Pigeons process actor-action configurations more readily than bystander-action configurations

Muhammad A. J. Qadri¹ · Robert G. Cook¹

Published online: 10 February 2020
© The Psychonomic Society, Inc. 2020

Abstract

Behavior requires an actor. Two experiments using complex conditional action discriminations examined whether pigeons privilege information related to the digital actor who is engaged in behavior. In Experiment 1, each of two video displays contained a digital model, one an actor engaged in one of two behaviors (Indian dance or martial arts) and one a neutrally posed bystander. To correctly classify the display, the pigeons needed to conditionally process the action in conjunction with distinctive physical features of the actor or the bystander. Four *actor-conditional* pigeons learned to correctly discriminate the actions based on the identity of the actors, whereas four *bystander-conditional* birds failed to learn. Experiment 2 established that this failure was not due to the latter group's inability to spatially integrate information across the distance between the two models. Potentially, the colocalization of the relevant model identity and the action was critical due to a fundamental configural or integral representation of these properties. These findings contribute to our understanding of the evolution of action recognition, the recognition of social behavior, and forms of observational learning by animals.

Keywords Action recognition · Behavior · Actor · Pigeon

The recognition, discrimination, and classification of behavior by humans, computers, and nonhuman animals have received increasing attention (Blake & Shiffrar, 2007; Poppe, 2010). Behavior, of course, requires an actor. To recognize social behaviors and other actions, animals need to observe and visually process the dynamic positioning of the articulated limbs and body of other agents over an extended period of time. Extensive studies of human action recognition, largely conducted using point-light displays (PLD), have focused on the information contained in the motion components of articulated bodies (Blakemore & Decety, 2001; Johansson, 1973; Klüver, Hecht, & Troje, 2016).

The examination of action recognition mechanisms in animals has much of its modern origins in the research of Stephen Lea and his colleagues (W. H. Dittrich & Lea,

1993, 2001; W. H. Dittrich, Lea, Barrett, & Gurr, 1998; Ryan & Lea, 1994). Using the presentation of short video clips, for example, they found that pigeons could learn to discriminate moving from still images of behaviors and to classify different categories of behaviors (e.g., walking vs. pecking). These seminal papers in action recognition represent just a small part of Lea's considerable contributions toward understanding categorization, classification, and discrimination learning by animals. Employing a wide range of different stimuli and categories, his research has contributed over the years substantially to our theoretical understanding of how animals process and represent information (e.g., Lea, 1984, 2010; Lea & Harrison, 1978; Lea et al., 2018; Lea, Slater, & Ryan, 1996; Osthaus, Lea, & Slater, 2005).

Nonetheless, limited evidence has since been collected about how nonhuman animals visually process and represent the actions of other organisms. For example, biologists have determined that static form and color properties of stimuli readily contribute to visually identifying conspecifics in many species (Rowe & Skelhorn, 2004). However, few studies have directly examined how animals process the actions of other animals. This is in large part because controlling the behavior of animals to make appropriate stimuli is challenging. Presentation of short videos has helped to move this forward. When enough support is provided through auditory and visual

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13420-020-00416-7>) contains supplementary material, which is available to authorized users.

✉ Muhammad A. J. Qadri
Muhammad.Qadri@tufts.edu

¹ Department of Psychology, Tufts University, 490 Boston Ave, Medford, MA 02155, USA

information, video stimuli can elicit courtship behaviors from pigeons (Partan, Yelda, Price, & Shimizu, 2005; Shimizu, 1998), zebra finches (Galoch & Bischof, 2007), and Bengalese finches (Takahasi, Ikebuchi, & Okanoya, 2005; Takahasi & Okanoya, 2013). Beyond identifying conspecifics in video-recorded displays, budgerigars may recognize video-recorded conspecific behavior to the point of social contagion (Gallup, Swartwood, Militello, & Sackett, 2015). Increasingly, researchers are turning to newer techniques such as robots and computer animation to control the exhibited behaviors of the stimulus models (Frohnwieser, Murray, Pike, & Wilkinson, 2016).

