A comparison of form processing involved in the perception of biological and nonbiological movements

Steven M. Thurman

Department of Psychology, University of California, Los Angeles, CA, USA ✉

Hongjing Lu

Department of Psychology, University of California, Los Angeles, CA, USA
Department of Statistics, University of California, Los Angeles, CA, USA ✉

Although there is evidence for specialization in the human brain for processing biological motion per se, few studies have directly examined the specialization of form processing in biological motion perception. The current study was designed to systematically compare form processing in perception of biological (human walkers) to nonbiological (rotating squares) stimuli. Dynamic form-based stimuli were constructed with conflicting form cues (position and orientation), such that the objects were perceived to be moving ambiguously in two directions at once. In Experiment 1, we used the classification image technique to examine how local form cues are integrated across space and time in a bottom-up manner. By comparing with a Bayesian observer model that embodies generic principles of form analysis (e.g., template matching) and integrates form information according to cue reliability, we found that human observers employ domain-general processes to recognize both human actions and nonbiological object movements. Experiments 2 and 3 found differential top-down effects of spatial context on perception of biological and nonbiological forms. When a background does not involve social information, observers are biased to perceive foreground object movements in the direction opposite to surrounding motion. However, when a background involves social cues, such as a crowd of similar objects, perception is biased toward the same direction as the crowd for biological walking stimuli, but not for rotating nonbiological stimuli. The model provided an accurate account of top-down modulations by adjusting the prior probabilities associated with the internal templates, demonstrating the power and flexibility of the Bayesian approach for visual form perception.

Introduction

The human visual system is exquisitely tuned to perceive and understand the actions of others. The ability to rapidly and accurately process biological motion information serves an important functional role to promote biological fitness and social acumen. Behavioral studies on biological motion perception reveal that motion and form-based representations each play a role in human action recognition (Casile & Giese, 2005; Giese & Poggio, 2003; Lange & Lappe, 2006; Thurman, Giese, & Grossman, 2010; Thurman & Grossman, 2008; Thurman & Lu, 2013a). Human fMRI studies also confirm a distinct contribution from brain regions involved in motion and form-based processing (Jastorff & Orban, 2009; Jastorff, Vanduffel, Popivanov, Orban, & Vogels, 2012; Thompson & Baccus, 2011; Vangeneugden, Peelen, Tadin, & Battelli, 2014).

Although it is under debate what factors influence the relative contributions from form and motion, the involvement of form processing in biological motion perception is well established in the literature. For example, even for point-light displays depicting human actions using only discrete joints in motion, researchers have found that form processing plays a featured role in perception (de Lussanet et al., 2008; Lange & Lappe, 2006; Lu, 2010; Theusner, de Lussanet, & Lappe, 2014). Several psychophysical studies have demonstrated that action discrimination is robust even when stimuli lack veridical or reliable local motion information (Beintema, Georg, & Lappe, 2006; Beintema & Lappe, 2002; Thurman & Lu, 2014b). Computational models that employ strictly form-based analysis, such as posture-based template matching, have also been shown to account quantitatively for a variety of behavioral and neurophysiological findings, without analyzing image

motion features (Lange & Lappe, 2006; Theusner et al., 2014; Thurman & Lu, 2014b). Human brain-imaging studies reveal regions in occipito-temporal cortex that are selectively activated by static images of the human body (Downing, Jiang, Shuman, & Kanwisher, 2001; Taylor, Wiggett, & Downing, 2007; Weiner & Grill-Spector, 2011), and regions in superior temporal cortex that respond selectively to human bodies in action (Grossman, Jardine, & Pyles, 2010; Grossman et al., 2000). Single-cell electrophysiological studies with nonhuman primates have also discovered cells in temporal cortex that are tuned to individual body postures (Oram & Perrett, 1994; Singer & Sheinberg, 2010; Vangeneugden et al., 2011).

However, our visual system engages in form recognition constantly, not just for biological entities, but also for nonbiological objects of all shapes and sizes. It is unclear whether the human brain employs generic computational machinery for this task, or whether there might be specialized mechanisms for processing biological form information. This is an important question because it is commonly understood that biological motion perception, owing to its unique ecological significance and complexity due to articulation, is privileged and specialized in terms of neural processing. While there is evidence from developmental studies with humans (Bardi, Regolin, & Simion, 2013; Fox & McDaniel, 1982; Simion, Di Giorgio, Leo, & Bardi, 2011; Simion, Regolin, & Bulf, 2008) and other species (Lucia, Luca, & Giorgio, 2000; Vallortigara & Regolin, 2006; Vallortigara, Regolin, & Marconato, 2005) for early, and perhaps innate, sensitivity to biological motion per se, little experimental work has sought to examine the specialization of biological form processing in isolation from motion.

A hurdle in examining this issue is that most studies employing computational models of form perception have focused directly on the fundamental case of recognizing static objects in isolation (Kersten, Mamassian, & Yuille, 2004; Riesenhuber & Poggio, 1999, 2000). Yet, we inhabit a rich and dynamic visual environment where our goal is not just to recognize static objects, but also to extract meaningful dynamic properties from moving objects and their relationship to surrounding objects within the environment. This has obvious importance for biological action perception, where dynamics play a chief role in defining and distinguishing actions, but is also relevant for nonbiological objects (e.g., moving cars or machinery, leaves, or objects blowing in the wind). Another key feature of form perception is that it involves both bottom-up processing and top-down modulatory influences (Bar, 2004). In the natural environment objects rarely appear in isolation, but exist alongside rich contextual information from which we learn to make predictions and generate expectations about the likelihood of

objects within a particular visual scene (Bar et al., 2006). Many studies demonstrate that object recognition is facilitated when objects are placed in a familiar context; for instance, when a loaf of bread is easier to be identified in a kitchen scene than in an outdoor scene (Davenport, 2007; Palmer, 1975).

A primary goal of the current study was to provide a systematic investigation of dynamic form perception comparing biological to nonbiological objects, and to present a computational framework to account for both bottom-up and top-down influences on perception. In the first experiment, we focus on bottom-up or stimulus-driven processing in which low-level visual features are systematically controlled and stimuli are presented in isolation. We used the classification image technique to investigate critical features used by human observers in recognizing biological and nonbiological objects, and compared with a Bayesian observer model using generic computational principles that visual form cues are combined according to their associated reliabilities (Thurman & Lu, 2014b). In the second and third experiments, we examined differential top-down effects of spatial context on perception of biological and nonbiological forms. Human performance was compared with the observer model, which is equipped to account quite naturally for contextual effects by adjusting the prior probabilities associated with internal templates. Prior expectations, perhaps based on long-run observed statistics of the object and its surrounding context in the natural world (Schwartz, Hsu & Dayan, 2007), can exert their influence as a top-down modulation to lower level stages of processing. We found that the proposed model was able to provide a good fit and parsimonious account of behavioral data from all three experiments for both biological and nonbiological objects. These results demonstrate that domain-general form processes are capable of supporting recognition of rigid nonbiological objects and nonrigid human actions, and demonstrate the power and flexibility of the Bayesian approach to account for bottom-up and top-down effects in dynamic form perception.

Experiment 1

The first experiment was designed to investigate how human observers process dynamic form-based stimuli and resolve ambiguity originated from conflicting form cues for biological and nonbiological objects. In a recent study (Thurman & Lu, 2014b), we created hybrid stimuli by putting two form cues, position and orientation, into conflict using Gabor patches such that positions were sampled from one target (e.g., leftward facing walker) and orientations were sampled from the opposite target (e.g., rightward facing walker). A limited-lifetime meth-

od, which randomly samples elements sparsely according to the skeleton of the shapes, was employed to selectively target form-based processing, and to limit the contribution of local motion cues to dynamic form perception (Beintema et al., 2006; Beintema & Lappe, 2002). We found that the spatial frequency of the oriented Gabor patches played a significant role in modulating the contribution of orientation information to the overall appearance of the global shape (Day & Loffler, 2009; Thurman & Lu, 2014b). By adjusting spatial frequency according to individual performance, like a dial, we can generate ambiguous stimuli that are perceived to be walking both leftward and rightward, or rotating clockwise and counter-clockwise, to an equal degree in the sense that across many trials the stimulus will be perceived as one or the other 50% of the time. This type of stimulus is useful because it presumably activates opposing internal representations (e.g., leftward and rightward walking) to roughly the same degree on average, but in a particular trial certain information in the randomly sampled stimulus may bias an observer to make a particular response or the other.

