The relative weight of shape and non-rigid motion cues in object perception: A model of the parameters underlying dynamic object discrimination

Quoc C. Vuong Institute of Neuroscience, Newcastle University, UK

Alinda Friedman Department of Psychology, University of Alberta, Canada

Jenny C. A. Read Institute of Neuroscience, Newcastle University, UK

Shape and motion are two dominant cues for object recognition, but it can be difficult to investigate their relative quantitative contribution to the recognition process. In the present study, we combined shape and non-rigid motion morphing to investigate the relative contributions of both types of cues to the discrimination of dynamic objects. In Experiment 1, we validated a novel parameter-based motion morphing technique using a single-part three-dimensional object. We then combined shape morphing with the novel motion morphing technique to pairs of multipart objects to create a joint shape and motion similarity space. In Experiment 2, participants were shown pairs of morphed objects from this space and responded "same" on the basis of motion-only, shape-only, or both cues. Both cue types influenced judgments: When responding to only one cue, the other cue could be ignored, although shape cues were more difficult to ignore. When responding on the basis of both cues, there was an overall bias to weight shape cues more than motion cues. Overall, our results suggest that shape influences discrimination more than motion even when both cue types have been made quantitatively equivalent in terms of their individual discriminability.

Keywords: relative weight, object perception, motion cue

Citation: Vuong, Q. C., Friedman, A., & Read, J. C. A. (2012). The relative weight of shape and non-rigid motion cues in object perception: A model of the parameters underlying dynamic object discrimination. *Journal of Vision, 12*(3):16, 1–20, http://www.journalofvision.org/content/12/3/16, doi:10.1167/12.3.16.

Introduction

Motion is important for visual perception and object recognition in both human and non-human animals (Gibson, 1979; Johansson, 1973; Lettvin, Maturana, McCulloch, & Pitts, 1959; Tinbergen, 1951; Vernon, 1952). For instance, the "slithering of a snake" conjures up an image of the bending, twisting, and stretching of a serpentine body; the "flittering of a butterfly" conjures up a very different image. Clearly, there are different types of motion in the environment (see Aggarwal, Ciao, Liao, & Sabata, 1998, for a discussion of different types of motion), but only some types of motion have been systematically investigated with respect to object recognition. Moreover, few studies to date have systematically examined the relative contribution of shape and motion cues to recognition and how these cues interact (e.g., Lander & Bruce, 2000; Newell, Wallraven, & Huber, 2004; Pilz, Thornton, & Bulthoff, 2006; Stone, 1998; Vuong & Tarr, 2006). The evidence from these studies further suggests that observers rely predominantly on shape cues for everyday recognition. Thus, how motion is represented more generally in the service of object recognition remains unclear. To test the relative contribution of shape and motion cues to recognition, we created novel "dancing" objects for which we could independently morph their shape and their motion, creating a two-dimensional shape/motion space. Morphing thus allowed us to tightly control the motion and shape similarity between two objects. We then tested observers' ability to discriminate morphed objects on the basis of shape and motion cues either singly or in combination and modeled their performance. In this way, we measured how motion and shape interacted with each other to contribute to the recognition process.

Most studies of the role of motion in object recognition in human (e.g., Liu & Cooper, 2003; Newell et al., 2004; Schultz, Chuang, & Vuong, 2008; Stone, 1998; Vuong & Tarr, 2006) and non-human animals (e.g., Cook & Katz, 1999; Friedman, Vuong, & Spetch, 2009; Spetch, Friedman, & Vuong, 2006) investigated rigid motion of novel three-dimensional (3D) objects, such as translations and rotations. Some studies have also investigated semi-rigid, articulatory motion (e.g., Bassili, 1978; Jastorff, Kourtzi, & Giese, 2006; Johansson, 1973; Kellman, 1993; Pyles, Garcia, Hoffman, & Grossman, 2007; Setti & Newell, 2010; Vuong, Friedman, & Plante, 2009). This is the type of motion produced by humans and other animals whose "parts" move at their joints (e.g., walking or galloping). Other studies have focused on highly

doi: 10.1167/12.3.16 Received August 20, 2011; published March 16, 2012 ISSN 1534-7362

familiar facial motions (e.g., expressions, speech; Hill & Johnston, 2001; Knappmeyer, Thornton, & Bülthoff, 2003; Lander & Bruce, 2000; Pilz, Bülthoff, & Vuong, 2009; Pilz et al., 2006; Watson, Hill, Johnston, & Troje, 2005).

Aggarwal et al. (1998) classified articulations (e.g., body movements) and deformations (e.g., facial motions) as non-rigid motion. It is important to study non-rigid motion for at least three reasons. First and most critically, non-rigid motion changes the 3D shape of an object. This change poses a strong challenge to existing theories of object recognition that are based on shape, irrespective of their assumptions about the underlying representation of shape (i.e., as image-dependent views or as structural descriptions; Biederman, 1987; Tarr, 1995). Second, non-rigid motion may provide a source of information about object identity, as in the snake vs. butterfly example given above. Thus, both shape and motion may contribute to object identity independently or interactively. Third, studies of facial and body motion suggest that observers are highly sensitive to non-rigid and semi-rigid motion even in the absence of shape cues (Bassili, 1978; Johansson, 1973). However, both faces and bodies are highly familiar stimuli, so the extent to which non-rigid motion may be used more generally to recognize unfamiliar stimuli is unclear. Furthermore, because we ourselves make facial and body movements, there may be implicit motor influences on the visual perception of facial and body movements (e.g., Casile & Giese, 2006). Although articulations of unfamiliar objects have been studied (e.g., Jastorff et al., 2006; Pyles et al., 2007; Setti & Newell, 2010; Vuong et al., 2009), few studies to date have looked at non-rigid deformations of novel objects (e.g., Chuang, Vuong, Thornton, & Bülthoff, 2006; Mayer & Vuong, 2012). Thus, in the present study, we focus on this type of non-rigid motion.

However, it is important to bear in mind that in most situations, there is a shape bias in object recognition because shape is typically the most diagnostic cue for many everyday situations (Biederman, 1987). The contribution of motion to the recognition process becomes evident in performance when shapes are visually similar, somehow degraded, or ambiguous (e.g., Johansson, 1973; Knappmeyer et al., 2003; Lander & Bruce, 2000; Liu & Cooper, 2003; Spetch et al., 2006; Stone, 1998; Vuong & Tarr, 2006). Thus, having a means to quantify the relative contribution of shape and motion cues across changes in shape and motion similarity is important for understanding both the shape bias and how both cues interact. We therefore modeled observers' performance when they discriminated objects that underwent both shape and motion changes and quantified the relative weight that observers assigned to each cue. In addition, we used the model fits to generate *discrimination contours* through a joint shape and motion space to visualize the relative contribution of both types of cues to the task.

Combining shape and motion cues

The physical structure of an object can provide biomechanical constraints on the type of motion that is possible for it to make. For example, the bending of a snake is linked to its cylindrical structure and musculature. Similarly, the extent of facial deformations is constrained by the underlying skull and facial muscles. These constraints further highlight the importance of shape and motion interactions for recognition. In previous studies, several groups have contrasted how shape and motion interacted for both novel articulated motion and for familiar biological motion. The differences between the results of these studies illustrate how shape and motion can interact. For example, Jastorff et al. (2006; see also Kellman, 1993) used simple sinusoidal motion of points that were attached or not attached to an "invisible" underlying novel skeleton to test the extent to which global shape could constrain motion recognition. When the points were attached to a skeleton, these artificial point-light movements mimicked point-light displays of human actions (Johansson, 1973). Jastorff et al. found that observers' performance on a categorization task with the skeleton version of the novel motion was comparable to their performance with point-light human motion. In contrast, observers performed poorly on the task when the novel motion had no underlying skeleton.

In another study, Pyles et al. (2007) used "creatures" built from connected rod-like parts and endowed them with a nervous system. They simulated the motion of these creatures through an environment to generate animal-like biological motion and, unlike Jastorff et al. (2006), embedded both the creatures and the human point-light displays in noise points. The creatures had different shapes than human bodies. Pyles et al. found that observers were much better able to segregate point-light human actions compared to point-light creature actions when each was embedded in noise. The authors suggested that observers could use the highly familiar human shape to group signal points together (i.e., those points that form part of the global human shape) and use the grouped display to detect the human point-light displays. They could not group the points together for the novel shapes, even though the actions of the grouped points were coherent.

The studies by Jastorff et al. (2006) and Pyles et al. (2007) highlight a strong interaction between shape and motion. These studies focused on articulated motion; other studies have examined how global object motion and local articulatory part motion may influence object recognition (Setti & Newell, 2010) or how local articulatory part motion could facilitate generalization to novel viewpoints (Vuong et al., 2009). As mentioned in the Introduction section, however, few studies have tested the role of deformation in object recognition more generally (Chuang et al., 2006). More critically, to date, no studies have parametrically manipulated *both* shape and motion. As we

and others have found (e.g., Cutzu & Edelman, 1996; Giese, Thornton, & Edelman, 2008; Jastorff et al., 2006; Lawson & Bülthoff, 2008; Pyles et al., 2007; Schultz et al., 2008; Vuong et al., 2009), perceptual similarity in the shape or motion domain can affect how well observers recognize objects.

Parameter-based morphing in the shape and motion domain

The main challenge in examining dynamic object recognition is to find a means to systematically manipulate both shape and motion to create a joint parametric space in both dimensions. One efficient way to both represent and synthesize complex classes of static 3D objects is by the *linear combination of prototypes* (Giese & Poggio, 2000; see Ullman, 1998, for a linear combination of 2D views to represent a 3D object). The prototypes serve as stored examples of an object class for which a 3D description is available. In the simple case, this description is the 3D position (i.e., *x*-, *y*-, and *z*-coordinates) of the set of vertices that define the 3D object model. New object models are synthesized by taking a linear combination of the prototypes, that is, by taking a weighted average of the 3D position at each corresponding vertex between prototypes. This linear combination is also referred to as *morphing*. On this view, given a set of prototypes ($P_1 \ldots P_n$), a morph, M, can be created by the linear combination: $M = c_1P_1 + \ldots + c_nP_n$, where the weight, c, represents the contribution of each prototype to the morph and with the constraint that $c_1 + \ldots + c_n = 1$. Giese and Poggio (2000) extended the linear combination of prototypes to human actions (e.g., walking, running, marching), which added a temporal dimension to shape. In their spatiotemporal motion morphing technique, prototype actions were represented by trajectories of key parts and joints of the human body (e.g., forehead, shoulders, elbows, wrists, hips, pelvis, knees, and feet).

Prototypes can be used to define a multidimensional space to represent objects, which has been a useful concept to test different aspects of object recognition (Edelman, 1999). Any object can be represented by its weight vector ($c_1, \ldots, c_n$) within the space spanned by the prototypes. That is, the prototypes define the physical dimensions of the space and a weight vector defines a point (i.e., object) within it. The similarity between any two objects can then be defined as the Euclidean distance between their weight vectors on the specific dimensions that are represented. Using such morphing techniques, behavioral experiments have shown that similarity in the parametric space maps to perceived similarity and is important for behavior in both the shape and motion domains (e.g., Cutzu & Edelman, 1996; Giese et al., 2008; Jastorff et al., 2006; Lawson & Bülthoff, 2008; Schultz et al., 2008; Troje, 2002). Thus, the linear combination framework allows researchers to create perceptually meaningful spaces, that is, observers are sensitive to systematic variations in this type of space.

