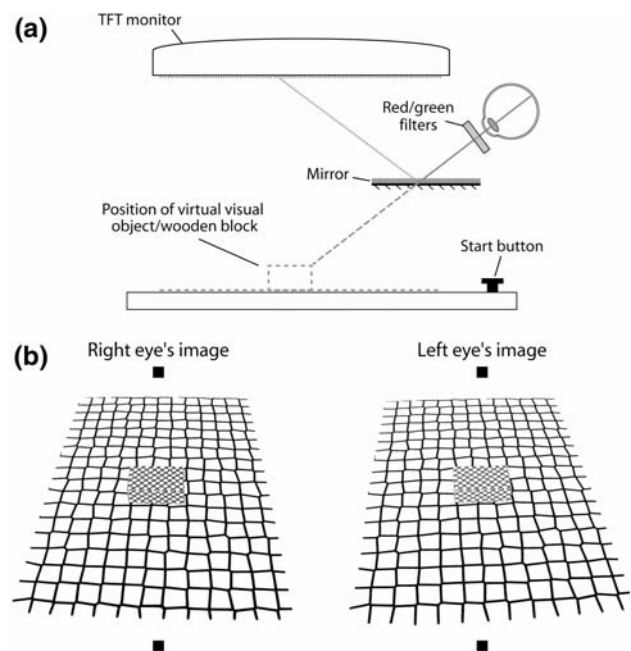


Fig. 2 **a** A diagram of the experimental setup. **b** A stereogram representative of the visual stimuli. Cross-fuse to view. In monocular conditions only the right eye’s image was presented

position stimulus elements with subpixel accuracy. Participants wore anaglyph glasses throughout the experiment, but in the monocular conditions only the right eye's image was drawn, and a patch covered the left eye. Both experiments were completed in the dark, so only the stimulus was visible. Participants grasped real wooden blocks, the relevant dimensions of which exactly matched the depth and distance of the visual objects.

Procedure

On each trial participants held down the start button with their thumb and index finger. The visual stimulus was then displayed, and after 2 s there was an audible beep, which was the signal to grasp the object. Participants were instructed to pick up the object front-to-back, quickly and naturally, using their thumb and index finger only. Releasing the start button extinguished the visual stimulus (the hand was never visible). Movements initiated before the start signal, or more than 600 ms afterwards, were considered void and were repeated at the end of the block.

The apparatus, stimuli, and procedure were identical in the small and randomised stimulus-set conditions with the exception of differences in the exact stimulus parameters used, and the numbers of repetitions. Different participants completed the small stimulus-set, and randomised stimulus-set conditions.

Small stimulus-set condition

The small stimulus set consisted of three object distances (200, 350, and 500 mm from the start button) and three object depths (30, 45, and 60 mm). The visual objects were always 60 mm wide and 25 mm high, and were presented along the body midline. Participants completed six repetitions of each object–distance combination, blocked by viewing condition (binocular or monocular) ($6 \times 3 \times 3 \times 2 = 108$ trials). Trial order was randomised within blocks.

Randomised stimulus-set condition

The randomised stimulus set consisted of six 'base' object distances (200, 250, 300, 350, 400 and 450 mm), and on each trial a random distance (uniform distribution) between 0 and 30 mm was added to the base distance. It was not practical to fully randomise object depth because an equivalent real object was to be grasped. Instead we used a (large) set of seven object depths (30, 35, 40, 45, 50, 55 and 60 mm). We also randomised the width of the visual object on each trial in the range 60 ± 20 mm, to prevent the retinal image size of the object, or its aspect ratio, serving as a simple cue to distance and/or depth. In this condition,

participants completed just one repetition of each combination, and trials were again blocked by viewing condition ($1 \times 7 \times 6 \times 2 = 84$ trials). Therefore, in addition to the randomising of object parameters, they were never presented with the same object/distance combination twice. Trial order was randomised within blocks.

Movement recording

The x , y , and z positions of infrared reflective markers, placed on the nail of the index finger and thumb, were recorded using a ProReflex motion tracking system (sampling at 240 Hz). The raw position data were low-pass filtered (Butterworth filter, 12 Hz cut-off) before computing movement kinematics.

