Perception, 2009, volume 38, pages 1628 – 1648

doi:10.1068/p6430

Modulation of viewpoint effects in object recognition by shape and motion cues

Quoc C Vuong
Institute of Neuroscience, University of Newcastle, Newcastle upon Tyne NE2 4HH, UK;
e-mail: q.c.vuong@ncl.ac.uk

Alinda Friedman, Courtney Plante
Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2M7, Canada

Received 6 April 2009, in revised form 3 June 2009; published online 28 October 2009

Abstract. In three experiments, we examined the role of structural similarity and different types of motion on the efficiency of performing same – different shape judgments across changes in viewpoints. In all experiments, participants judged whether two novel, multi-part objects were structurally identical, and they were to ignore any viewpoint or motion differences between the objects. In experiment 1, participants were affected by viewpoint differences more for structurally similar than structurally distinct objects, but this interaction was mitigated by rigid motion. In experiments 2 and 3, we used only structurally similar objects that moved only some of their parts, either in a similar way between objects within a pair or in distinctive ways. Participants' recognition performance was facilitated by this articulated motion relative to both static and scrambled controls. We conclude that coherent motion facilitates generalisation across different views of dynamic objects under some conditions.

1 Introduction

Despite the apparent ease with which observers recognise and interact with objects in a dynamic environment, there are measurable effects of changing viewing conditions (eg viewing distance, perspective, lighting) on the speed and/or accuracy of recognition performance. One important parameter that affects recognition is the perspective viewpoint from which an object is first (and then subsequently) encountered. For example, observers are sometimes slower or less accurate at recognising objects they have previously seen when they are subsequently encountered from an unfamiliar viewpoint (eg Tarr 1995). In addition, this recognition deficit tends to increase with the physical distance between the familiar viewpoint and the new viewpoint. The monotonic relation between performance and angular disparity between two viewpoints of the same object is often referred to as the 'viewpoint effect', and it has been used extensively to investigate the representations and mechanisms that underlie observers' robust — if not perfect — ability to recognise objects under unfamiliar viewing conditions (see Peissig and Tarr 2007, for a review).

Until recently, researchers have typically focused on how observers recognise static three-dimensional (3-D) objects from new viewpoints (eg Bülthoff and Edelman 1992; Hayward and Williams 2000; Khan and Humphrey 1992; Lawson and Humphreys 1996; Tarr 1995; Tarr et al 1998). The projected two-dimensional (2-D) shape of the same 3-D object (eg onto the 2-D retinal array) can be drastically different if the viewpoint changes, thereby making it computationally difficult to simply match, even at a perceptual level, the projected 2-D shapes of 3-D objects. It has been well established that factors that contribute to the recognition of a 3-D object, such as its 3-D geometric structure and colour, can also modulate viewpoint effects (Biederman and Gerhardstein 1993). Lawson et al (2003), for example, have shown that viewpoint changes affected picture – picture matching of static novel objects with similar 3-D structures, but did not affect the efficiency of matching objects with dissimilar structures (see also Lawson and Bülthoff 2006, 2008). Hayward and Williams further showed that the availability of other static visual features, such as colour, can dramatically decrease viewpoint effects even for structurally similar objects.

In contrast to shape, the manner in which motion information interacts with viewpoint has not been systematically investigated. Given that many objects in our environment can move and are often encountered from different perspective viewpoints, our main goal in the present study was to examine whether and how motion influences the recognition of objects encountered from disparate viewpoints. The evidence to date suggests that motion plays an important role in recognition (eg Knappmeyer et al 2003; Lander and Bruce 2000; Liu and Cooper 2003; Newell et al 2004; O'Toole et al 2002; Stone 1998, 1999; Vuong and Tarr 2004, 2006; Watson et al 2005). For example, some types of motion (eg rigid rotation in depth) may help observers estimate an object's 3-D structure which, in principle, is viewpoint-invariant (Todd 2004; Ullman 1979). Thus, rigid rotation in depth may diminish the recognition deficit caused by presenting objects at larger viewpoint disparities because it facilitates this estimation. Alternatively, motion may bias certain viewpoints by allowing observers to predict upcoming views (eg Friedman et al 2009; Stone 1998, 1999; Vuong and Tarr 2004) or by reducing observers' attention to certain views (eg Harman and Humphrey 1999). This view bias may potentially modulate viewpoint effects because some views may be more efficiently processed than others.

Our recent work further suggested that coherent rigid rotation of objects in depth, in which changes occur in a temporally smooth fashion, facilitated mechanisms that allowed both human and pigeon observers to generalise from learned viewpoints to new viewpoints (Friedman et al 2009). In that study, we trained observers to recognise a moving target object from two different training viewpoints. We then tested how quickly and accurately they recognised the target from novel viewpoints that were either between the two training viewpoints (interpolated views) or beyond the range of the two training viewpoints (extrapolated views; see also Bülthoff and Edelman 1992). Importantly, the key role of motion was underscored by the finding that extrapolated views that followed the learned rotation direction of the objects during training were recognised more accurately than extrapolated views that preceded the training views, although both extrapolated views were the same angular distance to a trained view.

An important difference between the present study and our previous work (Friedman et al 2009) is that the effects of coherent motion on recognising objects from new viewpoints found previously were presumably based on representations of 3-D objects that were acquired during training and stored in long-term visual memory. Here, in contrast, we examined the role of shape and motion in modulating the viewpoint effect during immediate perception. Across three experiments, observers made identity judgments based on the 3-D structures of object pairs presented simultaneously at different viewpoints. In experiment 1, we manipulated the structural similarity of the object pairs. In experiments 2 and 3, we used only pairs with similar 3-D structure. To investigate motion cues, the objects underwent either rigid rotation in depth (experiment 1) or a type of motion called 'articulation' (experiments 2 and 3), in which only parts of the objects move rigidly (see Aggarwal et al 1998, for a formal distinction between these qualitatively different types of motion). This motion is similar to familiar human movements, in which body parts move rigidly at the joints (eg bending the knee). We also varied the similarity between the articulated motions in experiments 2 and 3 by morphing the motions in a manner similar to the way we morphed shapes to vary their perceived structural similarity (Schultz et al 2008). Finally, in experiment 3, we scrambled the temporal order of the articulation sequence to disrupt coherent motion information while preserving shape information (eg the total number of views seen on a given trial—Vuong and Tarr 2004).

As discussed above, motion can both reveal more information about objects and/or bias specific views of objects (eg O'Toole et al 2002). Thus, we predict that motion conditions should be responded to more efficiently than static conditions in experiments 1 and 2. Furthermore, our previous studies suggested that temporally smooth motion can facilitate recognition (Friedman et al 2009; Vuong and Tarr 2004). Thus, we also predict that the coherent motion condition should be responded to more efficiently than the scrambled condition in experiment 3. What has not been investigated in previous research is (i) the role of motion in mitigating the effect of structural similarity and viewpoint changes, and (ii) what happens to structure-based recognition performance when objects undergo similar motions. It might be expected that, in experiment 1, because rigid motion can facilitate recognition from long-term memory, this type of motion could also mitigate the effects of structural similarity and viewpoint when objects are available to perception. It might also be expected that, when structurally similar objects undergo similar motions, the overall effect could be to make the objects seem even more similar to each other, and thus more difficult to differentiate, especially at larger angular disparities. This expectation predicts that in experiments 2 and 3 there should be a larger viewpoint effect for motion on trials when the motions are similar. However, motion can also draw attention to particular parts or features; on this account we would expect that motion may generally improve performance relative to static images (eg Harman and Humphrey 1999). Furthermore, although motion similarity might affect recognition adversely, the coherence and predictability of coherent motion may help reduce viewpoint effect under some conditions (eg Friedman et al 2009; Harman and Humphrey 1999; Lawson et al 1994; Liu 2007; Vuong and Tarr 2004). Thus, it is not necessarily clear, a priori, what the effect of motion will be when structurally similar objects undergo similar motions: experiments 2 and 3 were designed to examine this issue.

Different models have been proposed to account for the patterns of viewpoint dependence that we and others have observed (eg Biederman and Gerhardstein 1993; Friedman et al 2009; Hayward and Williams 2000). For instance, in a view-combination approach (Bülthoff and Edelman 1992; Edelman 1999; Poggio and Edelman 1990; Ullman 1998), recognition is construed as a form of generalisation between points in a multidimensional shape space. Notably, these points can represent different views of an object because they are defined by the metric values of the object's features (eg colours, edges, textures, and so on) measured from a given viewpoint. Recognition occurs when an input view, whether familiar or not, activates one or more prototypes stored in visual memory beyond some threshold of activation. Thus, in this model, depending on training and experience, performance can be viewpoint-dependent or viewpoint-independent.

View combination has been contrasted with a structural-description model that posits representations that are much less, if at all, sensitive to viewpoint changes (eg Biederman 1987; Hummel and Biederman 1992; Marr and Nishihara 1978). This model assumes that recognition relies on the extraction of qualitative features that are available from a wide range of viewpoints (eg non-accidental properties—Biederman 1987) and thus predicts little to no cost of viewpoint changes on recognition performance under most circumstances. Viewpoint effects are attributed to the need to discriminate highly similar stimuli that have only metric rather than qualitative differences between them (eg faces), or to post-recognition artifacts such as retrieving a name label or making a motor response (Biederman and Gerhardstein 1993).

