STEREO DISPARITY FACILITATES VIEW GENERALISATION DURING SHAPE RECOGNITION FOR SOLID MULTI-PART OBJECTS

Filipe Cristino[1], Lina Davitt[1],

William G. Hayward[2] & E. Charles Leek[3]

[1] School of Psychology, Bangor University, Bangor, UK

[2] School of Psychology, University of Auckland, Auckland, New Zealand

[3] Wolfson Centre for Clinical and Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, UK

Running Head: STEREO-DISPARITY AND VIEW GENERALIZATION

Keywords: Object Recognition; View generalisation; Stereo disparity

Corresponding Author: Dr Filipe Cristino

School of Psychology, Brigantia Building, Bangor University, Bangor, Wales, UK LL57 2AS

Tel: (44) 1248 383478

ABSTRACT

Current theories of object recognition in human vision make different predictions about whether the recognition of complex, multi-part, objects should be influenced by shape information about surface depth orientation and curvature derived from stereo disparity. We examined this issue in five experiments using a recognition memory paradigm in which observers (N=134) memorised and then discriminated sets of 3D novel objects at trained and untrained viewpoints under either mono- or stereo viewing conditions. In order to explore the conditions under which stereo-defined shape information contributes to object recognition we systematically varied the difficulty of view generalisation by increasing the angular disparity between trained and untrained views. In one series of experiments objects were presented from either previously trained views or untrained views rotated (15˚, 30˚ or 60˚) along the same plane. In separate experiments we examined whether view generalisation effects interacted with the vertical or horizontal plane of object rotation across 40˚ viewpoint changes. The results showed robust viewpoint dependent performance costs: observers were more efficient in recognizing learned objects from trained relative to untrained views, and recognition was worse for extrapolated relative to interpolated untrained views. We also found that performance was enhanced by stereo viewing but only at larger angular disparities between trained and untrained views. These findings show that object recognition is not based solely on 2D image information but that it can be facilitated by shape information derived from stereo disparity.

Word count: 234 words

The human visual system is remarkably adept at recognising three-dimensional (3D) objects across changes in viewpoint, although there is considerable evidence that recognition is not entirely viewpoint invariant (e.g., Arguin & Leek, 2003; Bülthoff & Edelman, 1992; Harris, Dux, Benito, & Leek, 2008; Hayward, 2003; Leek, Atherton, & Thierry, 2007; Tarr & Pinker, 1990). Other work has shown that viewpoint dependency in recognition is influenced by a variety of factors including, for example, shape discriminability (e.g., Hayward & Williams, 2000), complexity (e.g., Bethell-Fox & Shepard, 1988), object geometry (e.g., Leek & Johnston, 2006; Tarr & Pinker, 1990), familiarity (e.g., Leek, 1998a & 1998b) and the level of image classification required by the task (e.g., Hamm & McMullen, 1998).

Of interest in the current study is the potential contribution to view generalisation of shape information derived from stereo disparity. Stereo disparity potentially provides useful information about shape that may facilitate view generalisation and recognition, including for example, surface orientation, surface curvature gradients and polarity, as well as 3D aspect ratio (Norman, Todd, & Phillips, 1995; Norman, Swindle, Jennings, Mullins, & Beers, 2009; Welchman, Deubelius, Conrad, Bülthoff, & Kourtzi, 2005; Wismeijer, Erkelens, Ee, & Wexler, 2010). However, it has been proposed that stereo disparity plays little, if any, role in the perception and recognition of object shape (Li, Pizlo, & Steinman, 2009; Li & Pizlo, 2011; Pizlo, 2010). For example, Pizlo and colleagues have suggested that 3D shape can be recovered from mono-ocular two-dimensional (2D) perceptual input alone when image reconstruction follows geometric simplicity constraints (e.g. Li, Pizlo, & Steinman, 2009; Li & Pizlo, 2011; Pizlo, 2010). Similarly, current appearance- or image-based models of recognition have proposed that view generalisation is accomplished via interpolation among viewpoint-specific templates that comprise solely 2D image features (e.g. Bülthoff & Edelman, 1992; Reisenhuber & Poggio, 1999; Serre, Oliva and Poggio, 2007). In contrast, a key assumption of structural description models is that recognition involves the segmentation of complex objects into constituent 3D parts or shape primitives (e.g., Biederman 1987; Hummel & Stankiewicz, 1996; Marr & Nishihara, 1978; Leek, Reppa & Arguin, 2005; Leek, Reppa, Rodriguez & Arguin, 2009). Part segmentation has been assumed to be supported by the recovery of local surface depth orientation and concave curvature minima which arise at part boundaries (Biederman, 1987; Hoffman & Richards, 1984; Marr & Nishihara, 1978) – a hypothesis that is supported by a growing body of evidence highlighting the special functional status of concave curvature minima in shape perception (e.g., Cohen, Barenholtz, Singh & Feldman, 2005; Davitt, Cristino, Wong & Leek, 2014; Hoffman & Richards, 1984; Leek, Cristino, Conlan, Patterson, Rodriguez & Johnston, 2012; Lim & Leek., 2012). Thus, these contrasting theoretical models of shape representation make different predictions about the use of certain kinds of shape information from stereo disparity. Notably, structural description models, unlike 2D image-based models, predict that the recovery of information about local surface depth orientation and concave curvature minima from stereo input should facilitate the derivation of a 3D object model to support view generalisation during recognition[1].

Current studies investigating stereo effects on view generalisation and recognition provide an incomplete picture (Bennett & Vuong, 2006; Burke, Taubert, & Higman, 2007; Burke, 2005; Chan, Stevenson, Li, & Pizlo, 2006; Edelman & Bülthoff, 1990; Liu, Ward, & Young, 2006; Humphrey & Khan, 1992; Lee & Saunders, 2011; Pasqualotto & Hayward, 2009; Rock & DiVita, 1987). For example, Burke (2005) presented stereo images of deformed

[1] We do not claim that stereo disparity provides the only source of shape information about local surface orientation and curvature. Indeed, cues such as shadow and shading gradients, texture and non-accidental properties of edge features that are computable from mono input could also be used to infer surface properties. However, surface depth orientation and part boundaries at surface intersections are enhanced under conditions of stereo viewing due to binocular disparity.

paperclips in a sequential matching task and reported reduced viewpoint costs when stereo information about object structure was available (see also Bennett & Vuong, 2006; Lee & Saunders, 2011). Similar results were earlier found by Edelman & Bülthoff (1992) also using 'bent paperclip' and amoeba stimuli (see also Burke, 2005; Chan et al., 2006; Lee & Saunders, 2011; Rock & DeVita, 1987). However, a limitation of studies using these types of stimuli is that the results are not readily generalizable as the stimuli cannot be decomposed into parts. Thus, they do not provide the appropriate conditions to test predictions about the potential role of stereo-defined cues to part segmentation in complex 3D objects. However, there are two reports of studies contrasting mono- vs. stereo viewing of multi-part object stimuli. Humphrey and Khan (1992, Exp 3) contrasted effects of viewpoint change between learning and test views for novel 3D clay objects under conditions of mono- and stereo-viewing (manipulated by viewing the actual 3D models in either mono- or binocular viewing through a shutter). Training and test views differed by 40 or 80 degrees through rotation about a vertical axis perpendicular to the line of sight. Unfortunately, the results were equivocal. Overall they showed a clear advantage for the recognition of objects at trained views – indicating viewpoint-dependent performance. However, the effects of presentation mode (stereo, mono) were unclear. The results showed an advantage for mono-ocular viewing in response times, but the opposite pattern for accuracy. On the one hand, this could reflect a difference in the way in which shape information from stereo input is processed – where we might expect slower (but more accurate) processing due to the requirement for resolving stereo correspondence. Alternatively, as noted by Humphrey and Khan (1992, p. 186/187), it could simply reflect a strategic bias by observers to favour accuracy over speed with the stereo displays (i.e., a speed-accuracy trade-off). Unfortunately, it is not possible to distinguish between these two possibilities from the results presented by Humphreys and Khan (1992).