Results

Many kinematic indices can be computed to characterise grasping movements, and they are often highly correlated. Here we restrict our analysis primarily to 'peak grip aperture'—the maximum opening of the finger and thumb as the participant moved towards the object—because our predictions refer specifically to this measure. We also computed 'peak velocity' of the movements (by measuring the three-dimensional velocity of the thumb marker), since this is a primary measure of the transport, or reach component of the movement (Jeannerod 1984, 1988), and has sometimes, though not always, been found to reduce significantly when binocular information is removed (Servos et al. 1992; Jackson et al. 1997; Watt and Bradshaw 2000; Loftus et al. 2004; Melmoth and Grant 2006). We found no significant effects of viewing condition on velocity in either stimulus-set condition. Also, examination of individual velocity profiles, in all conditions, revealed almost no occurrences of 'online corrections' to the movements. We therefore present only grip aperture data here (peak velocity results can be seen in the Supplementary Material). Estimates of kinematic parameters for 2 of the 14 participants in the small-set condition were extremely unreliable and so these participants' data were excluded from the analyses.

Peak grip aperture analysis

To account for differences in grip apertures across individuals (due to differences in hand sizes, for example) we normalised peak grip apertures by expressing them as a proportion of each participant's average peak grip aperture across all trials. For ease of interpretation, and to allow comparisons with previous work, we multiplied these proportion values by the average peak grip aperture across participants, within the small-set, and randomised

conditions, to give normalised grip apertures in millimetres.

Figure 3a, b plot the mean peak grip aperture for all participants in binocular and monocular viewing conditions for the small stimulus set, and randomised set, respectively, as a function of object depth (collapsed across object distance). The solid lines denote the best-fitting linear regressions in each case. It can be seen that in all conditions peak grip apertures increased linearly with object depth in the normal fashion (Jeannerod 1984, 1988). This confirms that grasping movements remained reliable under all conditions. It can also be seen, however, that removing binocular information lead to a clear change in peak grip apertures in both the small stimulus-set, and randomised conditions, and that as predicted, this effect was larger in the randomised conditions.

To test our specific predictions statistically, we computed the best-fitting linear regressions to each participant's grip aperture data, and carried out statistical analysis on the fit parameters. Peak grip apertures are linearly related to object size (Jeannerod 1984, 1988), and using these 'grip aperture scaling functions' we can reliably characterise the overall pattern of each individual's grip apertures, and directly compare grasps made to the different stimulus sets. Figure 3 shows that in both the small and the randomised conditions the effect of removing binocular information was largest for small object sizes, and progressively decreased to zero with increasing object size. This pattern of data has been observed previously (e.g. Watt and Bradshaw 2000), and presumably reflects the fact that, as grip aperture increases overall, progressively increased effort is required to open it yet wider. This effect is also likely to be more pronounced during open-loop grasping, because grip apertures are larger overall when visual feedback is prevented (Jakobson and Goodale 1991). Therefore, although one might expect the principal effect of removing binocular

information to be an increase in the intercept of the grip aperture scaling function (i.e. larger grip apertures), it should also be expected to lead to a reduction in its slope.

Figure 4a, b plot the mean slopes and intercepts, respectively, of the linear regression fits to each individual's data in the small-set and randomised-set conditions. Planned pair-wise t tests confirmed that removing binocular information had a significant effect on grip aperture scaling functions in both stimulus-set conditions. In the small-set condition, the slope of the grip scaling function was significantly less under monocular viewing [$t(11) = 6.33$, $P < 0.001$, one-tailed], and the intercept was significantly higher [$t(11) = 5.50$, $P < 0.001$, one-tailed]. The same pattern of results was observed for the randomised-set condition [slope: $t(13) = 4.54$, $P < 0.001$; intercept: $t(13) = 5.49$, $P < 0.001$; both one-tailed]. Figure 4 also shows that, in line with our predictions, removing binocular information resulted in a greater *change* in grip aperture scaling functions in the randomised condition. To test this statistically we computed the difference in (1) slope, and (2) intercept in the monocular and binocular conditions for each participant (i.e. the effect on each participant of removing binocular information). We then ran independent-measures t tests comparing this effect in the small and randomised conditions, for each regression parameter. This analysis confirmed that removing binocular information resulted in a significantly larger change in both slopes [$t(24) = 1.74$, $P < 0.05$, one-tailed] and intercepts [$t(24) = 2.01$, $P < 0.05$, one-tailed] of the grip aperture scaling functions in the randomised stimulus-set condition.