To examine the relationship between stimulus characteristics and perceptual responses in a trial-by-trial manner, we employed the classification image (CI) technique to dynamic stimuli (Ahumada, 2002; Keane, Lu, & Kellman, 2007; Knoblauch & Maloney, 2008; Lu & Liu, 2006; Murray, 2011) to assess how randomly sampled element locations on each trial influence the perceived walking or moving direction of the ambiguous stimulus. These spatial-temporal CI images have the potential to reveal subtle and behaviorally relevant differences in the weighting of two form cues, position and orientation, across space and time. An observer model is introduced to cope with the ambiguity by combining two distinct visual cues in a bottom-up fashion. The observer model employs the Bayesian framework to integrate local visual information of position and orientation by incorporating knowledge about the reliability, or uncertainty, associated with the visual samples and featural knowledge about the underlying shape templates (Thurman & Lu, 2014b). Simulations of the observer model yield a decision of recognizing dynamic forms in individual trials, which can be used to derive spatial-temporal CIs to compare with those derived from human performance. A significant agreement between human and model CIs would provide strong support for basic computational-level characteristics commonly shared by humans and the observer model.

Participants

Fifty-two participants were recruited through the Department of Psychology subject pool at the Univer-sity of California, Los Angeles (UCLA), and were given course credit for their participation. All participants reported normal or corrected vision and gave informed consent approved by the UCLA Institutional Review Board. All participants were naïve to the stimuli and to the purpose of the study. Participants were assigned to one of two groups that performed the discrimination task either with biological human walker stimuli ($n = 24$, 17 female and seven male, mean age $= 19.8 \pm 1.7$), or with nonbiological rotating square stimuli ($n = 28$, 17 female and 11 male, mean age $= 19.7 \pm 1.2$).

Materials and method

Stimuli were created using Matlab (MathWorks, Natick, MA) and displayed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) on a calibrated monitor with a gray background (60 Hz, background luminance 16.2 c/m^2) and powered by a Dell PC running Windows XP. Experiments were conducted in a dark room with a chin rest to maintain a constant viewing distance (35 cm).

Motion capture data of a human walking was obtained from the Carnegie Mellon Graphics Lab Motion Capture Database, available free online (http://mocap.cs.cmu.edu). The BioMotion toolbox was used to convert the raw motion capture files to point-light format, with 11 points representing the head, mid-shoulder, elbows, wrists, knees, and feet (van Boxtel & Lu, 2013). Horizontal translation of the actor was subtracted so that the walker appeared to walk on a treadmill, and the stimulus was trimmed to comprise a loopable walking cycle consisting of 60 frames. The stimulus of nonbiological form was created by rotating a rigid square shape in-plane by increments of 6° per frame so that the animation would complete one full rotation over the course of 60 frames. Leftward and rightward walkers were created by reflecting the stimulus across the vertical meridian, and clockwise and counterclockwise rotating square stimuli were generated by playing the sequence in forward or reverse temporal order. Biological and nonbiological stimuli were equated in terms of vertical size to subtend approximately 9°. Stimuli were presented for 1 s on each trial to complete one full gait cycle or rotation cycle.

On each trial of the experiment, stimulus frames were constructed by randomly sampling sparse locations along the skeletal shape of the walker or the square (Figure 1). The locations of elements were resampled independently on each frame of the animation; hence, the apparent motion induced by these element shifts were inconsistent with the local trajectories of the actual global object movements (Beintema

1. Choose *n* random spatial samples on each frame

2. Determine orientation from template with opposite direction

3. Gabors represent incongruent position + orientation cues.
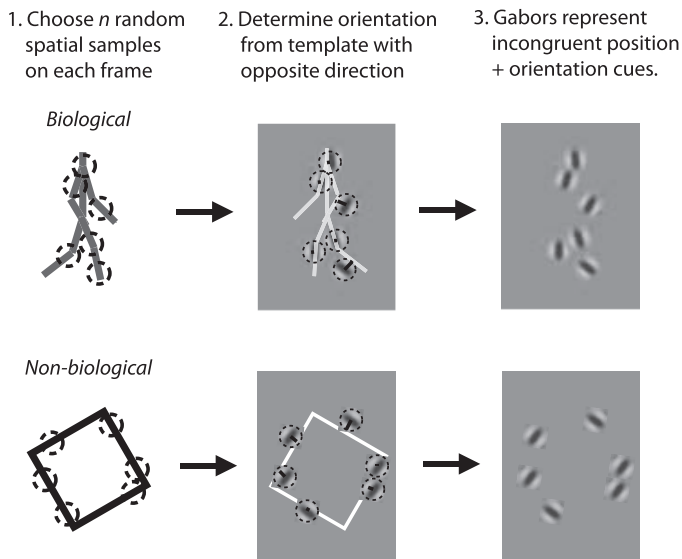
Biological

Non-biological

Figure 1. Schematic of stimulus construction for biological (human walkers) and nonbiological (rotating squares) stimuli. In step 1, on each frame of the animation sequence we randomly select *n* spatial samples from the contour of the underlying stimulus. In the schematic we illustrate a single example frame. In step 2, we derive orientation from the limb angle of the nearest point on the stimulus moving in the opposite direction. In step 3, we illustrate how a single frame of this stimulus would appear, with conflicting information from position and orientation cues. Supplemental movie is included for demonstration.

& Lappe, 2002), forcing the perceptual system to rely predominantly on the structural cues provided at each moment (Lange & Lappe, 2006). Two elements were randomly sampled per frame along the skeleton of a walker for the biological stimulus. Four elements were sampled along the contour of a square for the nonbiological stimulus. Discrimination of a sparsely sampled square is severely underconstrained without at least one sample from each of the four edge segments based on our pilot data, whereas walking direction has been shown to be discriminable with as few as two points per frame (Beintema & Lappe, 2002; Thurman & Lu, 2014b). The elements were high-contrast Gabor patches (Michelson contrast = 0.5, spatial sigma = 0.84°, sine phase), where the orientation of the patch was manipulated to be incongruent with its location sampled from a target. Specifically, for each element we determined the nearest point on the figure moving in the opposite direction and made orientation consistent with the corresponding limb on the opposing figure (e.g., position sampled from front ankle of leftward walker, but orientation derived from the back ankle of a rightward walker). As such, orientation and position provided conflicting information about the dynamic shape, where perception would occasionally follow the structure implied by either position or orientation cues.

Participants performed a block of 36 trials to provide an estimate for their individualized point of subjective equality (PSE) by manipulating spatial frequency to equate the contribution of position and orientation cues to the task. We used a one-up-one-down staircase procedure (initiated at 0.3 c/°, step size = 0.05 c/°) to vary spatial frequency according to subject performance, where a response consistent with position cues would lead an increase in spatial frequency on the following trial and a response with orientation cues would lead a decrease in spatial frequency. The staircase stabilized around the PSE where participants produced approximately 50% of their responses consistent with each type of form cue.

Participants next performed direction discrimination on 480 trials of ambiguous stimuli using their individualized level of PSE for spatial frequency. For biological stimuli, participants discriminated the walking direction (left vs. right), and for nonbiological stimuli participants discriminated rotation direction (clockwise vs. counterclockwise). Observer responses were collected in terms of whether the perceived direction was consistent with the position or orientation-defined stimulus direction.