Our contributions to action recognition research use highly controlled digitally generated models to produce animated videos with various actors or agents engaged in simple locomotion behaviors or in complex human action sequences (Asen & Cook, 2012; Qadri, Asen, & Cook, 2014a; Qadri & Cook, 2017; Qadri, Sayde, & Cook, 2014b). The utility of these artificial displays lies in the regimented control of the scenes, high-quality motion information, and the relative “completeness” of the final, rendered agent (i.e., as compared with point-light displays; see Qadri & Cook, 2015, for a more complete discussion). This research has found that pigeons can discriminate and categorize types of repetitive locomotive actions from digitally animated animal models (walking vs. running) and more complex sequences of behaviors (martial arts vs. Indian dance) from digitally animated humans (Asen & Cook, 2012; Qadri, Asen, et al., 2014; Qadri, Sayde, et al., 2014). In each case, the pigeons demonstrated transfer of this discrimination to novel exemplars of each class of behavior. Further, the pigeons show a dynamic superiority effect in which dynamic video presentations of complex actions are discriminated better than static presentations of single frames from the same videos (similar to Cook & Katz, 1999; Koban & Cook, 2009), as well as the ability to respond divergently to both acting and static displays (Qadri & Cook, 2017), suggesting they independently process action cues and pose cues.

One challenge of all of this research is the persistent question of whether animals see these as “true” behaviors that are being generated by an animate being or agent, potentially with intentions and goals (Lea & Dittrich, 2000). Some authors have questioned the appropriateness of the medium by which this information is being transmitted (D’Eath, 1998; Fleishman & Endler, 2000; Patterson-Kane, Nicol, Foster, & Temple, 1997). While animacy detection has been a large topic within studies of human action recognition (for a review, see Scholl & Tremoulet, 2000), it has not been well explored in animals. For instance, the human brain seems to retain specialized processes for the detection and processing of animate and agentive objects in the environment (Castelli, Happé, Frith, & Frith, 2000). Perhaps this deep connection between action and animacy is what causes us to so readily see an “actor” in the swirl of lights that form a point-light

display (Bertenthal, Proffitt, & Cutting, 1984). Similarly, this connection may result in our strong tendency to impute motives and intentions to the interaction of even small shapes (e.g., W. H. Dittrich & Lea, 1994; Heider & Simmel, 1944). At the heart of this research is the core idea that behaviors originate from an intentional actor or agent. In support of this possibility, humans show rapid and spontaneous processing of actors and the action roles from scenes with two humans, one acting and one receiving the action (e.g., hitting; Hafri, Papafragou, & Trueswell, 2013; Hafri, Trueswell, & Strickland, 2018), suggesting that actions and actors are automatically processed together (cf. integral dimensions in Garner & Felfoldy, 1970).

The question of animacy and agency in computerized displays is intriguing and challenging to answer in non-verbal animals. Still if behavior requires an actor; do the animals recognize that an actor is acting? In the current experiment, we begin attempting to tackle this notion by looking at the critical unity between actor and action. In our prior experiments, a single actor or model has always been used to create the behavioral action that the pigeons discriminate. Here, we sever this connection for the first time by using two models while conveying this information, and we then examine if the pigeons can learn to discriminate this two-body display. The pigeons must critically learn to ignore competing contextual information that is naturally salient and highly similar to the relevant discriminative information. Specifically, one group of pigeons must conditionally process the action of one model depending on the identity of a second model.

For these experiments, we created displays containing two digital human models placed side by side that looked and dressed differently. One engaged in a complex action (martial arts or Indian dance), while the adjacent second model adopted a neutral static pose (see Fig. 1). Two conditions were tested that required the pigeons to identify conditionally both a digital model and the behavior in the display. A similar conditional procedure was employed to separate the contribution of action and pose information (Qadri & Cook, 2017). Two groups of four pigeons each were tested with the two versions of this conditional discrimination.

