Pigeons and humans use action and pose information to categorize complex human behaviors

Muhammad A.J. Qadri *, Robert G. Cook

Department of Psychology, Tufts University, USA

ARTICLE INFO

ABSTRACT

The biological mechanisms used to categorize and recognize behaviors are poorly understood in both human and non-human animals. Using animated digital models, we have recently shown that pigeons can categorize different locomotive animal gaits and types of complex human behaviors. In the current experiments, pigeons (go/no-go task) and humans (choice task) both learned to conditionally categorize two categories of human behaviors that did not repeat and were comprised of the coordinated motions of multiple limbs. These "martial arts" and "Indian dance" action sequences were depicted by a digital human model. Depending upon whether the model was in motion or not, each species was required to engage in different and opposing responses to the two behavioral categories. Both species learned to conditionally and correctly act on this dynamic and static behavioral information, indicating that both species use a combination of static pose cues that are available from stimulus onset in addition to less rapidly available action information in order to successfully discriminate between the behaviors. Human participants additionally demonstrated a bias towards the dynamic information in the display when relearning the task. Theories that rely on generalized, non-specific visual mechanisms involving channels for motion and static cues offer a parsimonious account of how humans and pigeons recognize and categorize behaviors within and across species.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

How humans and computers recognize and classify behaviors has been of increasing theoretical focus (Blake & Shiffrar, 2007; Dittrich, 1993; Poppe, 2010). Because behaviors are temporally extended, dynamic, and organized series of semi-rigid, articulated motions by an agent, they have been challenging to create as stimuli, test as cues, and model or represent symbolically. Most behaviors have both form cues (poses) and motion cues (actions). The form cues consist of the 2-dimensional or 3-dimensional representation of the spatial configuration of an agent's body parts. The motion cues can be either the ordering of the poses or the directional motion of the agent's body parts, likely using the same 2-dimensional or 3-dimensional space in which the poses are represented. Given this relationship, simultaneously presenting both pose and action cues is almost unavoidable when testing realistic behavior stimuli, such as in videos.

Previous investigations in humans have concentrated on the integrality of motion to the perception of behavior (Decety & Grèzes, 1999; Johansson, 1973; Troje, Westhoff, & Lavrov, 2005). The majority of studies on action recognition in humans has used point-light displays (PLDs; Johansson, 1973), for example, because they ostensibly eliminate pose information (although see Thirkettle, Benton, & Scott-Samuel, 2009). Consequently, the extensive investigation of separate pathways and structures has primarily focused on the relationship between behavior recognition and motion processing (Oram & Perrett, 1994; Ptito, Faubert, Gjedde, & Kupers, 2003). The contribution of static features, such as poses, has not been investigated nearly as much. Even when static agents are presented with implied motion, the investigations have centered on the activation or perception of motion (Kourtzi & Kanwisher, 2000; Shiffrar & Freyd, 1990). Contrastingly, the study of behavior recognition by computers often relies on processing poses or integrating the poses over time (Poppe, 2010) because of its relative simplicity. Altogether, these different approaches suggest that both action cues and pose cues play important roles in the recognition of complex behaviors.

The computational visual models that describe human action recognition are of two major types. One computational line of work

---

* Corresponding author at: Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, USA.

E-mail address: Muhammad.Qadri@tufts.edu (M.A.J. Qadri).

has operated using neural networks modeled after primate vision, focused on using both form and motion cues (Fleischer, Caggiano, Thier, & Giese, 2013; Giese & Poggio, 2003). In these models, separate pathways process the form and motion information independently, extract pose and action information from each separately, and then the separate results are combined to solve the specific task at hand. These motion-and-form mixture models have had success in modeling human performance in canonical biological motion tasks (Giese & Poggio, 2003). Consistent with these mixture models, studies of human and non-human primates have suggested that action and pose cues are processed in both shared and distinctive neural regions (Jastorff & Orban, 2009; Jastorff, Popivanov, Vogels, Vanduffel, & Orban, 2012; but see Vangeneugden, Peelen, Tadin, & Battelli, 2014).

A set of form-only models has focused on people's abilities to discriminate point light displays by matching the static point-lights to possible static form templates. To solve tasks of order or direction, the relative activation of these static templates over time is integrated. Thus, any action or motion concepts are derived from pose changes instead of from motion features in the environment. These form-only models have been successful at imitating human performance, despite noise in PLDs so severe as to prohibit any sort of true motion perception (Lange, Georg, & Lappe, 2006; Lange & Lappe, 2006; Theusner, de Lussanet, & Lappe, 2014; although see Casile and Giese (2005) for an alternative motion-based explanation). Both the motion-and-form and the form-only models have been developed to explain the visual aspect of action recognition in humans, and both models represent behaviors as series of snapshots of the pose and/or action.

The recognition and categorization of behavior is of equal importance to non-human animals (Asen & Cook, 2012; Dittrich, 1993). Yet, the cognitive and neural mechanisms of action recognition in animals have received far less attention than in computers and humans. The difficulty of controlling and displaying behavioral stimuli to investigate such questions in animals has seriously hampered their investigation. For example, the testing of biological motion stimuli in non-humans has produced at best only mixed results regarding the perception of action (Blake, 1993; Dittrich, Lea, Barrett, & Gurr, 1998; Parron, Deruelle, & Fagot, 2007; Qadri, Asen, & Cook, 2014; Regolin, Tommasi, & Vallortigara, 2000; Troje & Aust, 2013; Vallortigara, Regolin, & Marconato, 2005). The reason for these mixed results is debated (Qadri & Cook, 2015), but one potential source of animals' failure to discriminate is a cognitive difficulty in grouping or simultaneously processing separated elements on the display. Regardless, these outcomes suggest that we need an alternative approach to investigating action recognition in animals.

We have successfully examined behavior recognition in pigeons by using controlled digital models to depict different types of actions. This research has suggested that these highly visual birds can learn to categorize types of repetitive locomotive actions (walking vs. running) by different digitally animated animals as well as different kinds of complex behaviors (martial arts vs. Indian dance) as depicted by digitally animated humans (Asen & Cook, 2012; Qadri, Asen, & Cook, 2014; Qadri, Sayde, & Cook, 2014). By using full-featured, connected, digitally rendered models, these investigations avoided the difficulty of connecting separated elements which may have critically limited prior investigations using PLDs.

In each study, the pigeons learned to discriminate among the contrasting behaviors and demonstrated transfer of this discrimination to novel exemplars of each class of behavior. Furthermore, we found that both pigeons and humans show a dynamic superiority effect (DSE), in which dynamic video presentations of actions are discriminated better than static presentations of single frames from the same videos (similar to Cook & Katz, 1999; Koban & Cook,

2009). Further, we have attempted to identify whether the pigeons learn to process either the sequences of motion in these dynamic conditions as global actions or alternatively the more local motion flow features in the stimuli. Using different type of tests, the pattern of results have been more consistent with them perceiving the generalized motion of the agent's body as the basis of discrimination (Asen & Cook, 2012; Qadri, Asen, & Cook, 2014).

In the current experiment, we examine the hypothesis that both humans and pigeons use a combination of immediately available static pose information and subsequently perceived dynamic action cues to discriminate behavioral categories (i.e., as in Qadri, Sayde, & Cook, 2014). To verify this possibility, we isolated and separated the contribution of pose and action cues in the current experiments by placing them in conflict. We developed and tested a conditional cuing paradigm in which both the pigeons and humans needed to perform different conditional responses depending on whether the actions displayed were being presented dynamically (i.e., in motion/pose sequence) or statically. Thus, for instance, when Indian dance was presented dynamically (i.e., in motion/pose sequence), both species performed one response (e.g., pigeons – a go response, humans – a left choice), and when presented statically, the opposite response (pigeons – a no-go response, humans – a right choice). These response contingencies were reversed for displays containing the martial arts behavior. Consequently, this design puts the cues in competition, so successful performance requires the separation and recognition of pose and action cues for each behavior. Thus, the exclusive use of only pose information or only action information could not successfully be employed to solve the task, because exclusive use of one cue or the other would cause interference on the trials with the reversed contingencies. Using this conditional procedure, we examined the same two classes of complex behaviors (martial arts vs. Indian dance) as investigated by Qadri, Sayde, and Cook (2014) with both pigeons and humans.