Modeling the weighting of shape and motion cues

The parameter-based morphing technique we used allowed us to systematically manipulate both shape and motion cues in a similar way. We next modeled how these types of cues were combined when observers discriminated pairs of dynamic objects under the different task constraints (i.e., discrimination based on motion alone, shape alone, or motion and shape). In line with Bayesian cue combination models (e.g., Landy, Maloney, Johnston, & Young, 1995), our general assumption is that observers independently estimate the motion and shape of each object presented, take a weighted combination of these estimates, and then use this sum to discriminate between the objects (i.e., determine whether they are the same or different). We quantify this according to the equation below:

$$C^2 = w^2S^2 + (1 - w^2)M^2, \quad (1)$$

where C is the combined difference estimate between two shapes and motions, S is the shape difference estimate, M is the motion difference estimate, and w is the relative weighting of shape and motion differences. The objects are considered different if the sum is greater than some decision threshold θ. The relative weight, w, ranges from 0 to 1: As observers rely more and more on the shape cue, w^2 approaches 1; conversely, as they rely progressively more on the motion cue, w^2 approaches 0; and if they use both cues equally, $w^2 = 0.5$. For convenience, we also define $w_s = \sqrt{w^2} = w$ (the relative weight assigned to shape differences) and $w_m = \sqrt{(1 - w^2)}$ (the relative weight assigned to motion differences).

To derive this model, each object can be characterized by a motion and a shape signal. We assume that an observer can extract these signals but that the signals are subject to some noise. We further assume that this noise is Gaussian and applies independently to the motion and shape signals, with different standard deviations affecting each signal. On a given trial, observers are shown two objects. Therefore, we can collapse the two motion signals from each object into a single "motion difference" signal. Similarly, we can collapse the two shape signals into a single "shape difference" signal. Ideally, we can model an observer's proportion of "different" responses as a function of the motion and shape (difference) signals with the following parameters: θ that determines the amount of difference required for the observer to give a "different" response (i.e., the observer's decision threshold), σ_m and σ_s that determine the reliability of the motion and shape signals (i.e., the variance or noise in the estimates M and S), respectively, and w that determines the relative weight

assigned to the motion and shape signals. Thus, our full model has four parameters (θ, σ_m, σ_s, and w). The Deriving a model of shape and motion discrimination section in Appendix A provides a derivation of our model (Equations A4 to A9).

To summarize, we used a novel parameter-based morphing technique described more fully in the Stimuli section, which allowed us to parametrically manipulate both motion and shape information. We then tested and modeled observers' ability to discriminate the motion of a single-part brick-like object (Experiment 1) and their ability to discriminate the motion, the shape, or the shape motion of different multipart objects (Experiment 2).

Experiment 1

The purpose of Experiment 1 was to validate the parameter-based morphing technique for motion discrimination described in the Stimuli section. Following previous work (e.g., Cutzu & Edelman, 1996; Giese et al., 2008; Schultz et al., 2008), we used a *same–different* discrimination task. We expected that if participants were sensitive to the parametric motion space spanned by the three prototype motions we used, their ability to discriminate morphs within this space would be monotonically related to the *motion distance* between two prototypes (Giese et al., 2008). Furthermore, we designed the experiment to reduce reliance on image changes *per se* by changing the viewpoint at which each of the two objects was presented.

Participants

Twenty volunteers from the University of Alberta undergraduate pool (11 females) participated in this study for course credit. All participants provided informed consent and were naive to the purpose of the experiment.

Stimuli

We started with three readily recognizable non-rigid motions: *bending, twisting,* and *stretching* (recall the snake). To relate these motions to the linear combination framework, we refer to them as the "prototype" non-rigid motions (see below). These different motions can be considered to be global *deformation fields* that smoothly warp the 3D position of all the vertices on the surface of a 3D model (Barr, 1984; Watt & Watt, 1992). By smooth deformations, we mean that there are no sharp changes or discontinuities to the resulting 3D geometry. These deformations are illustrated in Figure 1. For example, a single- or multipart object (Figure 1, left and right panels, top row) can be bent, twisted, or stretched (Figure 1, left and right panels, middle row). In Experiment 1, we used the single-part object illustrated in the left panel, and in Experiment 2, we used the multipart object shown in the right panel; however, their creation and morphing were identical so we describe the stimulus creation for both experiments together. Prototype motions can be combined and the resulting deformation field mapped to the object. Thus, the technique of using deformation fields to morph motions on any shape offers tremendous freedom because it is not necessary to find either corresponding points on the surfaces of two different 3D shapes or corresponding points in time.

In our study, non-rigid motions are defined by five time-varying parameters; these *temporal profiles* are illustrated in Figure 2. The five parameters are: bend angle (0° [straight] to 180° [fully bent]), bend direction (arbitrary range in degrees, 0° to 270° used in the present study), twist angle (−90° to +90°), twist bias (arbitrary range in degrees, 7° to 15° used in the present study), and stretch amount (−1 to 1, in arbitrary units). The bend direction and twist bias affect the initial direction of bending or twisting relative to some arbitrary starting position (0°), respectively. For the stretch amount, positive values stretch the shape longer (see Figure 1) and negative values compress the shape shorter (not illustrated).

From these five parameters, we first created the deformation field for each prototype motion. For example, a prototype bending is characterized by temporal profiles in which bend angle and direction vary over time, while the temporal profiles for the remaining parameters are "flat" and thus do not vary over time (left column in Figure 2). There were 101 time points (frames) in each profile. The initial temporal profiles for the prototype motions were based on stimuli used by Mayer and Vuong (2012). These profiles were created by adding "key frames" at different arbitrary time points and setting the values for each of the five parameters within each key frame. For example, to create a prototype bending motion, key frames could be added on frames 1, 30, 67, and 101, and then setting the bend angle and bend direction at each key frame. The software then smoothly interpolates the parameter values between the key frames to provide values at all 101 frames, as illustrated in Figure 2. The only constraint we made was that the parameter value for the first and last frames was the same so that the motion was cyclic and could be played in an endless loop without any temporal discontinuities (see open circles in Figure 2). Thus, the 3D shape of an object underwent smooth deformations over time based on these parameters.

We then morphed the deformation fields between two prototype motions by taking a weighted average at each time point separately for each temporal profile (which defined the deformation field), as shown in Figure 3. Note that the only difference between the prototype motions

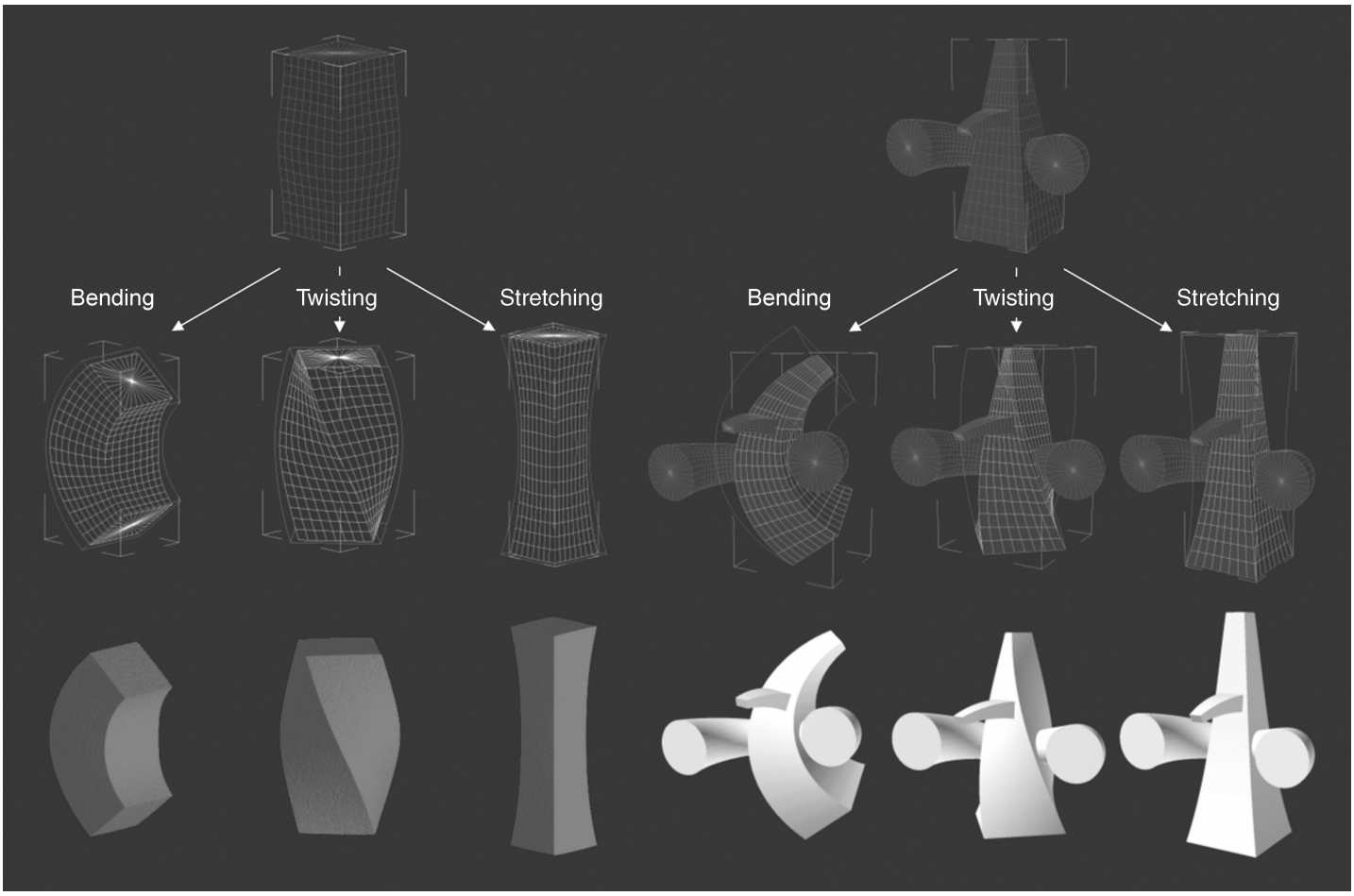


Figure 1. We conceptualize a non-rigid motion as a deformation field that smoothly deforms the 3D geometry of an object (i.e., the 3D position of each point on its surface). The underlying object geometry is illustrated as a rectangular mesh so that the effect of the deformation field can be visualized. The same deformation fields can be mapped to (left) single-part or (right) multipart objects. The bottom row shows a fully rendered "snapshot" of the prototype deformations. In Experiment 1, we used the single-part object, and in Experiment 2, we used multipart objects.

(i.e., bending, twisting, and stretching) and the morphed motions (e.g., some linear combination of two prototypes) is that prototypes only have varying temporal profiles for the parameters that define the pure motion (see Figure 2). In contrast, morphs have varying temporal profiles for the parameters that define different proportions of two of the pure motions. For the current study, we created morphs between all pairwise combinations of the three prototype motions (bending–twisting, bending–stretching, and twisting–stretching, with 0% representing the first prototype in the pair). For each pair, we created morphs from one prototype in the pair to the other prototype in 5% steps. Thus, for example, a 20% morph between bending and twisting would contain 80% of the bending parameters and 20% of the twisting parameters. There were 19 morphs between the two prototypes for each morph pair.