In a more recent study, Pasqualotto and Hayward (2009) contrasted mono versus stereo viewing of 3D computer-generated models of real-world objects in a sequential matching (rather than recognition) task. They found that when observers were required to match objects across large (180 degree) viewpoint differences stereo viewing produced a performance cost (in both RTs and accuracy). The cost was attributed to a mismatch at large angular disparities between the information computed from mono and stereo-input. This finding appears to challenge the prediction of parts-based models that stereo-defined shape information should facilitate view generalisation and object recognition. However, such a conclusion would be premature. In the first place the task involved sequential matching rather than recognition, which requires indexing a long-term memory representation of object shape. In addition, the absence of a stereo advantage for view generalisation may have arisen because observers were able to efficiently match familiar features of the objects across views from 2D input alone – particularly given that the stimuli were highly common ‘real-world’ objects. Indeed, if stereo-disparity can potentially facilitate recognition by making explicit 3D image features (e.g., such as part-boundaries at regions of concave surface curvature minima), we might expect to only observe stereo effects when mono-ocular cues to shape alone are insufficient, or ambiguous. That is, a stereo advantage in view generalisation is more likely when shape equivalence cannot be determined on the basis of 2D image similarity. A key factor determining 2D image similarity is viewpoint disparity: 2D image similarity for any given object will be larger for small angular differences in viewpoint. At the same time, the earlier work by Bülthoff and Edelman (1992) showed that effects of viewpoint disparity on recognition efficiency are non-linear – in part because of familiarity: we are more familiar with some views of known objects than others. Notably, viewpoint costs in shape matching are not equivalent for horizontal and vertical axis rotations in depth. Thus, potential effects of stereo disparity on view generalisation must also take into account the axis of depth rotation.

Our goal in this paper was to address these limitations of previous studies in order to examine whether shape information derived from stereo disparity facilitates view generalisation and object recognition as predicted by parts-based structural description accounts. To control prior experience and familiarity with specific views our studies used a set of CAD generated surface rendered, solid, multi-part, novel objects. The complexity and part structure of the object set was carefully controlled. In addition, we used a recognition memory task that allowed us to examine how the availability of stereo information influences performance when observers must match a perceptual description of object shape to a long-term memory representation at increasing disparities between trained and test views. This also allowed us to contrast performance between horizontal and vertical axis rotations in depth.

In a series of five experiments, observers were asked to memorise a subset of four-part novel objects from either mono or stereo visual displays. In a subsequent test phase they discriminated targets (previously learned objects) from visually similar distracters at either trained or novel viewpoints in either mono or stereo. In Experiment 1, we replicated previous studies by contrasting performance between mono or stereo displays across trained and untrained interpolated viewpoints around the horizontal plane. In Experiment 2a and 2b we examined whether stereo viewing effects were modulated by task difficulty. In these two experiments, trained and interpolated viewpoints were on the same axis while extrapolated views were on the orthogonal axis. In Experiment 2a, the trained viewpoints were in the horizontal axis while in Experiment 2b the trained viewpoints were on the vertical axis. In Experiments 3a and 3b we examined the boundary conditions for stereo viewing effects by reducing the angular disparity between trained and untrained views.

EXPERIMENT 1

METHODS

Participants

Participants were 30 students (22 females age *M*= 20.73, *SD*= 1.96, one left handed) from Bangor University who took part in exchange for course and printer credits. All participants reported normal or corrected to normal vision and normal colour vision. Prior to the experiment participants were tested for normal stereoscopic vision. During this test, the participants were presented with a stereo anaglyph image and were asked to describe to the experimenter what they were seeing. The protocol was approved by the local Research Ethics and Governance Committee. Participation was voluntary and prior informed consent was obtained. All the participants tested were included in the analysis.

Stimuli

The stimulus set comprised 24 novel objects (see Figure 1), each consisting of a unique spatial configuration of four volumetric parts (see Leek et al., 2012).

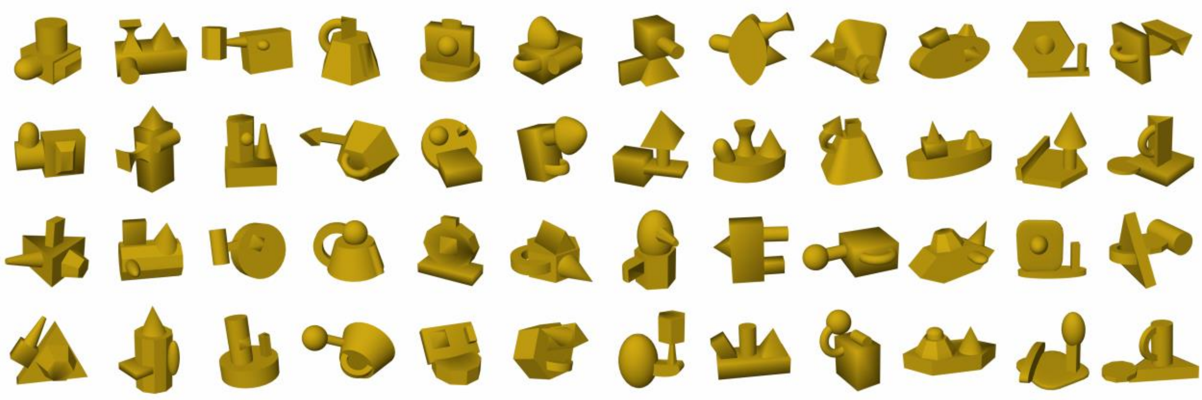

Figure 1: Stimuli used throughout the experiments. The top row was used as targets while the other three rows were used as distractors. Please visit the online version of this Journal to view this figure in colour.

The parts were uniquely defined by variation among non-accidental properties (NAPs) comprising: Edges (Straight vs. Curved), symmetry of the cross section, tapering (co-linearity) and aspect ratio (Biederman, 1987). For each stimulus a baseline 'canonical' three-quarter view was determined. The models were designed in Strata 3D CX software (Strata Inc, USA) then imported and rendered in Matlab (Mathworks Inc, USA) using custom code. Objects were rendered with a smooth surface and no texture cues in a yellow mustard colour (RGB: 227, 190, 43) using a stereoscopic camera with an Inter-Pupillary Distance (IPD) of 67mm. Each rendered image pair were converted to a red-cyan anaglyph. Stimuli were scaled to have the same rendered size for the canonical viewpoint (X=0˚, Y=0˚, Z=0˚) sustaining on screen an average size of 15˚ by 15˚ and were displayed on a white background.

Apparatus and Materials

Stimulus presentation and data collection were performed with *e-Prime v1.2* (Psychology Software Tools, USA) software. Stimuli were presented on a CRT 22 inch monitor with a screen resolution of 1024 x 768 pixels at 100 Hz refresh rate. In each experiment a chin rest was used to stabilise the participant's head at a 60 cm viewing distance and a standard USB keyboard was used for response collection.

Design and procedure

We used a 2 (Display: mono, stereo) x 2 (Viewpoint: trained, interpolated) mixed design with Display as the between-subjects factor. The study was divided in three phases; learning, verification and test. In the learning phase (see Figure 2) the participants were instructed to memorise the shapes of 12 out of 24 novel 3D objects presented on the computer screen. Each trial started with a fixation cross in the centre of the monitor for 500ms, followed by blank screen for 500ms Inter-Stimulus Interval (ISI) and then the name of the object in that trial (i.e. Object 1, Object 2 etc.) for 2s. This was followed by an ISI for 500ms and then the object was displayed from the first viewpoint for 3s, an ISI of 500ms, the same object from the second viewpoint for 3s, ISI (500ms), the same object from the third viewpoint (3s), and then a blank screen for 500ms before the next trial began. This trial order was sequentially repeated three times to ensure each object and its associated object number was well learned for the verification task.