However, it remains uncertain how observers recognise dynamic objects from unfamiliar viewpoints in either of these models. For example, different studies on either faces or novel objects suggest that different types of motions interact differently with viewpoint effects. With faces, Watson et al (2005) found that non-rigid facial motions were likely to be encoded in a viewpoint-invariant manner whereas rigid head motions (eg nodding or shaking) were likely to be encoded from specific views. With novel objects, Chuang et al (2006) found that non-rigid deformations modulated the viewpoint effect when observers discriminated amoeba-like shapes from different angular disparities. Several researchers have incorporated motion or other dynamic cues into models of object recognition to try to account for some of these findings, but they focused on how the models learn dynamic object representations from specific viewpoints or from a large range of views of an object rather than on the immediate perception of dynamic objects (eg Bülthoff et al 2002; Földiák 1991; Giese and Poggio 2003; Stone and Bray 1995; Stone and Harper 1999; Wallis and Bülthoff 2001). By investigating how structural similarity and different types of motion interact with the viewpoint effect during perception, we aim to further support or constrain how object recognition models must deal with the contribution of motion to the recognition process.

2 Experiment 1

The main goal of experiment 1 was to investigate whether and how structural similarity and rigid motion interact with the viewpoint effect. Lawson and her colleagues have shown that viewpoint changes are more detrimental to performance for both structurally similar novel and familiar objects (Lawson and Bülthoff 2006, 2008; Lawson et al 2003). Similarly, we have shown that changing viewpoint was more detrimental to performance if dynamic objects were structurally similar than when they were distinct (Friedman et al 2009; Vuong and Tarr 2004). However, we used drastically different types of objects—the structurally similar objects were amoeba-like shapes with little part-structure, whereas structurally distinct objects were multi-part geometric volumes. Many intrinsic differences between these two sets of objects, rather than similarity per se, may have contributed to the data we observed (Hayward and Williams 2000; Tjan and Legge 1998). For example, the amoeba-like shapes lacked part-structure that may contribute to viewpoint-invariant performance (Biederman 1987).

To address this issue, observers in experiment 1 discriminated structurally similar or distinct pairs of multi-part objects from different viewpoints in depth. Importantly, these two types of object pairs were drawn from the same shape space; that is, we selected multi-part objects from a shape parameter space in which the structural similarity among the objects had been parametrically varied and their perceived similarity was known (Schultz et al 2008). For half of the observers, the pairs rotated in depth in unison. For the other observers, the pairs were presented as static images. Both structural similarity and rigid rotation should affect the viewpoint effect, though perhaps in opposite directions. Furthermore, there is often a bias to process shape information in object recognition (eg Biederman 1987; Spetch et al 2006; Vuong and Tarr 2006). Thus, we expect that structural similarity should amplify the viewpoint effect (eg Lawson et al 2003), whereas rigid rotation should modulate it. It is not clear, however, how these factors may interact.

2.1 *Method*

2.1.1 *Participants.* Forty volunteers (twenty-five women, fifteen men) were drawn from the University of Alberta undergraduate participant pool, receiving partial credit for a course requirement. Participants were randomly assigned to either the rotation or static condition. Data from one participant in the static condition were removed because of near-chance performance (55% correct).

2.1.2 *Stimuli and apparatus.* Figure 1 illustrates examples of the structurally similar and distinct object pairs used in experiments 1–3 (dynamic versions can be found on the *Perception* website at http://dx.doi.org/10.1068/p6430). These objects were drawn from the shape space created by Schultz et al (2008). Briefly, the objects were multi-part volumetric primitives in which each part was defined by three continuous parameters: shape of the cross-section (from square to circle), curvature of the elongation axis of the cross-section, and amount of tapering along that axis. By varying the values of these parameters for each part, the similarity (in parameter space) can be defined between pairs of objects (Edelman 1999). Figure 2 illustrates this systematic manipulation of structural similarity.

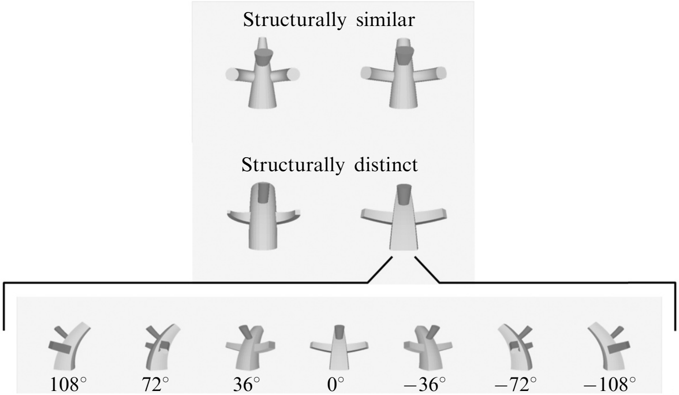

Figure 1. Examples of a structurally similar and structurally distinct object pair used in experiments 1 – 3. The objects are shown at the 0° viewpoint (frontal view). Each object contained two arms and a nose. Different viewpoints of the bottom-right object are also shown.

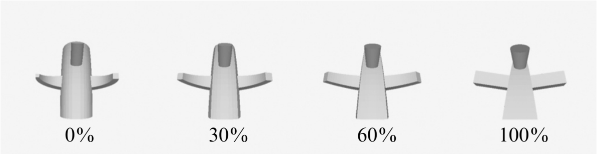

Figure 2. An example of a morph continuum. The end-point prototype objects are represented by the 0% object and the 100% object. Intermediate morphs are similarly represented by the 30% object and the 60% object. The 30% morph is more similar to the 0% prototype than the 100% prototype. By comparison, the 60% morph is more similar to the 100% prototype than the 0% prototype.

Three pairs of objects were chosen from the shape space so that objects in each pair were structurally similar; and three pairs were chosen so that objects in each pair were structurally distinct. The structural similarity of each pair was based on the psychometric curve obtained in Schultz et al's (2008) study (see their figure 4). For structurally similar pairs, participants in Schultz et al's study responded "different" on approximately 60% of the trials (slightly above chance). For structurally distinct pairs, participants would nearly always respond "different", based on extrapolating the psychometric curve. One similar and one distinct pair were used only in practice trials. These practice pairs were the same for all participants.

To render images of the objects, a virtual camera was placed in front of the geometric centre of each object. An image was rendered by rotating the object 0° (front of the object), ±36°, ±72°, and ±108° about the vertical axis for a total of seven static views. These views are shown for the bottom right object in figure 1. The objects were rendered in a matte grey colour against a uniform yellow background. The images were 500 × 500 pixels, subtending approximately 12.2 deg of visual angle. The object themselves were approximately 9 deg of visual angle, and centred within the image.

We also rendered 42-frame movies of each object oscillating in depth ±10° about each of the seven viewpoints used to render the static images (ie 0°, ±36° ±72°, and ±108°). Thus, there were seven rendered movies for each object. The movies began with the object initially rotated 10° counterclockwise of the static viewpoints. The object then rotated clockwise 20°, and then counterclockwise 20° to return to its initial position. All movies therefore began and ended on the same view of the object. For example, a movie began with the 26° view of an object; the object then rotated 20° clockwise to the 46° view (passing through the static 36° view); and then counterclockwise back to the initial 26° view. A full oscillation (42 frames) took 1.4 s so that objects oscillated at 30° s^{-1}.

The stimuli were presented in E-Prime (PST Software 2002) on a Samsung SyncMaster 940BF monitor (1024 × 768 pixel resolution; 60 Hz refresh rate; 2 ms grey-to-grey time). The computer for running the experiments had an Intel Core 2 CPU 6300 (2 GB RAM; 1.86 GHz) and an NVidia GeForce 7600GS video card (256 MB of video memory) for accurate timing of the stimuli. Participants sat approximately 68 cm from the monitor.

2.1.3 *Design.* A mixed design was used with motion type (rotation, static) as a between-subjects factor, and trial type (same, different), structural similarity (similar, distinct), and angular disparity (ie the viewpoint difference between the pair of objects: 0°, 36°, 72°, 108°) as within-subjects factors. In the rotation condition, participants were presented with pairs of movies. In the static condition, participants were presented with pairs of static images from the rendered static views (ie 0°, ±36°, ±72°, and ±108° views).

In each group, the order of presentation for the 32 different conditions (2 objects × 2 same/different trials × 2 structurally similar/distinct pairs × 4 angular disparities) was randomised and shown 10 times for a total of 320 trials. Each participant saw one of the two structurally similar pairs and one of the two structurally distinct pairs used on experimental trials (ten participants per experimental pair). The different pairs across participants were counterbalanced for any object-specific idiosyncrasies because we had arbitrarily set the shape parameters (see Schultz et al 2008). Each object in a pair was presented with itself on same trials and with the other member of the pair on different trials. Each angular disparity was seen 10 times with the constraint that every possible combination of views giving rise to that disparity difference was presented in roughly equal numbers. The side of presentation of the objects was randomly determined on each trial.

2.1.4 *Procedure.* Participants first learned the same/different task during a short practice session consisting of 32 trials, in which each experimental condition was presented once. The order of these trials was randomised. On each trial, participants saw a black fixation cross for 1000 ms, followed immediately by a stimulus pair presented side-by-side and separated by either 0°, 36°, 72°, or 108° of rotation in depth. For the participants in the rotation group, both objects started the oscillation from the first frame of the movie and continued to oscillate in synchrony until a response was made. For those in the static group, the objects were displayed as static images until a response was made. Participants were instructed to decide whether the stimuli were the same or different objects, ignoring any viewpoint differences, by pressing one of two buttons on a button-box. The response mapping was counterbalanced across participants. After the participants responded, the stimuli were removed and the feedback "correct" was displayed for 500 ms or "incorrect" for 750 ms. Following the feedback, there was a short 500 ms interstimulus interval before the next trial began. Observers were instructed to respond as quickly and as accurately as possible.

Participants were then tested on the 320 experimental trials. The same procedural sequence was used on test trials as on training trials except that no feedback was provided. The test phase consisted of two blocks of 160 trials, between which participants took a self-timed break. The experiment took approximately 20 min to complete.