Analysis

To compute classification images, multiple linear regression was performed with subject decisions serving as the response variable and the locations of the elements on each trial serving as the predictor variable. Element locations for trials with rightward facing walkers were mirror reversed across the vertical meridian, and element locations for trials with counterclockwise rotating shapes were time-reversed, prior to analysis to place-sampled locations across all trials within a common reference frame (e.g., a leftward-facing walker or clockwise rotating square reference frame). The result of this analysis was a set of CIs—a spatio-temporal map representing the correlation between each point along the structure of the underlying stimulus and human responses across trials. In the CIs, positive values indicate a correlation between perceptual responses consistent with position cues, and negative values indicate a correlation with orientation cues. Classification images were computed for each subject and then normalized (*z*-scored) before performing group averaging by summing the *z*-scored images across all participants and dividing by the square root of the number of subjects. Finally, the group CIs were averaged across identical postures (1 gait cycle = 2 step cycles, each with roughly identical postures) or identical square templates (each frame repeats four times during a full rotation, e.g., 0°, 90°, 180°, 270°) to increase signal-to-noise, and spatially
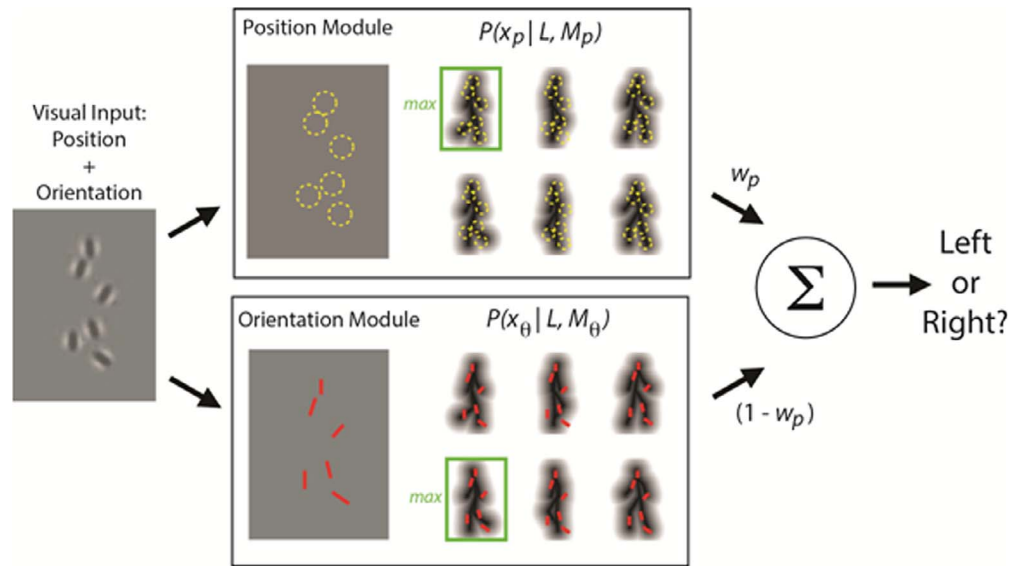
Figure 2. An illustration of the Bayesian observer model. Two modules that implement local Bayesian inference based on either position cues alone, (i.e., position module) or orientation cues (i.e., orientation module). The posterior probabilities from each module are weighted and linearly combined to derive the final decision on the facing direction.

smoothed with a two-dimensional Gaussian filter (sigma = 8 pixels). Statistical significance was evaluated by identifying pixels with a *z*-score greater than ± 1.65 ($p = 0.05$, uncorrected).

In addition to identifying stimulus features with significant correlation to behavior, we were chiefly interested in comparing the spatio-temporal patterns in the CIs derived from human observers and the computational model respectively. Next we provide a brief description of the model, where more details can be found in our previous article (Thurman & Lu, 2014b).

Bayesian observer model

The observer model (Figure 2) contains two modules that implement local Bayesian inference based on either position cues alone, (i.e., position module) or on orientation cues (i.e., orientation module). Of note, the orientation module performs template matching between individual limb angles of the template and the orientation of the nearest Gabor element. While orientation information is primary to this computation, positional information does play a role through this spatial constraint in anchoring the orientation to a particular region of space and, hence, in *carrying* the orientation signal (Thurman & Lu, 2014b). Input data is compared frame-by-frame to templates (eight equidistant frames from the walker or square sequences), which are implemented as probabilistic likelihood functions centered on the true underlying stimulus template with certain variability modeled by a Gaussian function.

Each module produces a posterior probability of the walking direction given the data, which is computed according to Bayes rule by combining the likelihood of the observed positions or orientations given a template with the prior probability of the walking direction $P(L)$. For the position module, M_p the recognition of the stimulus with sparsely sampled elements is based on the posterior probability of walking direction (i.e., L indicating the left walking direction) conditional on perceived locations x_p, $P(L \mid x_p, M_p)$. The orientation module, M_θ, computes the posterior probability of walking direction conditional on perceived orientations x_θ, $P(L \mid x_\theta, M_\theta)$. Next, the posterior probabilities from each module are weighted and linearly combined according to Equation 1 below.

$$P(L \mid x_p, x_\theta) = \{ w_p P(x_p \mid L, M_p) \\ + (1 - w_p) P(x_\theta \mid L, M_\theta) \} P(L) \quad (1)$$

The weight represents the relative contribution of the orientation and position modules to the final discrimination decision. In our prior work, these weights were estimated for each subject individually from psychometric performance on low-level position and orientation discrimination tasks (Thurman & Lu, 2014b). In the current simulations, we estimated this weight as a parameter in order to equate the contribution of the position and orientation cues to fit the average accuracy of human responses. That is, for each stimulus type we ran several thousand trials varying the weight parameter systematically until a weight would produce approximately 50% responses consistent with position and orientation cues across trials. This is analogous to the behavioral experiment in which we
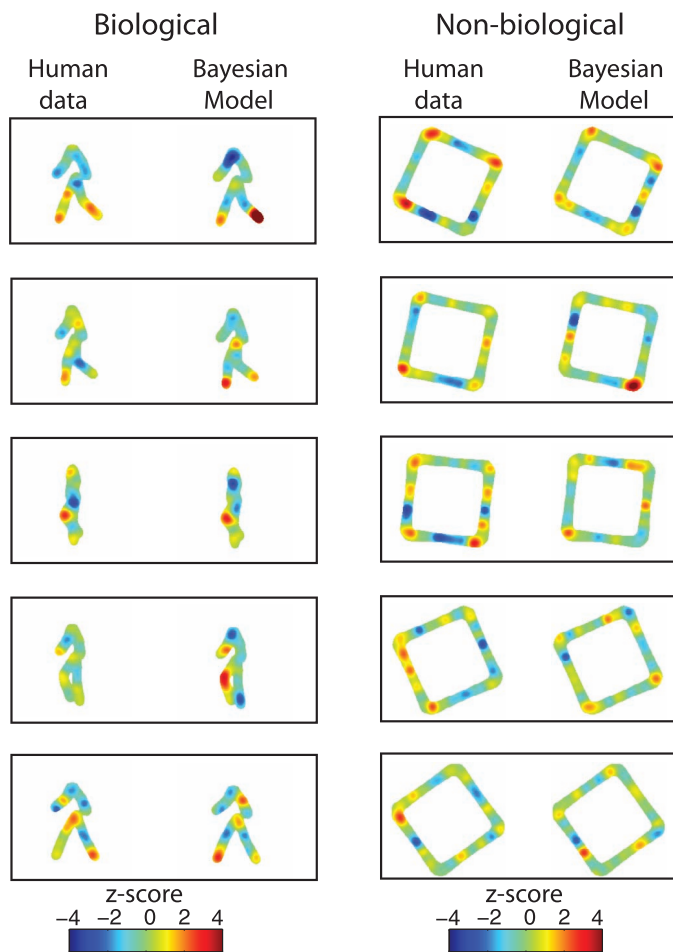
Figure 3. Classification image results for Experiment 1 for biological and nonbiological stimuli, with human data presented in the left column, and data from simulations of the Bayesian observer model presented in the right column. In separate rows identified with a rectangular outline, we show select time points, or frames, from the classification image sequence. The color maps represent z-scores for the classification images.

estimated the PSE (e.g., the observer's internal weight) by manipulating spatial frequencies in the stimulus to equate the contributions from the two types of cues to human perception. Lastly, the template probabilities are integrated across frames with a temporal weighting function that penalizes sequences that are in the incorrect order, a procedure analogous to the template matching model of Lange and colleagues (Lange & Lappe, 2006). The decision criterion of the model is to select the direction with the highest posterior probability over time.

In simulations for Experiment 1, the prior probability $P(L)$ was set to 0.5 for each facing direction to match the probability in the actual experiment. In simulation for Experiment 2, the prior probability $P(L)$ was changed depending on spatial context.

We tested the model by simulating 10,000 trials for each task (biological and nonbiological), generating ambiguous stimuli with randomly selected positional samples as in the behavioral experiment and deriving the model responses. The model data were analyzed with the same processing pipeline as the behavioral data to compute CIs for direct comparison with CIs derived from human decisions.