For the pigeons in the *actor-conditional* assignment, the critical identity information and the action were embodied in the same model (see Table 1). For example, if the white-shirt model was engaged in martial arts, the display was positive, with pecks being reinforced, and if instead she was engaged in Indian dance, the display was negative, with pecks contributing to a time-out. These reward contingencies would be reversed if the actor was the black-shirt model. Thus, the pigeons needed to conditionally process the actor’s identity and its action to successfully learn the discrimination. The static presence of the green-shirt and purple-shirt models were simply bystanders that were irrelevant context to the

Fig. 1 Two sample displays from the Experiment 1. Each of the four digital models are depicted. The lines superimposed on the models trace out the position of one foot, one leg, and the head for the Indian dance

action (white shirt) and the martial arts action (black shirt). Besides their behavior and identity, displays varied in spatial position (i.e., left panel vs right panel), camera distance, height, and angle (i.e., azimuth)

discrimination. Two representative examples of these displays are shown in Fig. 1.

For the pigeons in the *bystander-conditional* assignment, in contrast, the critical identity information and the action were embodied in different models (see Table 1). In this condition, the identity of the actor was the irrelevant context information and the identity of the bystander was the relevant information. Thus, to correctly determine the reward outcomes for a given

display, the pigeons had to attend to and identify the physical characteristics of the adjacent bystander as well as the actor's ongoing behavior. Put another way, the adjacent green-shirt and purple-shirt models were now critical to successfully learning how to respond to the actions. For example, if green-shirt model was next to an actor engaged in martial arts, the display was positive, and if instead the actor was engaged in Indian dance, the display was negative. From a purely

Table 1 Illustrative sample contingencies for the actor and bystander conditional discriminations

	Behavior	Actor	Bystander	Peck outcome
Actor conditional	Martial arts	White shirt	Purple/green shirt	Grain reward
		Black shirt	Purple/green shirt	Dark time-out
	Indian dance	White shirt	Purple/green shirt	Dark time-out
		Black shirt	Purple/green shirt	Grain reward
Bystander conditional	Martial arts	Purple/green shirt	White-shirt	Grain reward
		Purple/green shirt	Black-shirt	Dark time-out
	Indian dance	Purple/green shirt	White-shirt	Dark time-out
		Purple/green shirt	Black-shirt	Grain reward

Note. This table shows how outcomes differ among different presentation conditions based on the behavior the actor, the actor's identity, and the bystander's identity for each group of pigeons. This table only shows a subset of the possible displays that were yoked and counterbalanced across models and spatial positions (see [Method](#) section). Note also that the models differed in other physical characteristics besides their shirt color

computational perspective, the two conditions are equivalent because bystanders contain all the same type of form information as the actor (color, clothes, human articulation, faces, etc.). This latter condition, however, required the pigeons to simultaneously process actions in conjunction with behavior-relevant characteristics not embodied by the behaving actor.

The major question of interest was whether the conjoint processing of model identity and action when embodied in the same model was better, equivalent, or worse than when they were separated across two different models. Two experiments were conducted. The first experiment tested the acquisition of the actor-conditional and the bystander-conditional conditions in different groups of pigeons. The second experiment tested just the pigeons in the bystander-conditional group, but with solid colored panels replacing the acting models to investigate how processing action and different spatial factors may have affected how they processed the prior bystander-conditional discrimination.

Method

Animals

Eight male pigeons (*Columba livia*) were tested. Each was maintained at 80%–85% of their free-feeding weight with ad libitum grit and water. They were individually housed in a colony kept on a 12:12 light–dark cycle. Four of these pigeons had previously been in action discriminations (Qadri & Cook, 2017; Qadri, Sayde, et al., 2014). Two had been pilot tested in an action recognition choice task. The other two had completed an experiment examining dimensional shape and form unrelated to action discriminations. These prior experiences were distributed across the two groups. All procedures were approved by Tufts University Institutional Animal Care and Use Committee (Protocols M2011-11 and M2013-124).