In this experiment, humans and pigeons were tested with dynamic video and static image presentations of the two different actions depicted by the same digital human model. The pigeons were tested in a go/no-go task and the humans in a two-alternative forced-choice task. For both species, the experiment had two phases. In the first phase, the action and pose information were *consistent cues*. Both species learned to discriminate behaviors in which the dynamic and static presentations of each were mapped to the same responses (i.e., as in Qadri, Sayde, & Cook, 2014). Each cue independently indicated the correct discriminative response. In the second phase, action and pose cues were made into *conditional cues*. This was done by reversing the responses required for the dynamic (or static) presentations. The details of these stimulus-response assignments for each phase are in Table 1. Thus, for both species, identifying both the behavior depicted and the cues (action or pose) conveying that information was needed for successful discrimination. If both species learn the task, the result would be most consistent with mixture models in which dynamic and static cues independently contribute to the computation of action recognition (i.e., Giese & Poggio, 2003).

2. Methods

2.1. Participants

Three male pigeons (*Columba livia*) and 36 Tufts University undergraduates (*Homo sapiens*; 12 females) were tested. The pigeons were housed and tested at 80–85% of their free-feeding weights, with ad libitum grit and water in their home cage. These pigeons had previously been in a study examining the time-course of learning the basic action recognition task (unpublished). The

Table 1

Reinforcement contingencies. For conditional cuing, the underlined values are reversed when compared to consistent cuing. For half of the human participants, the category assignment for static trials was reversed, while for the remainder the dynamic trial assignment was reversed. For two of the pigeon subjects and approximately half of the human participants, Behavior A corresponded to "Indian dance" and Behavior B corresponded to "martial arts." For the third pigeon and the remaining human participants, this was reversed.

	Behavior	Trial Mode	Pigeon Category	Human Category Dynamic Reversed	Human Category Static Reversed
Consistent Cuing	A	Dynamic	S+	Left	Left
		Static	S+	Left	Left
	B	Dynamic	S−	Right	Right
		Static	S−	Right	Right
Conditional Cuing	A	Dynamic	<u>S−</u>	<u>Right</u>	Left
		Static	S+	Left	<u>Right</u>
	B	Dynamic	<u>S+</u>	<u>Left</u>	Right
		Static	S−	Right	<u>Left</u>

humans participated for course credit and provided signed consent prior to the experimental session. All had normal or corrected vision. All animal (IACUC) and human (IRB) procedures were approved by Tufts University.

2.2. Apparatus

Pigeons: A touchscreen-equipped (EZ-170-WAVE-USB) operant chamber was used to present video stimuli and record peck responses. Stimuli were displayed on an LCD computer monitor (NEC LCD 1525X; 1024 × 768, 60 Hz refresh rate) recessed 8 cm behind the touchscreen. Mixed grain reward was delivered via a central food hopper positioned beneath the touchscreen. A 28 V houselight in the ceiling was constantly illuminated, except during timeouts.

Humans: Testing was conducted using a computer and LCD monitor (Dell 1906 FPt; 1280 × 1024, 60 Hz refresh). Participants were seated in front of the monitor and permitted to adjust the position of the monitor and seat. Choice responses were recorded using a USB mouse recording at 500 Hz.

2.3. Stimuli

In these stimuli, a digital human model (motioncapturesociety.com) engaged in either an "Indian dance" or "martial arts" sequence of poses (available from mocap.cs.cmu.edu), rendered to PNG via digital software (Poser 8, SmithMicro). The digital model was presented within a 118 mm blue frame such that the figure subtended an average estimated visual angle of 3.4° × 6.9° for humans and 12.4° × 25.0° for pigeons. Each AVI video consisted of 600 frames that were assembled using MATLAB (MathWorks) with the Cinepak codec. Each class of behavior was depicted with

Fig. 1. Complex action figures with motion depiction. Each panel shows a single frame from each of the behavior categories. The different lines trace out the path of motion for the head, one hand, and one foot. A version of this figure also appeared in Qadri, Sayde, and Cook (2014).

four different exemplars involving different pose sequences (see Fig. 1). Each exemplar was rendered from twelve different camera perspectives (two distances × two elevations × three directions). On dynamically presented trials, the video started from a randomly selected point within the first 300 frames and was played continuously at approximately 33 frames per second over the 20-s trial duration. Starting from a random frame prevented early item-specific cues, but restricting the start frame to within the first half of the video ensured that the discontinuity from frame 600 to frame 1 occurred late in the trial so it could not be used as a cue for discrimination early in the presentation. On static presentation trials, a single randomly selected frame from these videos was displayed for the 20 s trial duration.

2.4. Procedure

Pigeons: The pigeons were first trained to discriminate the stimuli using the consistent cue condition. For two of the pigeons, the "Indian dance" behavior was designated the S+ so that pecks to these dynamic or static behaviors were reinforced with 2.9 s access to the food hopper on a concurrent VI-10 and FI-20 schedule (i.e., regular, post-trial reinforcement). For these two pigeons, the "martial arts" behavior was designated the S− so that pecks to these dynamic or static stimuli contributed to a variable timeout, and no within-trial or post-trial food reinforcement was given. The third pigeon had the reverse contingency (i.e., "martial arts" was the S+ and "Indian dance" was the S−). Each trial was initiated by a single peck to a circular, 2.5 cm white ready signal. This signal was then replaced by the static or dynamic display for 20 s. After this 20-s duration, the video was removed and post-trial food reinforcement was given for S+ trials or a variable timeout (0.5 s per peck) was given for S− trials, both which were then followed by a 3-s inter-trial interval.

In each session, the pigeons saw each of the four pose sequences for both behaviors 12 times each. These were presented as six static and six dynamic trials, each with a randomly selected perspective. Thus, sessions were comprised of 96 total trials, 48 S+ and 48 S− trials, and half were presented in the static condition and half were presented in the dynamic condition. To obtain peck rates for the S+ trials without the food hopper interruption, 12 randomly selected S+ trials were designated to be probe trials in which the hopper was not activated. The peck rates from these probe S+ trials were used to evaluate pecking in the S+ condition.

First, the pigeons were trained to discriminate the depicted behaviors regardless of the static or dynamic presentation condition. In this *consistent cues discrimination*, the dynamic features and static features in the display cued the same response. We report the final eight sessions of this consistent training to evaluate performance prior to the partial reversal. After obtaining estimates of performance with this consistent cuing, the contingencies for
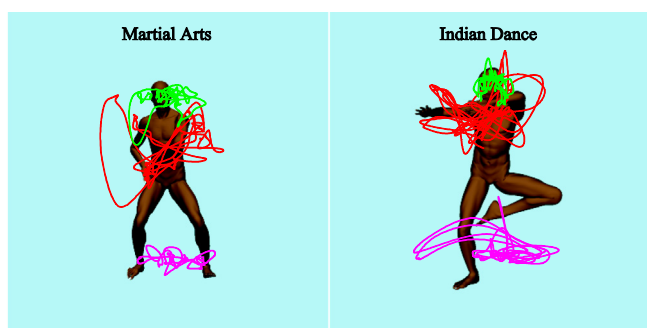
the dynamic trials were reversed as depicted in Table 1. For the two pigeons who had "Indian dance" as the S+, this partial reversal meant that during dynamic trials "martial arts" was the S+ and "Indian dance" was the S−, while the assignment for static presentations remained the same. In this *conditional cues discrimination*, the dynamic and static features in the display indicate conflicting responses. The composition of the session remained the same, with 48 S+ (12 of these were randomly selected to be non-reinforced probes) and 48 S− trials. The pigeons were trained in this discrimination for 60 sessions.