Human results

The mean proportion of responses consistent with position cues was not significantly different from 0.5 for either task (biological = 0.49, SD = 0.03; nonbiological = 0.49, SD = 0.01), demonstrating the success of using the staircase procedure to estimate the PSE for each subject. Group CIs for both tasks are displayed in the left columns of Figure 3. The analysis revealed several regions with a significant correlation to behavioral responses. The regions with positive z-score (orange-red) represent body regions that, when sampled, contributed to perception consistent with position cues. The regions with negative z-score (green-blue) represent body regions contributing to perception according to orientation cues. For the biological stimulus, many of the positive regions are located at the terminal of the extremities, particularly near the feet, and also at the leading knee in the swing phase of the gait cycle. This result is consistent with prior studies demonstrating the high relative importance of limb extremities for action perception (Mather, Radford, & West, 1992; Thurman, Giese, & Grossman, 2010; van Boxtel & Lu, 2015), a result that is predicted from position-based template models (Lange & Lappe, 2006). Negative regions show a markedly different pattern, tending to occur on the torso or at midlimb positions. For the nonbiological stimulus, the positive regions occur predominantly at the corners of the square shape, whereas the negative regions typically occur at intermediate positions along the edge segments.

Model results

The Bayesian observer model provides a probabilistic framework for optimally combining information from position and orientation cues on a frame-by-frame basis and integrating this information across time to produce a perceptual discrimination (Thurman & Lu, 2014b). Classification images computed from responses of the Bayesian observer model are shown in Figure 3. The model exhibited similar CIs to the human observers, for instance, showing significant correlations with positional cues (e.g., positive z-scores) at extreme locations on the limbs and the corners of the square shape for the biological and nonbiological stimuli, respectively.

To quantify these relationships, we computed the correlation coefficient between human and model classification images, and performed a permutation test to create a null distribution for assessing statistical significance. In the permutation test, we randomly scrambled the mapping between subject responses and stimulus data, and processed this permuted data through the same pipeline as the experimental data to derive permuted group CIs. This procedure simulates a sample of observers using the same stimulus trial data, but with random responses. We computed the correlation coefficient for each of 100 randomly permuted CIs and used this null distribution to convert the experimental correlations to z-scores. This analysis revealed a statistically significant relationship between human and the model CIs for both the biological stimulus ($z = 3.14$, $p < 0.05$), and nonbiological stimulus ($z = 3.53$, $p < 0.05$). This suggests that humans process these sparse stimuli in a bottom-up fashion that is consistent with the Bayesian observer model, for instance, relying on probabilistic template matching frame-by-frame and the assessment of cue reliability. The model accounted well for data in both tasks, casting doubt on the idea that form processing is unique or specialized for biological actions. Instead, this result supports the theory that dynamic form processing involves generalized neural mechanisms regardless of the biological nature of the stimulus.

We performed supplemental post hoc analyses to help explain why the classification images exhibited this specific spatio-temporal pattern (see Supplemental Material). We found that patterns revealed in the human and model classification images can be largely explained by distinctive feature differences (e.g., position and orientation) in the underlying templates when compared on a single frame basis. For example, human participants and the Bayesian observer model each tended to rely on position cues when spatially distinct object locations were sampled, and rely on orientation cues when indistinct, or spatially overlapping, regions were sampled. When comparing opposing shape templates, the biological limb extremities and the corners of the square shape tend to be the most spatially distinct and, hence, correlate with perceptual decisions that correspond to stronger use of position cues in relation to orientation. This finding provides further support that humans engage weighted template matching of individual frames in analyzing the dynamics of each shape sequence.

Experiment 2

When we track moving objects in the environment, such as a human walking, we move our eyes to keep the object centered on the fovea. This causes the (static) background that surrounds the object to appear to flow in the opposite direction. Related to this effect, Fujimoto and Sato (2006) discovered a relatively strong visual illusion called the *backscroll illusion*. When a sinusoidal grating is presented with counterphase flicker, its motion direction is completely ambiguous and the grating appears to move both leftward and rightward with equal strength. However, if a human walker stimulus is superimposed on top of such a grating, it suddenly appears to move unambiguously in the direction opposite to the walking direction of the human figure (Fujimoto & Sato, 2006; Fujimoto, Yagi, & Sato, 2009). Interestingly, the effect also occurs for other naturalistic objects that have a characteristic shape that we learn through experience to imply a particular movement direction (e.g., for car stimuli, but not an isolated wheel spinning).

In Experiment 2, we investigated the inverse of this effect. We hypothesized that a background with high certainty in terms of its motion direction could influence the perceived direction of a dynamic object with ambiguous global movements. Such spatial contextual effects could be explained in terms of a top-down modulation, or bias, due to prior knowledge and learned expectations with moving objects in the environment. In order to account for such contextual influences on form perception, we fit behavioral data with the Bayesian observer model by changing the prior probabilities associated with the underlying templates representing each direction of object movement.

Participants

Twenty-four participants were recruited through the Department of Psychology subject pool at the University of California, Los Angeles (UCLA), and were given course credit for their participation. All participants reported normal or corrected vision and gave informed consent approved by the UCLA Institutional Review Board. All participants were naïve to the stimuli and to the purpose of the study. Participants were assigned to one of two groups that performed the discrimination task either with biological human walker stimuli ($n = 12$, seven female and five male, mean age $= 20.3 \pm 1.1$), or with nonbiological rotating square stimuli ($n = 12$, seven female and five male, mean age $= 20.6 \pm 0.9$).

Materials and methods

The experimental setup and testing conditions were similar to Experiment 1, including the procedure for generating hybrid stimuli. The experiment began with a
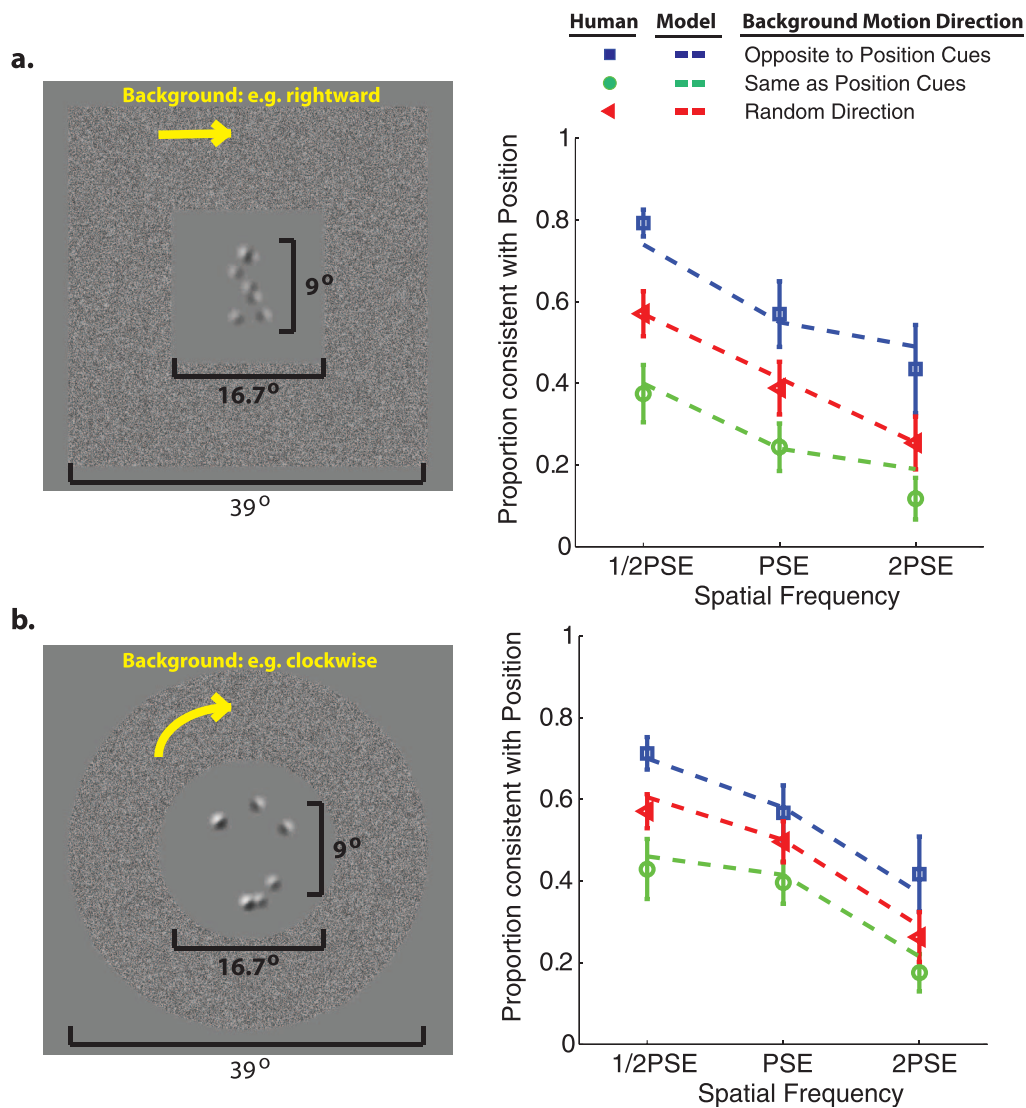
Figure 4. Illustration of stimulus design and spatial parameters for Experiment 2 (left), and subplots displaying human and model data for biological and nonbiological stimuli (right). For illustration purposes, the contrast is greatly increased in the schematic, and more elements are displayed as samples from the underlying shapes as compared to the parameters used in the actual experiment. Human data are represented by symbols and model data are represented as dashed lines. The module weights of the model were fit to the red data points in the random motion condition, and a single prior parameter was adjusted to best fit the remaining six data points (e.g., blue and green) in which a directional background motion signal was presented.