2.2 *Results*

Correct reaction times (RTs) greater than 2.5 standard deviations from each participant's overall mean correct RTs were trimmed from the data. These trimmed trials were counted as errors in the accuracy analyses to apply a similar exclusion criterion to the accuracy data and to permit comparisons between measures taken on the average of identical trials. This exclusion of trimmed trials from the RT and accuracy analyses was used in all three experiments. For experiment 1, the error rate was 13.6% and the percentage of trimmed error trials was 2.4%, for an overall error rate of 16.0%.

We conducted mixed-design analyses of variance (ANOVAs) on both the accuracy and RT data, with trial type (same, different), structural similarity (similar, distinct), and angular disparity (0°, 36°, 72°, 108°) as repeated measures. Motion type (rotation, static) was a between-subjects factor. In this and subsequent experiments, we adopted $p < 0.05$ as our criterion for statistical significance and used η_{p}^2 as the measure of effect size. Furthermore, there were no indications of speed–accuracy trade-offs in any of the experiments.

Following previous studies, we also regressed RTs onto angular disparity to calculate the slope, which is a measure of the magnitude of the viewpoint effect (eg Cohen and Kubovy 1993; Shepard and Cooper 1982; Tarr 1995). The RT slope represents the change in response time per degree of rotation. The larger the slope value, the larger the viewpoint effect. Significant modulations of the viewpoint effect are captured by the interaction between a factor and the linear component of angular disparity; thus, where appropriate, we present RT slopes to represent this modulation. The RT slopes for all experiments are presented in table 1.

Table 1. Mean RT slopes (ms per degree) with standard errors in parentheses as a function of experiment, motion condition, trial type, and structural (experiment 1) or motion (experiments 2 and 3) similarity.

Experiment	Motion	Same trial type		Different trial type	
		similar	distinct	similar	distinct
1	Rigid	5.14 (0.79)	3.73 (0.60)	1.59 (0.85)	0.16 (0.24)
	Static	5.29 (0.95)	4.45 (0.96)	1.34 (1.39)	1.56 (0.48)
2	Articulation	1.17 (0.21)	1.13 (0.22)	0.69 (0.24)	0.55 (0.23)
	Static	2.47 (0.38)	1.48 (0.37)	0.70 (0.29)	1.27 (0.44)
3	Coherent	1.30 (0.24)	1.00 (0.14)	0.20 (0.20)	0.38 (0.15)
	Scrambled	1.29 (0.28)	1.30 (0.23)	0.90 (0.13)	0.35 (0.15)

2.2.1 *Accuracy.* Figure 3 shows the accuracy data averaged across participants. For the between-subjects factor, there was a main effect of motion type ($F_{1,37} = 5.10$, $\mathrm{MSE} = 511.24$, $p = 0.03$, $\eta_{\mathrm{p}}^2 = 0.12$). Observers were more accurate with rotating stimuli than with static stimuli (86.0% versus 81.9%).

There were main effects of trial type ($F_{1,37} = 5.51$, $p = 0.02$, $\mathrm{MSE} = 936.89$, $\eta_{\mathrm{p}}^2 = 0.13$), structural similarity ($F_{1,37} = 96.85$, $p < 0.001$, $\mathrm{MSE} = 369.64$, $\eta_{\mathrm{p}}^2 = 0.72$), and angular disparity ($F_{3,111} = 37.26$, $p < 0.001$, $\mathrm{MSE} = 82.88$, $\eta_{\mathrm{p}}^2 = 0.50$). Importantly, however, these factors interacted: structural similarity interacted with both trial type ($F_{1,37} = 78.93$, $p < 0.001$, $\mathrm{MSE} = 374.80$, $\eta_{\mathrm{p}}^2 = 0.68$) and angular disparity ($F_{3,111} = 4.73$, $p = 0.04$, $\mathrm{MSE} = 64.49$, $\eta_{\mathrm{p}}^2 = 0.11$). There was also an interaction between trial type and angular disparity ($F_{3,111} = 9.59$, $p < 0.001$, $\mathrm{MSE} = 101.47$, $\eta_{\mathrm{p}}^2 = 0.20$). Observers were more accurate overall on same trials than on different trials (86.8% versus 81.1%). However, they were much more accurate on same trials than on different trials for structurally similar pairs (86.2% versus 66.6%) compared with a reversal of this pattern

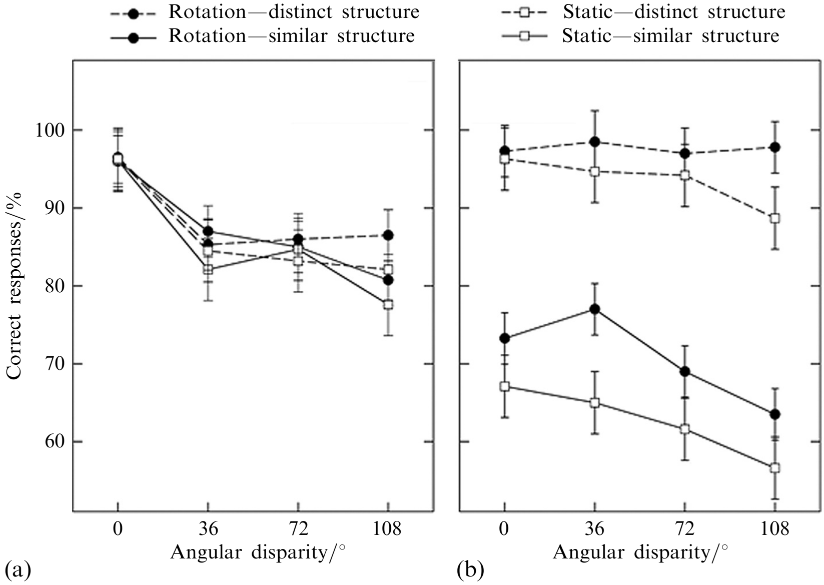

Figure 3. The accuracy data for experiment 1 across the different conditions. Trial type: (a) same; (b) different. Error bars in this and subsequent figures are 95% confidence intervals, computed within subjects as suggested by Loftus and Masson (1994).

for structurally distinct pairs (87.5% versus 95.6%). Likewise, the effect of viewpoint differed from structurally similar versus distinct object pairs, and it differed across trial type (see figure 3).

There was a trend towards a 3-way interaction between motion type, structural similarity, and the linear component of angular disparity ($F_{1,37} = 2.16$, $p = 0.14$, $\eta_p^2 = 0.05$). To explore this trend further, we analysed the rotation and static groups separately to reduce between-subject variability. For the rotation group, there was a significant interaction between structural similarity and the linear component of angular disparity ($F_{1,19} = 12.52$, $p = 0.002$, $\mathrm{MSE} = 62.04$, $\eta_p^2 = 0.39$), indicating a significant difference in slopes between structural similarity conditions. In contrast, for the static group, this interaction was not significant ($F_{1,18} = 1.21$, $p = 0.28$, $\mathrm{MSE} = 79.98$, $\eta_p^2 = 0.06$).

We also conducted separate ANOVAs for same and different trials. The main rationale for conducting separate analyses is that it has been noted that different processes may occur to make same versus different judgments (eg Hayward and Williams 2000; Shepard and Cooper 1982). For example, on different trials, observers may discriminate the object pairs based on a single local difference between the two stimuli. By comparison, on same trials, observers may try to match the percepts of the two stimuli in a holistic manner; certainly they must determine that all the visible parts match. In the present case, for same trials, there was only a significant effect of angular disparity ($F_{3,111} = 31.56$, $p < 0.001$, $\mathrm{MSE} = 101.15$, $\eta_p^2 = 0.46$). By comparison, for different trials, there was a significant effect of structural similarity ($F_{1,37} = 106.61$, $p < 0.001$, $\mathrm{MSE} = 611.87$, $\eta_p^2 = 0.74$), angular disparity ($F_{3,111} = 10.43$, $p < 0.001$, $\mathrm{MSE} = 83.19$, $\eta_p^2 = 0.22$), their interaction ($F_{3,111} = 3.99$, $p = 0.01$, $\mathrm{MSE} = 59.05$, $\eta_p^2 = 0.09$), and the linear component of the interaction ($F_{1,37} = 8.02$, $p = 0.007$, $\mathrm{MSE} = 72.34$, $\eta_p^2 = 0.17$). These differences between same and different trials are evident in figure 3. Overall, the accuracy data provided new evidence that both rotation and structural distinctiveness between objects can reduce the viewpoint effect during perception (see Friedman et al 2009 for effects of rotation following training).

2.2.2 *Reaction times.* Figure 4 shows RTs averaged across participants for the different experimental conditions. Unlike the accuracy data, there was no main effect of motion type ($F_{1,37} = 1.18$, $p = 0.28$). Thus, although rotation improved overall accuracy, it did not speed responses.

Like the accuracy data, there were main effects of trial type ($F_{1,37} = 22.69$, $p < 0.001$, MSE = 707 423.45, $\eta_\mathrm{p}^2 = 0.38$), and angular disparity ($F_{3,111} = 38.95$, $p < 0.001$, MSE = 85 430.71, $\eta_\mathrm{p}^2 = 0.51$), and an interaction between these factors ($F_{3,111} = 25.73$, $p < 0.001$, MSE = 59 044.34, $\eta_\mathrm{p}^2 = 0.41$). The interaction between trial type and the linear component of angular disparity was also significant ($F_{1,37} = 47.34$, $p < 0.001$, MSE = 65 016.03, $\eta_\mathrm{p}^2 = 0.56$). Observers responded more slowly on same trials than on different trials (2002 ms versus 1681 ms) and they were more affected by angular disparity on same trials than on different trials (4.65 ms per degree versus 1.16 ms per degree).