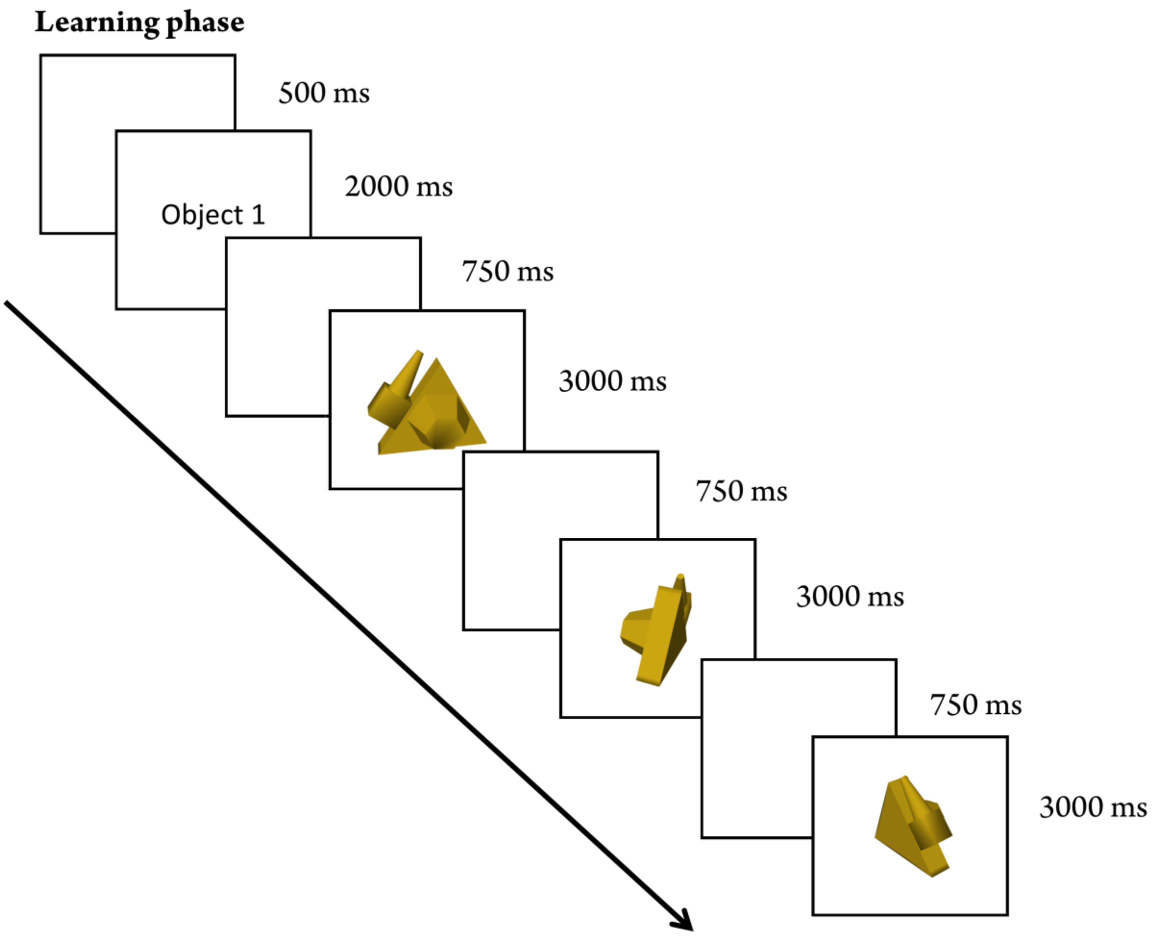

Figure 2: Main trial structure during the learning phase. Please visit the online version of this Journal to view this figure in colour.

Following the learning phase, participants completed a short verification task with feedback after each trial to ensure they had learned the target object set and could recognise them. Participants were shown a fixation cross in the centre of the monitor display for 500ms, followed by an ISI for 500ms and then one of the 12 learned objects from one of the three trained viewpoints. The task was to indicate as accurately as possible the object identification number using the function keys. Following the response accuracy feedback (i.e. correct/incorrect) was displayed for 1s. The verification task started with 72 trials (12 objects x 3 viewpoints x 2 repetitions) and participants had an accuracy criterion of 80% which they had to achieve in order to continue to test phase. If this criterion was not obtained the verification task was repeated to ensure participants would be able to correctly identify the target objects in the test phase. If participants did not reach an accuracy level of 80% after the verification stage they were asked to forgo the rest of the experiment.

The design of the test phase was similar to the verification task with the exceptions that no accuracy feedback was provided and participants had to decide whether the object displayed was one of the original trained objects (regardless of viewpoint changes) from the learning phase (target) or if it was previously unseen object (distractor). Responses were made using the 'K' for target and 'D' for distractors keys. During the experimental session participants wore either a pair of cyan/cyan glasses for the mono viewing condition or red/cyan glasses for the stereo viewing condition (for both learning and test phases). The testing phase consisted of 576 randomised trials (288 trained object trials: 12 objects * 6 viewpoints * 4 repetitions and 288 untrained object trials: 12 objects * 6 viewpoints * 4 repetitions) in a single block.

Analysis of Results

Trained and untrained object trials were analysed, signal detection theory sensitivity index's *d*' was used as the primary dependent measure (Brophy, 1986). Additionally analyses of RTs were also undertaken, but for ease of exposition are only reported where relevant. Response Times (RTs) were first trimmed, removing any trials with very fast (<180ms) or slow (>2500ms) responses and reported for the target trials only. Incorrect responses and RTs more than two standard deviations above and under the trimmed mean were classed as outliers and excluded from the reaction time analysis. The same criteria for pre-processing of the data were applied for all the reported studies.

RESULTS

Mean d' is shown in Figure 3.

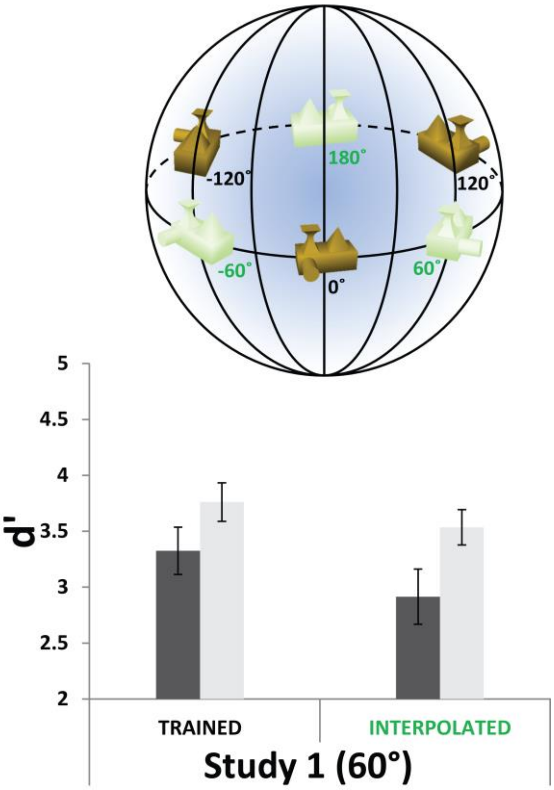

Figure 3: Experiment 1 schematic rotations and mean performance (d') as a function of display type (dark grey – 2D; light grey – 3D) and viewpoints (Trained X axis: -120º, 0º, 120º ;Interpolated X axis:

-60º, -180º, -60º). Error bars represent Standard Errors (SE). Please visit the online version of this Journal to view this figure in colour.

A 2 (Display: Mono, stereo) x 2 (Viewpoint: trained, interpolated) mixed ANOVA on the d' scores revealed a significant main effect of Viewpoint *F* (1, 28) =10.744, *p*= .003, ηp^2= 0.277, a marginally significant effect of Display *F* (1, 28) =3.956, *p*= .056, ηp^2= 0.124; but no interaction. Planned comparisons were performed using t-tests to contrast viewpoint effects for mono and stereo stimulus presentation. There was a significant difference for the Mono viewing condition between trained (*M*= 3.32, *SE*= 0.21) and interpolated views (*M*= 2.92, *SE*= 0.25; *t* (14) = 2.78, *p*= .015), but not for the Stereo display condition (trained *M*=3.76, *SE* = 0.17, interpolated views *M*=3.54, *SE*=0.16; *t*(14)=1.79, *p*=0.09) There was also a significant difference in the interpolation viewpoint condition between mono (*M*=2.92, *SE*= 0.25) and stereo viewing (*M*= 3.54, *SE*= 0.16) groups *t* (28) = -2.11, *p*= .004.