block of 36 trials using a one-up-one-down staircase (initiated at 0.3 c/°, step size = 0.05 c/°) to adjust spatial frequency until the PSE was estimated for equating the contributions from position and orientation cues for each subject. The estimated PSE was used subsequently as a benchmark to manipulate spatial frequency across three conditions in the main experiment. The three conditions represented ratios of the PSE including half of PSE, PSE, and double PSE. Based on results of our prior study (Thurman & Lu, 2014b), stimuli presented with half PSE are expected to produce more than 50% of responses consistent with position cues due to lower reliability (i.e., high uncertainty) of orientation cues, and those with double PSE to produce fewer than 50%

of responses with position cues due to relatively higher reliability (i.e., low uncertainty) of orientation cues.

In the main experiment, we introduced a field of luminance white noise (Michelson contrast = 0.1) generated from a uniform distribution to surround the centrally presented ambiguous stimulus (Figure 4). The central region presented walker or rotating square stimuli on a solid background surrounded by the noise field. For the biological stimulus, the random noise comprised a square region (width = 39°), and for the nonbiological stimulus the random noise comprised a circular region (diameter = 39°). Leftward and rightward motion signals were introduced to the random noise field by shifting a random subset of pixels (75%)

in one direction or the other on each frame at a rate of 6.7°/s, while the remaining pixels (25%) were generated randomly to have no inherent motion signal. Clockwise and counterclockwise motion was introduced to the noise field for nonbiological stimuli by rotating a subset of pixels (75%) on each frame at a rate of 240°/s. Random background motion served as a baseline for each condition, in which we generated white noise fields independently on each frame with no coherent motion signal. Participants were told explicitly to ignore the noise pattern surrounding the figure because it was unrelated and uninformative for the central discrimination task.

Participants performed a left versus right discrimination task on the walker stimuli in the biological group, and clockwise versus counterclockwise discrimination task in the nonbiological group. From subject responses, we computed the proportion of trial in which the perceived direction was consistent with position cues, where values less than 0.5 indicate perceptual reversals consistent with orientation cues. The experiment had a mixed design with 3×3 within-subject factors including three levels of spatial frequency (half PSE, PSE, double PSE), and three background motion directions for the biological stimulus (left, right, or random) and for the nonbiological stimulus (clockwise, counterclockwise, or random). Task type (biological vs. nonbiological) served as a between-subjects factor. In total, participants completed 180 trials, randomized and counterbalanced, for a total of 20 trials per condition.

Bayesian observer model

We used the same model as described in Experiment 1, but here we manipulated the priors associated with each stimulus direction during simulations. The priors introduced a slight bias for templates manifesting a particular movement direction. We modeled group-level data by first simulating trials without changing the prior parameter in order to estimate the appropriate weights given to position/orientation cues to match behavioral performance for each spatial frequency condition in which the background motion was random (e.g., where there were no coherent motion signals from the contextual surround). Once these weights were estimated, we adopted these weights but systematically adjusted the priors associated with each movement direction of the dynamic forms. We chose the best fitting prior weight that minimized the squared errors between behavioral data and model performance. Hence, once the three free weight parameters were fit to a subset of the data where no bias was present (e.g., due to random background motion), we fit only a single free parameter (the prior) to account for the remaining

six data points in which the background motion direction was manipulated.

Results

Mean behavioral results are displayed in Figure 4, in which the results of model simulations are represented as dashed lines. The direction of the moving surround exerted a significant influence on the perception of both biological and nonbiological stimuli. When the background moved in the direction opposite to position cues, participants were more likely to report the direction consistent with position cues, and when the background moved in the same direction as position cues, participants were more likely to report the other direction, consistent with orientation cues. This effect induced a relatively constant and systematic bias to perception across all spatial frequencies. The significance of these results was confirmed with a 3×3 within-subject analysis of variance (ANOVA), with stimulus type (biological vs. nonbiological) serving as a between-subjects factor. This analysis revealed significant main effects of spatial frequency, $F(2, 44) = 29.9$, $p < 0.001$, and background direction, $F(2, 44) = 28.5$, $p < 0.001$, and a nonsignificant main effect of the between-subjects factor stimulus type (biological vs. nonbiological), $F(1, 22) = 0.27$, $p = 0.61$. There were also no significant interaction effects between any within-subject factor and stimulus type (all p-values > 0.3), demonstrating that background motion exerted a comparable influence on perception of biological walkers and nonbiological square stimuli.

We fit the mean behavioral data with the Bayesian observer model by adjusting the prior probabilities associated with templates for each stimulus direction. Fitting this one parameter provided a strong fit to behavioral data, evaluated by computing the root mean squared errors (RMSE; biological $= 0.037$, nonbiological $= 0.029$). We found that a change in the prior probability from 0.5 (an uninformative prior) to 0.51 in favor of the direction opposite to background motion provided the best fit to behavioral data in the biological task, and that a prior of 0.507 provided the best fit for the nonbiological task. It is worth mentioning that, in the model, the prior is multiplied by the template likelihoods on each of the 60 frames presented on each trial, so the cumulative influence of the prior across the entire stimulus sequence is much greater than the magnitude of the fitted prior weights would seem to suggest. Taken together, these data clearly show that the contextual cues (e.g., background motion) exerted a strong and significant top-down influence on perception, causing a systematic bias in the perception of ambiguous dynamic forms.

Experiment 3

The second experiment revealed that background motion modulated perception to a similar degree for biological and nonbiological stimuli, suggesting that the learned priors were of a similar magnitude. In the natural environment, we usually see people walking along a path in a particular direction and often in groups of more than one person. Crowds are defined by a group of people walking together with a common direction. Based on this social experience, we might expect the brain to adopt a prior stating that humans tend to walk together in groups toward the same general direction. In fact, previous work demonstrates that human observers are keenly sensitive to social information provided by human crowd movements (Sweeny, Haroz, & Whitney, 2012). In contrast, we have much less experience with groups of rotating rigid objects, and probably much weaker priors stating that objects in groups tend to rotate together. Hence, in the biological motion case, we predict that the presence of a crowd of people should influence the perceived walking direction of an individual walker with ambiguous movements. However, in the rotating square case, a crowd of rotating objects should have a much weaker impact on the perceived direction of an object with ambiguous rotation.

In Experiment 3, we introduced a new environmental context to investigate the influence of a crowd of entities in dynamic form perception. We surrounded an ambiguous walker by a set of eight other point-light walkers sharing a common walking direction. All of the background walkers faced the same direction, either left or right, and this was either consistent or inconsistent with the direction implied by position/orientation cues in the ambiguous walker. Likewise, ambiguous rotating square stimuli were surrounded by point-light versions of the square objects all rotating in the same direction. We examined whether the presence of the crowd would exert a top-down influence on perception, and whether the strength of this influence would differ between biological and nonbiological objects.

Participants

Twenty-four participants were recruited through the Department of Psychology subject pool at the University of California, Los Angeles (UCLA), and were given course credit for their participation. All participants reported normal or corrected vision and gave informed consent approved by the UCLA Institutional Review Board, and were naïve to the stimuli and to the purpose of the study. Participants were assigned to one of two groups that performed the discrimination task either with biological human walker stimuli ($n = 12$, 7 female and 5 male, mean age = 20.8 $\pm$ 2.0), or with nonbiological rotating square stimuli ($n = 12$, 7 female and 5 male, mean age = 20.8 $\pm$ 3.7).