Apparatus

Two touch-screen-equipped (EZ-170-WAVE-USB) operant chambers were used to present video stimuli and record peck responses, with two pigeons from each group tested in each chamber. Stimuli were displayed on LCD computer monitors (NEC LCD 1525X and NEC LCD 51vm; 60 Hz refresh rate) recessed 8 cm behind the touch screen. One chamber permitted a monitor resolution of $1,280 \times 1,024$, while the other was limited to a resolution of $1,024 \times 768$. Mixed grain reward was delivered via a central food hopper positioned beneath the touch screen. A 28 V houselight in the ceiling of each chamber was constantly illuminated, except during time-outs.

Stimuli

Four digital human models were used in this experiment (available in Poser software package, smithmicro.com). Each model was visually distinct due to both natural human characteristics (i.e., each model had an apparent sex, age, height, body shape, etc.) and their apparel (clothing color; see Fig. 1) that reliably identified the model. For ease of discussion, shirt color will be used to identify and reference the separate models in specific videos (white-shirt, purple-shirt, black-shirt, and green-shirt models). The models were paired off in two ways during the experiment, purple-shirt with green-shirt (and therefore white-shirt with black-shirt) and purple-shirt with white-shirt (and therefore green-shirt with black-shirt). Both models within a pair were designated as either *actors* or *bystanders*. Two pigeons (one in each group) experienced each pair as actors while two other pigeons (again, one in each group) experienced the same pair as bystanders. Thus, the assignments of the models as actors and bystanders was effectively yoked across and counterbalanced within groups. For a complete listing, see Supplemental Table S1.

From these models, two types of videos were created: *acting videos* and *neutral videos*. In acting videos, a digital model was configured using motion-captured sequences of “Indian dance” or “martial arts” (four distinct sequences of each; available from mocap.cs.cmu.edu), and it was rendered using digital software (Poser 8, SmithMicro) to create 600-frame AVI videos (MATLAB, The MathWorks, Natick, MA) using the Cinepak codec. In neutral videos, a digital model was configured in a single, neutral, arms-outward pose (see Fig. 1). Each model, pose, and action sequence was rendered from 32 different camera perspectives (two distances [close/far] $\times$ two elevations [high/low] $\times$ eight directions [eight different equally spaced camera placements around the model]) to create a wide variety of perspectives. Assuming an average viewing distance of 3 cm behind the touchscreen, the average viewing angle for the full video frame was approximately 54 degrees of visual angle (dva; 51.3 dva for one chamber, 56.7 dva for the other). The size of the figure within the panel varied depending on the actor and perspective, but the horizontal extent of the actor’s waist with a front-facing figure averaged 2.6 dva with a far perspective and 5.7 dva with a close perspective (vertical extents or “heights” of the figures averaged 15.1 dva and 32.4 dva with far and close perspectives, respectively).

On each trial, one acting video and one neutral video were randomly selected and presented side by side. The distance, elevation, and perspective of the camera varied independently for both videos. The selected videos were centrally positioned

in the display with no gap in between them (see Fig. 1). For the acting videos, playback started at a randomly selected frame from the first half of the video and continued at approximately 33 frames per second over the 20-s duration of a trial (this playback rate resulted in a total video duration of 18.2 s; the initial 1.8 s of video were repeated after the 18.2 s elapsed). For the neutral video, the single frame was repeatedly displayed for the same 20-s trial duration. Within each session, the acting and neutral videos appeared equally often in the left or right positions.