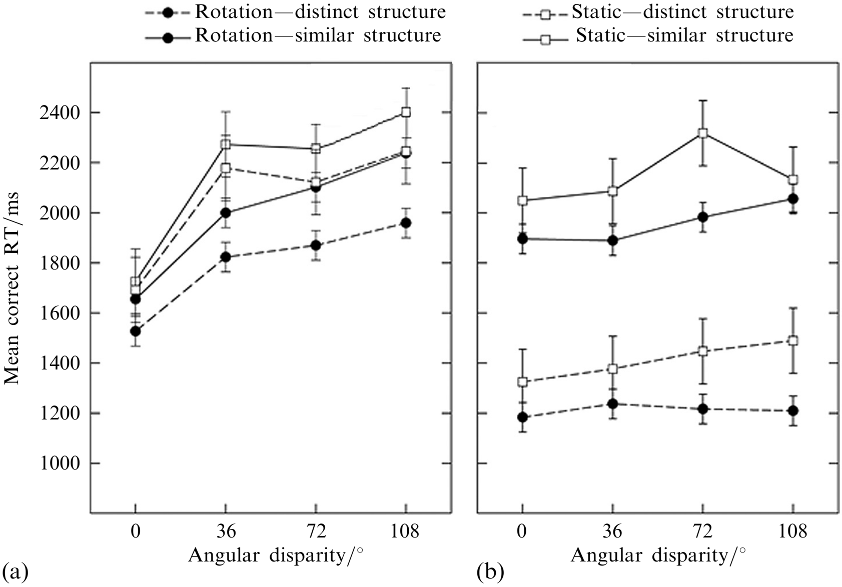

Figure 4. The reaction time (RT) data for experiment 1 across the different conditions. Trial type: (a) same; (b) different.

The effects of structural similarity on RTs showed a similar pattern to its effects on accuracy. There was a main effect of structural similarity ($F_{1,37} = 58.80$, $p < 0.001$, MSE = 524 519.40, $\eta_\mathrm{p}^2 = 0.61$) and a significant interaction between trial type and structural similarity ($F_{1,37} = 67.48$, $p < 0.001$, MSE = 201 369.01, $\eta_\mathrm{p}^2 = 0.64$). On different trials, structurally similar pairs had much longer reaction times than structurally distinct pairs (2051 ms versus 1311 ms), but on same trials, the difference between similar and distinct pairs was much smaller (2077 ms versus 1927 ms). Again, this is likely because different trials can be responded to on the basis of local features.

For comparability with the accuracy data, we further analysed the rotation and static group separately. For the rotation group, there was a significant interaction between structural similarity and the linear component of angular disparity ($F_{1,19} = 6.39$, $p = 0.02$, MSE = 41 010.02, $\eta_\mathrm{p}^2 = 0.25$). That is, there was a larger viewpoint effect for the structurally similar object pairs than for the structurally distinct pairs (3.37 ms per degree versus 1.95 ms per degree). By comparison, this interaction was not significant for the static group ($F < 1.0$, $p = 0.70$) (3.31 ms per degree versus 3.01 ms per degree for structurally similar and distinct pairs, respectively). Indeed, overall, the structurally distinct pairs that were moving had only a very slight viewpoint effect, compared with the other conditions.

Finally, we also analysed same and different trials separately. For same trials, there was a 2-way interaction between structural similarity and the linear component of angular disparity ($F_{1,37} = 7.54$, $p = 0.009$, MSE $= 21\,040.12$, $\eta_{\mathrm{p}}^2 = 0.16$). By comparison, no interactions were significant on different trials (Fs < 1.14, ps > 0.43). Thus, consistent with the accuracy data, these findings provide further evidence that rotation and structural similarity reduced the viewpoint effect.

2.3 *Discussion*

In experiment 1, observers in the rotation group were more accurate than those in the static group but both groups responded equally fast. Consistent with previous studies (eg Bülthoff and Edelman 1992; Hayward and Williams 2000; Khan and Humphrey 1992; Lawson and Humphreys 1996; Tarr 1995; Tarr et al 1998), we found a robust viewpoint effect for both groups. Observers' errors and RTs systematically increased as the two objects were shown from increasingly larger differences in viewpoint. We further found that observers were more affected by viewpoint (ie there was a larger slope) for structurally similar pairs than for structurally distinct pairs for both measures (see also Friedman et al 2009; Lawson et al 2003; Lawson and Bülthoff 2006, 2008; Vuong and Tarr 2004). Importantly, however, rigid rotation modulated the interactions between structural similarity and viewpoint (see also Friedman et al 2009). In particular, the observed modulation of the viewpoint effect by structural similarity was driven mostly by observers in the rotation group. Observers in the static group showed very little difference in the slope of the RT functions for similar and distinct objects.

3 Experiment 2

In experiment 1, we found that both structural similarity and rigid rotation modulated the viewpoint effect. To further extend these results, in experiment 2 we used articulatory motion, which is qualitatively different from rotation. Using different articulated motions of the same 3-D structures had the additional advantage of allowing us to examine the role of motion similarity while holding shape similarity constant. For this purpose, we morphed the motion trajectories between the two articulations we used to vary their similarity. In previous work in which some form of motion morphing has also been used (eg Giese and Lappe 2002; Giese et al 2008), highly degraded point-light displays of human actors were employed in which only the joints of the actors were visible. In contrast, we used fully shaded objects.

It has been found that observers can use motion cues to recognise individuals from their unique articulations (eg the way they walk—Cutting and Kozlowski 1977). In addition, there is some evidence that dynamic articulation cues are used to recognise unfamiliar objects. Newell and Setti (2006), for example, recently found that articulations primed the recognition of static images of learned objects only when these motions were related to the global body movements (eg if an object translated upwards, then priming occurred if the parts moved downwards in a propelling motion). The results of experiment 1 and previous work made us expect that articulated motion would facilitate performance and reduce the viewpoint effect. We also expected that the similarity between the articulations should affect the size of the viewpoint effect, such that more similar motions should make discrimination more difficult.

3.1 *Method*

3.1.1 *Participants.* Forty-eight new volunteers (thirty-one women, seventeen men) were drawn from the same subject pool as experiment 1. They were randomly assigned to either the articulation or static condition.

3.1.2 *Stimuli and apparatus.* For this and the subsequent experiment, four new structurally similar pairs of objects were used as experimental stimuli. A fifth new structurally similar pair was used only in practice trials for all participants in experiments 2 and 3.

Figure 5 shows how the objects in experiment 2 were animated (example videos can be found on the *Perception* website at http://dx.doi.org/10.1068/p6430), and figure 6 illustrates this 'motion' space. Each object either 'waved' its arms by pivoting them upwards in the $X-Y$ plane at the 'shoulder joints' (see the left panel of figure 5), or 'hugged' its arms by pivoting the arms inwards in the $X-Z$ plane from the same joints (see the middle panel of figure 5). The arms pivoted 85° in unison at 85° s^{-1}, starting from the extreme outward position (ie perpendicular to the main body) and then returning to the start position.

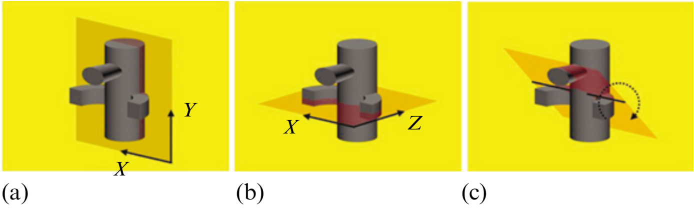

Figure 5. [In colour online, see http://dx.doi.org/10.1068/p6430] An illustration of how the wave and hug actions were created. (a) A pure wave action. The transparent red (in the colour figure) plane represents the articulation plane in which the arms rotated along. (b) A pure hug action. In (c) it is shown how intermediate actions were created. The articulation plane is rotated about the X axis, thereby morphing between the pure wave and pure hug actions.

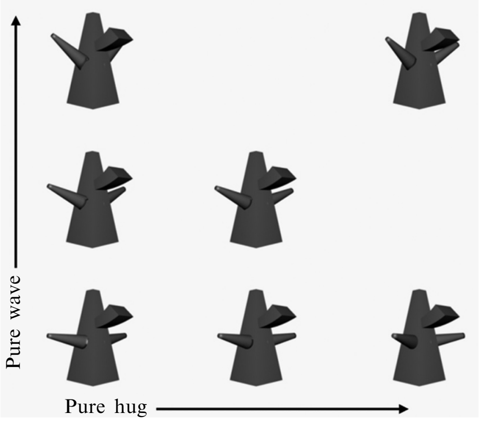

Figure 6. An illustration of the motion space used in experiments 2 and 3. Static frames extracted from a pure wave, pure hug, and intermediate actions. Note that all actions had the same starting position (lower left corner).

New articulations were systematically created by interpolating between the pure wave and pure hug actions. As shown in figure 5c this interpolation was accomplished by rotating the articulation plane through which the arms pivoted. This plane was rotated about the X axis. For instance, an articulation that was half-way (50%) between a wave and hug was created by rotating the articulation plane 45° about the X axis. The diagonal in figure 6 illustrates this interpolated motion. Thus, we were able to parametrically vary motion similarity in a comparable way to structural similarity (see Giese and Lappe 2002, and Giese et al 2008, in the domain of biological-motion perception). For similar motion pairs, one object of a pair had a pure articulation (either 100% wave or 100% hug), whereas the other had 67% of the pure articulation. By comparison, for distinct motion pairs, one object had a pure articulation and the other had 33% of the pure articulation. These proportions are similar to what we have used when investigating the effects of structural similarity (eg Schultz et al 2008).