In Experiment 1, we mapped the 60 (=19 morphs × 3 morph pairs + 3 prototypes) unique non-rigid motions to a slightly bulging box (see Figure 1, left panel), thereby smoothly deforming its geometry over time. The animated object was rendered with a pinkish, bumpy texture and centered on a black background. In its initial starting position, the object subtended 6.4° (width) × 9.1° (height) of visual angle. In Experiment 2, these motion morphs were mapped to different multipart shapes. All animations were rendered as AVI videos for three animation cycles (25 frames/s). Each video was approximately 9 s in length (3 s/cycle).

Design and procedure

Participants performed a same–different motion discrimination task on two simultaneously presented videos of non-rigidly moving objects. They were given the following instructions, which emphasized accurate motion

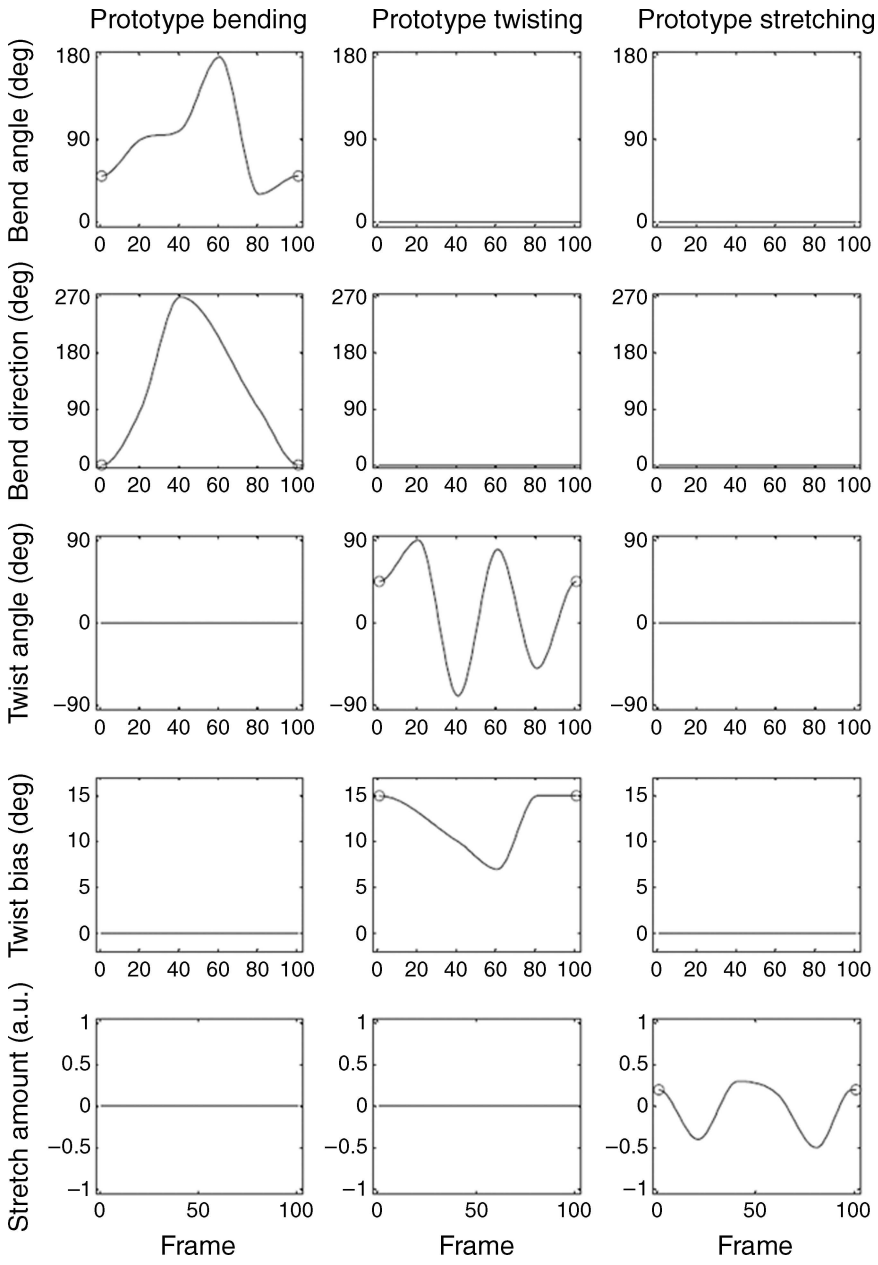
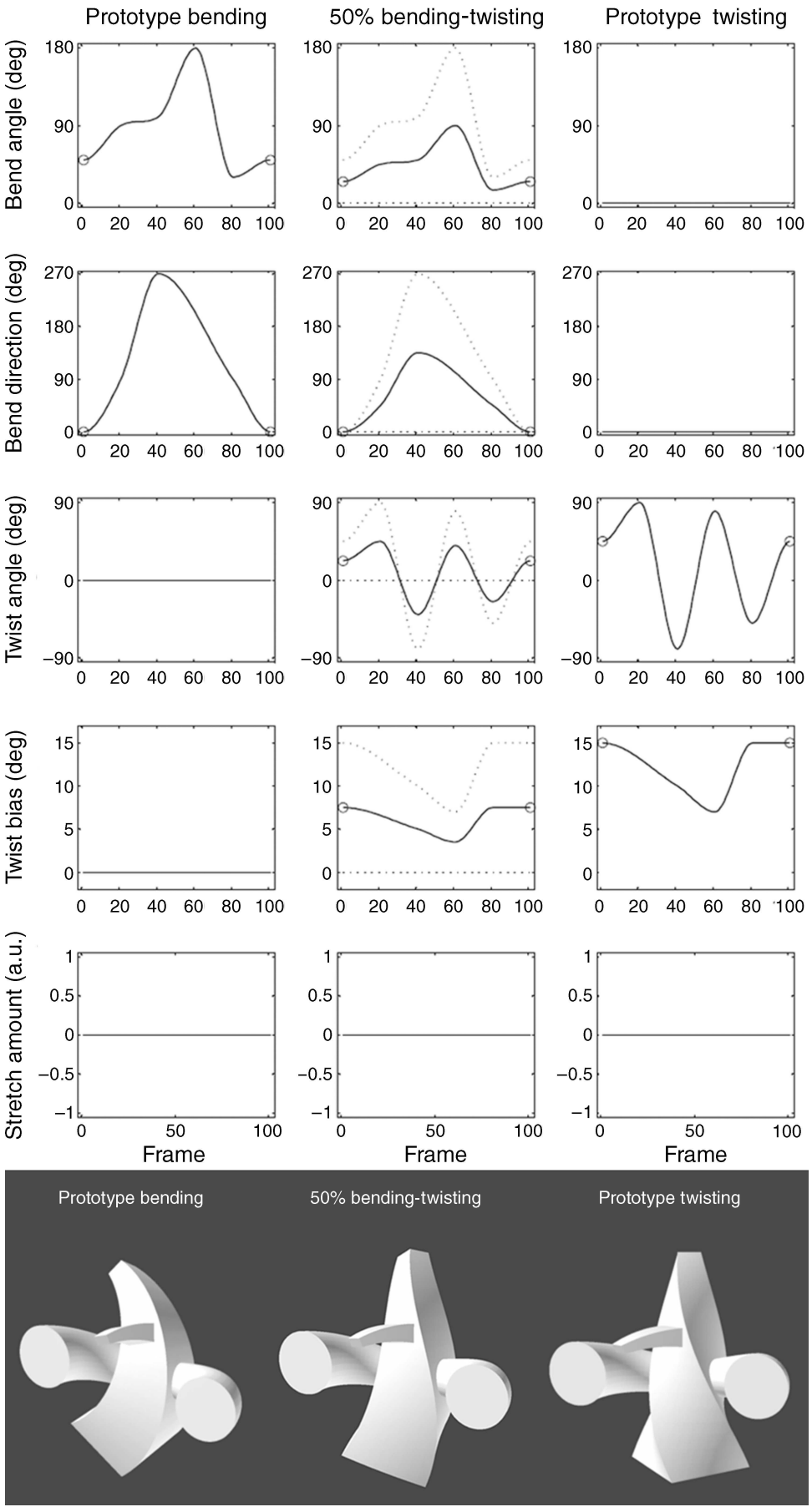


Figure 2. The five temporal profiles (rows: Bend Angle, Bend Direction, Twist Angle, Twist Bias, and Stretch Amount) for the three prototype motions (columns: Bending, Twisting, and Stretching) are illustrated. Each plot illustrates the variation of the parameter value (*y*-axis) over time (*x*-axis; frame). Note that the Bend Angle and Bend Direction vary over time for the prototype bending; the Twist Angle and Twist Bias vary over time for the prototype twisting; and the Stretch Amount varies over time for the prototype stretching. The circles indicate that the first and last frames of temporal profile have the same value so that the animation can be looped.

discrimination: "In this experiment, your task is to decide as accurately as possible whether two objects have the same movements. Both objects have the same shape, but they may differ in how they move around. This difference in motion may be subtle so try your best to make this judgment. There is no time pressure." Participants were tested with three different morph pairs as described above. For each morph pair, there were six motion differences between the morphs of the two videos (0% [same], 10%, 20%, 30%, 40%, and 50%). There were 12 trials in each of the 3 $\times$ 6 conditions for a total of 216 trials. All the trials were run in a random order for each participant.

For each motion difference, we first randomly selected a morph for one of the videos. For the second video, we selected the morph that provided the corresponding motion difference. For example, suppose we randomly select a video that was a 35% morph between prototypes A and B. In this case, for a 10% motion difference, we would then


Prototype bending
50% bending-twisting
Prototype twisting
Bend angle (deg)
Bend direction (deg)
Twist angle (deg)
Twist bias (deg)
Stretch amount (a.u.)
Frame
Prototype bending
50% bending-twisting
Prototype twisting

select a 25% morph between A and B (or equally possible, a 45% morph between A and B). For a 0% motion difference (i.e., same object), we would select the same 35% morph between A and B. We randomly selected the stimuli with the constraint that each participant saw the entire morph continuum for each morph pair (from 0% to 100% in 5% increments). That is, each participant saw the entire motion morph space.

To avoid low-level image matching, we presented the objects from slightly different viewpoints in each rendered video. The virtual camera was positioned at approximately 19° elevation and 120° azimuth (camera distance = 160 arbitrary units) on the viewing sphere for one viewpoint and approximately at 30° elevation and 150° azimuth for the second viewpoint (camera distance = 165 arbitrary units). Thus, even when the motions were identical, there were still image differences between the two videos. In this way, observers could not use image differences *per se* to perform the task. The location (left or right of fixation) and viewpoint of the two selected videos were randomly determined on each trial.

Each trial began with a white fixation cross shown for 500 ms at the center of the screen. After the fixation cross disappeared, two stimuli appeared simultaneously side by side. One video was shifted 9.3° to the left of fixation, and the other video was shifted by the same amount to the right. The videos were shown for three complete cycles (9 s). Both videos always began at the beginning of a cycle. Participants could respond at any time following the presentation of the stimuli. If they did not respond within 9 s, the stimuli were removed from the screen. However, participants still had to respond before advancing to the next trial.

Participants pressed one button on a response box if they thought the two stimuli had the same movement and a different button if they thought the two stimuli had different movements. No feedback was provided. The response mapping was counterbalanced across participants.

Prior to beginning the experiment, participants were given 18 practice trials in which feedback was provided (a brief tone for incorrect responses). Each of the 18 conditions was presented once on practice trials. The experiment lasted approximately 30–40 min. The experiment used E-Prime (PST Software 2002) to present the videos and control the experiment. Participants sat approximately 68 cm from a Samsung SyncMaster 940BF monitor (1024 × 768 pixel resolution; 60-Hz refresh rate; 2-ms gray-to-gray time).