The pigeons' discrimination across the experiment was measured using ρ (rho; i.e. as in Herrnstein, Loveland, & Cable, 1976). This measure of discrimination accuracy can be computed by normalizing the Mann-Whitney U score. The total count of pecks produced over the final 10 s of stimulus presentation were evaluated, because behavior during this portion of the trial typically best represents asymptotic discrimination (e.g. see Cook, Beale, & Koban, 2011; Cook & Roberts, 2007; Koban & Cook, 2009). A ρ of 1 indicates perfect discrimination, indicating that all S+ probe trials contained more pecks in the terminal 10 s than all of the S− trials, while a ρ of 0.5 indicates chance-level performance with the different conditions not separated by rank.

Humans: Human participants were tested with the same action discrimination as the pigeons. Participants were tested with eight blocks of 64 trials (512 total trials), in which the four pose sequences depicting both behaviors were displayed equally in dynamic and static conditions while perspective randomly varied (2 behaviors × 4 exemplars × 2 presentation conditions × 4 repetitions = 64 trial block). Trials were initiated by moving the mouse cursor over a white circle at the center of the screen. This circle was then replaced by the stimulus, which was displayed for up to 20 s. Two white square choice keys were also displayed, vertically centered and equidistant from the center of the display on both sides of the stimulus. For half of the participants, dynamic and static stimuli depicting "martial arts" were initially assigned to the left key, and those depicting "Indian dance" were initially assigned to the right key. For the remaining participants, this contingency was reversed. After the participants made a choice, a green or red box was displayed to indicate accuracy, with a green box representing correct responses and red representing incorrect responses.

All participants were trained for the first four blocks (256 total trials) to perform the *consistent cues discrimination*, in which dynamic and static cues independently and accurately identified the correct response. If the participants performed with at least 75% accuracy for both dynamic and static trials (evaluated separately) in the last 48 trials of the fourth block (trials 209–256 inclusive), their contingencies were then partially reversed as described in Table 1. For these participants, the remaining four blocks (256 trials) of the session trained the *conditional cues discrimination*. The first half of the participants tested had their dynamic contingencies reversed and the remaining half had their static contingencies reversed. Participants who did not meet the 75% accuracy criterion were instead given four additional blocks of the *consistent cue discrimination*, but that data is omitted. Sessions typically lasted from 30 min to 1 h.

In order to best imitate the experience of the pigeons, participants received minimal instructions. They were told "In this study, you will be asked to make choices about stimuli." They were then told to move the mouse cursor over the white circle at the center of the screen (i.e. the same shape as the pigeons' 2.5 cm ready signal) to advance the session to the stimulus, and they were told that they would make a response by moving the mouse cursor over one of the two white squares located on either side of the stimulus. Finally, they were told that a red (incorrect) or green (correct) box would appear after their choice to indicate whether they got the

trial right or wrong. Thus, the participants were given a goal and information regarding how to advance the session, but no information regarding the purpose, nature, or content of the displays or the experiment.

3. Results

3.1. Pigeons

Consistent Cues Discrimination: The pigeons knew the categorical discrimination already because of their prior training. This can be seen in the above-chance discrimination in the results left of the dashed vertical line in Fig. 2. This section depicts discrimination in the last eight sessions of the consistent cues training phase. In this phase, the pigeons were readily discriminating between the actions (across bird $t(2) = 7.1$, $p = 0.019$, $d = 4.1$; all tests use $\alpha = 0.05$). All three pigeons individually evidenced discrimination in both the dynamic and static conditions over these eight sessions (using individual sessions as independent observations; $ts(7) > 3.1$, $ps < 0.019$, $ds > 1.1$).

These pigeons also exhibited the same DSE as reported previously, in which the dynamic condition (Fig. 2, black symbols) supports superior discrimination compared to the static condition (Fig. 2, white symbols). Over the terminal eight sessions in the consistent cues phase, mean discrimination in the dynamic condition ($\rho_{dynamic} = 0.97$, SE = 0.03) was better than in the static discrimination ($\rho_{static} = 0.82$, SE = 0.08). Because one pigeon performed much more poorly on the static discrimination than the other two, the group statistic for the static discrimination only neared significance ($t(2) = 3.6$, $p = 0.064$), but all three pigeons exhibited better discrimination in the dynamic condition than the static condition (all three within-subject $ts(7) > 2.4$, $ps < 0.045$, $ds > 0.9$).

While overall session-wise peck rates are traditionally informative about the pigeons' ability to discriminate different conditions, the examination of peck rates within a trial further reveal how the pigeons process the displays and extract the digital actor's behavior. The left panel of Fig. 3 shows standardized peck rates during the first 10 s of a trial for the positive S+ and negative S−, dynamic and static conditions during consistent cues training. Standardized peck rates were used to remove individual differences among the pigeons in overall rate of pecking. Peck rates were standardized by dividing each bird's peck rates by the maximum pecks per bin
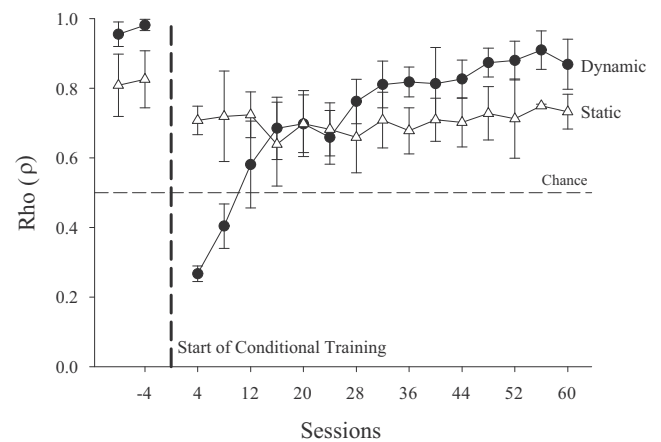
Fig. 2. Pigeons' overall discrimination. The pigeons' discrimination (as indexed by ρ) is depicted as a function of training. These data are reported relative to the start of the conditional cue training (i.e. relative to the start of the partial reversal training) in 4-session blocks. Negative time values indicate pre-reversal session blocks. The vertical dashed line indicates the start of conditional cue training. Error bars depict standard error.
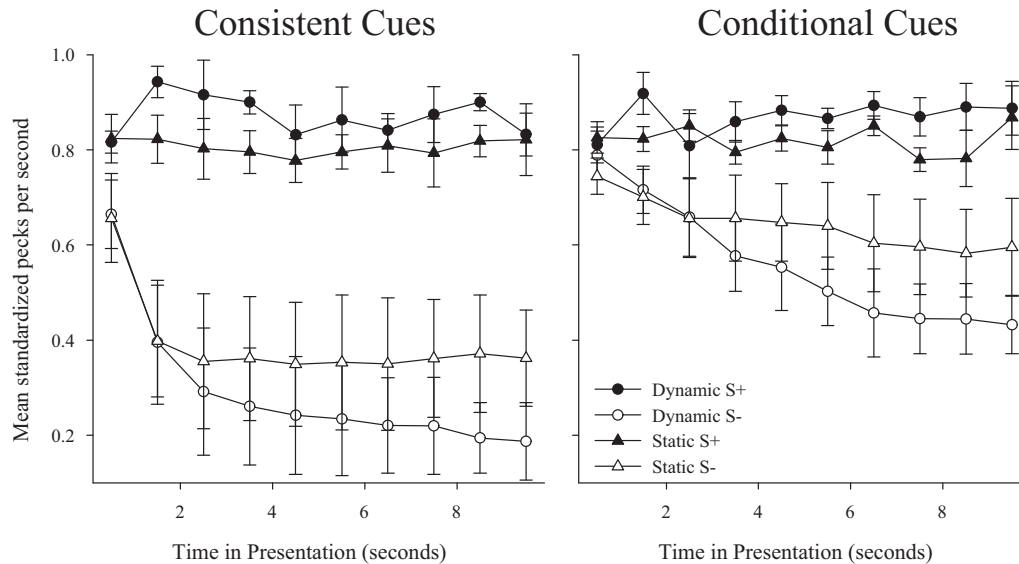
Fig. 3. Pigeons' within-trial peck rates. The left panel shows data from the last eight sessions of consistent cue training, and the right panel shows data from the last eight sessions of conditional cue training. Each panel depicts peck rates for the first ten seconds of the trial, standardized to accommodate individual differences in overall peck rates. The filled symbols correspond to peck rates on S+ trials and the open symbols correspond to peck rates on S− trials, while the circle symbols indicate performance on dynamic trials and the triangle symbols indicate performance on static trials. Error bars depict standard error.