<u>Analyses of RTs</u>

The mean RTs (SE) across conditions are shown in Table 1. The same patterns of significant differences were found in the analyses of RTs as in the analyses of d'. A main viewpoint effect *F* (1, 28) = 6.059, *p*= .002, ηp^2 = 0.178 was found with no other main or interaction effects. Post-hoc analysis using a paired t-test showed again a difference in the Mono viewing condition where trained viewpoint RTs (*M*=952.03, *SE*= 57.53) where slower that interpolated (*M*=1005.53, *SE* = 58.74) viewpoints significantly *t* (14) = -2.64, *p*= .0019. This effect was not present during the stereo condition t(14)=-0.80, p=0.43.

Table 1: Experiment 1 mean Reaction Times (ms) and Standard Errors (ms)

		Trained viewpoints		**Interpolated Viewpoints**	
		2D	3D	2D	3D
Exp1	*RT M(ms)*	952	890	1006	906
	SE (ms)	58	26	59	28

Discussion

Experiment 1 showed that observers were more efficient at recognising objects from trained views over untrained interpolated views – replicating results from previous studies (Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Leek, 1998b; Tarr & Pinker, 1990; Wong & Hayward, 2005). In contrast, while there was only a weak main effect of the mono- vs. stereo display, planned comparisons showed higher sensitivity for stereo over mono presentation during view interpolation. This finding is consistent with the hypothesis that stereo-defined depth information can constrain view generalisation (that is, reduce uncertainty about shape equivalence) for solid, multi-part, objects. However, given the relatively weak effects of stereo vs. mono presentation (and the mixed results reported in previous studies), in Experiment 2 we examined whether stronger evidence for the use of depth cues derived from stereo disparity would be observed for harder discriminations between trained and untrained views. In order to increase difficulty we included out-of-plane rotations requiring extrapolation from trained views along an orthogonal axis as well as adding extra distractors. This is motivated by prior evidence showing an advantage for view generalisation based on within-plane interpolation between trained views over out-of-plane extrapolation from trained views (e.g., Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992).

EXPERIMENT 2

METHODS

Participants

50 students from Bangor University participated in this experiment in exchange of course or printer credits. None of the participants had taken part in Experiment 1. All participants were screened for normal stereo vision (see E1, above). Two participants were excluded from further analyses as their performance was at chance level when recognising objects from trained viewpoints. Experiments 2a and 2b included 24 participants each (2a: 16 females age $M = 20.08$, $SD = 3.73$, three left handed; 2b:18 females age $M = 22.53$, $SD = 7.55$). The protocol had received approval from the local Research Ethics and Governance Committee. Participation was voluntary, and all participants gave prior informed consent.

Stimuli apparatus and procedure

The stimulus set was the same as in Experiment 1. Each stimulus was rendered from a different viewpoint each differing by 40 degrees around the horizontal and vertical axis (see figure 4). The object set for Experiments 2a and 2b consisted of the same 12 target objects as in E1 but with 36 distracters. We used the same apparatus as in Experiment 1.

Design and Procedure

We used a 2 (Display: mono, stereo) x 3 (Viewpoint: trained, interpolated, extrapolated) mixed design with Display as a between-subjects factor. For Experiment 2a (see Figure 4a) the trained viewpoints (-80 ^{o}, 0^o; +80^o) were rotated on the horizontal axis. In the test phase the interpolated viewpoints were rotated ±40^o on the same horizontal axis falling within the trained viewpoints. Extrapolated viewpoints were rotated ±40^o outside of the trained viewpoints on the vertical axis from the +80^o location. For Experiment 2b (see Figure 4b) the trained viewpoints (-80 ^{o}, 0^o; +80^o) were rotated around the vertical axis. In the test phase the interpolated viewpoints were rotated ±40° on the same axis as the trained viewpoints (-40°; +40°). Extrapolated viewpoints were rotated orthogonally on the horizontal axis ±40°. We used the same paradigm as in experiment 1, where subjects had to learn 12 objects at 3 different viewpoints followed by a verification stage. The testing phase consisted of 504 randomized trials (252 trained object trials: 12 objects * 7 viewpoints * 3 repetitions and 288 untrained object trials: 36 objects * 7 viewpoints) in a single block.

RESULTS

Exp. 2a – Interpolated viewpoints along horizontal axis/Extrapolated vertical axis

Mean d' are shown in Figure 4a. A 2 (Display: mono, stereo) x 3 (Viewpoint: trained, interpolated, extrapolated) mixed ANOVA on the d' scores showed significant main effects of Viewpoint, F (2, 52) = 53.382, p< .0001, ηp^2= 0.672, and Display, F (1, 26) = 4.947, p= .0035, ηp^2 = 0.160. There was no significant interaction. Planned comparisons using Bonferroni corrected t-tests showed significant differences for the mono display condition (Trained: M = 3.24, SE= 0.29; Interpolated: M = 2.90, SE= 0.27; Extrapolated: M = 2.46, SE = 0.25) between all viewpoint comparisons (all ps < .002). The same pattern of effects was found in the stereo display condition (Trained: M= 3.94, SE= 0.24; Interpolated: M = 3.66, SE = 0.24; Extrapolated: M = 3.42, SE= 0.28) between all viewpoint comparisons (all ps < .002). Bonferroni corrected independent t-tests between stereo and mono viewing conditions showed a significant advantage for the stereo viewing group for both the interpolation t (26) = -2.113, p= .0044 and extrapolation conditions t (26) = -2.561, p = .0016.

Analyses of RTs

Mean RTs per condition are shown in Table 2. RT analyses showed a significant main effect of Viewpoint F (2, 52) = 30.340, p < .0001, ηp^2 = 0.539, but no other main effects, and no interaction. Bonferroni corrected post-hoc analyses using paired t-tests showed a significant difference in the mono viewing condition (Trained: M = 843.92, SE = 46.72; Interpolated: M = 893.46, SE = 55.58; Extrapolated: M = 924.77, SE = 53.72) where trained viewpoints were recognised significantly faster than interpolated t (13) = -3.587, p = .0009 and extrapolated views t (13) = -5.428, p< .0001. In the stereo condition (Trained: M = 953.22, SE = 46.76; Interpolated: M = 981.66, SE = 53.11; Extrapolated: M =1022.98, SE =54.51), RTs to extrapolated viewpoints were significantly slower than the trained t (13) = -4.648, p= .0001 and interpolated views t (13) = -5.262, p< .0001. No other contrasts were significant.

Table 2: Experiment 2a and 2b mean Reaction Times (ms) and Standard Errors (ms)

		Trained viewpoints		**Interpolated Viewpoints**		**Extrapolated Viewpoints**	
		2D	3D	2D	3D	2D	3D
Exp2a	*RT M(ms)*	844	953	893	982	925	1023
	SE (ms)	47	47	56	53	54	55
Exp2b	*RT M(ms)*	880	943	912	968	898	961
	SE (ms)	59	35	59	38	65	35

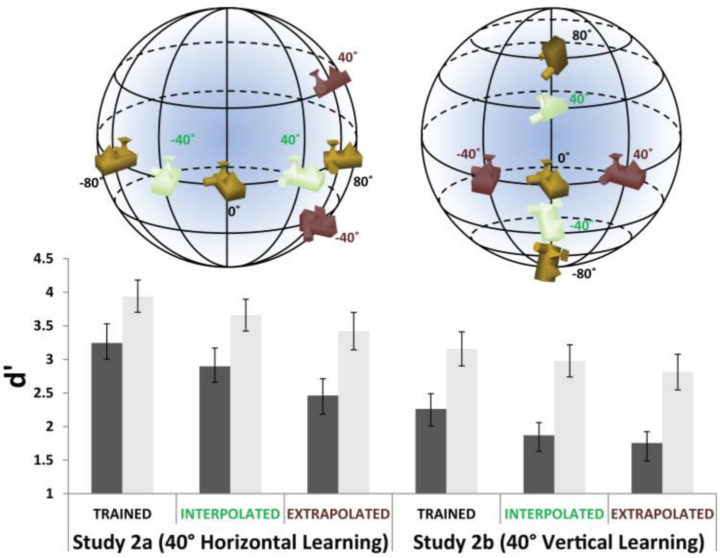

Figure 4: Experiment 2a (left) and 2b (right) schematic rotations and mean performance (d') as a function of display type (dark grey – 2D; light grey – 3D) and viewpoints (*Exp2a*: Trained X axis: -80º, 0º, 80º ;Interpolated X Axis: -40º, 40º; Extrapolated Y axis: -40º, 40º - *Exp2b*: Trained Y axis: -80º, 0º, 80º ;Interpolated Y Axis: -40º, 40º; Extrapolated X axis: -40º, 40º). Error bars represent Standard Errors (SE). Please visit the online version of this Journal to view this figure in colour.