The magnitude of the effects of removing binocular information

The analysis of the regression parameters, above, demonstrates that removing binocular information caused a

Fig. 3 Mean peak grip apertures as a function of object depth (collapsed across distance) in **a** the small stimulus-set condition, and **b** the randomised stimulus set condition. The *grey circles* denote the binocular condition and the *black squares* denote the monocular condition. The *solid lines* are the best-fitting linear regression to the data in each case. The *error bars* indicate ± 1 SEM

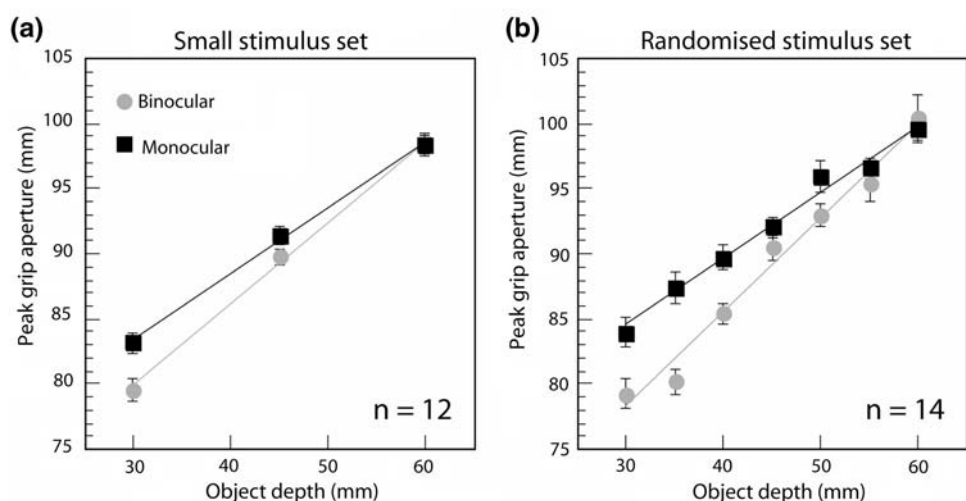
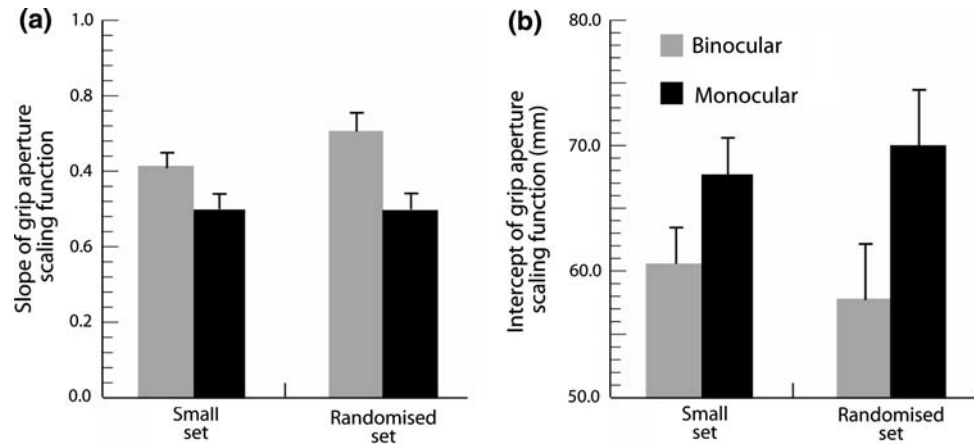


Fig. 4 **a** The mean of the slopes for the best-fitting linear regressions to each participant's grip aperture data, under each experimental condition. **b** The mean intercept of the same fitted curves. Grey bars denote binocular viewing and black bars denote monocular viewing. The error bars are ± 1 SEM



significantly larger change in grip aperture scaling functions when the stimulus could not be learned. How does this relate to actual grip apertures? We used the fit parameters of each individual's grip aperture scaling function to reliably estimate their peak grip aperture for a 30 mm object depth. In the small-set condition the average increase in grip apertures with removal of binocular information was 3.58 mm, compared to 6.27 mm, in the randomised condition. That is, removing binocular information had a 75% greater effect on grip apertures when the stimulus set could not be learned.