prior to averaging across birds. Unlike behavior over the last 10 s of the trial, which exhibits stable peck rates and best indexes the quality of the discrimination as captured by our discrimination index ρ, the behavior over the first 10 s reflects the time course of the discrimination as the pigeons process the contents of the display.

Three major features are evident in the left panel of Fig. 3. First, peck rates on S+ and S− trials separated rapidly, indicating that the discriminative information is available within the first seconds of presentation. Second, there was little difference between pecks to the dynamic and static S− presentations over the first few seconds of rapid initial suppression. Third, the DSE emerges over the latter portions of the trial as pecking during dynamic S− trials suppressed to a lower level than during static S− trials. There is little evidence of a comparable difference on the S+ trials.

These patterns were confirmed by a repeated measures (RM) ANOVA (Presentation mode × Time in trial) evaluating change in pecking over time for the S− trials. It confirmed the visible main effect of time in trial ($F(9,18) = 10.6$, $p < 0.001$, $\eta_p^2 = 0.842$) which documents that peck rates suppress over this initial interval. The ANOVA also confirmed the DSE in a main effect of presentation mode ($F(1,2) = 32.8$, $p = 0.029$, $\eta_p^2 = 0.943$) and reported a significant interaction between presentation mode and time in trial ($F(9,18) = 12.9$, $p < 0.001$, $\eta_p^2 = 0.865$). The significant interaction between presentation mode and time in trial suggests that the DSE emerges over the course of the trial. Qadri, Sayde, and Cook (2014) explained an identical pattern of results by suggesting that pose information controlled initial discrimination in both presentation conditions, while action information emerged with continued viewing of the motion. Despite some small visible differences between dynamic and static S+ trials, an analogous RM ANOVA (Presentation mode × Time in trial) found no effects of presentation mode or the time within the presentation on S+ peck rates; presentation mode $F(1,2) = 5.3$, $p = 0.147$, time $F(9,18) = 0.9$, $p = 0.560$, interaction $F(9,18) = 1.2$, $p = 0.333$.

Conditional Cues Discrimination: Not unexpectedly, reversing the mapping of reinforcement contingencies for the dynamic condition visibly impacted the pigeons' behavior. The pigeons initially performed their established behavior on the reversed dynamic condition causing it to fall to below chance levels for a number of

sessions. However, all three pigeons learned to properly respond to the new conditional contingencies over the next 60 sessions. The learning of this conditional cue over sessions can be seen in the rightward portion of Fig. 2.

During the first four-session block of conditional training, the discrimination of dynamic presentations was significantly below chance ($t(2) = 10.6$, $p = 0.008$, $d = 6.1$) as the pigeons initially responded according to their previous training. Over the next 20 sessions, the pigeons adjusted to the conflicting contingencies of the dynamic and static conditions. A one-way repeated measures ANOVA of discrimination in the dynamic condition revealed a significant main effect of block ($F(14, 28) = 12.2$, $p < 0.001$, $\eta_p^2 = 0.860$) as $\rho_{Dynamic}$ gradually increased with training. While the pigeons differed in the rate at which they learned this new contingency, by session 32 the pigeons as a group were significantly performing above chance with the new reversed contingency ($t(2) = 4.6$, $p = 0.044$, $d = 2.7$). This discrimination continued to improve over the 60 total sessions of training, with $\rho_{Dynamic}$ well above chance over the final 4-session block ($t(2) = 5.1$, $p = 0.036$, $d = 3.0$).

The pigeons' discrimination of the established static condition was also affected during the learning of this reversed contingency, though not as strongly. Each pigeon showed brief periods of disruption in static discrimination as they began to accommodate the new dynamic contingencies. However, a one-way RM ANOVA evaluating change in ρ_{Static} over time found no significant effect of block, $F(14, 28) = 0.3$, $p = 0.984$. On the final 4-session block of training, ρ_{Static} remained well above chance ($t(2) = 4.6$, $p = 0.044$, $d = 2.7$). Interestingly, the DSE re-emerged after the extensive training for two of the three pigeons. Examining the last eight sessions, two of the pigeons show a strong DSE ($ts(7) > 4.1$, $ps < 0.004$, $ds > 1.4$) while the third did not ($t(7) = 1.5$, $p = 0.19$).

Thus, by the end of training, all three pigeons were correctly and conditionally responding to the conflicting behavioral categories in both the dynamic and static conditions. Comparing the last eight sessions of the consistent cues discrimination and the last eight sessions of the conditional cues discrimination suggested the latter phase was more difficult, as the pigeons' performance was better when the dynamic and static conditions had the same assigned response contingencies ($t(2) = 5.7$, $p = 0.03$, $d = 3.3$).

We again examined the within-trial peck rate curves from the final eight sessions of the conditional cues training to examine how the pigeons now processed the dynamic and static displays. The right panel of Fig. 3 depicts the peck rates over the first ten seconds of the trial from this period of training. Again, changes in peck rates on S− trials are the principal indicators of discrimination. Similar to the consistent cues phase, peck rates on dynamic S− presentations had an extended phase of suppression as compared to static trials. Examining peck rates for S− trials using an RM ANOVA (Presentation mode × Time in trial) confirmed that the pigeons' peck rates suppressed as the trial progressed by a main effect of time in trial ($F(9,18) = 11.7$, $p < 0.001$, $\eta_p^2 = 0.85$) and that the degree of this suppression depended on the presentation mode (interaction $F(9,18) = 4.3$, $p = 0.004$, $\eta_p^2 = 0.68$). As before, the S+ peck rates were fairly uniform within a trial, with little or no difference based on presentation mode or time (RM ANOVA; Presentation mode × Time in trial; all $p > 0.204$).

A comparison of the two panels in Fig. 3 shows that the biggest change to within-trial peck rates caused by the new training was the elimination of a fast initial decline in pecking right after stimulus onset. To examine this, we conducted an RM ANOVA (Cue phase × Time in trial × Presentation mode) on standardized peck rates over the first three time bins. This confirmed an effect by way of an interaction between cue phase and time in trial ($F(2,4) = 11.3$, $p = 0.023$, $\eta_p^2 = 0.85$). This change in behavior likely reflects the increased relevance of dynamic information to the discrimination of the different types of displays. Although the pigeons might be able to rapidly extract the form information from the displays, processing motion information in conjunction with this pose information seems to have delayed the onset of discriminative suppression. Repeating this ANOVA using the terminal three time bins of S− trials found only an effect of cue phase ($F(1,2) = 39.7$, $p = 0.024$, $\eta_p^2 = 0.95$) reflecting the overall reduction in discriminative performance.