All objects were animated with the open-source Blender software (Version 2.41 2006, Blender Foundation), and rendered as 30-frame movies with the first frame depicting the arms in their starting position (ie perpendicular to the body—see lower left corner of figure 6). The objects were rendered in a matte grey colour against a 500 × 500 pixel yellow background from the same seven views as in experiment 1.

For movie presentations, both objects performed their actions continuously and in synchrony until participants responded. Both movies were looped, starting from the first frame, going to the 30th, and then reversing from the 30th to the first. This entire animation took 2 s to complete. For static presentations, the particular frame shown in a given trial was randomly selected from frames 2 to 29 without replacement, for each object, on each trial. We excluded frames 1 and 30 because they represent the turning points in the action, and were very similar to frames 2 and 29, respectively. Thus, both static and motion groups received roughly the same structural information by the end of the experiment.

3.1.3 *Design and procedure.* The design and procedure were similar to those in experiment 1. A mixed design was used with motion type (articulation or static) as a between-subjects factor, and trial type (same or different), motion similarity (as opposed to structural similarity; similar or distinct), and angular disparity as within-subjects factors. As in experiment 1, there were roughly equal numbers of the different-view combinations for each angular disparity tested. In the articulation group, participants were presented with pairs of movies. In the static group, participants were presented with pairs of static images. For both groups, each participant was shown only one of the four possible experimental pairs (six participants per experimental pair). An object in the pair was paired with itself on same trials, and with the other member on different trials. As described above, in each trial one of the objects in a pair always made a pure wave or pure hug action while the other object in the pair made either a similar or distinctive articulated motion relative to the pure action. Each object was shown an equal number of times with a pure wave action or a pure hug action across the other conditions. Thus, there were 64 conditions in total (2 objects × 2 hug/wave actions × 2 same/different trials × 2 similar/distinct articulated motions × 4 angular disparities). Each of these conditions was repeated 7 times for a total 448 trials. These trials were presented in a random order for each participant. For observers in the static group, a different frame from the movie was selected in each trial (see section 3.1.2). The side of presentation of the objects was randomly determined in each trial.

Participants were run in a practice session consisting of 64 trials with feedback, with the same procedural trial sequence as in the first experiment. Following the practice session, they were tested on the 448 experimental trials. The test phase consisted of two blocks, between which participants took a self-timed break. The experiment took approximately 30 min to complete.

3.2 *Results*

3.2.1 *Accuracy.* In experiment 2, the error rate was 6.9% and the percentage of trimmed trials was 2.9%, for an overall error rate of 9.8%. The accuracy data were submitted to a mixed-design ANOVA with motion type (articulation, static) as a between-subjects factor, and trial type (same, different), motion similarity (similar, distinct), and angular disparity (0°, 36°, 72°, 108°) as repeated measures.

Figure 7 shows the accuracy across the different conditions. Overall, observers were very accurate on the task ($> 84\%$ on average for all conditions). Thus, in this experiment we focus our discussion on RTs and only report the omnibus ANOVA for accuracy. There was a main effect of angular disparity ($F_{3,138} = 13.34$, $p < 0.001$, $\mathrm{MSE} = 43.06$, $\eta_\mathrm{p}^2 = 0.22$), and a significant interaction between trial type and angular disparity ($F_{3,138} = 8.49$, $p < 0.001$, $\mathrm{MSE} = 40.08$, $\eta_\mathrm{p}^2 = 0.15$). There was a main effect of motion similarity ($F_{1,46} = 4.65$, $p = 0.03$, $\mathrm{MSE} = 28.16$, $\eta^2 = 0.09$) and a 3-way interaction between motion type, trial type, and motion similarity ($F_{1,46} = 6.12$, $p = 0.01$, $\mathrm{MSE} = 46.44$, $\eta_\mathrm{p}^2 = 0.11$).

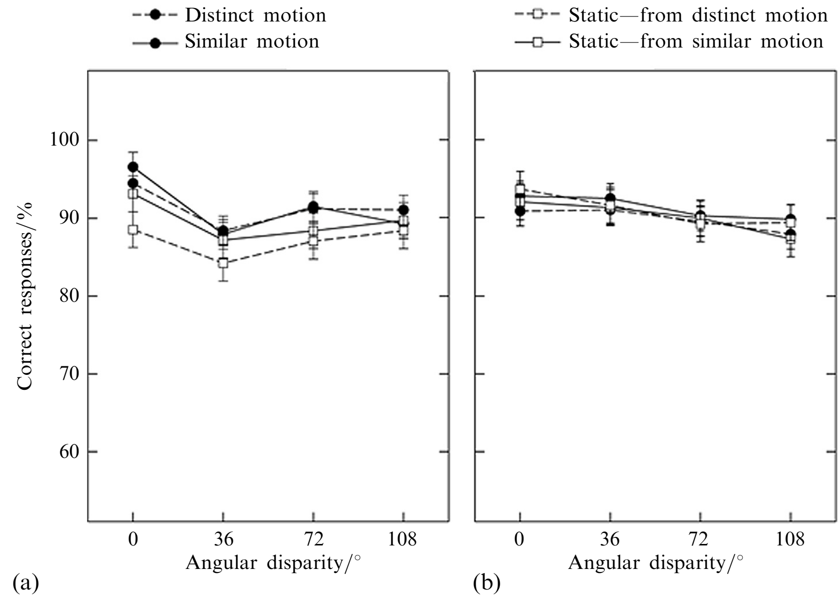

Figure 7. The accuracy data for experiment 2 across the different conditions. Trial type: (a) same; (b) different.

3.2.2 *Reaction times.* Figure 8 shows the correct trimmed RTs across the different conditions. There was a significant effect of angular disparity ($F_{3,138} = 39.35$, $p < 0.001$, $\text{MSE} = 19\,608.25$, $\eta^2 = 0.46$). However, this viewpoint effect was modulated by trial type ($F_{3,138} = 16.90$, $p < 0.001$, $\text{MSE} = 10\,543.90$, $\eta_p^2 = 0.26$), and by the linear component of the trial type by angular disparity interaction ($F_{1,46} = 32.16$, $p < 0.001$, $\text{MSE} = 5589.95$, $\eta_p^2 = 0.41$). Thus, observers had a larger magnitude of the viewpoint effect for same trials than for different trials (1.56 ms per degree versus 0.80 ms per degree).

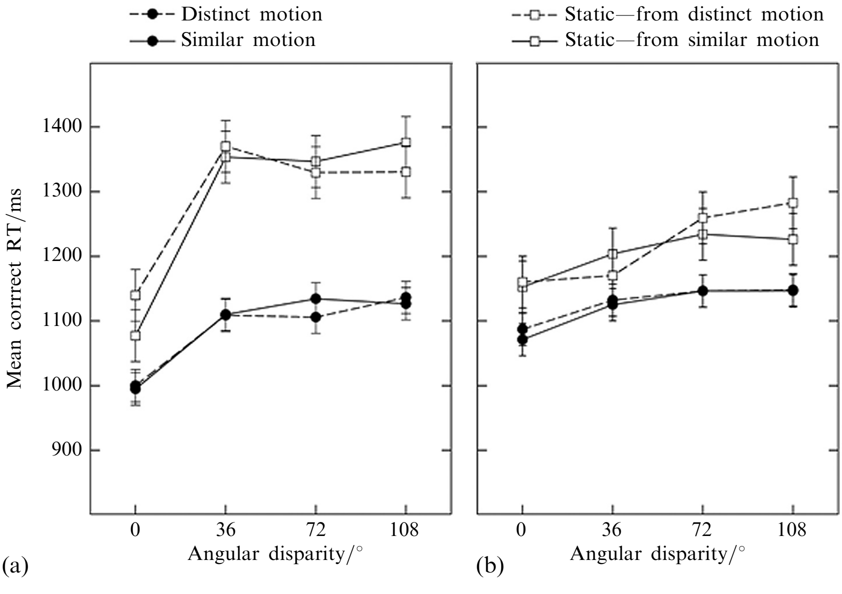

Figure 8. The reaction time (RT) data for experiment 2 across the different conditions. Trial type: (a) same; (b) different.

Importantly, there was an overall effect of motion type ($F_{1,46} = 3.74$, $p = 0.059$, $\text{MSE} = 1\,052\,487.06$, $\eta_\text{p}^2 = 0.07$), with observers in the articulation condition responding more quickly than those in the static condition (1107 ms versus 1250 ms). Motion type also interacted with several of the other factors. There was a significant interaction between motion type and trial type ($F_{1,46} = 4.29$, $p = 0.04$, $\text{MSE} = 147\,941.02$, $\eta_\text{p}^2 = 0.08$), and between motion type and angular disparity ($F_{3,138} = 2.67$, $p = 0.05$, $\text{MSE} = 19\,608.25$, $\eta_\text{p}^2 = 0.05$), a 3-way interaction between motion type, trial type, and angular disparity ($F_{3,138} = 4.89$, $p = 0.003$, $\text{MSE} = 10\,543.90$, $\eta_\text{p}^2 = 0.09$), and, most importantly, a small but significant 4-way interaction between all factors ($F_{3,138} = 2.98$, $p = 0.03$, $\text{MSE} = 6746.68$, $\eta_\text{p}^2 = 0.06$). The linear component of the 4-way interaction was also significant ($F_{1,46} = 8.99$, $p = 0.004$, $\text{MSE} = 6024.66$, $\eta_\text{p}^2 = 0.16$), suggesting that the slope of the viewpoint effect was modulated across all of the remaining conditions (none of the higher-order polynomial trends was significant, $F\text{s} < 1.0$).