Isolating the effect of learning the stimulus set

As described above, increasing object size, and so increasing grip apertures overall, leads to a reduction in the effects of removing information (see Fig. 3). Similar effects are to be expected with increasing object distance, because grip apertures also generally increase with object distance (Loftus et al. 2004; Watt and Bradshaw 2000), presumably to compensate for increased variability in movement endpoint with larger motor responses (Harris and Wolpert 1998). An analysis of grasps made to the subset of near and small objects should therefore yield the least biased estimate of the effect of removing binocular information, allowing us to quantify the effects of learning per se more accurately.

Figure 5 plots mean peak grip apertures for grasps made to the 'near/small' subset of trials. For the small-set condition, we included data from all grasps to the smallest object (30 mm depth) at the near distance (200 mm). For the randomised condition, in order that a sufficient number of trials contributed to the analysis, we included grasps to 30, 35 and 40 mm object depths at a distance of <250 mm (mean object distance with binocular and monocular viewing was very similar: 217 and 215 mm, respectively). Note that the average object depth and distance in the randomised condition (35 and 216 mm, respectively) were

therefore both slightly larger than in the small-set condition.

The overall pattern of data in Fig. 5 is consistent with our predictions. In the small-set condition, removal of binocular information resulted in a small increase in grip apertures but this did not reach statistical significance [$t(11) = 1.68$, $P = 0.06$, one-tailed]. In the randomised set, however, grip apertures were significantly larger under monocular viewing [$t(13) = 3.13$, $P < 0.01$, one-tailed]. Moreover, an independent t test comparing monocular-minus-binocular 'difference scores' in each condition showed that the effect of removing binocular information was significantly larger in the randomised condition [$t(24) = 1.80$, $P < 0.05$, one-tailed]. Therefore, when the

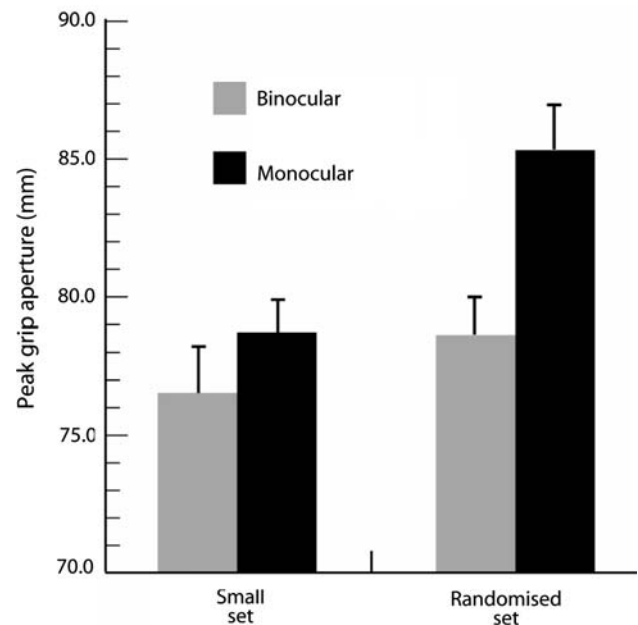


Fig. 5 Mean peak grip apertures for grasps made to the subset of small/near objects. See main text for details. Error bars denote ± 1 SEM

confounding effects of large overall grip apertures were minimised, the effect of learned stimulus information was substantial. Removing binocular information resulted in a three times larger increase in grip apertures with the randomised set than with the small set (6.70 vs. 2.19 mm).

Testing the assumption that learning occurred

The above analysis assumes that the differences between conditions are attributable to learning the stimulus set. This seems reasonable, given that the experimental conditions differed only in terms of the small versus randomised stimulus sets used. Nonetheless, we also looked for direct evidence of learning, which we would expect to see in our small-set condition (but not in our randomised condition). Learning the stimulus set should reduce uncertainty, leading to reduced grip apertures. Under monocular viewing, when learning should have the largest effect, we would therefore expect to see a reduction in grip apertures during the experiment.