Exp. 2b – Interpolated viewpoints along vertical axis/Extrapolated horizontal axis

Mean d' are shown in Figure 4b. A 2 (Display: mono, stereo) x 3 (Viewpoint: trained, interpolated, extrapolated) mixed ANOVA on the *d'* scores revealed a significant main effect of Viewpoint $F(2, 44) = 26.204$, $p < .0001$, $\eta p^2 = 0.544$. There was also a significant main effect of Display with participants significantly better at recognising objects shown in stereo than mono $F(1, 22) = 10.616$, $p = .0004$, $\eta p^2 = 0.325$. No interaction effects were found. Bonferroni corrected post-hoc analyses by t-test showed significant differences for the mono display condition (Trained: $M = 2.26$, $SE = 0.23$; Interpolated: $M = 1.87$, $SE = 0.19$; Extrapolated: $M = 1.75$, $SE = 0.17$) between the trained and interpolation viewpoints $t(11) = 4.488$, $p = .0003$, and between the trained and extrapolation viewpoints $t(11) = 6.496$, $p < .0001$. In the stereo display condition (Trained: $M = 3.16$, $SE = 0.25$; Interpolated: $M = 2.98$, $SE = 0.24$; Extrapolated: $M = 2.81$, $SE = 0.26$), there was a significant difference between the extrapolation and the trained viewpoints $t(11) = 3.601$, $p = .0012$.

Analysis of RTs

Mean RTs per condition are shown in Table 2. There was a significant main effect of viewpoint, $F(2, 44) = 4.479$, $p = .0017$, $\eta p^2 = 0.169$ but no other main effects, and no significant interaction. Post-hoc analyses comparing within and between factors showed no significant effects.

Discussion

As in Experiment 1 we found a highly significant and robust effect of viewpoint in both the analyses of sensitivity and RTs. Participants were more efficient at recognising trained views, and least efficient with extrapolated views on an orthogonal axis – for both vertical and horizontal out-of-plane rotations. Overall, participants showed higher sensitivity when learning objects on rotations around the horizontal axis. Of most relevance to the current study was the finding that in both experiments we found a significant benefit for stereo presentation in the analysis of sensitivity – which was equivalent for interpolated and extrapolated untrained views. This pattern of results suggests that depth information derived from stereo disparity is computed by the visual system, and facilitates view generalisation –particularly when generalisation among trained and untrained views is difficult. We further tested this hypothesis in Experiment 3 using relatively small angular disparities between trained and untrained interpolated and extrapolated views of 15° and 30 degrees° which were smaller than those used in either Experiment 1 (60°) or Experiments 2a and 2b (40°). We predicted that under these conditions view generalisation would not be influenced by the availability of stereo-defined depth cues.

EXPERIMENT 3

METHODS

Participants

In this last experiment we tested 60 participants, four participants were excluded from further analyses as their performance was at chance level when recognising objects from trained viewpoints. Experiments 3a and 3b included 32 participants (11 females age *M*= 20.42, *SD*= 2.41) and 24 participants (11 females age *M*= 24.43, *SD*= 5.33, one left handed) respectively. The protocol had received approval from the local Research Ethics and Governance Committee. Participation was voluntary, and all participants gave prior informed consent.

Stimuli apparatus and procedure

Stimuli, apparatus and procedure were the same as in Experiments 1 and 2. The angular disparity between trained and untrained viewpoints (see Figure 5) was 15˚ for Experiment 3a and 30˚ for Experiment 3b around horizontal (Y) axis.

Design and Procedure

Experiments 3a and 3b used the same 2 (Display: mono, stereo) x 3 (Viewpoint: trained, interpolated, extrapolated) mixed design as Experiment 2a and 2b. We uses the same stimuli than in the two previous experiment. In experiment 2a we used 12 target and 12 distractors while in experiment 3b 12 target and 36 distractors were used.

RESULTS

Exp. 3a (view generalisation across 15° disparity)

Mean d' are shown in Figure 5a. A 2 (Display: mono, stereo) x 3 (Viewpoint: trained, interpolated, extrapolated) mixed ANOVA showed a main significant effect of viewpoint, F (2, 44) = 3.998, p = .0025, $\eta p^2 = 0.154$, but no effect of Display, and no interaction. Bonferroni corrected post-hoc analyses by t-test showed no significant differences for the mono display condition (Trained: M = 4.08, SE = 0.27; Interpolated: M = 4.09, SE = 0.26; Extrapolated: M = 3.97, SE = 0.27). In the stereo display condition (Trained: M = 4.18, SE = 0.23; Interpolated: M = 4.27, SE = 0.19; Extrapolated: M = 3.99, SE = 0.22), there was a significant difference between the interpolation and the extrapolated viewpoints t (11) = 3.83, p = .008. Between subjects post-hoc analyses showed no significant effects.

Analyses of RTs

Mean RTs per condition are shown in Table 3. As with the analyses of d', RTs showed a significant effect of viewpoint, F (2, 44) = 10.455, p < .0001, $\eta p^2 = 0.322$ with no other main or interaction effects. Bonferroni corrected post-hoc analyses by t-test showed no significant differences for the mono display condition. In the stereo display condition there was a significant difference between the trained (M = 825.79, SE = 61.55) and the interpolated (M = 849.91, SE = 62.88) viewpoints t (11) = -4.08, p = .005. Between subjects post-hoc analyses showed no significant effects.

Table 3: Experiment 3a and 3b mean Reaction Times (ms) and Standard Errors (ms)

		Trained viewpoints		Interpolated Viewpoints		Extrapolated Viewpoints	
		2D	3D	2D	3D	2D	3D
Exp3a	*RT M(ms)*	970	826	1002	850	1055	893
	SE (ms)	43	62	49	63	53	68
Exp3b	*RT M(ms)*	750	783	773	786	780	826
	SE (ms)	21	32	20	32	19	34