Figure 3. A 50% morph (middle column) between the prototype bending motion (left column) and prototype twisting motion (right column). The prototype motions are the same as in Figure 2. For each parameter, we take the average at each time point to create the morph (solid lines). The parameter values of the two prototypes are superimposed as dotted lines in the plots of the middle column. The bottom part of the figure represents the three different deformations mapped to a multipart object. We used the multipart objects from Experiment 2 to clearly illustrate the effects of the morphing technique (although Experiment 1 used the single-part object). Examples of the dancing multipart objects can be found at www.staff.ncl.ac.uk/q.c.vuong/VuongFriedmanRead.html.

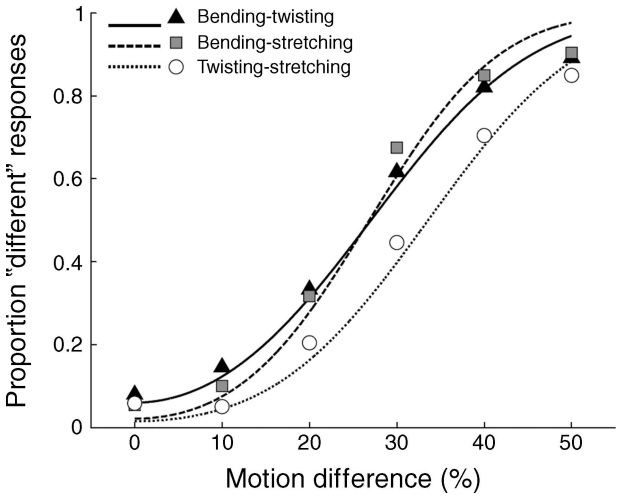


Figure 4. The proportion of "different" responses as a function of the morph pair and motion difference in Experiment 1. The symbols represent the data for each morph pair averaged across participants. The curves show the average fit of the model to the data for each morph pair.

Results and discussion

We were mainly interested in how likely participants responded "different" as a function of the monotonic objective motion difference between two videos. We therefore tested for significant linear trends in an analysis of variance (ANOVA). We also fitted psychometric functions to the data to estimate participants' 75% motion discrimination thresholds (see Psychometric functions section).

Proportion "different" responses

Figure 4 (symbols) shows the mean proportion of "different" responses, averaged across participants in each group. We submitted these data to a 3 (morph pair types) × 6 (motion difference) repeated measures design. As expected, there was a significant linear trend in the data ($F(1,19) = 461.1$, $\eta_p^2 = 0.60$), which suggests that participants' responses were increasing as a function of motion difference. There was also a main effect of morph pair ($F(2,38) = 16.1$, $\eta_p^2 = 0.45$). Lastly, there was a significant interaction between the two factors ($F(10,190) = 3.65$, $\eta_p^2 = 0.16$). Post-hoc comparison using Tukey's Honestly Significant Difference (with the within-subjects error term from the significant interaction) showed that the

twisting–stretching pair was different from both the bending–twisting and bending–stretching pairs ($ps < 0.05$). However, qualitatively, the curves are the same.

Psychometric functions

We next fitted the model described in the Modeling the weighting of shape and motion cues section to each participant's data for each morph pair. Participants in Experiment 1 only discriminated motion so we could set $w_s = 0$ (i.e., all the weight is assigned to motion differences), making the value of σ_s immaterial. We could then fit the two remaining parameters (θ and σ_m) with Equation A9 using maximum likelihood fitting. The Maximum likelihood fitting section in Appendix A describes the maximum likelihood procedure we used. We assessed the relative fit for each participant's data across the different morph pairs using the root mean square error (RMSE) between the actual data and the predicted data (from the participant's individual fitted parameters). The RMSE was similar across the three morph pairs (bending–twisting: $M = 7.6\%$, $SE = 0.6\%$; bending–stretching: $M = 6.5\%$, $SE = 0.8\%$; and twisting–stretching: $= 7.6\%$, $SE = 0.6\%$). We then averaged the parameters across participants to compute a population psychometric function for each morph pair. These functions are illustrated in Figure 4 (see Figure A2 in Appendix A for fits for each participant).

From the psychometric functions, we computed a 75% motion threshold to qualitatively compare the current results with our work on shape discrimination using a similar parameter-based approach (Schultz et al., 2008). To make this comparison, we collapsed across the three morph pairs to yield 36 trials for each motion difference per participant. We found that the mean motion threshold was 35.1% ($SEM = 1.8\%$). This threshold for motion discrimination was similar to the threshold obtained in our previous study using shape morphs of multipart objects (Schultz et al., 2008). In that study, we found that the 75% shape threshold was 41.3% ($SEM = 2.0\%$; $N = 15$).

Experiment 2

Having established the validity of a linear combination approach to morphing between non-rigid motion prototypes that used deformation fields, in Experiment 2 we used multipart objects and simultaneously combined motion and shape morphing. This allowed us to compare performance on three possible tasks. One group of observers were required to discriminate shape, another motion, and a third group had to discriminate both shape and motion. Different groups of observers were tested in each task to avoid practice effects. We then used predictions from the model to determine the relative weightings of shape and motion cues when they were judged either alone or together, as well as to determine whether one type of cue influenced discriminations made on the other type of cue when the first cue was irrelevant.

Participants

Sixty volunteers from the University of Alberta undergraduate pool (36 females; 21 males; missing sex information for 3 volunteers) participated in this study for course credit. All participants provided informed consent and were naive to the purpose of the experiment. An equal number of participants were randomly assigned to the three discrimination tasks (*shape only, motion only,* and *shape + motion*).

Stimuli

The stimuli consisted of multipart objects similar to the ones we have used previously (Schultz et al., 2008). Briefly, each object was defined by a large body, a small central part (defining the front of the object), and two small lateral parts (see Figures 1 and 3). Each part was defined by three shape parameters: the shape of its cross section (from circle to square), the amount of bending, and the amount of tapering. There were four shape prototypes defined based on fixed values of these 12 parameters (3 shape parameters × 4 parts). We arbitrarily paired two of the shape prototypes forming two morph pairs. For each pair, we then morphed between all parameters along an "identity vector" (see Schultz et al., 2008, for more details). For the motion of these multipart objects, we only used the bending and twisting prototype motions from Experiment 1. The multipart objects were rendered with a matte gray surface (see Figures 1 and 3) and subtended approximately 10.1° (width) × 9.0° (height) of visual angle.

Design and procedure

The design and procedure used in Experiment 2 was similar to those used in the Design and procedure section of Experiment 1. Participants performed a same–different discrimination task on two simultaneously presented videos of non-rigidly moving multipart objects. Participants in the shape-only group discriminated only the shape of the objects while ignoring the motion (which could be the same or different), those in the motion-only group discriminated the motion while ignoring the shape of the objects (which also could be same or different), and those in the shape + motion group made their discriminations based on both the shape and motion of the objects. That is, participants in the last group responded *same* if and only if both stimuli had the same shape and same motion (disregarding changes in viewpoint).

All participants were given instructions appropriate for their task condition, which emphasized accurate discrimination. For example, in the shape + motion condition: "Your task is to decide as accurately as possible whether the two objects have the same three-dimensional shape and motion." In addition, all three groups were informed that: "Because of the motion, the objects' shape will become distorted but the objects will maintain their identity. Imagine your face making different exaggerated expressions. Although the shape changes, it's still you making the change. The differences in the shape/motion/shape and motion may be subtle so try your best to make this judgment. There is no time pressure." There were five shape differences and five motion differences between morphs of the two videos. Based on the results of Experiment 1 and from the results of Schultz et al. (2008), we used the same range of shape/motion differences (0% [same], 10%, 20%, 30%, and 40%). Within each group, there were 10 trials in each of the 5 × 5 conditions for a total of 250 trials. All the trials were run in a random order for each participant. For each shape and motion difference, we used a similar procedure as in the Design and procedure section to select the two stimuli. Lastly, we presented the two dynamic objects from two different viewpoints to avoid low-level image matching. Prior to beginning the experiment, participants were given 25 practice trials (one per condition) in which feedback was provided (a brief tone for incorrect responses). Each of the 25 conditions was presented once on practice trials. Other than these changes, all other procedural aspects were the same as in the Design and procedure section.

Results and discussion

For each group, we first submitted the proportion "different" responses to a 5 shape differences × 5 motion differences ANOVA to investigate the overall pattern of effects. As in Experiment 1, we focused on linear trends for the shape and motion differences.

Proportion "different" responses

Figure 5 shows the mean proportion "different" responses for the three groups. For the shape-only (Figure 5a) and motion-only (Figure 5b) groups, the *x*-axis represents the task-relevant cue. There was a significant linear effect of the task-relevant cue (shape only: $F(1,19) = 405.4$, $\eta_p^2 = 0.96$; motion only: $F(1,19) = 92.2$, $\eta_p^2 = 0.83$). From the point of view of interference from

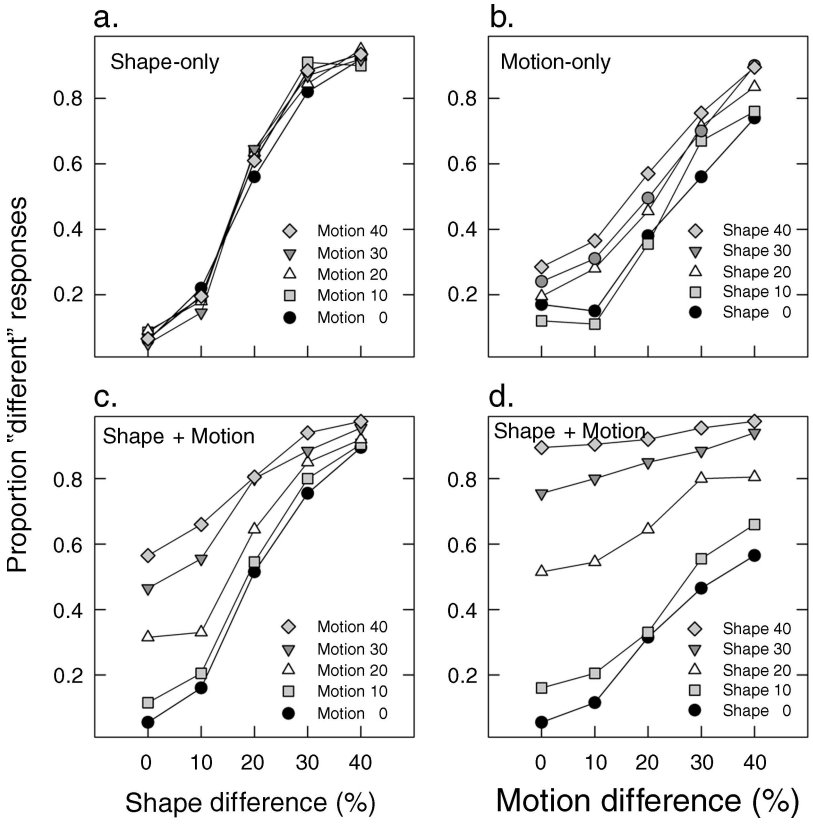


Figure 5. The proportion of "different" responses as a function of the shape and motion difference in Experiment 2, averaged across participants in each group. (a) Results for the shape-only group. (b) Results for the motion-only group. (c) Results for the shape + motion group. In this plot, the shape difference is plotted on the *x*-axis. (d) Results for the shape + motion group. The plot shows the same data as (c) but with motion difference on the *x*-axis.

the task-irrelevant cue, for the motion-only group there was a small but significant linear effect of the task-irrelevant shape ($F(1,19) = 5.0$, $\eta_p^2 = 0.21$). Otherwise, there was generally no effect of the task-irrelevant cue or interactions with it for either group. In other words, participants were generally able to make their discriminations based on only the shape or motion cue and could ignore the task-irrelevant cue, although participants in the motion-only group were slightly affected by shape differences.