As in experiment 1, and to further investigate these interactions, we conducted separate ANOVAs for same and different trials. There was a main effect of motion type on same trials ($F_{1,46} = 4.71$, $p = 0.03$, $\text{MSE} = 820\,819.75$, $\eta_\text{p}^2 = 0.09$), but not on different trials ($F_{1,46} = 1.86$, $p = 0.17$). Observers in the articulation group were faster than those in the static group on same trials (1090 ms versus 1290 ms), but not on different trials though there was a tendency in that direction (1125 ms versus 1211 ms).

Importantly, for same trials, there was a significant interaction between motion type and the linear component of angular disparity ($F_{1,46} = 4.79$, $p = 0.03$, $\text{MSE} = 22\,305.14$, $\eta_\text{p}^2 = 0.09$), a significant interaction between motion similarity and the linear component of angular disparity ($F_{1,46} = 7.50$, $p = 0.009$, $\text{MSE} = 5510.77$, $\eta_\text{p}^2 = 0.14$), and a significant 3-way interaction between motion type, motion similarity, and the linear component of angular disparity ($F_{1,46} = 6.46$, $p = 0.01$, $\text{MSE} = 5510.77$, $\eta_\text{p}^2 = 0.12$). Observers in the articulation group showed a smaller magnitude of the viewpoint effect than those in the static group on same trials, which was confirmed by a two-sample t-test (1.15 ms per degree versus 1.98 ms per degree; $t_{23} = 2.25$, $p = 0.03$). By comparison, observers in the two groups did not differ in their magnitude of viewpoint effect on different trials (0.62 ms per degree versus 0.99 ms per degree; $t_{23} = 0.19$, $p = 0.84$). Furthermore, only observers in the static group showed a larger viewpoint effect for similar articulation compared to distinct articulation on same trials (2.47 ms per degree versus 1.48 ms per degree). For different trials, these interactions were not significant.

3.3 *Discussion*

We again found a viewpoint effect for both accuracy and RTs in experiment 2. In addition, observers in this experiment were much more accurate overall than those in experiment 1. Although there was no overall accuracy difference between the two groups in experiment 2, observers in the articulation condition responded more quickly than those in the static condition. Importantly, articulatory motion decreased the viewpoint effect for RTs, relative to the static condition in some conditions (ie there was an interaction between motion type, trial type, and angular disparity). For example, observers in the articulation group showed no difference in the magnitude of the viewpoint effect for similar and distinct motion. By comparison, observers in the static group had a larger viewpoint effect for similar articulations compared with distinct articulations on same trials but not on different trials. This difference for the static group must reflect image differences because a pair of static images should necessarily be more different in the distinct motion condition than in the similar motion condition. For example, the static position of the arms would be more different with static images extracted from distinct motion than those extracted from similar motion. Thus, when the objects were moving, the motion similarity manipulation made the task relatively easy (and was in itself ineffective), whereas when the objects were static, there was an effect of similarity.

4 Experiment 3

In experiment 2, we found evidence that articulation reduced the magnitude of the viewpoint effect under some conditions. However, as noted earlier, observers in the articulation group may have been more attentive than those in the static group, given the nature of the changing stimuli (Harman and Humphrey 1999). To address this issue, in experiment 3 we compared coherent articulations with scrambled versions of these articulations. The scrambled versions randomised the frame order of the coherent articulation sequences to control for attention and to equate for the availability of image changes (Friedman et al 2009; Lawson et al 1994; Liu 2007; Vuong and Tarr 2004).

4.1 *Method*

4.1.1 *Participants.* Forty-eight new volunteers (twenty-nine women, nineteen men) were drawn from the same subject pool as in the previous experiments. They were randomly assigned to either the coherent or scrambled condition. Two participants in each group performed near chance (55%) so their data were not analysed further.

4.1.2 *Stimuli and apparatus.* The object pairs from experiment 2 were used in experiment 3. The only difference was that we rendered coherent waves and hugs (as in experiment 2) or scrambled these actions. Following previous studies (eg Spetch et al 2006; Vuong and Tarr 2004), we grouped the frames that comprised the videos into 10 3-frame subsets for both ascending frame order and descending frame order (recall that movies were looped). Thus, there were 20 such subsets, differing only in the direction of the action. We then scrambled the order of these 20 subsets while maintaining the frame order within subsets. This scrambling procedure was carried out on each trial (examples of scrambled versions of these animations can be found on the *Perception* website at http://dx.doi.org/10.1068/p6430).

4.1.3 *Design and procedure.* The design and procedure were identical to those in experiment 2, except that the static condition was replaced by the scrambled articulation condition.

4.2 *Results*

4.2.1 *Accuracy.* In experiment 3, the error rate was 5.1% and the percentage of trimmed trials was 2.8%, for an overall error rate of 7.9%. The accuracy data were submitted to a mixed design ANOVA with motion type (coherent, scrambled) as a between-subjects factor, and trial type, motion similarity, and angular disparity as repeated measures.

Figure 9 shows the accuracy across the conditions. As in experiment 2, observers were very accurate overall (> 87% on average for all conditions) so we also focused on RTs in this experiment and only report the omnibus ANOVA for accuracy. Consistent with experiments 1 and 2, there was a significant main effect of angular disparity ($F_{3,126} = 15.33$, $p < 0.001$, $\mathrm{MSE} = 35.14$, $\eta_\mathrm{p}^2 = 0.26$), and a significant interaction between trial type and angular disparity ($F_{3,126} = 6.96$, $p < 0.001$, $\mathrm{MSE} = 39.61$, $\eta_\mathrm{p}^2 = 0.14$). Importantly, there was a significant main effect of motion type ($F_{1,42} = 4.53$, $p = 0.03$, $\mathrm{MSE} = 268.0$, $\eta_\mathrm{p}^2 = 0.09$). Observers in the coherent-articulation group were more accurate than those in the scrambled-articulation group (93.5% versus 90.8%). There was no main effect of motion similarity ($F < 1.0$, $p = 0.52$), but there was again a significant interaction between trial type and motion similarity ($F_{1,42} = 3.99$, $p = 0.052$, $\mathrm{MSE} = 17.86$, $\eta_\mathrm{p}^2 = 0.08$).

4.2.2 *Reaction times.* Figure 10 shows the correct trimmed RTs across the different conditions. It is notable that all of the motion conditions were responded to relatively fast (compare figures 8 and 10). In addition to the main effect of angular disparity ($F_{3,126} = 34.30, p < 0.001$, $\mathrm{MSE} = 8921.99$, $\eta_\mathrm{p}^2 = 0.45$), there was a significant interaction between trial type and angular disparity ($F_{3,26} = 20.25$, $p < 0.001$, $\mathrm{MSE} = 4240.60$,

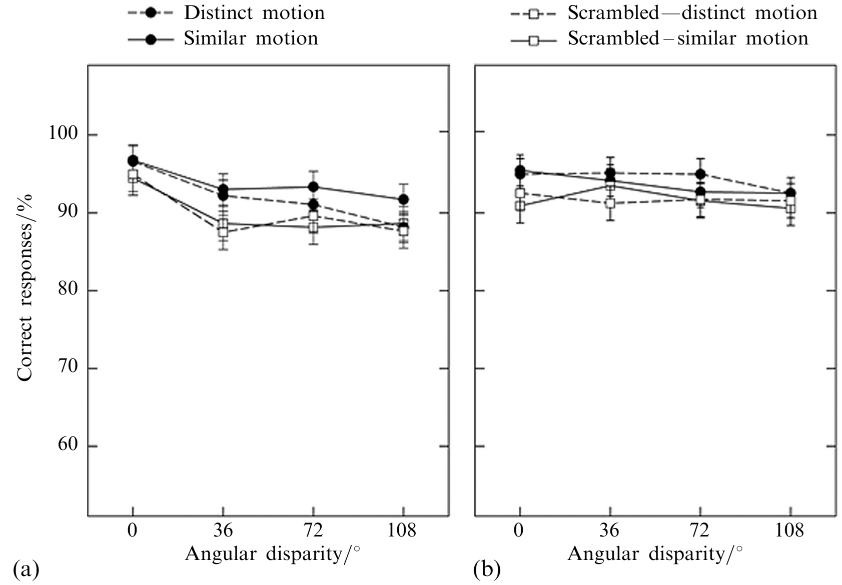

Figure 9. The accuracy data for experiment 3 across the different conditions. Trial type: (a) same; (b) different.

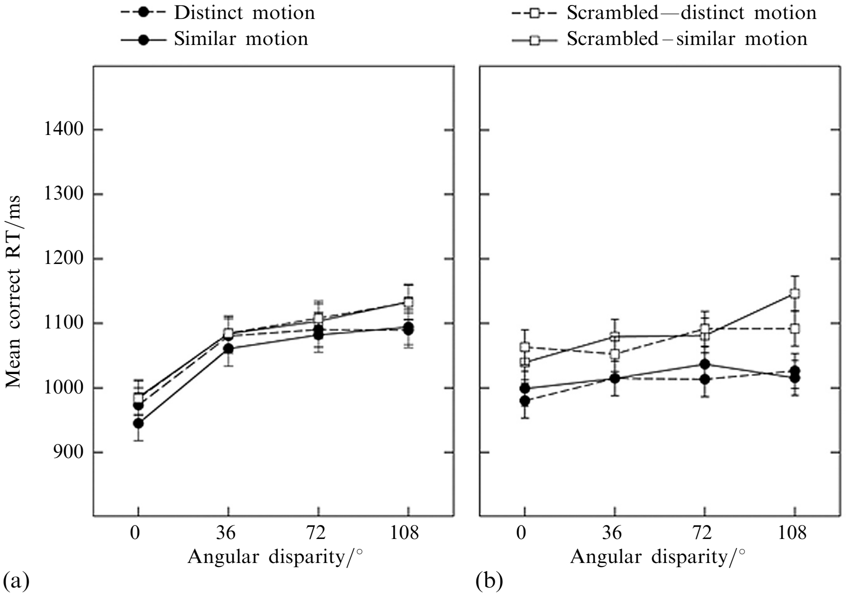

Figure 10. The reaction time (RT) data for experiment 3 across the different conditions. Trial type: (a) same; (b) different.