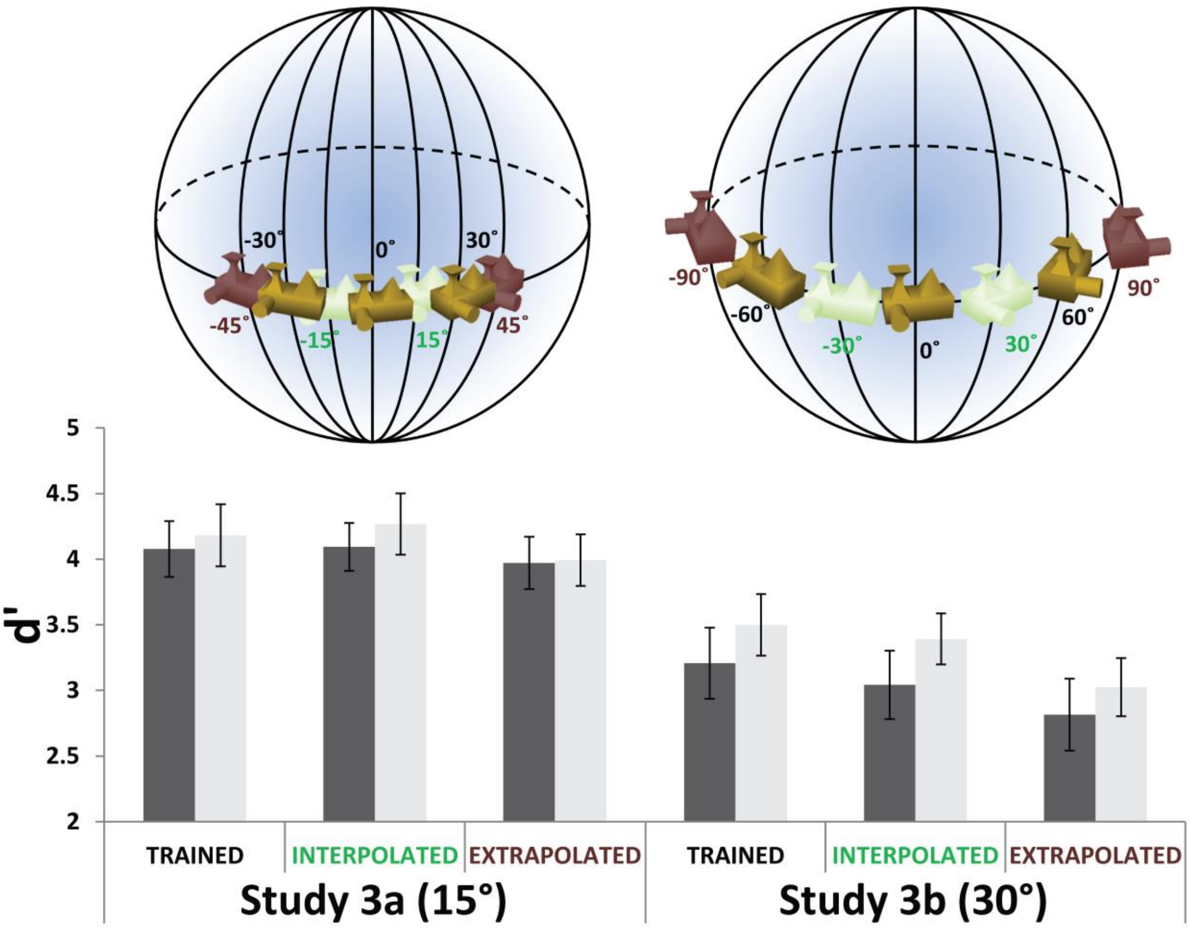

Figure 5: Experiment 3a (left) and 3b (right) schematic rotations and mean performance (d') as a function of display type (dark grey – 2D; light grey – 3D) and viewpoints (*Exp3a*: Trained X axis: -30º, 0º, 30º ;Interpolated X Axis: -15º, 15º; Extrapolated Y axis: -45º, 45º -- *Exp3b*: Trained X axis: -

60º, 0º, 60º ;Interpolated X Axis: -30º, 30º; Extrapolated X axis: -90º, 90º). Error bars represent Standard Errors (SE). Please visit the online version of this Journal to view this figure in colour.

Exp. 3b (view generalisation across 30˚ disparity)

Mean d’ are shown in Figure 5b. A 2 (Display: mono, stereo) x 3 (Viewpoint: trained, interpolated, extrapolated) mixed ANOVA showed a main significant effect of viewpoint, F (2, 60) = 27.198, p< .0001, ηp² = 0.476, but no effect of Display, and no interaction. No further analyses were conducted. Bonferroni corrected post-hoc analysis using a paired t-test showed a significant difference for the mono display condition (Trained: M = 3.21, SE = 0.21; Interpolated: M = 3.04, SE = 0.18; Extrapolated: M = 2.82, SE= 0.20) between the trained and extrapolated viewpoint t (15) = 3.700, p = .0006 as well as between the interpolated and extrapolated viewpoints t (15) = 2.804, p = 0.04. Similar effects were found in the stereo condition (Trained: M = 3.50, SE = 0.24; Interpolated: M = 3.39, SE = 0.23; Extrapolated: M = 3.02, SE = 0.20) where accuracy is worst for the extrapolated viewpoint than the interpolated t (15) = 4.800, p < .0001 and trained viewpoint t (15) = 4.896, p < .0001. Between subjects post-hoc analyses found no significant differences.

Analyses of RTs

Mean RTs per condition are shown in Table 3. Analyses of RTs showed a similar pattern with a main effect of viewpoint F (2, 60) = 22.393, p < .0001, ηp² = 0.427, but no effect of Display, and no interaction. Bonferroni corrected post-hoc analysis using a paired t-test showed a difference in the mono display condition (Trained: M = 749.88, SE = 20.73; Interpolated: M = 772.94, SE = 20.39; Extrapolated: M = 780.13, SE = 18.60) where trained viewpoint were recognised significantly faster than interpolated t (15) = -3.351, p = .0013 and extrapolated ones t (15) = -3.370, p= .0012. In the stereo display condition (Trained: M = 783.23, SE = 31.59;

Interpolated: *M*= 785.58, *SE* = 32.05; Extrapolated: *M* = 825.85, *SE* = 33.77), responses to the extrapolated viewpoints were significantly slower than the interpolated t (15) = -4.995, p < .0001 and trained viewpoints t (15) = -5.720, p < .0001. Between subjects post-hoc analyses found no significant differences.

Discussion

Experiments 3a and 3b showed that even at relatively small angular disparities between trained and untrained views of 15˚ and 30˚ we found significant time costs in view generalisation for both interpolated and extrapolated relative to trained views. In contrast, and unlike Experiments 1, 2a and 2b, we did not find an effect of display, showing that observers performed the task just as efficiently under mono and stereo viewing conditions. This is consistent with the hypothesis that depth cues to shape derived from stereo disparity are computed when available, but only facilitate recognition when the task is made more difficult through greater angular disparity between views. In order to examine this further we conducted a meta-analysis of the data across all five experiments by using the learnt viewpoints as the baseline.

Meta-Analysis: Generalisation effect modelling

For this analysis we wanted to capture the degree of change in accuracy between the trained and untrained viewpoints across all five experiments by performing a regression analysis on the difference scores. View interpolation was expressed as the *d'* percentage change between the trained and interpolated viewpoints, and view extrapolation as the *d'* percentage change between the trained and extrapolated viewpoints. A negative interpolation effect of, for example

-25%, would mean that a participant is 25% worse at performing the recognition task when viewing an object from an interpolated view than from a trained viewpoint. This approach allowed us to compare the data across experiments.

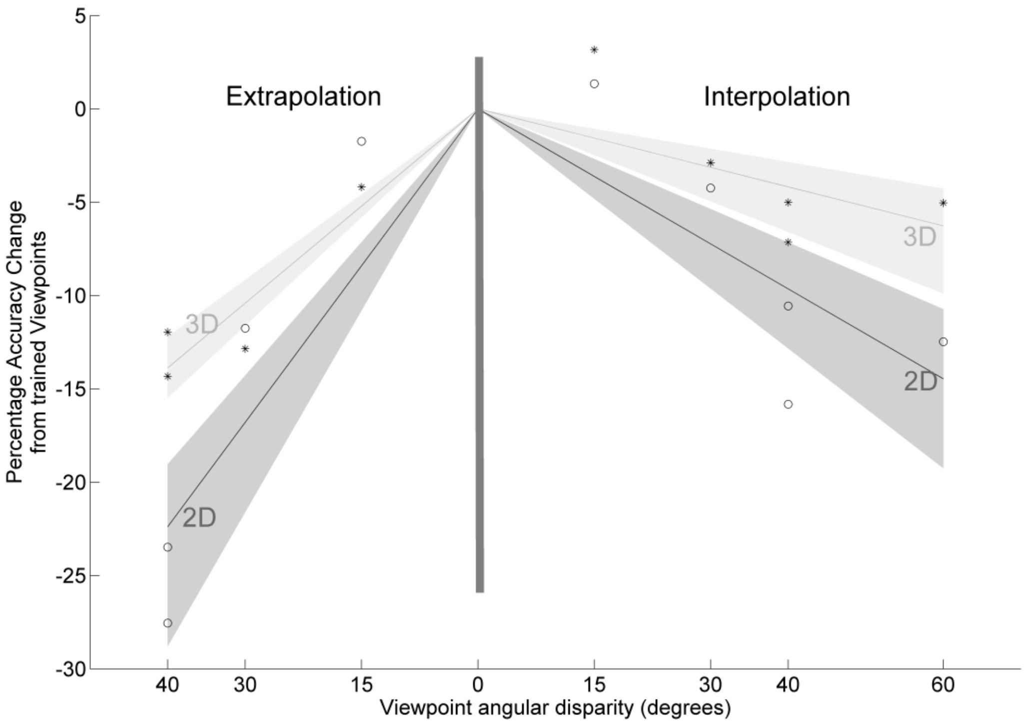

Figure 6: Meta-analysis across all 5 experiments. Circles indicate the 2D mean percentage accuracy change across participants either between the learnt and interpolated viewpoints on the right hand side or between the learnt and extrapolated viewpoints on the left hand side. Stars indicate mean percentage accuracy change in 3D. Light grey designates the 3D viewing condition; while dark grey indicate the 2D viewing condition. The shaded areas are the bootstrapped Confidence Intervals (CI) computed over 10000 permutations of the linear regression.