To examine this we compared peak grip apertures at the beginning and end of the experiment in both monocular conditions. Given the dependence of the experimental effects on object size we again analysed grasps to the small objects (defined as above). In order to include enough trials to reliably estimate grip apertures, we analysed the first six, and last six grasps to the small objects, across all distances. In the monocular/small-set condition peak grip apertures were significantly smaller at the end of the experiment than at the beginning [mean reduction of 1.8 mm; $t(11) = 2.40$, $P < 0.05$], suggesting that uncertainty in estimates of object properties reduced following repeated exposure to the same stimuli. In the monocular/randomised-set condition there was no significant change in grip apertures across the experiment [mean increase of 1.2 mm; $t(13) = 0.82$, $P = 0.43$]. Taken together, these results suggest that learning of the stimulus set did occur in the small-set condition, but not in the randomised condition.

Discussion

According to the model of sensory integration set out in the “[Introduction](#)”, since closing an eye removes information, it should always result in increased uncertainty in estimates of object properties. This in turn would lead to increased grip apertures, in order to build an increased ‘safety margin’ into the movement (Schlicht and Schrater 2007). All else being equal, however, removing one of two cues will have a greater effect than removing one of three. So we expected to see a larger effect of removing binocular information when learned information was unavailable (leaving only monocular cues) than when it was available

(leaving monocular cues *plus* learned stimulus information). The pattern of data we observed was consistent with these predictions. In both small and randomised stimulus-set conditions, removing binocular information led to significant changes in grip aperture scaling functions (larger grip apertures for grasps to small objects, and reduced grip aperture scaling with object size). However, this effect was significantly larger in the randomised condition, when the stimuli could not be learned, than in the conventional small-set condition. We also found more direct evidence that participants learned the small stimulus set through the course of the experiment: in the monocular/small-set condition only, grip apertures were significantly smaller at the end of the experiment than at the beginning, consistent with a reduction in uncertainty about object properties. Taken together, these results suggest that participants do learn the properties of small stimulus sets in typical grasping experiments, and that this (uncontrolled) information contributes to grasping performance. Similar to previous findings (e.g. Watt and Bradshaw 2000), effects of removing binocular information reduced with increasing object size (and distance), and were absent for grasps to the largest objects. We attribute this to the increased effort required to open the grasp yet wider, as overall grip aperture increases. The analysis of grasps to near and small objects, which should best isolate the effect of learning during the experiment (Fig. 5), suggests that the contribution of learned stimulus information to grasping performance can be substantial.

One obvious implication of our findings is that the effects of removing binocular information may previously have been underestimated. Another way to state this is that ‘monocular’ grasping performance has been overestimated, because monocular cues were not presented in isolation. To draw this conclusion definitively it is useful to consider the magnitude of our effects in comparison to similar studies. In their Experiment 3, Jackson et al. (1997) measured grasps to one 22.5 mm object presented at three distances, and found a grip aperture increase of ~ 3.2 mm when binocular information was removed.³ In Loftus et al.’s (2004) Experiment 4, participants grasped a 25 mm object at three distances, and grip apertures increased by 3.0 mm when binocular information was removed. The effect we observed—an increase of a 3.58 mm for grasps to the 30 mm object (across all distances)—was therefore not atypical. By comparison, in our randomised condition, in which participants could not learn the stimulus set, grip apertures at 30 mm object size increased by 6.27 mm

³ Jackson et al. (1997) reported peak grip apertures in terms of the angle formed by the finger, wrist and thumb markers. We estimated the conversion into millimetres using the average hand size of our observers.

(a 75% greater increase), suggesting that previous studies have indeed underestimated the effect on grip aperture of removing binocular information.

The effect observed in our randomised condition clearly suggests that binocular information does make a significant contribution to the initial estimate of object properties used to programme a grasp, and not just to online control (see Melmoth and Grant 2006). It is important to note, however, that this finding does not necessarily imply a binocular specialism for grasping. Reliability-based cue-weighting predicts that removing *any* reliable source of information will affect grasp kinematics. So observing that removing a given cue leads to a decrement in performance does not, on its own, constitute evidence for that cue having a particular or unique role over any others. Furthermore, since we restricted the available depth cues—information was available only from texture/perspective, binocular disparity and convergence—we may have observed a greater effect of removing binocular information than would be expected when the normal range of cues is available. For example, information from motion parallax has been shown to contribute to the control of natural grasping (Marotta et al. 1998), but was unavailable in our experiment because head movements were prevented.