Procedure

Experiment 1: Conditional action acquisition

For each group, the reinforced S+ and nonreinforced S− stimuli were defined by the combination of the behavior in the acting video and the identity of one of the digital models on the display, as laid out with examples in Table 1 (for full and complete assignments, see Supplemental Table S1). For the *actor-conditional* group, the reinforcement contingencies were defined by the combination of the behavior and the behaving actor. For example, an S+ could be a white-shirt model engaged in Indian dance or a black-shirt model engaged in martial arts, regardless of the bystander's identity. The corresponding S− in this example would be the reverse: white-shirt model engaged in martial arts or black-shirt model engaged in Indian dance, regardless of bystander identity. Thus, the pigeons in this group would be required to process conditionally both the identity of the actor and the behavior to respond correctly.

For the *bystander-conditional* group, these contingencies were defined by the combination of the behavior and the bystander. Following the same example, regardless of whether white-shirt or black-shirt model was performing the action, the S+ would be Indian dance when the green-shirt model was the bystander and martial arts when the purple-shirt model was the bystander. Again, the contingencies for the S− stimuli were reversed, requiring the pigeons in this group to process conditionally both the identity of the bystander and the actor's behavior to respond correctly.

Pigeons initiated each trial by emitting a single peck to a circular, 2.5 cm white circle. This circle was then replaced by the display for 20 s, after which the videos were removed. For all of the pigeons, pecks to S+ displays were reinforced with 2.9 s access to the food hopper on a concurrent VI-10 and FI-20 schedule (i.e., variable interval reinforcement during the trial plus regular, posttrial reinforcement). After each video presentation, post-trial food reinforcement was given for S+ trials or a variable timeout was given for S− trials (0.5 s per peck). A 3-s intertrial interval then followed.

Sessions were comprised of 96 total trials (48 S+ & 48 S− trials). Each session contained six repetitions of each combination of behavior, actor, bystander, and actor location (i.e., actor-left or actor-right locations). The camera perspective for each video in a trial was randomly selected from the 32 possible positions, resulting in the actor and bystander rarely being presented from the same perspective. Eight randomly selected S+ trials were designated to be probe trials, which allowed for the measurement of S+ peck rates without interruptions by food reinforcement. Peck rates from these probe S+ trials were used to evaluate pecking in the S+ condition. Each pigeon completed 110 sessions of training for this experiment.

Experiment 2: Conditional color acquisition

Three of the four pigeons from the bystander-conditional group were tested in Experiment 2 (the results from the fourth pigeon could not be used because of a programming error). The critical modification in this experiment was replacing the acting video with a solid red or green panel (i.e., a 54-dva colored square). This removed the two potential stimulus dimensions of actor identity and actor behavior, and it also caused the displays to be fully static and without motion. In this experiment, the reinforcement contingencies for the displays were now defined by the combination of the solid color of the panel and the identity of the adjacent, neutrally posed “bystander” model. Continuing the example from above, the S+ would be indicated by the green-shirt bystander if the green color panel was present and the purple-shirt bystander if the red color panel were present. The remaining aspects of daily testing were otherwise identical to Experiment 1. The pigeons completed 80 sessions of this modified training.

Analysis

Like in our previous action recognition research, the pigeons' performance was measured using the discrimination index ρ (rho; i.e., as in Herrnstein, Loveland, & Cable, 1976), which is computed as a normalized Mann–Whitney U . Because the pigeons' peck rates at the end of the trial best represents the pigeons' ability to discriminate a display, ρ was calculated from the number of pecks made over the last 10 s of each 20-s stimulus presentation of positive-probe trials and negative trials. A chance discrimination index value of .5 indicates that S+ probe trial peck rates had no differential ranking relative to peck rates on S− trials, while a ρ of 1 indicates perfect discrimination, with all the S+ probe trials receiving more pecks within a session than all of the S− trials.