3.2. Humans

Consistent Cues Discrimination: Most of the participants were able to learn the consistent discrimination, effectively replicating the previous investigation with humans (Qadri, Sayde, & Cook, 2014). Of the 36 human participants in this experiment, 26 reached the 75% accuracy criterion by the end of the fourth block. These participants' accuracy during the two 64-trial blocks prior to the conditional cues training is shown left of the dashed vertical line in each panel of Fig. 4. The time within the session is reported in exposures to each condition relative to the start of the conditional cue discrimination, with negative values indicating trials during the consistent cue discrimination. Overall, the participants were above chance with both dynamic and static presentations (dynamic $t(25) = 27.6$, $p < 0.001$, $d = 5.4$; static $t(25) = 20.5$, $p < 0.001$, $d = 4.0$). Like the pigeons, humans discriminated the dynamic presentations significantly better than the static presentations ($t(25) = 7.9$, $p < 0.001$, $d = 1.5$). No differences were found between the individuals that were assigned to receive dynamic or static reversal during conditional cue training ($ts(24) < 1.1$, $ps > 0.27$). The data for the 10 participants who failed to reach criterion has been omitted from the analysis. Post-experimental debriefing suggested that these individuals often got captured attending to visual or "social" features of the actor that were not relevant to the behavioral categories (facing direction, pointing direction, etc.).

In order to gain some understanding of the time course of how humans processed these displays, we further examined their response times to see how long they viewed the displays prior to making a choice. Response time was recorded as the time between stimulus onset and the recording of a choice response, and responses that took >10 s were discarded (for the period depicted, less than 0.5% of responses). For the period of consistent cue training, these data are on the left of the dashed line in the two panels in Fig. 5. They depict that response time was on average 363 ms faster for static presentations (1567 ms) as compared to dynamic presentations (1930 ms; $t(25) = 3.6$, $p = 0.001$, $d = 0.7$), and again there were no differences between the individuals assigned to the different reversal conditions ($ts(24) < 0.8$, $ps > 0.4$).

Conditional Cues Discrimination: The human participants were largely able to accommodate the change in reinforcement contingencies. After a brief period of poor performance, most people learned to treat the dynamic and static displays differently. Of the 26 participants who learned the initial consistent cues discrimination, 12 participants received the reversed dynamic contingencies (like the pigeons) and 14 participants received the reversed static contingencies. In the first block of conditional cues training, all participants show a large decrement in accuracy with the stimuli within the reversed condition. To provide a more fine-grained perspective on their accuracy with these reversed contingencies, the first 64-trial block of conditional cues training is broken apart into two-exposure blocks in Fig. 4. The participants with reversed dynamic contingencies were well below chance for the first two video displays ($M_{Dynamic} = 4.2\%$, SE = 4.2%; $t(11) = 11$, $p < 0.001$, $d = -3.2$), and the participants with reversed static contingencies were below chance with the first two static displays ($M_{Static} = 21.4\%$, SE = 6.9%; $t(13) = 4.2$, $p = 0.001$). However, the non-reversed condition differed between the groups. For every human participant with the reversed dynamic contingencies, static performance in the first two-exposure block of the conditional cue discrimination was worse than it had been prior to the conditional cues training. This is reflected in the chance-level accuracy of the first two static trials after the partial reversal ($M_{Static} = 54.2\%$, SE = 11.4%; repeated measures $t(11) = 2.6$, $p = 0.025$). In contrast, those participants with reversed static contingencies generally maintained the dynamic discrimination in those first two trials ($M_{Dynamic} = 82.1\%$, SE = 8.5%; repeated measure $t(13) = 2.1$, $p = 0.054$) and throughout the remainder of training. Only 8 of the 14 participants in the static condition had worse dynamic performance in the first conditional cuing block compared to the last consistent cuing block, suggesting that the participants' performance was likely equivalent prior to and after the partial reversal.

By the end of training, the vast majority of the participants were able to learn the conditional cuing conditions, using the same 75% criterion as employed at the end of consistent cues training. These 21 participants reached high levels of accuracy in dynamic ($M_{Dynamic} = 96.6\%$, SE = 0.91%; $t(20) = 51.0$, $p < 0.001$, $d = 11.1$) and static ($M_{Static} = 87.4\%$, SE = 1.62%; $t(20) = 23.0$, $p < 0.001$, $d = 5.0$) trials, and they – like pigeons – evidenced a strong dynamic superiority effect (11.2%, SE = 2.1%; $t(20) = 5.3$, $p < 0.001$, $d = 1.2$). For the two participants who received the reversed dynamic contingencies but could not learn the conditional cuing discrimination, one participant completely lost discrimination ($M_{Total} = 51.6\%$), and another one never resolved the static discrimination after the dynamic contingency reversal ($M_{Dynamic} = 90.6\%$, $M_{Static} = 53.1\%$). For the two participants who received the reversed static contingencies but could not learn the conditional cuing discrimination, two lost the discrimination entirely ($M_{Total} = 51.6\%$ and $M_{Total} = 46.9\%$), and one never recovered the static discrimination ($M_{Dynamic} = 96.9\%$, $M_{Static} = 56.3\%$).

Finally, the effect of the discrimination and partial contingency reversal on response times was examined. In general, response times increased as a result of the partial reversal, but the groups differed on which response times changed (see Fig. 5). For those with reversed dynamic contingencies, responses in both presentation modes were slowed, and dynamic trials (2423 ms) continued to maintain a slower RT than static trials (2124 ms). Contrastingly,
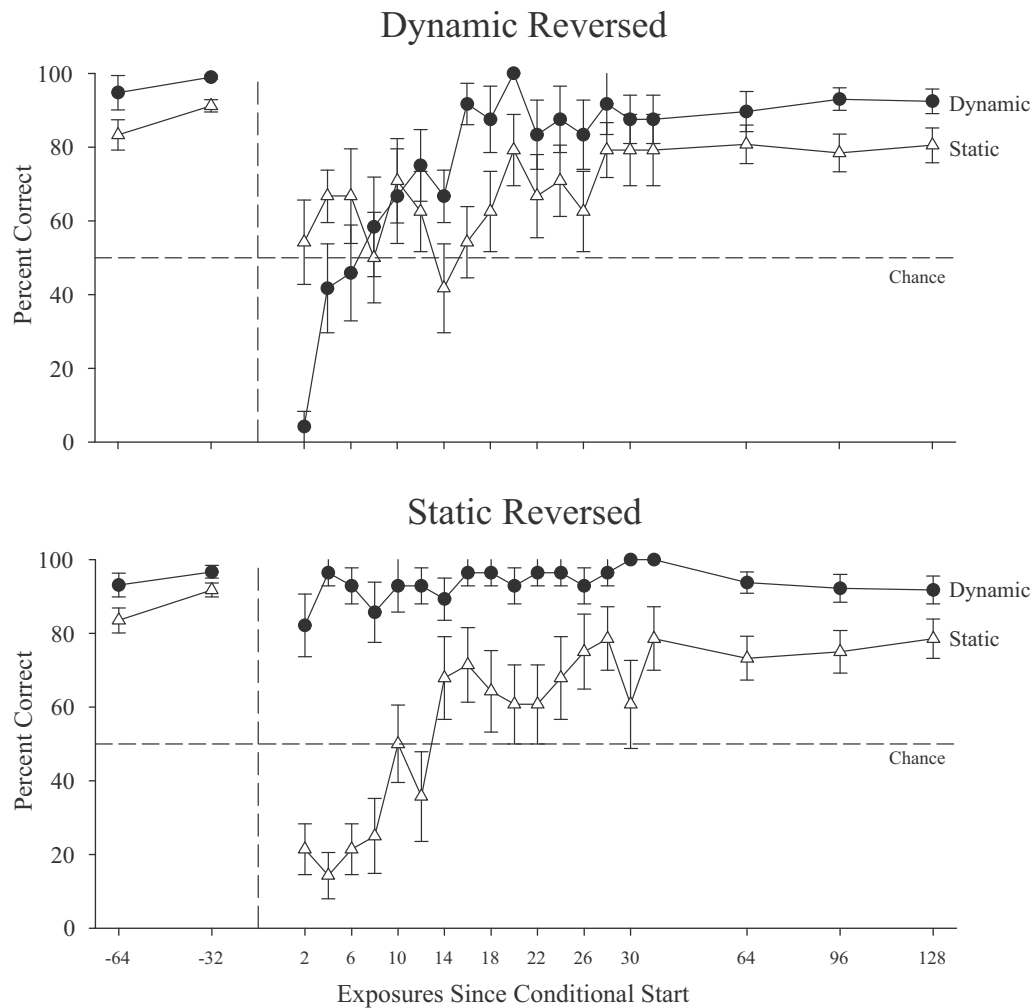
Dynamic Reversed