One potentially surprising aspect of our data was the finding that removing binocular information had no effect on peak velocities in either stimulus-set condition. This accords with previous findings (Jackson et al. 1997; Loftus et al. 2004; but see Bingham and Pagano 1998). In principle, however, one might expect increased uncertainty about object distance to lead to slower, more ‘cautious’ reaches (i.e. reduced peak velocity). The fact that we did not find this in our randomised condition suggests that the lack of effects in previous studies cannot be attributed to learning of distance information. It should be noted, however, that various other aspects of the stimuli were invariant across trials, possibly introducing further uncontrolled information. The stimuli were always presented along the midline, on the same table surface, and eye-height was fixed above the table. If these parameters were learned, variations in the height-in-scene of the objects could have been a very reliable cue to distance. In which case monocularly available cues to location may have been sufficiently reliable that overall precision was not affected significantly by removing binocular cues.

The effects we have reported indicate that learned stimulus information, acquired during the experiment, contributes to grasp planning. This finding is consistent with the more general idea that grasping is not planned solely on the basis of visual ‘measurements’ made at a given instant, but also on the basis of previously learned information about objects. Previous studies have examined this issue directly by examining the effects on grasping

movements of presenting conflicting information from familiar size and other ‘standard’ depth cues, under binocular and monocular viewing. In a study by Marotta and Goodale (2001), participants first completed a training phase in which they grasped repeatedly for plain, self-illuminated spheres of a constant size. This continued in the experimental conditions, but they were also presented with smaller or larger ‘probe’ spheres on a small number of trials. The conflicting familiar size information led to an increased number of online corrections, indicating reaches to the wrong location, under monocular viewing but not under binocular viewing. More recently, McIntosh and Lashley (2008) ran a similar ‘cue-conflict’ experiment, but using common household objects (matchboxes), which participants were already familiar with. They used actual-size, as well as scaled-up and scaled-down replica matchboxes, to manipulate the conflict between familiar size and the other available depth cues. Conflicting familiar size information resulted in biases in reach magnitude, and grip apertures, in the predicted direction. For example, participants over-reached to the scaled-down matchbox, suggesting that the smaller-than-normal retinal image caused an overestimate of object distance. These effects were present in both binocular and monocular conditions, but were larger when binocular cues were removed. These findings have important implications for understanding grasp control in real-world settings, because we very often grasp objects that are highly familiar to us (Melmoth and Grant 2006; McIntosh and Lashley 2008).

The results of this study are consistent with the theoretical idea (formalised within the reliability-based cue-weighting framework), as well as a growing body of literature, suggesting that the visuo-motor system acts as a near-optimal ‘integrator’, combining all sources of information available to it (Knill 2005; Greenwald and Knill 2008). According to this view, the visuo-motor system does not ‘switch’ strategies depending on which cues are available, but rather weights *all* the available information according to the relative reliabilities of the signals involved. The familiar size studies described above are also consistent with this idea. Reliability-based cue-weighting predicts that familiar size will have a greater influence when binocular information is unavailable (see “[Introduction](#)”). Furthermore, it predicts that familiar size will have an increasing effect the more reliable it is. It seems very likely that the learned representation of size was much more reliable in McIntosh and Lashley’s (2008) study, than in Marotta and Goodale’s (2001). In the former case people interacted with everyday, functional objects they had commonly encountered for many years, whereas in the latter the objects were arbitrary shapes, and participants were only exposed to them during the experiment. Thus the model provides a good account of why McIntosh and

Lashley (2008) found an effect of familiar size even when binocular cues were available, whereas Marotta and Goodale (2001) did not.

We have found that the effect of removing binocular information on grip aperture scaling is significantly larger when grasping to a randomised stimulus set than to a typical stimulus set comprised of a small number of fixed object sizes and distances. We suggest therefore that participants learn small stimulus sets, and that this introduces an additional, uncontrolled source of information, leading to misestimates of the effect of removing binocular depth cues. We have concentrated here on binocular vision but this argument applies, in principle, to manipulations of any source of depth information. Studies of visually guided grasping have almost all recognised the need to use several object/distance configurations to prevent ‘stereotyped’ movements. Our findings suggest, however, that even using three objects and three distances—a higher-than-typical number of stimulus configurations—is insufficient to prevent learning of the stimulus set, with resulting effects on grasp kinematics. Unfortunately, it seems likely that effects of stimulus learning are not limited to the manipulation used here, but may also be present in the data from a variety of studies of visuo-motor control. Fortunately, preventing them is relatively straightforward.