Figures 5c and 5d show the mean proportion "different" responses for the shape + motion group. The responses are plotted with respect to the shape difference (Figure 5c) or the motion difference (Figure 5d) on the *x*-axis; otherwise, the data are identical in the two panels. In contrast to the previous results, there were significant and large effects of shape and motion. For the main effects of shape and of motion, the linear effects were significant (shape: $F(1,19) = 117.1$, $\eta_p^2 = 0.86$; motion: $F(1,19) = 31.4$, $\eta_p^2 = 0.62$). There was also a significant interaction between shape and motion ($F(16,304) = 8.32$, $\eta_p^2 = 0.31$). In Figure 5c, it can be seen that when shapes were similar, the motion cues facilitated responses, but when shapes were very dissimilar at the largest shape difference (40%), the motion cues had little effect. By comparison, it is easy to see in Figure 5d that shape still has an effect on discrimination even at the largest motion difference tested (40%).

Quantifying the weights assigned to shape and motion cues

The results of the ANOVAs in the Proportion "different" responses section suggest that participants could generally attend to the task-relevant cues and that shape differences contributed more than the motion differences to the recognition process. Based on previous findings (e.g., Spetch et al., 2006; Vuong & Tarr, 2006), we further predicted relatively independent processing of shape and motion differences for the shape-only and motion-only groups. We also predicted a larger weighting and/or more reliable estimation of shape than motion cues in the shape + motion group.

To examine how participants behaved, we plotted their "iso-probability contours," along which the proportion of "different" responses remains constant, on the axes of motion and shape difference (Figure 6). The solid iso-probability contours mark $P_{diff} = 50\%$. In the area below

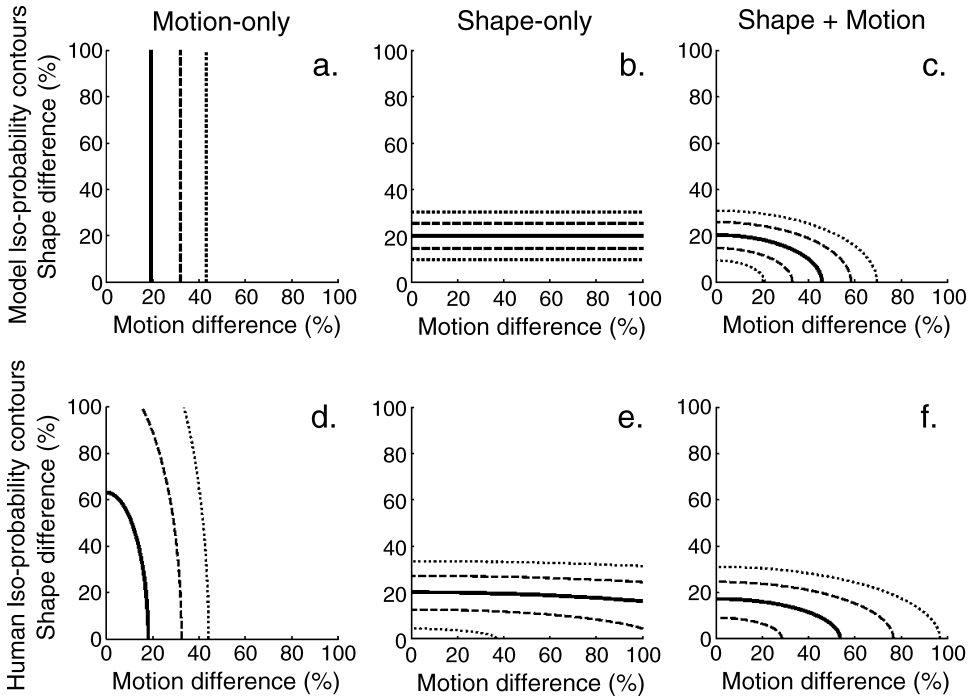


Figure 6. Model predictions of the (a–c) iso-probability contours for the three different discrimination tasks and the (d–f) fitted iso-probability contours averaged across participants in each task. In all plots, the solid lines mark $P_{diff} = 50\%$, dashed lines mark $P_{diff} = 25\%$ and 75%, and dotted lines mark $P_{diff} = 10\%$ and 90%. For the model observer, we used the same decision threshold and level of stimulus noise for all three tasks ($\theta = 20\%$, $\sigma_s = 8\%$, and $\sigma_m = 18\%$) and varied only the relative weights that the model observer assigned to the shape and motion cues for each task. We also assumed that motion cues are less reliable than shape cues ($\sigma_m > \sigma_s$). In (a), $w_s = 0.0$ and $w_m = 1.0$, i.e., shape differences are entirely discounted. In (b), $w_s = 1.0$ and $w_m = 0.0$, i.e., motion differences are entirely discounted. In (c), $w_s \propto 1/\sigma_s$ and $w_m \propto 1/\sigma_m$, i.e., the weights reflect the reliability of each cue. The fitted iso-probability contours from the data are similar to the model predictions for participants in the (d) shape-only group, (e) motion-only group, and (f) shape + motion group.

or to the left of this contour, stimuli are more likely to be judged "same"; in the area above or to its right, they are more likely to be judged "different." The dashed contours mark $P_{diff} = 25\%$ and $P_{diff} = 75\%$, and the dotted contours mark $P_{diff} = 10\%$ and $P_{diff} = 90\%$.

It is helpful to examine the behavior of an idealized model (Figures 6a–6c) before we turn to the behavior of the human participants (Figures 6d–6f). For this model observer, we assume that the parameters σ_m and σ_s are independent of the task, since they reflect a constant internal encoding noise. In Figures 6a–6c, we arbitrarily set $\sigma_m = 18\%$ and $\sigma_s = 8\%$. We further assumed that the decision boundary θ remained constant (at 20%). Differences in the model observer's performance between tasks therefore solely reflect changes in the weights it assigns to the different cues. In the shape-only and motion-only tasks, participants were instructed to completely discount the irrelevant cue. We therefore assumed that our idealized model observer achieves this perfectly, that is, the model observer sets $w_s = 0.0$ and $w_m = 1.0$ on the motion-only task (Figure 6a) and sets $w_s = 1.0$ and $w_m = 0.0$ on the shape-only task (Figure 6b). In each case, the iso-probability contours are straight lines parallel to the task-irrelevant axis. This is the signature of an observer who weights one cue to the total exclusion of the other.

In the shape + motion task, participants were instructed to use both cues. We assumed that our idealized model observer then weights each cue according to its reliability because this is the statistically optimal way to combine independent estimates (Landy et al., 1995). It is important to note that this statistical assumption is used to make model predictions and fit the model's behavior to participants' results. However, there is evidence from a range of different tasks that human observers weight cues according to their reliability (e.g., Ernst & Banks, 2002; Landy et al., 1995). In this example, motion is less reliable than shape ($\sigma_m > \sigma_s$), so the model gives more weight to the more reliable shape cue. The resulting iso-probability contours are shown in Figure 6c. They are ellipses centered on zero difference, with the semi-major axis of the ellipse parallel to the less reliable dimension (here motion). This is the signature of an observer who is using both cues but weighting one more than the other. The signature of an observer who is weighting both cues equally would be circular iso-probability contours (not shown).

In these plots, the intersection of a given iso-probability contour with the motion or shape axis is also meaningful. For instance, if we use the 75% iso-probability contour (i.e., the rightmost or topmost dashed lines in Figure 6), the intersection with the x-axis can be considered the 75% motion threshold (holding shape constant) and the intersection with the y-axis can be considered the 75% shape threshold (holding motion constant). For example, when both motion and shape cues are task-relevant, the model observer's 75% shape threshold is about half of its motion threshold (~25% vs. ~60%). That is, the model is better at discriminating shape than motion (not surprisingly, since motion is the less reliable cue by design).

Armed with this understanding of how the relative contribution of motion and shape cues can be quantified and visualized, we next examined how human performance compared to the idealized model. We fitted each participant's 2D data with a reduced version of our full model using the maximum likelihood method. That is, to avoid over-fitting our data, we constrained the full model by assuming that shape and motion cues were weighted by their reliability (Landy et al., 1995). This means that we only need to fit the decision threshold and reliabilities in the three conditions, not the weights as well. The Deriving a model of shape and motion discrimination section of Appendix A provides the derivation of this reduced model from the full model (Equation A6). This assumption may be conceptually flawed because it implies that the reliability of the motion and shape estimates for identical stimuli varies between tasks. This may or may not be the case, but our between-subjects design did not allow us to address this issue directly, given that changes in reliability and changes in weight can produce very similar predictions. However, Equation A6 provided reasonable fits so that we could meaningfully compare the relative weighting of shape and motion cues between the different groups (see Testing the assumption that w_m and w_s are inversely proportional to σ_m and σ_s section of Appendix A). We therefore fitted Equation A6 to each participant's data to estimate θ, σ_m, and σ_s and then averaged these parameters across participants in each group. As in Experiment 1 (see Psychometric functions section), the RMSE was similar for the three tasks (motion only: $M = 11.2\%$, $SE = 0.9\%$; shape only: $M = 10.3\%$, $SE = 0.6\%$; and shape + motion: $M = 10.2\%$, $SE = 0.6\%$). The parameter means were used to generate the iso-probability contours for each group shown in Figures 6d–6f.

Importantly, we could calculate w_s and w_m for each participant from his or her individually fitted parameters. Table 1 provides the means and standard error of the means (*SEM*) for these relative weights. In the shape + motion group, participants were able to use both shape and motion cues to discriminate pairs of objects (compare Figures 6c and 6f). However, they weighted the shape cue more than the motion cue to perform the task (paired-sample $t(19) = 3.07$, $p = 0.006$). This finding suggests that participants could detect differences in shape more reliably than differences in motion because the weight

	Motion only	Shape only	Shape + motion
w_s	0.34 (0.08)	0.99 (0.004)	0.80 (0.04)
w_m	0.84 (0.07)	0.13 (0.02)	0.50 (0.06)

Table 1. The mean (*SEM*) of the relative weights. Note that w_s and w_m are calculated from the fitted parameters σ_m and σ_s separately for each participant and then averaged.

given to each cue reflects its reliability (Landy et al., 1995). This is consistent with our expectation of a stronger weighting/more reliable estimation of shape cues (e.g., Spetch et al., 2006; Vuong & Tarr, 2006).

When participants were instructed to use a single cue to discriminate pairs of objects, they were generally able to do so (shape only: compare Figures 6a and 6d; motion only: compare Figures 6b and 6e). In both single-cue groups, participants weighted the respective task-relevant cue more than the task-irrelevant one (shape only: t-value approached infinity as $p < 10^{-18}$; motion only: $t(19) = 3.77$, $p = 0.001$). However, participants in the motion-only group had relatively more difficulty basing their decisions solely on motion differences and relied also on shape differences to some extent whereas those in the shape-only group were relatively better at ignoring the task-irrelevant cue. The weight assigned to the task-irrelevant cue between these two groups was significant (2-sample $t(19) = 2.93$, $p = 0.009$). Consistent with this, Figure 7 shows the 75% iso-probability contours individually for all 20 participants in each task. The colored thin contours (blue or red) represent a single participant; the black thick contour represents the group-averaged contour. The red (horizontal) contours represent those participants who assigned a very large relative weight to shape cues (which we chose to be $w_s > 0.9$). Not surprisingly, all participants in the shape-only group weighted shape cues relatively more than motion cues. However as evident in Figure 7, six of the 20 participants in the shape + motion group assigned a relatively large weight to shape cues and three of the 20 participants in the motion-only group also assigned a relatively large weight to shape cues. These individuals have likely set their own relative weights, despite the explicit task instructions (perhaps due to the difficulty of the task).