$\eta_p^2 = 0.32$), and between trial type and the linear component of angular disparity ($F_{1,42} = 33.25$, $p < 0.001$, $\mathrm{MSE} = 5017.0$, $\eta_p^2 = 0.44$). As in experiment 2, observers had a larger viewpoint effect for same trials than for different trials (1.22 ms per degree versus 0.46 ms per degree).

There was no main effect of motion type ($F < 1.0$, $p = 0.64$). However, as we anticipated on the basis of the results of experiment 2, the linear component of the 4-way interaction among the factors was significant ($F_{1,42} = 7.17$, $p = 0.01$, $\mathrm{MSE} = 2715.08$, $\eta_p^2 = 0.14$; none of the higher-order polynomial trends was significant, $F\mathrm{s} < 1.0$, $p\mathrm{s} > 0.37$).

We also conducted separate ANOVAs for same and different trials as in the previous experiments. There was no main effect of motion type in either ANOVA, nor did motion type interact with any other factors (Fs < 1.78). There was no main effect of motion similarity, although the effect approached significance on different trials ($F_{1,42} = 2.11$, $p = 0.15$, MSE $= 4019.53$, $\eta_{\mathrm{p}}^2 = 0.04$).

In contrast to experiment 2, we found a significant 3-way interaction between motion type, motion similarity, and angular disparity on different trials ($F_{3,126} = 4.01$, $p = 0.009$, MSE $= 3574.98$, $\eta_{\mathrm{p}}^2 = 0.08$), rather than on same trials. We further found a significant 3-way interaction between motion type, motion similarity, and the linear component of angular disparity on different trials ($F_{1,42} = 7.24$, $p = 0.01$, MSE $= 2588.50$, $\eta_{\mathrm{p}}^2 = 0.14$). For similar articulation on different trials, the viewpoint effect was smaller for the coherent-articulation group than for the scrambled-articulation group (0.20 ms per degree versus 0.90 ms per degree). In contrast, for distinct motion, the magnitude of the viewpoint effect was no different between the coherent-articulation and scrambled-articulation groups (0.38 ms per degree versus 0.35 ms per degree). None of these interactions was significant on same trials (Fs < 1.13, ps > 0.29). Overall, these results suggest that similar and distinctive motions had different influences on the viewpoint effect, depending on whether the articulation was coherent or scrambled, and that coherent articulation modulated the viewpoint effect as well as the effect of structural similarity, particularly on different trials.

4.3 *Discussion*

As in experiments 1 and 2, observers showed a clear viewpoint effect for both accuracy and reaction times. Observers in experiment 3 also responded more accurately and more quickly than those in experiment 1 but their level of performance was comparable to that of observers in experiment 2. Those in the coherent-articulation group responded more accurately than those in the scrambled-articulation group; this difference in performance was small but replicates the results of Friedman et al (2009). Importantly, coherent articulation reduced the magnitude of the viewpoint effect for some conditions (as in experiments 1 and 2).

As argued in section 1, coherent articulations lead to smooth and predictable motion, and we believe that this temporal smoothness may be what facilitated discrimination for some of the conditions in both experiments 2 and 3 (Friedman et al 2009; Vuong and Tarr 2004). Indeed, Stone and his colleagues (Stone and Bray 1995; Stone and Harper 1999) showed how the visual system may exploit temporal smoothness for visual perception. The results of experiment 3 are particularly important because coherent articulation modulated the complex interactions between the other factors, relative to scrambled articulation of the very same images. This finding thus helps rule out the possibility that the modulation effects observed across all three experiments are due only to differences in attention between dynamic versus static stimuli and it also helps rule out the possibility that the effects were due to differences in the availability of views (Harman and Humphrey 1999; Vuong and Tarr 2004). That is, the coherent-articulation and scrambled-articulation conditions were equated on the available dynamic cues that may draw attention and on the number of views on any single trial.

We note that the viewpoint effect was larger in experiment 1 than in experiments 2 and 3. This difference may reflect the fact that the task was more difficult in the first experiment than in experiments 2 and 3, judging from the accuracy data. However, we also note that task difficulty cannot be the only factor driving the modulation of the viewpoint effect by motion cues. For example, both coherent and scrambled articulation in experiment 3 reduced the viewpoint effect for distinct motions, relative to performance on the static views in experiment 2, even though accuracy was approximately equivalent across the two experiments (see table 1 and figures 7 and 9).

5 General discussion

In the present study, we examined the extent to which shape and motion cues modulated the viewpoint effect in the context of dynamic object recognition. For shape cues, we focused on the perceived similarity of 3-D structure. For motion cues, we focused on the type of motion cues available. Importantly, across the three experiments, we found that the viewpoint effect was modulated by structural similarity (experiment 1) as well as, though less so, by different types of motion cues, such as rigid rotation (experiment 1), articulation of parts (experiments 2 and 3), and scrambled articulation (experiment 3). In combination with our previous work (Friedman et al 2009; Vuong and Tarr 2004), these findings reinforce the view that motion is used not only to extract more features from a sequence of static images or to recover 3-D shape from that sequence but contributes, in itself, independent information to an object's representation.

The viewpoint effect has been central in the debate between different theoretical positions about how 3-D objects are represented and about the mechanisms that operate on these representations (see Biederman and Gerhardstein 1995; Tarr and Bülthoff 1995; and also Biederman and Bar 1999; Hayward and Tarr 2000). Our interpretation of the viewpoint effect found in all three experiments is that dynamic objects are encoded in a view-specific manner (eg Bülthoff and Edelman 1992; Hayward and Williams 2000; Khan and Humphrey 1992; Lawson and Humphreys 1996; Tarr et al 1998) but their 'view tuning' can be broadened by dynamic cues. That dynamic objects are encoded in a view-specific manner is interesting in light of experiment 1 because rotations in depth provide strong cues for an object-recognition system to extract view-invariant 3-D structure (Todd 2004; Ullman 1979) or other view-invariant visual features (Biederman 1987). Furthermore, we found a reliable overall viewpoint effect even though we used a relatively small number of objects over many trials, and we presented object pairs simultaneously.

The present results are generally consistent with a view-combination model (eg Edelman 1999). For instance, there were reliable viewpoint effects even for structurally distinctive multi-part object pairs. Such a finding is not consistent with a structural-description model of object recognition (eg Biederman 1987). More importantly, the pattern of results across the three experiments allows us to specify some additional constraints to a view-combination model, particularly in the context of dynamic objects. In conjunction with our previous results (Friedman et al 2009), it appears that coherent articulation helps observers generalise across view differences, thereby reducing viewpoint effects (at least in some conditions). However, it also appears that there is a limit to view generalisation even with dynamic stimuli. We also found a limit on generalisation for birds with static real objects (Friedman et al 2005) and for humans with static pictures of scenes (Friedman and Waller 2008). Thus, for instance, although motion may allow observers to predict upcoming views (Friedman et al 2009; Vuong and Tarr 2004), it appears that they are able to do so only within a limited range of angles. Furthermore, our results suggest that the extent of this view generalisation may depend on both the structural similarity and on the type of motion used. In a related work, for example, Wallis (2002) found that the structural similarity of faces can affect how different views of rotating faces are bound into a single representation. Thus, one interesting avenue for future research is to examine the kind of visual information that narrows or broadens the limits of generalisation.

On a final note, our study addresses an interesting issue raised by Hayward and Williams (2000). They suggested that the intrinsic geometry of objects determines the extent to which recognition may be affected by viewpoint in the absence of any diagnostic cues to the objects' identity (see also Tjan and Legge 1998). For example, objects that have very little part-structure (eg amoebas, stick objects, and even faces) tend to incur the largest viewpoint effects (eg Biederman and Gerhardstein 1993; see Tjan and Legge 1998, for a computational analysis). Hayward and Williams tested objects that consisted of a main body with smaller parts attached, much like the objects we used. With their objects, they found a similar magnitude of viewpoint effect irrespective of how discriminable the objects were from each other. Their interpretation was that all the objects were qualitatively similar across their discriminability manipulation, leading to the same magnitude of viewpoint effect independently of structural similarity. In contrast to Hayward and Williams (2000), however, in the present study we found larger viewpoint effects for structurally similar objects compared to structurally distinct objects even though both sets of objects were qualitatively of the same type (ie made of a distinctive part structure—see also Lawson and Bülthoff 2006, 2008; Lawson et al 2003). Notably, rather than generating our parts randomly, as Hayward and Williams did, we systematically varied the parts along structural dimensions (eg curvature of main axis.). We then selected pairs on the basis of the perceived similarity of the two objects. We therefore argue that structural similarity can contribute to the intrinsic complexity of objects. The differences between our study and that of Hayward and Williams may lie in the fact that *perceived similarity* may interact with geometry when determining complexity.

The present study further raises the interesting, yet speculative, idea that motion may contribute to the intrinsic complexity of objects and their representations. For example, the articulated motion condition in the present study led to shallower slopes than did rigid rotation, even though the objects were qualitatively similar in structure. Like shape, and following the work on human motion (eg Giese and Lappe 2002; Giese et al 2008), we parametrically varied the similarity of our articulation by averaging across a prototype hugging action and a prototype waving action (ie motion morphing). We found that, under some conditions, motion similarity modulated the viewpoint effect relative to static images. Thus, it will be interesting in future work to measure the relationship between intrinsic complexity and similarity for both shape and motion cues. In particular, it would be of importance to examine motion similarity psychophysically, as we did structural similarity, so as to have a basis for determining the sensitivity of the similarity manipulation for both dimensions.