Materials and methods

The experimental setup and testing conditions were nearly identical to Experiment 2. We used the same procedure to estimate the PSE signaling an equal contribution on average from orientation and position cues, and included the same three conditions for manipulating spatial frequency (half PSE, PSE, and double PSE). However, in this experiment we introduced a crowd of point-light stimuli to surround the central ambiguous stimulus. The dimensions of the stimuli are reported in Figure 5, and were similar in spatial extent to the contextual cues introduced in Experiment 2. The surrounding point-light walkers were all in phase and walked in the same direction, while the overall direction of the crowd (left or right) was randomized on each trial. Point-light square shapes were created by placing points at the corners and at the midpoint of each edge segment and, like the walker stimuli, they rotated together in phase in the same overall direction (clockwise or counterclockwise). Participants were told to ignore the point-light stimuli surrounding the central ambiguous figure because their movements were unrelated and nonpredictive of the target stimulus direction.

The experiment had a mixed design with 3×3 within-subject factors including three levels of spatial frequency, and three crowd direction conditions for biological stimuli (left, right, or no crowd) and for nonbiological stimuli (clockwise, counterclockwise, or no crowd). Task type (biological vs. nonbiological) served as a between-subjects factor. In total, participants completed 180 trials, randomized and counterbalanced, for a total of 20 trials per condition.

We ran simulations of the Bayesian observer model using the same procedure as Experiment 2 to fit the mean behavioral data. We estimated the best fitting prior weight to account for changes in perception as a result of contextual information provided by the surrounding crowd of point-light stimuli.

Results

Mean behavioral results are displayed in Figure 5, in which the results of model simulations are represented as dashed lines. Contextual information provided by the crowd of point-light walkers exerted a significant influence on the perception of biological stimuli, but
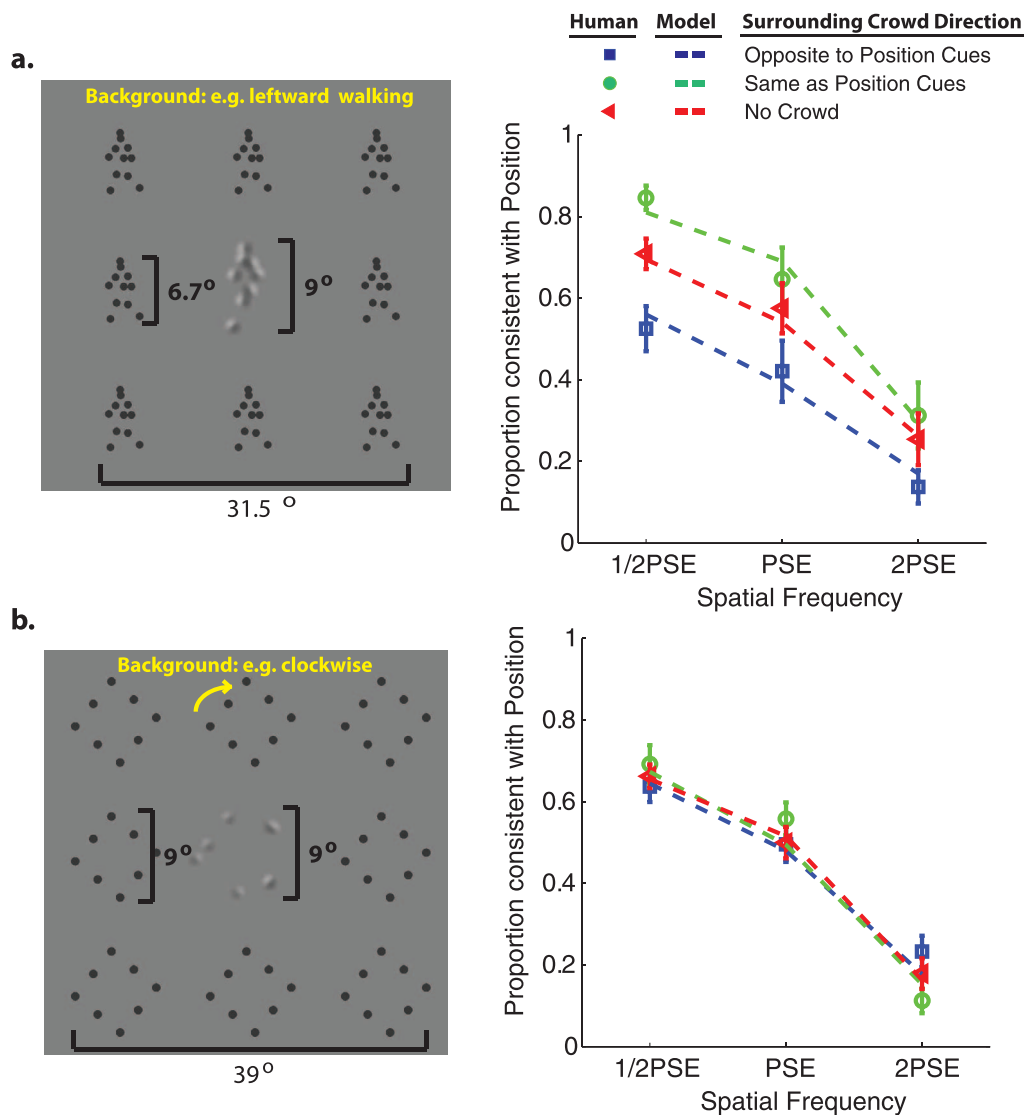
Figure 5. Illustration of stimulus design and spatial parameters for Experiment 3 (left), and subplots displaying human and model data (right) for each type of stimulus. For illustration purposes, the contrast is greatly increased in the images, and more elements are illustrated as samples from the underlying shapes as compared to those used in the actual experiment. Human data are represented by symbols and model data are represented as dashed lines. The module weights of the model were fit to the red data points and a single prior parameter was adjusted to best fit the remaining six data points (e.g., blue and green) in which a crowd direction signal was presented.

the surrounding crowd of rigidly rotating objects had no effect on perception of nonbiological stimuli. When the crowd walked in the same direction as position cues, participants were more likely to report the direction consistent with position cues, and when the crowd walked in the opposite direction to position cues, participants were more likely to report the other direction, consistent with orientation cues. Like Experiment 2, this effect induced a relatively constant and systematic bias to perception across all spatial frequencies. However, unlike Experiment 2, the directionality of the influence was reversed where perception was now biased toward the same direction as the contextual surround (e.g., the crowd direction). The

significance of these results was confirmed with a 3×3 within-subject ANOVA, with stimulus type (biological vs. nonbiological) serving as a between-subjects factor. This analysis revealed significant interaction effects between spatial frequency and direction, $F(4, 88) = 4.4$, $p = 0.004$, and between crowd direction and the between-subjects factor stimulus type (biological vs. nonbiological), $F(2, 44) = 8.2$, $p = 0.004$. The three-way interaction was nonsignificant, so we performed step-down 3×3 ANOVAs for each stimulus type separately to evaluate the source of the two-way interactions. This analysis revealed a significant main effect of crowd direction for the biological stimuli, $F(2, 22) = 12.9$, $p = 0.001$, but not for nonbiological stimuli, $F(2, 22) = 0.03$,

$p = 0.91$. In contrast to Experiment 2 in which contextual background motion modulated performance to a similar degree for both stimulus types, the presence of a social crowd had an asymmetric influence on perception, only impacting biological stimuli.

We fit the mean behavioral data with the Bayesian observer model by adjusting the prior probabilities associated with templates for each stimulus direction. Fitting this one parameter provided a strong fit to behavioral data, evaluated by computing the root mean squared errors (RMSE, biological = 0.03, nonbiological = 0.033). We found that a change in the prior probability from 0.5 (an uninformative prior) to 0.507 in favor of the same direction as the crowd provided the best fit to behavioral data in the biological task. As can be seen in the data presented in Figure 5b, the influence of the crowd on nonbiological stimuli was nonexistent, leading to an estimate for the prior that was practically uninformative (0.5002).