Because of the asymmetry in the use of the shape and motion cues, it is interesting to consider how much difference is needed in one cue to produce a "different" judgment when the other cue is identical in the two stimuli across the three groups of participants (i.e., by looking at the intersection of the 75% iso-probability contour with the x- and y-axes in Figures 6d–6f). We found that the shape thresholds were similar for the shape-only (27.2%) and shape + motion groups (24.6%). In contrast, the motion threshold for the motion-only group (32.4%) was almost 2.4 times smaller than the motion threshold for the shape + motion group (76.6%). Lastly, we found that the *shape* threshold for the shape-only group was similar to the *motion* threshold for the motion-only group (27.2% versus 32.4%, respectively). This finding suggests that participants were equally sensitive to linear changes along our shape or motion morph continuum, even though motion cues were generally less reliable and weighted less than shape cues when both shape and motion were task-relevant. Furthermore, it suggests that the viewpoint changes we used to avoid image matching did not drastically bias shape or motion processing.

It is important to stress that the higher motion discrimination threshold for participants in the shape + motion group relative to those in the motion-only group did not necessarily occur because participants in the former group attended to both cues rather than one *per se*. If this was the case, we would also expect to see an increase in the shape discrimination threshold for the shape + motion group relative to participants in the shape-only group, and there was none. In fact, our model can explain this increase in the motion threshold for the shape + motion group; exactly the same effect is seen in the simulated example plotted in Figures 6b and 6c. In these simulations, the noise affecting each cue was held constant in the motion-only and shape + motion tasks, and only w varied. When only the motion cue was task-relevant, the model observer assigned all the weight to that cue (i.e., $w_m = 1.0$). However when both shape and motion cues were task-relevant, the model observer weighted each cue proportionally to its noise (which is larger for motion than shape information for these simulations); therefore; much less weight was given to the less reliable motion cue. In this case, small motion differences are, in effect, *further* reduced because of the small weight assigned to the motion cue (i.e., $w_m < 1.0$). Consequently, the model observer's sensitivity to motion differences decreased in

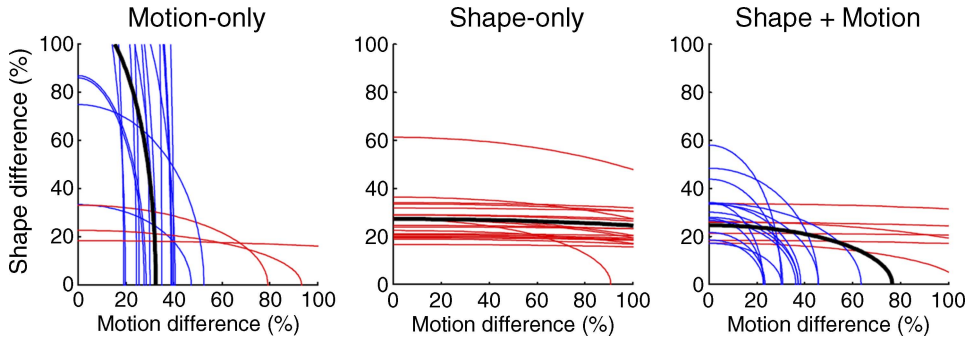


Figure 7. Individual discrimination contours for the motion-only, shape-only, and shape + motion groups. The colored thin contours are 75% iso-probability contours for each participant in each group. Participants who had $w_s > 0.9$ are highlighted by the red thin contours (horizontal lines). The thick black contour is the 75% iso-probability contour averaged across participants in each group.

the shape + motion task relative to the motion-only task. Thus, the model successfully captures this aspect of participants' discrimination performance.

Interestingly, the motion threshold for multipart objects from participants in the motion-only group was similar to the motion threshold of the single-part object from participants in Experiment 1 (32.4% versus 35.1%, respectively). This finding suggests, first, that the shape complexity *per se* may not influence motion discrimination and, second, helps further validate our parameter-based morphing technique for dynamic objects. Thus, overall, using a parametric manipulation of both shape and motion differences allowed us to model and visualize the relative contribution of shape and motion cues under different task demands.

General discussion

In the current study, we created a 2D parametric shape and motion space to investigate the joint contribution of these two cues for object perception and recognition. In Experiment 1, we validated a novel morphing technique when only motion differences were present in the stimuli. In Experiment 2, both shape and motion differences were present in the stimuli, but we manipulated which of these cues was relevant for the discrimination task. We found that observers were generally able to discriminate objects on the basis of their motion, shape, or both shape and motion. Importantly, however, there was a clear shape bias in our discrimination task: participants in the motion-only group weighted motion cues 2.5 times more than shape cues; those in the shape-only group weighted shape cues almost exclusively—7.6 times more than motion cues; and, critically, those in the shape + motion group weighted shape cues 1.6 times more than motion cues (see Table 1). We and others have found a similar bias for shape (e.g., Lander & Bruce, 2000; Spetch et al., 2006; Vuong & Tarr, 2006), but here we have quantified the relative weighting of shape and motion cues across different tasks.

It is of theoretical importance that non-rigid motions can be discriminated with the simple objects in Experiment 1 and the more complex multipart objects in Experiment 2 because it provides further evidence that dynamic features are independent of shape complexity. It is also of interest that the 75% threshold for discrimination was very similar for the two types of cues when they were responded to alone in Experiment 2 because this indicates that, to a certain extent, the cues require similar amounts of change in the shape and motion continuum (i.e., percentage of shape or motion changes) to be discriminated at the same level of accuracy. It would be important in future studies to investigate how changes along a morph continuum maps to "physical changes" (e.g., changes to the 3D position of object vertices) and the perceptual discriminability of these changes. In addition, we found evidence that when observers had to discriminate objects on the basis of both cues, motion cues helped discriminate between them when their shape was similar; contrariwise, motion cues played a relatively minor role when their shape was distinctive (Figure 5c). The type of non-rigid motion we tested here is typically associated with animate objects such as humans and non-human animals. However, the focus of this study is not about animacy *per se* but how non-rigid motion (articulations and deformations) can be used generally in the service of object recognition.

We note that different groups of participants were tested across the different discrimination tasks (shape only, motion only, and shape + motion). This design was used partly for practical reasons as our initial goal was to derive a complete psychometric function using the method of constant stimuli and to avoid practice effects. In future work, we hope to use adaptive methods to more quickly and more efficiently estimate discrimination contours for the different tasks for each participant (i.e., a within-subjects design for the task). In this way, we can directly estimate all four parameters without necessarily assuming that the relative weights assigned to motion and shape cues are proportional to the reliability of the individual cue.

A second contribution of the current study is that it provides a parameter-based motion morphing technique to create novel non-rigid motions using time-varying deformation fields. Importantly, these fields can be easily and meaningfully morphed to control the degree of physical (and perceived) similarity between motions and it is not necessary for the objects to have spatiotemporal correspondence (Giese & Poggio, 2000). We found that observers were sensitive to the parameters that defined the deformation fields, that is, their discrimination performance was an increasing function of the motion difference between two dynamic objects rather than a function of strictly low-level differences in the image sequences.

Overall, our results with parametric motion discrimination are similar to previous results with parametric shape discrimination (Cutzu & Edelman, 1996; Lawson & Bülthoff, 2008; Schultz et al., 2008). Thus, we argue that this motion morphing technique can yield interpretable perceptual results (Giese et al., 2008; Jastorff et al., 2006; Troje, 2002). Moreover, we believe that this technique can, in the future, be used to design hypothesis-driven experiments to generalize results from studies of the recognition of facial motion (e.g., Knappmeyer et al., 2003) and biological motion (see also Jastorff et al., 2006; Pyles et al., 2007). As shown here, the technique allowed us to apply the same set of motions to different objects and test how motion and shape cues combine.

The set of parameters defining the deformation fields was chosen partly for comparability to our earlier work in the shape domain (Schultz et al., 2008). More generally, however, any parameter (e.g., velocity) may have a temporal profile. Thus, parameters can be chosen to test

specific hypotheses. For example, in the shape domain, several groups, including us, have tested parameters thought to be important for recovering structural descriptions of objects, such as shape of cross section, amount of bending, amount of tapering, and so on (Biederman, 1987; Kayaert, Biederman, Op de Beeck, & Vogels, 2005; Schultz et al., 2008). With familiar biological motion, for example, Hill and Pollick (2000) showed that observers were sensitive to the temporal profile of wrist velocity in point-light displays of arm movements (i.e., how velocity varies over time).

Beyond testing specific parameters that might be critical to understand how motion may be more generally used in the service of object recognition and how motion may interact with shape, it is also important to understand the role of learning motions to acquire expertise. Observers can rapidly detect and extract information such as gender, emotion, and identity from highly shape-impoverished point-light displays (e.g., Bassili, 1978; Cutting & Kozlowski, 1977; Johansson, 1973). However, the extent to which this rapid processing is "innate" or develops through extensive learning and experience is unclear. Studies of perceptual expertise using static novel shapes have found that even discrimination training can alter neural responses to the trained shapes (e.g., Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). Relatively fewer studies have focused on learning motion (e.g., Chuang et al., 2006; Jastorff et al., 2006; Mayer & Vuong, 2012; Pilz et al., 2009). The results from studies with static objects suggest that perceptual expertise can lead to qualitative changes in how the objects of expertise are processed. For instance, observers may initially use individual parts to identify objects from an unfamiliar category but eventually shift to processing the objects more holistically (i.e., they may integrate the parts) as they acquire expertise with that object category. Perceptual expertise with particular motions could be valuable—for example, given our experiences with the information that can be derived from facial and body motions, it is likely that being able to understand what dynamic cues are learned for non-rigid motion is ecologically important (Jastorff et al., 2006; Pilz et al., 2009; Pyles et al., 2007).

The technique used here for creating a combined motion and shape parameter space allows for: (1) the tight control of the physical similarity of different shapes and motions; (2) the creation of a potentially large number of stimuli; (3) and the capacity to map the motion to arbitrary 3D objects. Thus, it is of interest to train observers to learn motion cues to object identity and then explore what happens once they acquire this expertise, for instance, how they generalize the motion to other objects and how the relative weighting of shape and motion cues might change with learning.

The parametric combination of both shape and motion allows an examination of how people extract a stable identity from changing shape. For instance, facial expressions deform a face's 3D shape. However, people can still extract a stable identity of a person. Similarly, animals (like a snake or shark) may have characteristic motions, like bending and slithering. Again, however, people can generalize over these shape changes to recognize the snake or shark. Computer vision algorithms have been developed to extract stable shape from non-rigid deformations (e.g., Torresani, Hertzman, & Bregler, 2003). We believe that the psychophysical results obtained in experiments like those conducted in the present study can help improve these algorithms. For instance, our results suggest that successful recognition requires a high-level representation of motion rather than low-level motion cues (e.g., simple leftward motion). Several researchers have proposed neurally inspired models in which high-level shape and motion detectors can be used to recognize dynamic objects (e.g., Cavanagh, LaBianca, & Thornton, 2001; Giese & Poggio, 2003). These complex dynamic feature detectors (e.g., "sprites," Cavanagh et al., 2001) can be derived from simpler shape and motion detectors. At present, we do not know what specific spatiotemporal information such a high-level representation might contain but that is clearly an important topic for further investigation.