Static Reversed

Fig. 4. Humans' overall discrimination. Accuracy on trials is plotted as a function of exposures to each presentation condition relative to the start of conditional cue training in 32-exposure blocks (i.e. 64-trial blocks). The vertical dashed line indicates the start of conditional cue training. Because of how rapidly the humans learn the reversal, the first 32-exposure block after the start of conditional training is divided into 2-exposure blocks. The top panel depicts performance for the 12 humans who received a reversal of the dynamic condition, while the bottom panel depicts performance for the 14 humans who received a reversal of the static condition. Error bars depict standard error.
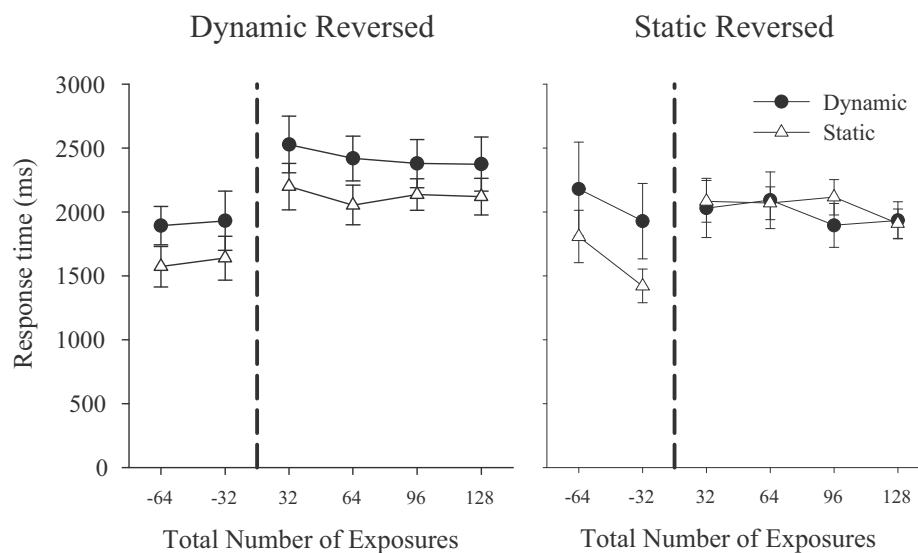
Dynamic Reversed Static Reversed

Fig. 5. Humans' response time. The average response time in milliseconds (ms) during the task as a function of 32-exposure blocks. Blocks are enumerated relative to the start of the conditional cue training. Filled circles depict response times during dynamic presentations while open triangles depict response times during static presentations. Error bars depict standard error.

for the participants with the reversed static contingencies, only the RT on static trials increased (static 2043 ms; dynamic 1984 ms). Because RTs within each time period of the experiment were largely similar, we collapsed across these data points in the graph for comparison. A mixed-factors ANOVA was conducted on RT testing the within-subject factors of presentation mode and time in session (i.e., pre- vs. post- reversal) and a between-subjects factor of reversal condition. The ANOVA confirmed this pattern in a three-way interaction between presentation mode, time in session, and reversal condition, $F(1,19) = 4.6$, $p = 0.044$, $\eta_p^2 = 0.20$. The ANOVA also captured the fact that static trials are responded to more quickly than dynamic trials and that this was affected over the course of the experiment: presentation mode $F(1,19) = 8.2$, $p = 0.010$, $\eta_p^2 = 0.30$, time in session $F(1,19) = 12.1$, $p = 0.003$, $\eta_p^2 = 0.39$, and their interaction $F(1,19) = 5.0$, $p = 0.038$, $\eta_p^2 = 0.21$ (although this two-way interaction largely reflects the triple interaction described first).

4. General discussion

In these experiments, both pigeons and humans learned to discriminate between two classes of behaviors, even when this classification was conditionally dependent on the type of cues available. Successful discrimination with this conditional assignment indicates that both pigeons and humans used a combination of separable features to discriminate the behaviors in the dynamic and static conditions. These features were the action cues, derived from the time course and motion of the human model's movements or the sequence of poses, and the pose or shape cues, derived from the structural relations of the model's articulated limbs relative to one another. It appears that both species used a combination of immediately available static pose information and later perceived dynamic action cues to discriminate among the conflicting behavioral categories. The similarity in how these species processed the displays suggests that their respective behavior recognition mechanisms operate similarly, likely on the same fundamental features of the stimuli.

Beyond this comparative similarity, the humans' performance showed a stronger link between static and dynamic processing that was absent in the pigeons. Reversing the dynamic condition's reinforcement contingencies greatly impaired people's performance with static displays, while reversing the contingencies for static presentations did not affect performance with dynamic presentations. One possible explanation for this asymmetric pattern of reversal suggests that the humans used their knowledge about the dynamic stimuli to respond to the static stimuli, but not vice versa. In this hypothesis, a single dynamic representation of the actions is used prior to conditional training and both the dynamic and static stimuli activate this single representation, although dynamic stimuli do so more broadly or more effectively than the static stimuli. Thus, when the contingencies reverse for the dynamic displays, accuracy with static displays is strongly affected because the processing of the static stimuli used the same shared representation. In contrast, reversing contingencies for the static displays does not affect performance on dynamic presentations because the static poses only weakly activate the shared representation making these contingency changes less impactful. While both of these types of conditions were not tested with pigeons, we did observe that, unlike humans, reversing the dynamic trials did not greatly impact static trial performance with the pigeons. This suggests that the pigeons may possibly lack a common representational linkage between the conditions (although one pigeon's performance may have shown hints of such a linkage). Testing pigeons with both types of reversals is clearly an interesting direction for the future.

Together, these results inform the computational models discussed previously in the introduction. The dynamic/static

asymmetry in human action representation possibly seen here can be added to the motion-and-form mixture models (i.e., Giese & Poggio, 2003) by simply adding unidirectional connections from the form pathway to the motion pathway, likely at the level of the snapshot representations. Psychologically, this corresponds to the static pose information of full-bodied agents being interpreted through the dynamic representation of the behavior. If this is an accurate model, it would suggest that behaviors are represented primarily according to their dynamic action features, and the static pose features are incidental, not diagnostic or definitional.