Discussion

In the current experiments, we created dynamic stimuli designed to tap specifically into form-based processes of visual analysis by limiting the usefulness of local motion cues. Using Gabor patches as elements we were able to place two form-based cues, spatial position and orientation, into direct conflict within a single stimulus, thus creating ambiguous stimuli that could be perceived as moving in either direction at once. This type of stimulus created a unique opportunity to study both how the brain resolves ambiguity about global shapes within the form-processing stream, and how contextual information and prior knowledge might bias perception via top-down modulation under such conditions of uncertainty.

One primary goal of the current study was to systematically compare dynamic form perception for biological stimuli (e.g., human walkers) to nonbiological stimuli (e.g., rotating squares), with the goal of understanding to what extent form processing may be specialized for human actions, or whether domain-general processes could account for perception of both types of stimuli. Experiment 1 was designed to address this issue by using the classification image technique to reveal how position and orientation cues are utilized across space and time in a bottom-up fashion to produce global percepts of dynamic shapes. Importantly, we implemented a Bayesian observer model that performed the same generic form-based computations for each type of stimulus to produce classification images for direct comparison to the human behavioral data. We discovered a strong qualitative agreement and a statistically significant correlation between the processing characteristics of human observers and the Bayesian observer model for both types of stimuli. This result provides compelling evidence that the human visual system engages a process analogous to probabilistic template matching on a frame-by-frame basis to analyze the time varying behavior of objects in motion, regardless of their biological nature.

These results reveal that when presented with uncertain and competing form information, the visual system combines these cues in an exquisite manner taking into account both the reliability of the low-level sensory cues (Thurman & Lu, 2014b), as well as feature-level differences between internal templates and the reliabilities associated with these feature differences in relation to the particular discrimination task at hand. These are both examples stimulus driven effects (e.g., bottom-up), in which specific aspects of the lower-level signals determine the globally perceived stimulus in a systematic and well-determined manner.

In the second and third experiments, we explored the possibility that global contextual information could also play a top-down role in modulating dynamic form perception. In Experiment 2, we found that perception of an ambiguous walking figure was biased toward the direction opposite to surrounding motion. Interestingly, we found that surrounding rotational motion exerted a similar degree of modulatory influence on nonbiological stimuli. This result suggests the existence of a generic prior stating that a background tends to move in the opposite direction to the movements of foreground objects, not just for translational motion of highly familiar objects (e.g., walkers) but also for rigidly rotating objects. Although the neural basis for this contextual bias is unknown, we suppose that it could be related to the center-surround antagonism observed for neurons in motion-sensitive regions of cortex such as the middle temporal area (MT). A study by Allman and colleagues (1985) revealed that MT neurons have a very large receptive field that is sensitive to antagonistic motion in the surround up to 100 times the size of the classic receptive field. Because neurons in area MT naturally integrate local stimulus information with surrounding contextual motion cues, they provide a viable candidate mechanism for explaining the effects observed in Experiment 2.

In Experiment 3, we explored another cue posited to have a modulatory influence on perception primarily due to social experience. We found that a crowd of human walkers facing in a particular direction could bias perception of an ambiguous walker toward the same direction as the crowd. This result could be due to prior knowledge of social cues in which humans tend to walk together as part of a group in the same general direction. Previous work has shown that human observers are highly sensitive to the global movement direction of human crowds (Sweeny et al., 2012).

Interestingly, we did not find a significant influence of crowd rotation direction on nonbiological objects, perhaps due to the fact that there is no social protocol for nonbiological objects to conform to a particular direction of group movement. Alternatively, the ability to perceptually group the movements of surrounding objects could be influenced by Gestalt laws of perceptual organization such as the law of common fate, which states that objects moving in the same direction tend to be grouped together to comprise a single perceptual unit. The effect of Gestalt grouping laws could have had a disproportionate effect on biological stimuli that appear to translate in a single direction, whereas nonbiological stimuli have a more complex rotational trajectory. It should be noted, however, that the walkers employed in the current study did not actually translate across the screen so Gestalt grouping laws would have to operate on higher-level representations such as the implied directionality of motion signatures or postural cues signaling the facing direction of the walkers.

Human brain imaging studies show that biological motion perception is supported by a large-scale cortical brain network that spans the ventral and dorsal processing streams, and extends from occipito-temporal cortex to high-level areas in frontal cortex (Grèzes et al., 2001; Grossman et al., 2000; Saygin, 2007; Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Thompson, Clarke, Stewart, & Puce, 2005). The top-down influence of socially relevant crowd information, which was specific to biological form processing, suggests that higher-level regions likely commute with perceptual representations in lower-level visual areas to presumably facilitate action understanding and effective social behaviors.

Because the model we developed to investigate bottom-up effects in form perception was cast in the framework of Bayesian inference (Thurman & Lu, 2014b), it has the inherent capacity to incorporate prior knowledge by changing the prior probabilities associated with internal representations. To account for data in Experiments 2 and 3, we fit a single parameter (e.g., the prior probabilities) to model the biases induced by contextual information. The model provided a robust fit to human behavioral data in both experiments, demonstrating the potency and flexibility of the Bayesian modeling approach to provide a comprehensive account of dynamic form perception. The current study demonstrates that this framework can generalize to multiple types of priors and to different types of dynamic objects including human actions with biological significance and rigidly moving objects with non-social significance.

It is generally assumed that human action perception involves neural pathways that are specialized for biological motion processing. In the current model,

human actions represent just one instance of a larger category of objects that manifest dynamic properties through their changes in shape and/or position over time. Although biological *motion* perception may be specialized in terms of motion-based processing (Thurman & Lu, 2013b; Thurman & Lu, 2014a; Troje & Westhoff, 2004; van Boxtel & Lu, 2012), the current study suggests that form-based processing of human actions engages the same generic computational mechanisms that handle nonbiological dynamic forms. Biological motion signals are indisputably unique and important in the natural environment, for instance because they provide pertinent social cues, but we believe that this type of specialized processing is more likely linked to motion-based systems. Recent studies have found that patients with brain damage to the ventral form-processing pathway, and who show generalized deficits in form perception, nonetheless retain the capacity to recognize point-light biological motion at normal levels (Gilaie-Dotan, Bentin, Harel, Rees, & Saygin, 2011; Gilaie-Dotan, Saygin, Lorenzi, Rees, & Behrmann, 2015). This suggests that although form analysis may be sufficient for recognizing human actions, the integrity of the form processing system is not a necessary condition for human action perception. However, studies have also shown that the integrity of the motion processing system is not necessary for recognizing human actions, likely due to compensation from form processing systems (McLeod, 1996; Vaina, Lemay, Bienfang, Choi, & Nakayama, 1990; Vangeneugden et al., 2014). Together, this shows that biological action perception involves processing within two networks that map roughly onto the dorsal and ventral processing streams. Under typical circumstances these systems complement each other and work together to provide the rich and robust processing capabilities for understanding and interacting with others. However, when one system is damaged, the other system appears able to compensate for this loss and retain functional abilities related to human action perception. It is critical that future work continues to investigate how these systems work together, while realizing that these systems differ significantly in terms of evolutionary, computational and functional properties. Hence, it may be just as important to characterize the properties of each system in isolation, and we believe that the current study provides a meaningful step in that direction.

Acknowledgments

Corresponding authors: Steven M. Thurman; Hongjing Lu.
Email: sthurman@ucla.edu; hongjing@ucla.edu.
Address: Department of Psychology, University of California, Los Angeles, CA, USA.

References

Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, 2(1):8, 121–131, doi:10.1167/2.1.8. [PubMed] [Article]

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, 14(2), 105–126.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, 5, 617–629.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, USA*, 103, 449–454.

Bardi, L., Regolin, L., & Simion, F. (2013). The first time ever I saw your feet: Inversion effect in newborns' sensitivity to biological motion. *Developmental Psychology*, 50, 986–993.

Beintema, J. A., Georg, K., & Lappe, M. (2006). Perception of biological motion from limited-lifetime stimuli. *Perception & Psychophysics*, 68, 613–624.

Beintema, J. A., & Lappe, M. (2002). Perception of biological motion without local image motion. *Proceedings of the National Academy of Sciences, USA*, 99, 5661–5663.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.

Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision*, 5(4):6, 348–360, doi:10.1167/5.4.6. [PubMed] [Article]

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, 35, 393–401.