6 Conclusion

Several studies have shown that motion is important for both face and object recognition (eg Chuang et al 2006; Friedman et al 2009; Liu and Cooper 2003; Knappmeyer et al 2003; Lander and Bruce 2000; Newell and Setti 2006; Newell et al 2004; Stone 1998, 1999; Vuong and Tarr 2004, 2006; Watson et al 2005). In line with these studies, our results are consistent with the idea that motion is encoded in the object representation even when it is not diagnostic of object identity. Importantly, the results further show that dynamic objects are likely to be encoded in a view-specific manner. These conclusions may seem surprising at first; however, in line with Hayward and Williams (2000), we would argue that a single mechanism, such as view combination, may suffice to recognise objects from a combination of static and dynamic cues that may be unpredictably present in the environment.

Acknowledgments. This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to Alinda Friedman. We thank Bernd Kohler for programming the experiments and for help collecting the data.

References

Aggarwal J K, Cai Q, Liao W, Sabata B, 1998 "Nonrigid motion analysis: Articulated and elastic motion" *Computer Vision and Image Understanding* **70** 142 – 156

Biederman I, 1987 "Recognition-by-components: A theory of human image understanding" *Psychological Review* **94** 115 – 147

Biederman I, Gerhardstein P C, 1993 "Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance" *Journal of Experimental Psychology: Human Perception and Performance* **19** 1506 – 1514

Biederman I, Gerhardstein P C, 1995 "Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995)" *Journal of Experimental Psychology: Human Perception and Performance* **21** 1506 – 1514

Biederman I, Bar M, 1999 "One-shot viewpoint invariance in matching novel objects" *Vision Research* **39** 2885 – 2899

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 of the USA* **89** 60 – 64

Bülthoff H H, Wallraven C, Graf A B A, 2002 "View-based dynamic object recognition based on human perception" *16th International Conference on Pattern Recognition* (Québec: IEEE) pp 768 – 776

Chuang L L, Vuong Q C, Thornton I M, Bülthoff H H, 2006 "Recognizing novel deforming objects" *Visual Cognition* **14** 85 – 88

Cohen D, Kubovy M, 1993 "Mental rotation, mental representations and flat slopes" *Cognitive Psychology* **25** 351 – 382

Cutting J E, Kozlowski L T, 1977 "Recognizing friends by their walk: Gait perception without familiarity cues" *Bulletin of the Psychonomic Society* **9** 353 – 356

Edelman S, 1999 *Representation and Recognition in Vision* (Cambridge, MA: MIT Press)

Földiák P, 1991 "Learning invariance from transformation sequences" *Neural Computation* **3** 194 – 200

Friedman A, Spetch M L, Ferrey A, 2005 "Recognition by humans and pigeons of novel views of 3-D objects and their photographs" *Journal of Experimental Psychology: General* **134** 149 – 162

Friedman A, Vuong Q C, Spetch M L, 2009 "View combination in moving objects: The role of motion in discriminating between novel views of similar and distinctive objects by humans and pigeons" *Vision Research* **49** 594 – 607

Friedman A, Waller D, 2008 "View combination in scene recognition" *Memory & Cognition* **36** 467 – 478

Giese M A, Lappe M, 2002 "Measurements of the generalization fields for the recognition of biological motion" *Vision Research* **42** 1847 – 1858

Giese M A, Poggio T, 2003 "Neural mechanisms for the recognition of biological movements" *Nature Reviews Neuroscience* **4** 179 – 192

Giese M A, Thornton I, Edelman S, 2008 "Metrics of the perception of body movement" *Journal of Vision* **8** 1 – 18

Harman K, Humphrey G K, 1999 "Encoding 'regular' and 'random' sequences of views of novel three-dimensional objects" *Perception* **28** 601 – 615

Hayward W G, Tarr M J, 2000 "Differing views on views: comments on Biederman and Bar (1999)" *Vision Research* **40** 3895 – 3899

Hayward W G, Williams P, 2000 "Viewpoint dependence and object discriminability" *Psychological Science* **11** 7 – 12

Hummel J E, Biederman I, 1992 "Dynamic binding in a neural network for shape recognition" *Psychological Review* **99** 480 – 517

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

Knappmeyer B, Thornton I M, Bülthoff H H, 2003 "The use of facial motion and facial form during the processing of identity" *Vision Research* **43** 1921 – 1936

Lander K, Bruce V, 2000 "Recognizing famous faces: Exploring the benefits of facial motion" *Ecological Psychology* **12** 259 – 272

Lawson R, Bülthoff H H, 2006 "Comparing view sensitivity in shape discrimination with shape sensitivity in view discrimination" *Perception & Psychophysics* **68** 655 – 673

Lawson R, Bülthoff H H, 2008 "Using morphs of familiar objects to examine how shape discriminability influences view sensitivity" *Perception & Psychophysics* **70** 853 – 877

Lawson R, Bülthoff H H, Dumbell S, 2003 "Interactions between view changes and shape changes in picture – picture matching" *Perception* **32** 1465 – 1498

Lawson R, Humphreys G W, 1996 "View specificity in object processing: evidence from picture matching" *Journal of Experimental Psychology: Human Perception and Performance* **22** 395 – 416

Lawson R, Humphreys G W, Watson D G, 1994 "Object recognition under sequential viewing conditions: evidence for viewpoint-specific recognition procedures" *Perception* **23** 595 – 614

Liu T, 2007 "Learning sequence of views of three-dimensional objects: The effect of temporal coherence on object memory" *Perception* **36** 1320 – 1333

Liu T, Cooper L A, 2003 "Explicit and implicit memory for rotating objects" *Journal of Experimental Psychology: Learning Memory and Cognition* **29** 554 – 562

Loftus G R, Masson M E J, 1994 "Using confidence intervals in within-subjects designs" *Psychonomic Bulletin & Review* **1** 476 – 490

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* **200** 269 – 294

Newell F N, Setti A, 2006 "The effect of rigid and non-rigid motion on object recognition" *Perception* **35** Supplement, 184

Newell F N, Wallraven C, Huber S, 2004 "The role of characteristic motion in object categorization" *Journal of Vision* **4** 118 – 129

O'Toole A J, Roark D A, Abdi H, 2002 "Recognizing moving faces: A psychological and neural synthesis" *Trends in Cognitive Sciences* **6** 261 – 266

Peissig J J, Tarr M J, 2007 "Visual object recognition: Do we know more now than 20 years ago?" *Annual Review of Psychology* **58** 75 – 96

Poggio T, Edelman S, 1990 "A network that learns to recognize three-dimensional objects" *Nature* **343** 263 – 266

Schultz J, Chuang L, Vuong Q C, 2008 "A dynamic object-processing network: Metric shape discrimination of dynamic objects by activation of occipito-temporal parietal and frontal cortex" *Cerebral Cortex* **18** 1302 – 1313

Shepard R N, Cooper L A, 1982 *Mental Images and their Transformations* (Cambridge, MA: MIT Press)

Spetch M L, Friedman A, Vuong Q C, 2006 "Dynamic object recognition in pigeons and humans" *Learning & Behavior* **34** 215 – 228

Stone J V, 1998 "Object recognition using spatiotemporal signatures" *Vision Research* **38** 947 – 951

Stone J V, 1999 "Object recognition: View-specificity and motion-specificity" *Vision Research* **39** 4032 – 4044

Stone J V, Bray A, 1995 "A learning rule for extracting spatio-temporal invariances" *Network* **6** 1 – 8

Stone J V, Harper N, 1999 "Temporal constraints on visual learning: a computational model" *Perception* **28** 1089 – 1104

Tarr M J, 1995 "Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects" *Psychonomic Bulletin & Review* **2** 55 – 82

Tarr M J, Bülthoff H H, 1995 "Is human recognition better described by geon-structural descriptions or by multiple-views? Comments on Biederman and Gerhardstein (1993)" *Journal of Experimental Psychology: Human Perception and Performance* **21** 1494 – 1505

Tarr M J, Williams P, Hayward W G, Gauthier I, 1998 "Three-dimensional object recognition is viewpoint-dependent" *Nature Neuroscience* **1** 275 – 277

Tjan B S, Legge G E, 1998 "The viewpoint complexity of an object recognition task" *Vision Research* **38** 2335 – 2350

Todd J T, 2004 "The visual perception of 3D shape" *Trends in Cognitive Sciences* **8** 115 – 121

Ullman S, 1979 *The Interpretation of Visual Motion* (Cambridge, MA: MIT Press)

Ullman S, 1998 "Three-dimensional object recognition based on the combination of views" *Cognition* **67** 21 – 44

Vuong Q C, Tarr M J, 2004 "Rotation direction affects object recognition" *Vision Research* **44** 1717 – 1730

Vuong Q C, Tarr M J, 2006 "Structural similarity and spatiotemporal noise effects in learning dynamic novel objects" *Perception* **35** 497 – 510

Wallis G, 2002 "The role of object motion in forging long-term representations of objects" *Visual Cognition* **9** 233 – 247

Wallis G, Bülthoff H H, 2001 "Effects of temporal association on recognition memory" *Proceedings of the National Academy of Sciences of the USA* **98** 4800 – 4804

Watson T, Johnston A, Hill H C H, Troje N F, 2005 "Motion as a cue for viewpoint invariance" *Visual Cognition* **12** 1291 – 1308

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 38 2009

www.perceptionweb.com