Appendix A

Deriving a model of shape and motion discrimination

We assume that observers can extract a single "motion difference" and "shape difference" signal from a pair of objects simultaneously presented on a given trial. We further assume that each signal is separately subject to Gaussian noise. Using the notation where $x \sim N(\mu, \sigma)$ means that x is a random deviate drawn from a normal (Gaussian) distribution with mean μ and standard deviation σ, the motion (difference) signal, m, and shape (difference) signal, s, estimated by the observer on any given trial are

$$M \sim N(m, \sigma_m) \text{ and } S \sim N(s, \sigma_s). \quad \text{(A1)}$$

In Experiment 1, $s = 0$ (i.e., the shape signal is zero as both objects on a given trial always had the same shape); in Experiment 2, both m and s were generally non-zero.

Our general model is that observers take a weighted combination of the estimated motion and shape signals and respond "different" if the sum of this combination exceeds some decision threshold. Thus, they judge "same" when $C < \theta$, where

$$C^2 = w^2 S^2 + (1 - w^2) M^2, \quad \text{(A2)}$$

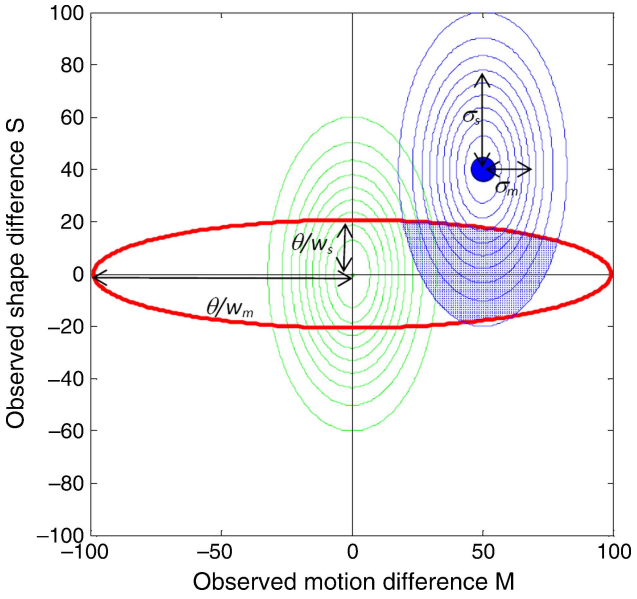


Figure A1. Behavior of the model of shape and motion discrimination. We assume that observers make noisy estimates (*M*, *S*) of the true stimulus differences (*m*, *s*). The thin green ellipses show iso-probability contours for (*M*, *S*) when (*m*, *s*) = (0, 0), i.e., there is no motion or shape difference. The thin blue ellipses show contours for (*M*, *S*) when (*m*, *s*) = (50%, 40%), marked with the blue dot. The contours are drawn such that the additional volume enclosed by each new contour represents 10% of the total volume under the distribution. The thick red ellipse shows the decision boundary: Combinations of (*M*, *S*) that fall inside the red ellipse are judged "same," while those falling outside are judged "different."

and w is in the range [0 1]. The parameter w is the relative weight assigned to the motion and shape signals. It is included because observers may explicitly choose to weight one cue relatively more than the other (e.g., because of task instructions) independently of how well they can estimate the motion or shape signals. This model therefore formally captures the different discrimination tasks that were implemented in Experiments 1 and 2 (motion only, shape only, and shape + motion tasks). For notational convenience, we define the relative weight assigned to the shape cue as $w_s = \sqrt{w^2} = w$ and the relative weight assigned to the motion cue as $w_m = \sqrt{(1-w^2)}$. Thus, we have $C^2 = w_s^2 S^2 + w_m^2 M^2$ and $w_s^2 + w_m^2 = 1$.

Figure A1 illustrates the discrimination behavior of the model. The thick red ellipse shows the decision boundary in the space (*M*, *S*). This boundary is dependent on the decision threshold (θ) and relative weighting of the two cues (w). The thin blue and green colored contours illustrate the distribution of (*M*, *S*) for two different sample stimulus differences (*m*, *s*). These signal distributions are dependent on the noise in the motion and shape estimation (σ_m and σ_s). The decision boundary and stimulus distributions extend into the negative range because we model a difference signal in the motion or shape dimension. We therefore based our fits on the sum of two cumulative Gaussian distributions. The use of two cumulative Gaussians allows us to ignore the sign of the motion/shape difference, e.g., if stimulus 1 was a 35% motion morph, then morphs of 25% and 45% for stimulus 2 would both be counted as a 10% motion difference.

The green contours in Figure A1 show the distribution of (*M*, *S*) under the null hypothesis that there is no difference in the stimulus (*m*, *s*). The blue contours show the distribution of (*M*, *S*) for a given non-zero pair of stimulus differences (*m*, *s*), marked by the blue dot. The aspect ratio of these contours depends on the relative standard deviations along the two axes, specifically, σ_m/σ_s. To maximize performance on a same–different task, we assume that the decision contour should also have an aspect ratio of σ_m/σ_s. We consider this case below. By definition, our model observer will judge the stimuli to be the same when (*M*, *S*) falls within the red ellipse.

In this general model, the probability of answering "same" when the true motion and shape difference is (*m*, *s*) is the volume of the stimulus distribution that falls within the red ellipse:

$$P_{\text{same}}(M,S|m,s) = \frac{1}{2\pi\sigma_m\sigma_s}\int_{-\theta/w_m}^{+\theta/w_m} dM \int_{-\frac{1}{w_s}\sqrt{\theta^2-M^2w_m^2}}^{+\frac{1}{w_s}\sqrt{\theta^2-M^2w_m^2}} dS \cdot \exp\left[-\frac{(M-m)^2}{2\sigma_m^2} - \frac{(S-s)^2}{2\sigma_s^2}\right]. \quad \text{(A3)}$$

To evaluate this integral, we change to polar coordinates r and α, defined by

$$Mw_m = r\cos\alpha \qquad Sw_s = r\sin\alpha \qquad dMdS = \frac{rdrd\alpha}{w_m w_s}$$

$$P_{\text{same}}(M,S|m,s) = \frac{1}{2\pi\sigma_m\sigma_s w_m w_s}\int_0^{\theta} rdr \int_0^{2\pi} d\alpha \cdot \exp\left[-\frac{(r\cos\alpha - w_m m)^2}{2w_m^2\sigma_m^2} - \frac{(r\sin\alpha - w_s s)^2}{2w_s^2\sigma_s^2}\right]. \quad \text{(A4)}$$

To fit this function to the data, we would need to fit the four parameters θ, σ_m, σ_s, and w to 25 data points (5 levels of shape differences × 5 levels of motion differences). We were reluctant to do this as the data did not allow all parameters to be well constrained; there is a trade-off between changes in relative weighting and changes in noise. We therefore limited our fits to the special case in which w is inversely proportional to stimulus noise. Because w provides a relative weighting of motion and shape cues, we can set the (relative) weight assigned to

each cue to be proportional to the inverse of its variance (Landy et al., 1995) so that the aspect ratio of the red decision boundary in Figure A1 matches the stimulus distributions shown in blue or green. That is, we set

$$w_m = \frac{\sigma_s}{\sqrt{\sigma_m^2 + \sigma_s^2}} \text{ and } w_s = \frac{\sigma_m}{\sqrt{\sigma_m^2 + \sigma_s^2}}. \quad \text{(A5)}$$

Equation A5 satisfies $w_s^2 + w_m^2 = 1$. In this case, we can do the integration over α analytically. Equation A4 now becomes

$$P_{\text{same}}(M, S|m, s) = A^2 \exp\left[-\frac{B^2}{2}\right] \int_0^\theta r dr \cdot \exp\left[-\frac{A^2 r^2}{2}\right] I_0(rAB), \quad \text{(A6)}$$

where $A^2 = \frac{1}{\sigma_m^2} + \frac{1}{\sigma_s^2}$ and $B^2 = \frac{m^2}{\sigma_m^2} + \frac{s^2}{\sigma_s^2}$, and I_0 is the modified Bessel function of the first kind. We used this function to fit the shape + motion data with three free parameters. We evaluated the integral numerically using the Matlab function QUAD, with the Matlab function BESSELI to evaluate the Bessel function.

In this 3-parameter model, iso-probability contours—i.e., values (m, s) along which the observer has a fixed probability of making a "same" judgment for a given set of parameters θ, σ_m, and σ_s—are constant values of B. That is, the contours are ellipses in (m, s) space whose semi-radii are $B\sigma_m$ and $B\sigma_s$ and where the value of B depends on the value of P_{same} along the contour. For example, to find the contour where the observer is equally likely to respond "same" or "different," we solve Equation A6 for B with $P_{\text{same}} = 0.5$. We used the Matlab function FZERO to solve this numerically.

Furthermore, we can evaluate P_{same} analytically when all the weight is assigned to one cue. For example, if $w_s = 0.0$ and $w_m = 1.0$, the probability of a "same" judgment becomes

$$P_{\text{same}}(M, S|m, s) = \frac{1}{2\pi\sigma_m\sigma_s} \int_{-\theta}^{+\theta} dM \int_{-\infty}^{+\infty} dS \cdot \exp\left[-\frac{(M - m)^2}{2\sigma_m^2} - \frac{(S - s)^2}{2\sigma_s^2}\right], \quad \text{(A7)}$$

and so the value of σ_s is immaterial:

$$P_{\text{same}}(M, S|m, s) = \frac{1}{2}\left\{\text{erf}\left(\frac{m + \theta}{\sqrt{2}\sigma_m}\right) - \text{erf}\left(\frac{m - \theta}{\sqrt{2}\sigma_m}\right)\right\}, \quad \text{(A8)}$$

and

$$P_{\text{diff}}(M, S|m, s) = 1 - \frac{1}{2}\left\{\text{erf}\left(\frac{m + \theta}{\sqrt{2}\sigma_m}\right) - \text{erf}\left(\frac{m - \theta}{\sqrt{2}\sigma_m}\right)\right\}. \quad \text{(A9)}$$

Maximum likelihood fitting

If the probability of answering "same" for some stimulus values (m, s) is P_{same}, then the probability of observing k "same" responses out of n trials is

$$\frac{n!}{k!(n - k)!} P_{\text{same}}^k (1 - P_{\text{same}})^{n-k}. \quad \text{(A10)}$$

To maximize the likelihood of obtaining the observed data, we maximize

$$X = \sum_{j=1}^{N} k_j \log P_{\text{same}}^j + (n_j - k_j) \log(1 - P_{\text{same}}^j), \quad \text{(A11)}$$

where the sum is over all different sets of stimulus values (m_j, s_j), n_j is the total number of trials performed with those stimulus values, and k_j is the number of trials on which the observer judged "same." P_{same}^j is the probability of getting a "same" response given the stimulus values (m_j, s_j) used on that trial and the particular fit parameters being tested. The fit parameters were adjusted using Matlab's FMINSEARCH function until the quantity X was maximized.