Acknowledgments Supported by an Economic and Social Research Council PhD studentship to Bruce Keefe, and by the Engineering and Physical Sciences Research Council. Thanks to Matthew Elsby for help with data collection. Thanks to Paul Hibbard, Kevin MacKenzie, and two anonymous reviewers for helpful comments on the manuscript, and to Llewelyn Morris for technical support. Part of this work was presented at the Vision Sciences Society annual meeting, in May 2008.

References

- Bingham GP, Pagano CC (1998) The necessity of a perception action approach to definite distance perception: monocular distance perception for reaching. *J Exp Psychol Hum Percept Perform* 24:145–168
- de Berg M, van Kreveld M, Overmars M, Schwarzkopf O (2000) *Computational geometry: algorithms and applications*, 2nd edn. Springer, New York
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433
- Greenwald HS, Knill DC (2008) A comparison of visuomotor cue integration strategies for object placement and prehension. *Vis Neurosci* (published online, Cambridge University Press. doi: [10.1017/S0952523808080668](https://doi.org/10.1017/S0952523808080668))
- Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. *Nature* 394:780–784
- Hillis JM, Watt SJ, Landy MS, Banks MS (2004) Slant from texture and disparity cues: optimal cue combination. *J Vis* 4:967–992
- Holway AH, Boring EG (1941) Determinants of apparent visual size with distance variant. *Am J Psychol* 54:21–37
- 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 Cogn* 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 (1984) The timing of natural prehension movements. *J Mot Behav* 16:235–254
- Jeannerod M (1988) *The neural and behavioural organization of goal-directed movements*. Clarendon Press, Oxford
- Knill DC (2005) Reaching for visual cues to depth: the brain combines depth cues differently for motor control and perception. *J Vis* 5:103–115
- Knill DC, Saunders JA (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vis Res* 43:2539–2558
- Künnapas T (1968) Distance perception as a function of available visual cues. *J Exp Psychol* 77:523–529
- Loftus A, Servos P, Goodale MA, Mendarozqueta N, Mon-Williams M (2004) When two eyes are better than one in prehension: monocular viewing and end-point variance. *Exp Brain Res* 158:317–327
- Marotta JJ, Goodale MA (2001) The role of familiar size in the control of grasping. *J Cogn Neurosci* 13:8–17
- Marotta JJ, Kruyer A, Goodale MA (1998) The role of head movements in the control of manual prehension. *Exp Brain Res* 120:134–138
- McIntosh R, Lashley G (2008) Matching boxes: familiar size influences action programming. *Neuropsychologia* 46:2441–2444
- Melmoth DR, Grant S (2006) Advantages of binocular vision for the control of reaching and grasping. *Exp Brain Res* 17:371–388
- O’Leary A, Wallach H (1980) Familiar size and linear perspective as distance cues in stereoscopic depth constancy. *Percept Psychophys* 27:131–135
- Oruç I, Maloney LT, Landy (2003) Weighted linear cue combination with possibly correlated error. *Vis Res* 43:2451–2468
- Schlicht EJ, Schrater PR (2007) Effects of visual uncertainty on grasping movements. *Exp Brain Res* 182:47–57
- Sedgwick HA (1986) Space perception. In: Boff KR, Kaufman L, Thomas JP (eds) *Handbook of perception and human performance. Sensory processes and perception*, vol 1. Wiley, New York, pp 1–57
- Servos P, Goodale MA, Jakobson LS (1992) The role of binocular vision in prehension: a kinematic analysis. *Vis Res* 3:1513–1521
- Watt SJ, Bradshaw MF (2000) Binocular cues are important in controlling the grasp but not the reach in natural prehension movements. *Neuropsychologia* 38:1473–1481
- Watt SJ, Keefe BD, Hibbard PB (2008) Visual uncertainty predicts grasping when monocular cues are removed but not when binocular cues are removed. *J Vis* 8:297
- Wing AM, Turton A, Fraser C (1986) Grasp size and accuracy of approach in reaching. *J Mot Behav* 18:245–260