Day, M., & Loffler, G. (2009). The role of orientation and position in shape perception. *Journal of Vision*, 9(10):14, 1–1, doi:10.1167/9.10.14. [PubMed] [Article]

de Lussanet, M. H. E., Fadiga, L., Michels, L., Seitz, R. J., Kleiser, R., & Lappe, M. (2008). Interaction of visual hemifield and body view in biological motion perception. *The European Journal of Neuroscience*, 27, 514–522. doi:10.1111/j.1460-9568.2007.06009.x.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293(5539), 2470–2473. doi:10.1126/science.1063414.

Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infants. *Science*, 218(4571), 486–487. doi:10.1126/science.7123249.

Fujimoto, K., & Sato, T. (2006). Backscroll illusion: Apparent motion in the background of locomotive objects. *Vision Research*, 46, 14–25.

Fujimoto, K., Yagi, A., & Sato, T. (2009). Strength and variability of the backscroll illusion. *Vision Research*, 49, 759–764.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, 4, 179–192. doi:10.1038/nrn1057.

Gilaie-Dotan, S., Bentin, S., Harel, M., Rees, G., & Saygin, A. P. (2011). Normal form from biological motion despite impaired ventral stream function. *Neuropsychologia*, 49, 1033–1043.

Gilaie-Dotan, S., Saygin, A. P., Lorenzi, L. J., Rees, G., & Behrmann, M. (2015). Ventral aspect of the visual form pathway is not critical for the perception of biological motion. *Proceedings of the National Academy of Sciences, USA, 112,* E361–E370.

Grèzes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *NeuroImage*, 13, 775–785.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12, 711–720.

Grossman, E. D., Jardine, N. L., & Pyles, J. A. (2010). fMR-adaptation reveals invariant coding of biological motion on the human STS. *Frontiers in*

Human Neuroscience, *4*, 15. doi:10.3389/neuro.09.015.2010.

Jastorff, J., & Orban, G. A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *The Journal of Neuroscience*, *29*, 7315–7329. doi:10.1523/JNEUROSCI.4870-08.2009.

Jastorff, J., Vanduffel, W., Popivanov, I. D., Orban, G. A., & Vogels, R. (2012). Integration of shape and motion cues in biological motion processing in the monkey STS. *NeuroImage*, *60*, 911–921. doi:10.1016/j.neuroimage.2011.12.087.

Keane, B. P., Lu, H., & Kellman, P. J. (2007). Classification images reveal spatiotemporal contour interpolation. *Vision Research*, *47*, 3460–3475.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Knoblauch, K., & Maloney, L. T. (2008). Estimating classification images with generalized linear and additive models. *Journal of Vision*, *8*(16):10, 1–19, doi:10.1167/8.16.10. [PubMed] [Article]

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *The Journal of Neuroscience*, *26*, 2894–2906. doi:10.1523/JNEUROSCI.4915-05.2006.

Lu, H. (2010). Structural processing in biological motion perception. *Journal of Vision*, *10*(12):13, 1–13, doi:10.1167/10.12.13. [PubMed] [Article]

Lu, H., & Liu, Z. (2006). Computing dynamic classification images from correlation maps. *Journal of Vision*, *6*(4):12, 475–483, doi:10.1167/6.4.12. [PubMed] [Article]

Lucia, R., Luca, T., & Giorgio, V. (2000). Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Animal Cognition*, *3*(1), 53–60. doi:10.1007/s100710050050.

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London B: Biological Sciences*, *249*(1325), 149–155.

McLeod, P. (1996). Preserved and impaired detection of structure from motion by a "motion-blind" patient. *Visual Cognition*, *3*, 363–392.

Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, *11*(5):2, 1–25, doi:10.1167/11.5.2. [PubMed] [Article]

Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of*

Cognitive Neuroscience, *6*(2), 99–116. doi:10.1162/jocn.1994.6.2.99.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519–526.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

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

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, *3*, 1199–1204.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain: A Journal of Neurology*, *130*, 2452–2461.

Saygin, A. P., Wilson, S. M., Hagler, D. J., Bates, E., & Sereno, M. I. (2004). Point-light biological motion perception activates human premotor cortex. *Journal of Neuroscience*, *24*, 6181–6188.

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, *8*, 522–535.

Simion, F., Di Giorgio, E., Leo, I., & Bardi, L. (2011). The processing of social stimuli in early infancy: From faces to biological motion perception. *Progress in Brain Research*, *189*, 173–193. doi:10.1016/B978-0-444-53884-0.00024-5.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences, USA*, *105*, 809–813.

Singer, J. M., & Sheinberg, D. L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. *Journal of Neuroscience*, *30*, 3133–3145.

Sweeny, T. D., Haroz, S., & Whitney, D. (2012). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 329–337.

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). fMRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology*, *98*, 1626–1633.

Theusner, S., de Lussanet, M., & Lappe, M. (2014). Action recognition by motion detection in posture space. *The Journal of Neuroscience*, *34*, 909–921. doi:10.1523/JNEUROSCI.2900-13.2014.

Thompson, J. C., & Baccus, W. (2011). Form and motion make independent contributions to the

response to biological motion in occipitotemporal cortex. *NeuroImage*, *59*, 625–634. doi:10.1016/j.neuroimage.2011.07.051.

Thompson, J. C., Clarke, M., Stewart, T., & Puce, A. (2005). Configural processing of biological motion in human superior temporal sulcus. *Journal of Neuroscience*, *25*, 9059–9066.

Thurman, S. M., Giese, M. A., & Grossman, E. D. (2010). Perceptual and computational analysis of critical features for biological motion. *Journal of Vision*, *10*(12):15, 1–14, doi:10.1167/10.12.15. [PubMed] [Article]

Thurman, S. M., & Grossman, E. D. (2008). Temporal "bubbles" reveal key features for point-light biological motion perception. *Journal of Vision*, *8*(3):28, 1–11, doi:10.1167/8.3.28. [PubMed] [Article]

Thurman, S. M., & Lu, H. (2013a). Complex interactions between spatial, orientation, and motion cues for biological motion perception across visual space. *Journal of Vision*, *13*(2):8, 1–18, doi:10.1167/13.2.8. [PubMed] [Article]

Thurman, S. M., & Lu, H. (2013b). Physical and biological constraints govern perceived animacy of scrambled human forms. *Psychological Science*, *24*, 1133–1141.

Thurman, S. M., & Lu, H. (2014a). Perception of social interactions for spatially scrambled biological motion. *PLoS One*, *9*(11), e112539.

Thurman, S. M., & Lu, H. (2014b). Bayesian integration of position and orientation cues in perception of biological and non-biological forms. *Frontiers in Human Neuroscience*, *8*, 91. doi:10.3389/fnhum.2014.00091.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: evidence for a "life detector"? *Current Biology*, *16*(8), 821–824.

Vaina, L. M., Lemay, M., Bienfang, D. C., Choi, A. Y., & Nakayama, K. (1990). Intact "biological mo-tion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience*, *5*, 353–369.

Vallortigara, G., & Regolin, L. (2006). Gravity bias in the interpretation of biological motion by inexpe-rienced chicks. *Current Biology*, *16*(8), R279–R280. doi:10.1016/j.cub.2006.03.052.

Vallortigara, G., Regolin, L., & Marconato, F. (2005). Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biology*, *3*(7), e208.

van Boxtel, J. J. A., & Lu, H. (2012). Signature movements lead to efficient search for threatening actions. *PLoS One*, *7*(5), e37085, 1–6. doi:10.1371/journal.pone.0037085.

van Boxtel, J. J. A., & Lu, H. (2013). A biological motion toolbox for reading, displaying, and ma-nipulating motion capture data in research settings. *Journal of Vision*, *13*(12):7, 1–13, doi:10.1167/13.12.7. [PubMed] [Article]

van Boxtel, J. J. A., & Lu, H. (2015). Joints and their relations as critical features in action discrimina-tion: Evidence from a classification image method. *Journal of Vision*, *15*(1):20, 1–17, doi:10.1167/15.1.20. [PubMed] [Article]

Vangeneugden, J., De Mazière, P. A., Van Hulle, M. M., Jaeggli, T., Van Gool, L., & Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. *Journal of Neuroscience*, *31*, 385–401.

Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *The Journal of Neuroscience*, *34*, 574–585.

Weiner, K. S., & Grill-Spector, K. (2011). Not one extrastriate body area: Using anatomical land-marks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipi-totemporal cortex. *NeuroImage*, *56*, 2183–2199.