Testing the assumption that w_m and w_s are inversely proportional to σ_m and σ_s

Ideally, a full model of the discrimination task would have four parameters (Equation A4). By making an additional assumption, we reduced this to three parameters to avoid over-fitting our data sets. This 3-parameter model is appropriate for the shape + motion data but has a potential conceptual flaw when applied to the shape-only and motion-only data. The differently shaped contours in Figures 6d–6f reflect different fitted values of σ_m and σ_s (albeit from different groups of participants in our study). However, it seems unlikely that an observer would generally encode shape and motion cues with differing reliability across the three tasks because the stimuli are identical. We therefore assumed that the reliability of the motion and shape estimates was the same across tasks, but the weights given to each cue changed as a result of the task instruction. This is what we modeled in the simulations in Figures 6a–6c. The values of θ, σ_m, and

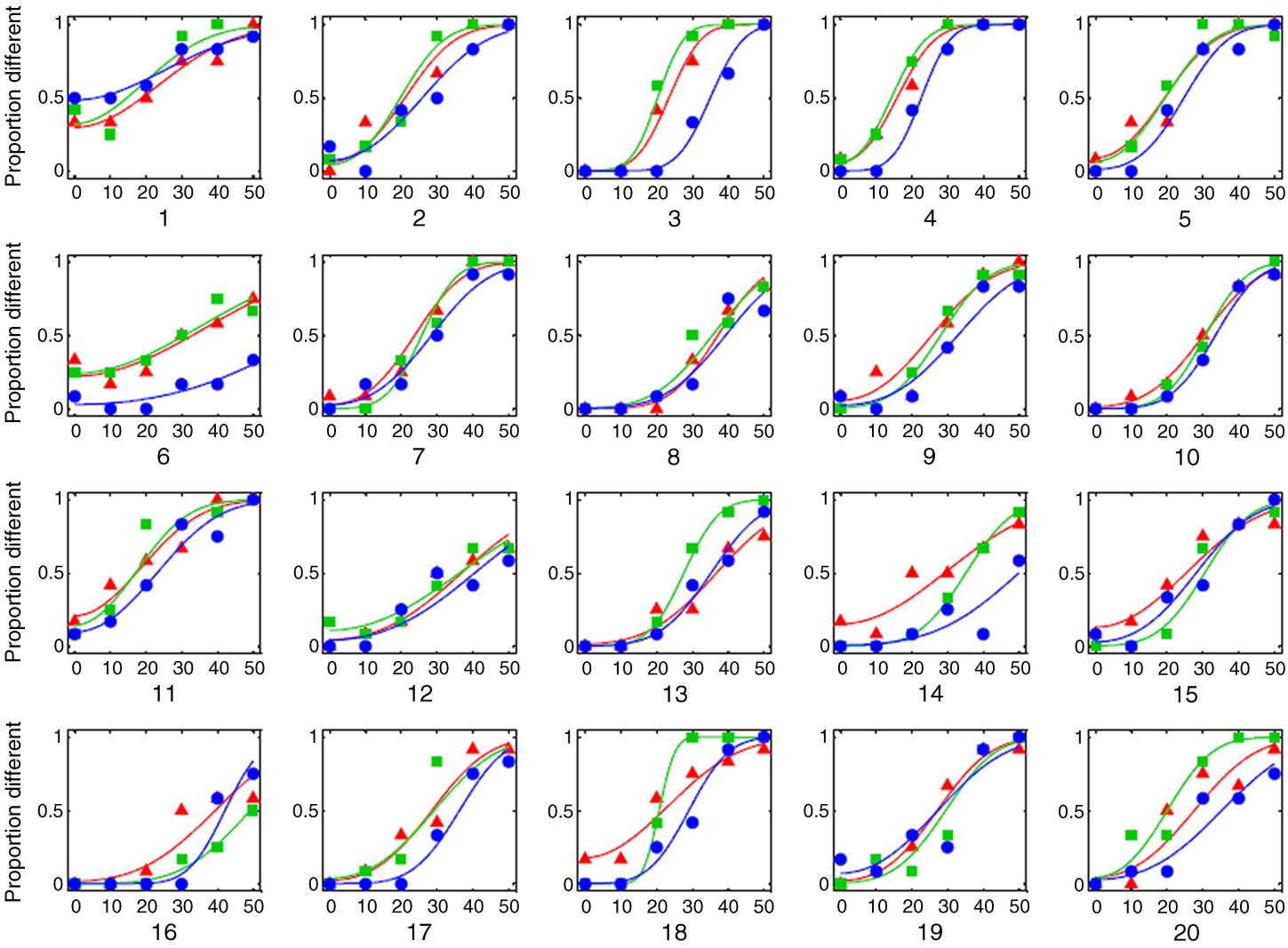


Figure A2. Individual fits for each of the 20 participants in Experiment 1. Red = bending–twisting morph pairs; green = bending–stretching morph pairs; blue = twisting–stretching morph pairs. The *x*-axis reflects the motion difference (%).

σ_s are kept constant in these three panels, and the different predicted performance is achieved by altering the relative weights w_m and w_s. However, despite its conceptual limitations, Equation A6 is adequate to describe our data. To check this, we fitted the shape-only data with a 2-parameter model that assumes that the observer uses only shape differences for the task (Equation A9). In this case, w_s = 1.0 and w_m = 0.0. The value of σ_m is then immaterial, leaving only two parameters to fit: θ and σ_s. We also fitted the two parameters, θ and σ_m, to the motion-only data in which w_m = 1.0 and w_s = 0.0. In each case, although the fits were slightly lower quality, the values of σ_m and σ_s were similar to those obtained with the 3-parameter model. This confirms that, in the shape-only task, the value of w_m, which is under-constrained by the data, does not greatly affect σ_s, which is much more constrained by the data. Similarly in the motion-only task, the precise value of w_s does not greatly affect σ_m. Therefore, in the main text, we report the results of fitting the same 3-parameter model to all three data sets, and then we estimated w_m and w_s based on σ_m and σ_s.

Individual data for Experiment 1

Figure A2.

Acknowledgments

We would like to thank Katja Mayer for the initial set of stimuli; Bernd Kohler for programming the experiment; and Tyler Austring, Michelle Foisey, and Geoff Hollis for data collection. There is no conflict of interest. This research was supported by a grant from the Natural

Sciences and Engineering Research Council of Canada to AF. JCAR was supported by Royal Society University Research Fellowship UF041260.

Author contributions: Quoc C. Vuong and Alinda Friedman contributed equally to this work.
Commercial relationships: none.
Corresponding author: Quoc C. Vuong.
Email: quoc.vuong@newcastle.ac.uk.
Address: Institute of Neuroscience, Henry Wellcome Building, Newcastle University, Newcastle upon Tyne NE2 4HH, UK.

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.

Barr, A. H. (1984). Global and local deformations of solid primitives. *Computer Graphics, 18,* 21–30.

Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 373–379. [PubMed]

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147. [PubMed]

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology, 16,* 69–74. [PubMed]

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition, 80,* 47–60. [PubMed]

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

Cook, R. G., & Katz, J. S. (1999). Dynamic object perception by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 25,* 194–210. [PubMed]

Cutting, J., & Kozlowski, L. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9,* 353–356.

Cutzu, F., & Edelman, S. (1996). Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Sciences, 93,* 12046–12050. [PubMed] [Article]

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. [PubMed]

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. [PubMed]

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Giese, M. A., & Poggio, T. (2000). Morphable models for the analysis and synthesis of complex motion patterns. *International Journal of Computer Vision, 38,* 59–73.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4,* 179–192. [PubMed]

Giese, M. A., Thornton, I. M., & Edelman, S. (2008). Metrics of the perception of body movement. *Journal of Vision, 8*(9):13, 1–18, http://www.journalofvision.org/content/8/9/13, doi:10.1167/8.9.13. [PubMed] [Article]

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology, 11,* 880–885. [PubMed]

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances the recognition of individuals from point light displays. *Psychological Science, 11,* 223–228. [PubMed]

Jastorff, J., Kourtzi, Z., & Giese, M. A. (2006). Learning to discriminate complex movements: Biological versus artificial trajectories. *Journal of Vision, 6*(8):3, 791–804, http://www.journalofvision.org/content/6/8/3, doi:10.1167/6.8.3. [PubMed] [Article]

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 201–211.

Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience, 22,* 212–224. [PubMed]

Kellman, P. J. (1993). Kinematic foundations of infant visual perception. In C. Granrud (Ed.), *Visual perception and cognition in infancy* (pp. 121–173). Hillsdale, NJ: Lawrence Erlbaum.

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. [PubMed]

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

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modelling of depth cue combination: In defense of weak fusion. *Vision Research, 35,* 389–412. [PubMed]

Lawson, R., & Bülthoff, H. H. (2008). Using morphs of familiar objects to examine how shape discriminability influences view sensitivity. *Perception & Psychophysics, 70,* 853–877. [PubMed]

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tell the frog's brain. *Proceedings of the IRE, 47,* 1940–1959.

Liu, T., & Cooper, L. A. (2003). Explicit and implicit memory for rotating objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 554–562. [PubMed]

Mayer, K. M., & Vuong, Q. C. (2012). The influence of unattended features on object processing depends on task demand. *Vision Research, 56,* 20–27. [PubMed]

Newell, F. N., Wallraven, C., & Huber, S. (2004). The role of characteristic motion in object categorization. *Journal of Vision, 4*(2):5, 118–129, http://www.journalofvision.org/content/4/2/5, doi:10.1167/4.2.5. [PubMed] [Article]

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *Journal of Neuroscience, 13,* 13025–13036. [PubMed] [Article]

Pilz, K. S., Bülthoff, H. H., & Vuong, Q. C. (2009). Learning influences the encoding of static and dynamic faces and their recognition across different spatial frequencies. *Visual Cognition, 17,* 716–735.

Pilz, K. S., Thornton, I. M., & Bülthoff, H. H. (2006). A search advantage for faces learned in motion. *Experimental Brain Research, 171,* 436–447. [PubMed]

Pyles, J. A., Garcia, J. O., Hoffman, D. D., & Grossman, E. D. (2007). Visual perception and neural correlates of novel 'biological motion'. *Vision Research, 47,* 2786–2797. [PubMed]

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. [PubMed] [Article]

Setti, A., & Newell, F. (2010). The effect of body and part-based motion on the recognition of unfamiliar objects. *Visual Cognition, 18,* 456–480.

Spetch, M. L., Friedman, A., & Vuong, Q. C. (2006). Dynamic object recognition in pigeons and humans. *Learning & Behavior, 34,* 215–228. [PubMed]

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research, 38,* 947–951. [PubMed]

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 and Review, 2,* 55–82.

Tinbergen, N. (1951). *The study of instinct*. Oxford, UK: Oxford University Press.

Torresani, L., Hertzman, A., & Bregler, C. (2003). Learning non-rigid 3D shape from 2D motion. *Advances in Neural Information Processing Systems, 16,* 1–8.

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision, 2*(5):2, 371–387, http://www.journalofvision.org/content/2/5/2, doi:10.1167/2.5.2. [PubMed] [Article]

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition, 67,* 21–44. [PubMed]

Vernon, M. D. (1952). *A further study of visual perception*. London: Cambridge University Press.

Vuong, Q. C., Friedman, A., & Plante, C. (2009). Modulation of viewpoint effects in object recognition by shape and two kinds of motion cues. *Perception, 38,* 1628–1648. [PubMed]

Vuong, Q. C., & Tarr, M. J. (2006). Structural similarity and spatiotemporal noise effects on learning dynamic novel objects. *Perception, 35,* 497–510. [PubMed]

Watson, T., Hill, H., Johnston, A., & Troje, N. (2005). Motion as a cue for viewpoint invariance. *Visual Cognition, 12,* 1291–1308.

Watt, A., & Watt, M. (1992). *Advanced animation and rendering techniques: Theory and practice*. New York: ACM Press.