Figure 6 shows the mean effects per experiment for the interpolated (5 studies) and extrapolated (4 studies) viewpoints as well as the best least squared fitted linear model passing through 0˚ (as the learning viewpoints where taken as the baseline). The regressions were performed using all the 134 participants. The shaded regions are the bootstrapped Confidence Intervals (CI) computed over 10000 permutations for the simple linear regression (actual values reported in table 4).

Table 4: Meta-analysis regression parameters. The 95% confidence intervals were computed using a 10000 permutation bootstrapping procedure.

	Interpolation Viewpoints		**Extrapolation Viewpoints**	
	2D	3D	2D	3D
B	0.239	0.103	0.548	0.339
$\mathbf{R^2}$	0.639	0.415	0.760	0.852
***b* 95% CI**	[0.165 0.296]	[0.051 0.139]	[0.416 0.632]	[0.298 0.379]

The regression slope *b* is an indicator of the viewpoint effect where a value close to 0 would indicate no viewpoint effect while a value close to one would indicate a very large effect. This analysis revealed several important results: (1) Participants are significantly more accurate when performing a recognition task while viewing the stimuli stereographically. This is true for both interpolated and extrapolated viewpoints (CI do not overlap). (2) The advantage for stereo viewing increases with difficulty – being greater for generalisation across large angular disparities than smaller disparities between trained and untrained views. (3) Recognition from extrapolated views was significantly harder than from interpolated views (CI do not overlap).

GENERAL DISCUSSION

The main results of the study can be summarised as follows: (1) Experiment 1 showed an advantage in terms of both sensitivity and RTs for the recognition of solid multi-part objects at trained relative to interpolated views differing by 60˚ in angular disparity around the horizontal rotation plane. Observers also showed an advantage in performance for stereo over mono displays for RTs (p = .004; sensitivity, p = .056). (2) In Experiments 2a and 2b we increased task difficulty by including both interpolated and orthogonal extrapolated untrained views differing by 40˚ in angular disparity from the trained view along either the vertical or horizontal axis. The results showed effects of viewpoint in both sensitivity and RTs, with observers showing least efficient performance for extrapolated relative to interpolated views regardless of rotation axis. Both experiments showed an advantage for stereo over mono displays in terms of sensitivity (p = .003 and p = .0004 respectively). (3) In Experiments 3a and 3b we tested the boundary conditions for the stereo viewing advantage by reducing the difficulty of discriminations to 15˚ and 30˚ angular disparity between trained and untrained views. The results showed an advantage for trained over untrained views but no benefit for stereo viewing in either sensitivity or RTs. We conducted a correlation analysis to check for a speed/accuracy trade off across all the experiments. We found no significant correlation between sensitivity (d') and reaction times for the mono condition r(187)=0.1457, p= 0.079 and for the stereo condition r(190)= -0.0132, p=0.85 across all the experiments. A meta-analysis of the data across all five studies confirmed the view generalisation effect and showed that the advantage for stereo viewing is modulated by the degree of angular disparity between trained and untrained views reflecting the ease of shape discrimination.

These results provide new evidence about the effects of stereo disparity on view generalisation during object recognition, and about the conditions under which effects of stereo are found. Stereo viewing enhances specific kinds of shape information about local surface depth orientation and curvature minima arising at part boundaries (Norman et al., 1995; 2009; Welchman et al., 2005; Wismeijer et al., 2010). We predicted that under structural description approaches which assume high-level part segmentation the availability of this information under conditions of stereo viewing should facilitate recognition and view generalisation (e.g., Biederman 1987; Hoffman & Richards, 1984; Hummel & Stankiewicz, 1996; Marr & Nishihara, 1978; Leek et al., 2005; Leek et al., 2009). In contrast, according to image-based models such as HMAX (Reisenhuber & Poggio, 1999; Serre et al., 2007) recognition across views is achieved via interpolation among stored 2D aspects. These models do not attribute special functional significance to 3D shape properties such as surface concavities at part boundaries that are enhanced by stereo viewing. The results show that – under some conditions, stereo viewing can facilitate view generalisation during object recognition. This finding is consistent with the predictions of structural description models that place emphasis on the important functional role of surface depth and curvature information for the representation of 3D object shape. They challenge models that do not attribute functional significance to these kinds of image features (e.g., Bulthoff & Edelman, 1992; Chan et al., 2006; Li & Pizlo, 2011; Li, Pizlo & Steinman, 2009; Pizlo, 2008; Reisenhuber & Poggio, 1999; Serre et al., 2007).

However, a further important finding was that the advantage for view generalisation from stereo viewing was only apparent at larger angular disparities between trained and novel views, and that the effect did not interact with axis of rotation. This suggests that shape information derived from stereo disparity may only facilitate object recognition over larger changes in viewpoint when 2D cues to shape equivalence become less reliable. One possibility is that shape information derived from stereo viewing is only used to facilitate view generalization when it provides additional useful information (or is only used when object classification from monocular cues alone is ambiguous). Thus, there may be sufficient monocular image information to support view generalisation when the task is relatively easy to perform (e.g., across relatively small angular disparities between trained and untrained views – see Exp. 3) but that 3D shape information enhances performance when the task becomes more challenging – see Exps 1, 2a and 2b. More broadly, this can be understood in terms of optimal cue combination approaches to perception (e.g. Hillis, Watt, Landy, & Banks, 2004; Jacobs, 2002; Knill & Saunders, 2003), and is also consistent with other proposals that recognition may involve both image-based and structural description representations (e.g., Foster & Gilson, 2002; Hummel, 2013; Hummel & Stankiewicz, 1996).

Finally, our findings add to previous work supporting a role for stereo-defined shape information in the perception of other classes of stimuli (Bennett & Vuong, 2006; Chan, Stevenson, Li & Pizlo, 2006; Edelman & Bülthoff, 1992; Burke, 2005; Burke et al., 2007; Humphrey & Khan, 1992; Lee & Saunders, 2011; Liu et al., 2006; Pasqualotto & Hayward, 2009; Rock & DiVita, 1987). Interestingly, the observation of Pasqualotto & Hayward (2009) of a stereo disadvantage during view generalisation across large angular disparities suggests that, on some occasions, the integration of monocular and stereo-based cues to shape can, under some conditions, disrupt recognition. We did not find evidence for that in the current study. A key issue for future work will be to further elucidate the conditions under which stereo information is beneficial or detrimental to shape perception.

In summary, this study examined whether shape information derived from stereo disparity influences recognition across viewpoints for solid, multi-part, 3D objects. Each study used a recognition memory paradigm in which observers memorised and then discriminated sets of 3D novel objects at trained and untrained interpolated and extrapolated viewpoints under either mono- or stereo viewing conditions. Task difficulty was manipulated by varying the magnitude of angular disparity between trained and untrained views, and by rotated around the horizontal and vertical axes. The results across all five studies including 134 participants showed robust viewpoint dependent performance costs: observers were more efficient in recognizing learned objects from trained relative to untrained views. In addition, performance was worse for extrapolated relative to interpolated untrained views. We also found evidence that viewpoint effects were reduced during stereo presentation – but only under conditions of relatively more difficult discrimination involving greater angular disparity between trained and untrained views. This pattern of results was confirmed in a meta-analysis of data across all five experiments. These findings show that, consistent with structural description models, object recognition in human vision is not based solely on 2D image information but that it can be facilitated by shape information derived from stereo disparity.