Results

Conditional action acquisition

Overall, all the pigeons in the actor-conditional group learned their conditional discrimination, while the pigeons in the bystander-conditional group did not. To evaluate this difference, we analyzed the differences in acquisition rates between the groups by considering (a) when the learning was first evident and (b) what level of performance was achieved. Figure 2's left panel depicts each group's average ρ across the experiment in 10-session blocks. The actor-conditional group (filled circles) started at chance levels and then increased almost linearly up to a discrimination index of 0.88 by the last 10-session block. Statistical analysis showed that by the third block of training, the actor-conditional group was significantly above chance, $t(3) = 1.3$, $p = .023$, $d = 2.1$. Three of the pigeons were considerably faster than the fourth bird in this group. The three faster birds were significantly above chance by the second (one bird) and fourth blocks (two birds) as detected by one-sample t tests within each block against chance for each pigeon, $ts(9) > 2.5$, $ps < .031$, $ds > 0.8$. These three birds were at discrimination levels of .75 or above by Block 5 (two birds) and Block 7. The fourth bird was much slower and was not above the chance level of 0.5 until Block 9, $ts(9) > 5.1$, $ps < .001$, $ds > 1.6$, but improved quickly over the last blocks of the experiment, getting to a discrimination index of .82 by the end of testing. During the final 10-session block, this group's average discrimination index was 0.88 ($SD = .098$, range: 0.76–.97), significantly above chance, $t(3) = 7.7$, $p = .005$, $d = 3.9$.

In contrast, the bystander-conditional group never rose significantly above chance, $ts(3) < 1.6$, $ps > .199$, over the entire

11 blocks. Examined individually, three of the four birds exhibited no signs of learning. One pigeon in the bystander-conditional group did learn to significantly discriminate the displays, doing so by the third block, $ts(9) > 3.1$, $ps < .012$, $ds > 1.0$. This one bystander-conditional pigeon performed with a discrimination index of 0.72 ($SD = .07$) in the final block of this training, numerically lower than every bird in the actor-conditional group. A mixed ANOVA (conditional assignment $\times$ 10-session block) using discrimination index ρ confirmed the expected significant interaction between block and group displayed in Fig. 2, $F(10, 60) = 7.7$, $p < .001$, $\eta_p^2 = .56$.

We also evaluated how the displays' characteristics affected the actor-conditional pigeons' successful discrimination over the last 20 sessions (i.e., two blocks) of training. The position of the actor in the display—to the left or to the right of the bystander—showed no effects in these analyses, $ts(19) < 1.4$, $ps > .171$. We found no clear or consistent influence of camera perspective on performance. Conclusions from such comparisons are made difficult, however, by the independent and randomized perspectives of the actor and the bystander across trials, which resulted in these different factors rarely being equivalent across the two models. Focusing on the actor video, one pigeon may have performed better with the lower camera elevation, difference of .08, $t(19) = 2.0$, $p = .056$, but none of the other birds showed any systematic effects, differences $< .02$, $ts(19) < 0.5$, $ps > .627$. Camera distance also did not greatly influence performance. For three of these pigeons, the distance of the focal actor made no difference, $ts(19) < 1.02$, $ps > .320$, but the fourth pigeon suffered a slight decrement of 0.12 when the actor was more distant—that is, smaller, $t(19) = 3.9$, $p < .001$, $d = 0.9$; same pigeon with the possible camera elevation effect. The identity of different models themselves made little or no difference for the successful actor-conditional birds. Three of the four pigeons learned to equivalent levels with their two acting models. The fourth bird showed a higher discrimination index with one model (.90) over the other (.66).