For the form-only models (i.e., Lange & Lappe, 2006), the current results would not alter the model structure in any fundamental fashion. The static/dynamic asymmetry could reflect the degree to which static poses and dynamic pose sequences separately activate the higher-level motion concept. In this computational model, responses prior to the reversal would be guided by activations of the higher-level motion concept by any type of static pose information. Responses after the reversal, however, would need to be guided by separable activations of the motion concept by static poses and sequential pose information. How the model would accommodate this shift in information use is unclear. Furthermore, if both the motion concept and the static features were initially available for making decisions, it is unclear whether this model would consistently produce a dynamic/static asymmetry. Developing models simulating the current human results with different types of representational asymmetry will be an informative course of investigation.

The critical aspect highlighted by the apparent asymmetry in human reversal learning in the present investigation is that the behavioral representations stored and used by humans are inherently dynamic. During static presentations, the static poses are activating a primary representation that is dynamic in character. Several investigations focusing on memory have come to a similar conclusion for dynamic mental representations more generally, partially supported by the concept of "representational momentum" (Freyd, 1987; Hubbard, 2005). In the line of "representational momentum" research, the memory for a given object is affected by its perceived dynamic properties. For example, when tasked with identifying the final position of an object that moved on the screen, observers exhibit greater difficulty in correctly rejecting an option which is a continuation of the object's motion. The results presented here highlight a similar link by demonstrating the dependence of static processing on the associations with the dynamic concept. Previous action recognition research has relied on the assertion that point-light displays removed pose cues, and therefore the motion cues were the fundamental representation of actions. The current experiment further establishes the active dominance of dynamic information in action representation by humans, even when pose cues were consistently available.

In the primate visual system, the extraction of form and motion relies on processing within the primary visual cortex and the lemnothalamic neural processing route (i.e., the ventral visual stream). Within these pathways, functional neuroimaging has highlighted the posterior superior temporal sulcus (pSTS) and extrastriate body area (EBA) as neural regions critical to the process of biological motion perception (Jastorff & Orban, 2009; Puce & Perrett, 2003; Vangeneugden et al., 2014). Giese and Poggio's (2003) model aimed to imitate this route's processing. In line with this model, but somewhat in conflict with the current results, are recent neuroimaging data that suggest that dynamic and static features are processed in distinct and separate locations in the human brain, ultimately making independent contributions to action recognition (Vangeneugden et al., 2014). Vangeneugden et al. used transcranial magnetic stimulation (TMS) to selectively disrupt discriminations that required participants to identify a walker's facing orientation or direction. The stimuli were like point-light displays, but Gabor

patches were used instead of dots in order to convey local orientation cues. Thus, these stimuli had both the structural direction as conveyed by the Gabor patch organization and the motion-based walking direction (forward vs. backward). When pSTS was disrupted, the motion discrimination suffered, and when EBA was disrupted, the static discrimination suffered. Our results do not necessarily contradict the proposition that the features are processed separately and independently, but they do raise the question regarding what information is represented and connected in these distinctly identified brain regions. Perhaps the critical function of the EBA in that study was processing the alignment of separated elements or the integration of information across space. Reconciling our cognitive results with this neuroimaging outcome may require testing the full-figured displays in an fMRI design to identify differences based on our dynamic and static stimulus conditions.

While the motion-and-form and the form-only computational models need only vision to successfully discriminate actions, it has also been suggested that there is a potential contribution from the motor system via mirror neuron activation in addition to the role of pose cues and motion cues (Decety & Grèzes, 1999; di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Grafton, 2009). For some theories of behavior recognition, the information from these visual processes would provide input for a more extended action observation network (AON) which would include mirror neurons and their embodied representations of action (Decety & Grèzes, 1999; di Pellegrino et al., 1992; Grafton, 2009). Investigations of action simulation look to join visual and motor cues using dynamic representations (Pezzulo, Candidi, Dindo, & Barca, 2013).

This possibility is complicated somewhat by the outcome that the pigeons in both this experiment and in the previous investigation stimuli showed little difference in basic performance as compared to the human participants (Qadri, Sayde, & Cook, 2014). Pigeons as well as humans benefit from dynamic displays versus static displays, and they showed differences in processing time based on the dynamic and static displays. Both species also learned the task readily, with pigeons perhaps more successful since all pigeons learned the task as compared to only 70% of humans. Regarding a hypothesized species-specific AON (i.e., a mirror neuron system), the pigeons' AON would quite likely not be activated by human behavior. Thus, the similarities across species suggests that the pigeons and humans are both computing pose and action features and solely using visual features to guide responding.

The exact functional neural areas and mechanisms used by the pigeons to process these displays are at the moment quite unclear. Simple motion and form seem to have specialized processing cells in the avian tectum, nucleus rotundus, and entopallium (Benowitz & Karten, 1976; Cook, Patton, & Shimizu, 2013; Laverghetta & Shimizu, 1999). Evolutionarily, these avian neural structures are homologous to the structures in humans' collothalamic processing route (i.e., dorsal visual stream; Cook, Qadri, & Keller, 2015). Evolutionary and comparative evidence suggests that although the mammalian and avian common ancestor had both processing routes, the extant members of these orders rely on different visual processing routes to engage with and perceive the world (Cook et al., 2015). The current investigation, however, highlights the possibility that both birds and mammals may classify behaviors using a combination of pose and action cues, regardless of neural differences. Comparing the ability and mechanisms of behavior recognition across species and determining the neural regions involved will be important to understanding the mechanisms of general, vision-based behavior recognition. If the avian collothalamic pathway is found to be relevant to this discrimination, it would provide further evidence that this pathway of birds is functionally homologous to the lemnothalamic pathway of mammals.

Having determined the general importance and potential dominance of dynamic information in behavior recognition in humans and pigeons, several more questions remain to be investigated to understand the complex relationship between pose and action as they relate to behavior recognition. Although the representation is fundamentally dynamic, what is the real contribution of pose features to social cognition and behavior recognition? While PLD investigations of behavior recognition are informative regarding the relevant motion features one could extract from sparse displays, humans generally do not recognize agents and behaviors from disjointed points. We are rarely (if ever) presented with a situation remotely similar to that tested with PLD stimuli. Instead, future behavior recognition research needs to determine what information can be extracted from the motions of conspecifics without the processing needed to derive structural information from the displays. Previous research with PLDs only reduced the availability of pose information, not eliminated it entirely as intended. In order to experimentally parse the separate contributions of pose and action, a fully rendered actor as used in this study could be used in functional imaging tasks. Given the joint use of the dynamic representation for pose and action processing, clever conditional designs can be used to force separated processing of pose and action. Furthermore, given that poses are understood through their dynamic counterpart, how does this affect models of behavior recognition deficits? Autism, for example, presents with social deficits, which have been studied and investigated with PLDs with varying success (Kaiser & Shiffrar, 2009). Perhaps the full-bodied stimuli from our experiment could reveal difficulties in the static aspect of social behavior recognition or deficits in connecting static and dynamic social information. Thus, studying behavior recognition with full-bodied displays would likely also be fruitful for the investigation of action recognition in clinical populations.

Author notes

References

Asen, Y., & Cook, R. G. (2012). Discrimination and categorization of actions by pigeons. *Psychological Science, 23*, 617–624. http://dx.doi.org/10.1177/0956797611433333.

Benowitz, L. I., & Karten, H. J. (1976). Organization of the tectofugal visual pathway in the pigeon: A retrograde transport study. *Journal of Comparative Neurology, 167*, 503–520.