ACKNOWLEDGEMENTS

This work was supported by an ESRC/EPSRC grant RES-062-23-2075 awarded to ECL. The authors wish to thank Candy Patterson and Pamela Arnold for they contribution to the study.

REFERENCES

Arguin, M., & Leek, E. C. (2003). Orientation invariance in visual object priming depends on prime-target asynchrony. *Perception & Psychophysics*, *65*(3), 469–477.

Bennett, D. J., & Vuong, Q. C. (2006). A stereo advantage in generalizing over changes in viewpoint on object recognition tasks. *Perception & Psychophysics*, *68*(7), 1082–1093.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 12–23. doi:10.1037/0096-1523.14.1.12

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*(2), 115.

Brophy, A. L. (1986). Alternatives to a table of criterion values in signal detection theory. *Behaviour Research Methods, Instruments, & Computers*, 18, 285-286.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences*, *89*(1), 60–64.

Burke, D. (2005). Combining disparate views of objects: Viewpoint costs are reduced by stereopsis. *Visual Cognition, 12*(5), 705–719. doi:10.1080/13506280444000463

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalizing across different views of faces. *Vision Research*, *47*(16), 2164–2169. doi:10.1016/j.visres.2007.04.018

Chan, M. W., Stevenson, A. K., Li, Y., & Pizlo, Z. (2006). Binocular shape constancy from novel views: the role of a priori constraints. *Perception & Psychophysics*, *68*(7), 1124–1139.

Cohen, E.H., Barenholtz, E., Singh, M. & Feldman, J. (2005). What change detection tells us about the visual representation of shape. *Journal of Vision, 5,* 313-321.

Davitt, L., Cristino, F., Wong, A. & Leek, E.C. (2014). Fixation preference for concave surface discontinuities during object recognition generalises across levels of stimulus classification. *Journal of Experimental Psychology: Human Perception & Performance*, *40*, 451-456.

Edelman, S., & Bülthoff, H. H. (1990). Viewpoint-Specific Representation in Three-Dimensional Object Recognition. Retrieved from http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.16.9627

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings. Biological Sciences / The Royal Society*, *269*(1503), 1939–1947. doi:10.1098/rspb.2002.2119

Gauthier, I, Hayward, W.G., Tarr, M.J., Anderson, A.W., Skudlarski, P. & Gore, J.C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron*, *34*, 161-171.

Hamm, J. P., & McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology. Human Perception and Performance*, *24*(2), 413–426.

Harris, I. M., Dux, P. E., Benito, C. T., & Leek, E. C. (2008). Orientation sensitivity at different stages of object processing: evidence from repetition priming and naming. *PloS One*, *3*(5), e2256. doi:10.1371/journal.pone.0002256

Hayward, W. G. (2003). After the viewpoint debate: where next in object recognition? *Trends in Cognitive Sciences*, *7*(10), 425–427.

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science*, *11*(1), 7–12.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: optimal cue combination. *Journal of Vision*, *4*(12), 967–992. doi:10:1167/4.12.1

Hong Liu, C., Ward, J., & Young, A. W. (2006). Transfer between two- and three-dimensional representations of faces. *Visual Cognition*, *13*, 51–64. doi:10.1080/13506280500143391

Hummel, J.E. (2013). Object recognition. In D. Reisburg (Ed.). *Oxford Handbook of Cognitive Psychology*, pp. 32-46. Oxford. Oxford University Press.

Hummel, J.E. & Stankiewicz, B.J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McClelland (Eds.). *Attention and Performance XVI: On information integration in perception and communication* (pp 93-121). Cambridge, MA. MIT Press.

Humphrey, G. K., & Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology*, *46*(2), 170–190.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*(8), 345–350.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558.

Lee, Y. L., & Saunders, J. A. (2011). Stereo improves 3D shape discrimination even when rich monocular shape cues are available. *Journal of Vision*, *11*(9). doi:10.1167/11.9.6

Leek, E. C. (1998a). Effects of stimulus orientation on the identification of common polyoriented objects. *Psychonomic Bulletin & Review*, *5*(4), 650–658. doi:10.3758/BF03208841

Leek, E. C. (1998b). The analysis of orientation-dependent time costs in visual recognition. *Perception*, *27*(7), 803–816.

Leek, E. C., Atherton, C. J., & Thierry, G. (2007). Computational mechanisms of object constancy for visual recognition revealed by event-related potentials. *Vision Research*, *47*(5), 706–713. doi:10.1016/j.visres.2006.10.021

Leek, E. C., Cristino, F., Conlan, L. I., Patterson, C., Rodriguez, E., & Johnston, S. J. (2012). Eye movement patterns during the recognition of three-dimensional objects: preferential fixation of concave surface curvature minima. *Journal of Vision*, *12*(1), 7. doi:10.1167/12.1.7

Leek, E. C., & Johnston, S. J. (2006). A polarity effect in misoriented object recognition: The role of polar features in the computation of orientation-invariant shape representations. *Visual Cognition*, *13*(5), 573–600. doi:10.1080/13506280544000048

Leek, E. C., Reppa, I., & Arguin, M. (2005). The Structure of Three-Dimensional Object Representations in Human Vision: Evidence From Whole-Part Matching. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(4), 668–684. doi:10.1037/0096-1523.31.4.668

Leek, E. C., Reppa, I., Rodriguez, E., & Arguin, M. (2009). Surface but not volumetric part structure mediates three-dimensional shape representation: Evidence from part–whole priming. *The Quarterly Journal of Experimental Psychology*, *62*(4), 814–830. doi:10.1080/17470210802303826

Li, Y., & Pizlo, Z. (2011). Depth Cues Versus the Simplicity Principle in 3D Shape Perception. *Topics in Cognitive Science*, *3*(4), 667–685. doi:10.1111/j.1756-8765.2011.01155.x

Li, Y., Pizlo, Z., & Steinman, R. M. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vision Research*, *49*(9), 979–991. doi:10.1016/j.visres.2008.05.013

Lim, I. S., & Leek, E. C. (2012). Curvature and the visual perception of shape: Theory on information along object boundaries and the minima rule revisited. *Psychological Review*, *119*(3), 668–677. doi:10.1037/a0025962

Marr, D., & Nishihara, H. K. (1978). Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294. doi:10.1098/rspb.1978.0020

Norman, J., Todd, J. T., & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception & Psychophysics*, *57*, 629–636. doi:10.3758/BF03213268

Norman, J. F., Swindle, J. M., Jennings, L. R., Mullins, E. M., & Beers, A. M. (2009). Stereoscopic shape discrimination is well preserved across changes in object size. *Acta Psychologica*, *131*(2), 129–135. doi:10.1016/j.actpsy.2009.03.009

Pasqualotto, A., & Hayward, W. G. (2009). A stereo disadvantage for recognizing rotated familiar objects. *Psychonomic Bulletin & Review*, *16*(5), 832–838. doi:10.3758/PBR.16.5.832

Pizlo, Z. (2010). 3D Shape. Its Unique Place in Visual Perception. *Literary and Linguistic Computing*, fqq025. doi:10.1093/llc/fqq025

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. doi:10.1038/14819

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, *19*(2), 280–293.

Serre, T., Oliva, A. & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization, *Proceedings of the National Academy of Sciences, 15*, 6424-6429.

Tarr, M. J., & Pinker, S. (1990). When Does Human Object Recognition Use a Viewer-Centered Reference Frame? *Psychological Science*, *1*(4), 253–256.

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience*, *8*(6), 820–827. doi:10.1038/nn1461

Wismeijer, D. A., Erkelens, C. J., Ee, R. van, & Wexler, M. (2010). Depth cue combination in spontaneous eye movements. *Journal of Vision*, *10*(6), 25. doi:10.1167/10.6.25