



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

Article history:

Received 26 April 2016

Received in revised form 14 September 2016

Accepted 16 September 2016

Keywords:

Behavior recognition

Poses

Dynamic representation

Representational momentum

Pigeon

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 re-learning 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

Download English Version:

<https://daneshyari.com/en/article/5705940>

Download Persian Version:

<https://daneshyari.com/article/5705940>

[Daneshyari.com](https://daneshyari.com)