Blake, R. (1993). Cats perceive biological motion. *Psychological Science, 4*(1), 54–57. http://dx.doi.org/10.1111/j.1467-9280.1993.tb00557.x.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*, 47–73. http://dx.doi.org/10.1146/annurev.psych.57.102904.190152.

Casile, A., & Giese, M. A. (2005). Critical features for the recognition of biological motion. *Journal of Vision, 5*(4). http://dx.doi.org/10.1167/5.4.6. 6–6.

Cook, R. G., Beale, K., & Koban, A. (2011). Velocity-based motion categorization by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 37*(2), 175–188. http://dx.doi.org/10.1037/A0022105.

Cook, R. G., & Katz, J. S. (1999). Dynamic object perception by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 194–210. http://dx.doi.org/10.1037/0098-7403.25.2.194.

Cook, R.G., Patton, T.B., & Shimizu, T. (2013). Functional segregation of the entopallium in pigeons. Philosophy (London, England), 130, 59–86.

Cook, R. G., Qadri, M. A. J., & Keller, A. M. (2015). The analysis of visual cognition in birds: Implications for evolution, mechanism, and representation. *Psychology of Learning and Motivation, 63*, 173–210. http://dx.doi.org/10.1016/bs.plm.2015.03.002.

Cook, R. G., & Roberts, S. (2007). The role of video coherence on object-based motion discriminations by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 33*(3), 287–298. http://dx.doi.org/10.1037/0097-7403.33.3.287.

Decety, J., & Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3*(5), 172–178. http://dx.doi.org/10.1016/S1364-6613(99)01312-1.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*(1), 176–180.

Dittrich, W. H. (1993). Action categories and the perception of biological motion. *Perception, 22*, 15–22.

Dittrich, W. H., Lea, S. E. G., Barrett, J., & Gurr, P. R. (1998). Categorization of natural movements by pigeons: Visual concept discrimination and biological motion. *Journal of the Experimental Analysis of Behavior, 70*(3), 281–299. http://dx.doi.org/10.1901/jeab.1998.70-281.

Fleischer, F., Caggiano, V., Thier, P., & Giese, M. A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. *Journal of Neuroscience, 33*(15), 6563–6580. http://dx.doi.org/10.1523/jneurosci.4129-12.2013.

Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review, 94*(4), 427–438. http://dx.doi.org/10.1037/0033-295X.94.4.427.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4*(3), 179–192. http://dx.doi.org/10.1038/nrn1057.

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Annals of the New York Academy of Sciences, 1156*(1), 97–117. http://dx.doi.org/10.1111/j.1749-6632.2009.04425.x.

Herrnstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 2*(4), 285–302.

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review, 12*(5), 822–851. http://dx.doi.org/10.3758/bf03196775.

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*(22), 7315–7329. http://dx.doi.org/10.1523/jneurosci.4870-08.2009.

Jastorff, J., Popivanov, I. D., Vogels, R., Vanduffel, W., & Orban, G. A. (2012). Integration of shape and motion cues in biological motion processing in the monkey STS. *Neuroimage, 60*(2), 911–921. http://dx.doi.org/10.1016/j.neuroimage.2011.12.087.

Johansson, G. (1973). Visual perception of biological motion and a model of its analysis. *Perception and Psychophysics, 14*(2), 201–211. http://dx.doi.org/10.3758/Bf03212378.

Kaiser, M. D., & Shiffrar, M. (2009). The visual perception of motion by observers with autism spectrum disorders: A review and synthesis. *Psychonomic Bulletin & Review, 16*(5), 761–777. http://dx.doi.org/10.3758/pbr.16.5.761.

Koban, A., & Cook, R. G. (2009). Rotational object discrimination by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 250–265. http://dx.doi.org/10.1037/a0013874.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12*(1), 48–55. http://dx.doi.org/10.1162/08989290051137594.

Lange, J., Georg, K., & Lappe, M. (2006). Visual perception of biological motion by form: A template-matching analysis. *Journal of Vision, 6*(8). http://dx.doi.org/10.1167/6.8.6. 6–6.

Lange, J., & Lappe, M. (2006). A model of biological motion perception from configural form cues. *The Journal of Neuroscience, 26*(11), 2894–2906. http://dx.doi.org/10.1523/jneurosci.4915-05.2006.

Laverghetta, A. V., & Shimizu, T. (1999). Visual discrimination in the pigeon (Columba livia): Effects of selective lesions of the nucleus rotundus. *Neuroreport, 10*, 981–985.

Oram, M., & Perrett, D. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience, 6*(2), 99–116. http://dx.doi.org/10.1162/jocn.1994.6.2.99.

Parron, C., Deruelle, C., & Fagot, J. (2007). Processing of biological motion point-light displays by baboons (*Papio papio*). *Journal of Experimental Psychology: Animal Behavior Processes, 33*(4), 381. http://dx.doi.org/10.1037/0097-7403.33.4.381.

Pezzulo, G., Candidi, M., Dindo, H., & Barca, L. (2013). Action simulation in the human brain: Twelve questions. *New Ideas in Psychology, 31*(3), 270–290. http://dx.doi.org/10.1016/j.newideapsych.2013.01.004.

Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing, 28*(6), 976–990. http://dx.doi.org/10.1016/j.imavis.2009.11.014.

Ptito, M., Faubert, J., Gjedde, A., & Kupers, R. (2003). Separate neural pathways for contour and biological-motion cues in motion-defined animal shapes. *Neuroimage, 19*(2), 246–252.

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358* (1431), 435–445. http://dx.doi.org/10.1098/rstb.2002.1221.

Qadri, M. A. J., Asen, Y., & Cook, R. G. (2014). Visual control of an action discrimination in pigeons. *Journal of Vision, 14*, 1–19. http://dx.doi.org/10.1167/14.5.16.

Qadri, M. A. J., & Cook, R. G. (2015). Experimental divergences in the visual cognition of birds and mammals. *Comparative Cognition & Behavior Reviews, 10*, 73–105. http://dx.doi.org/10.3819/ccbr.2015.100004.

Qadri, M. A. J., Sayde, J. M., & Cook, R. G. (2014). Discrimination of complex human behavior by pigeons (*Columba livia*) and humans. *PLoS One, 9*(11), e112342. http://dx.doi.org/10.1371/journal.pone.0112342.

Regolin, L., Tommasi, L., & Vallortigara, G. (2000). Visual perception of biological motion in newly hatched chicks as revealed by an imprinting procedure. *Animal Cognition, 3*(1), 53–60. http://dx.doi.org/10.1007/s100710050050.

Shiffrar, M., & Freyd, J. J. (1990). Apparent motion of the human body. *Psychological Science, 1*(4), 257–264. http://dx.doi.org/10.1111/j.1467-9280.1990.tb00210.x.

Theusner, S., de Lussanet, M., & Lappe, M. (2014). Action recognition by motion detection in posture space. *The Journal of Neuroscience, 34*(3), 909–921. http://dx.doi.org/10.1523/jneurosci.2900-13.2014.

Thirkettle, M., Benton, C. P., & Scott-Samuel, N. E. (2009). Contributions of form, motion and task to biological motion perception. *Journal of Vision, 9*(3), 1–11. http://dx.doi.org/10.1167/9.3.28.

Troje, N. F., & Aust, U. (2013). What do you mean with "direction"? Local and global cues to biological motion perception in pigeons. *Vision Research, 79*, 47–55. http://dx.doi.org/10.1016/j.visres.2013.01.002.

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception and Psychophysics, 67*(4), 667–675.

Vallortigara, G., Regolin, L., & Marconato, F. (2005). Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biology, 3* (7), 1312–1316. http://dx.doi.org/10.1371/journal.pbio.0030208.

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*(2), 574–585. http://dx.doi.org/10.1523/jneurosci.4032-13.2014.