We next evaluated the spatial distribution of pecks to the display by each group. Given the differences in the spatial arrangement of critical information in the actors and bystanders, examining the location of their pecking could be revealing, especially since pecking location is a potential measure of spatial processing (cf. L. Dittrich, Rose, Buschmann, Bourdonnais, & Güntürkün, 2010). Analyses from the last 20 sessions of discrimination testing are presented in the top two panels of Fig. 3 for Experiment 1. Each panel shows display-directed pecking during the trial by each group, with the contours depicting the locations concentrated on by each pigeon. Each panel represents the area of the two videos within the display. The x-axis has been standardized so that pecking in the left half of the panel represents pecks towards the bystander and pecking toward the right of the panel represents pecks towards the

Fig. 2 Discrimination performance as measured using ρ across Experiments 1 and 2. Filled circles depict performance by the actor-conditional group and open triangles depict performance by the bystander-conditional group. Error bars depict standard error

Fig. 3 The areas of the most concentrated pecking for each bird across conditions and experiments. The contoured areas enclose the 25% most targeted pecking area for each pigeon (for full-peck distributions for each bird, see [Supplemental Material](#)). The enclosed figure areas represent the total area of the two video panels within the display. This representation has been standardized so that pecking in the left half of the panel represents bystander-directed pecking and pecks toward the right half of the panel represents actor-directed pecking (though spatially this varied during the experiment). Colors and lines depicting each bird's spatial focus of pecking for birds in the bottom two panels are consistent, and the top panel refers to a second set of birds

actor. We used the data from all trials during these 20 sessions to produce the most accurate map possible. The general spatial characteristics of pecking distributions did not seem to be altered much by whether it was a positive or negative trial, although clearly positive trials contributed more pecks to the analysis. The spatial pecking distribution for each bird was smoothed in this standardized space

using a two-dimensional Gaussian kernel density estimation method (R function `kde2d` from package MASS; Venables & Ripley, 2002). Each panel shows the area of greatest pecking for each bird, defined as the smallest area containing 25% of the pecks. While pecking was widely distributed overall (see [Supplemental Fig. S1](#)), and all pecks were analyzed in the numeric analyses below, the 25% contours succinctly depict where the birds attended to within the display.

The top panel of [Fig. 3](#) shows the results for the four actor-conditional pigeons. For statistical analyses, the horizontal position of each peck was classified as being directed toward the actor's or the bystander's side or undirected toward the middle 10% of the display (all pecks were analyzed, not just those within the 25% contours). Of directed pecking, three of the four birds pecked most frequently toward the behaving actor rather than the bystander—mean proportions of pecks towards actor, 60%, 82%, and 73%; two-sided, one-sample *t* test comparing against a chance value of 0.5; $t_s(19) > 11.9$, $p < .001$. Similar to other findings, this directed behavior suggests these actor-conditional pigeons were processing the critical acting model. The fourth bird, who was also the bird who most slowly acquired the discrimination, equally distributed its pecks to the actor and bystander, $M = 49.7\%$, $t(19) = 0.2$, $p = .842$, and this pattern did not markedly change during the final training blocks when the discrimination was finally acquired.

The next panel of [Fig. 3](#) shows the results for the four pigeons in the bystander-conditional group. The patterns here were more varied. Two pigeons favored the actor's side of the display—mean proportions = 59% and 62%; $t_s(19) > 3.9$, $p < .001$. It was one of these two birds who learned the conditional discrimination. The other two birds showed less preference, distributing their pecks to both sides—one bird had a slight and detectable actor bias, $M = 54\%$, $t(19) = 4.1$, $p < .001$; other $M = 50\%$; $t(19) = 0.2$, $p = .991$.

Experiment 2: Conditional color acquisition

The three pigeons tested in Experiment 2 all learned to discriminate the conditional conjunction of color and bystander. This acquisition is shown in the right panel of [Fig. 2](#), which shows this group's average ρ across the eight 10-sessions blocks of the experiment. By the sixth block, the pigeons were significantly above chance, $t(2) = 7.2$, $p = .019$, $d = 4.2$. The one pigeon who had learned Experiment 1 learned this discrimination by the second block, $t_s(9) > 5.0$, $p < .001$, $d_s > 1.6$. The other two pigeons learned by the third, $t_s(9) > 5.0$, $p < .001$, $d_s > 1.6$, and sixth, $t_s(9) > 10.7$, $p < .001$, $d_s > 3.4$, block, respectively. By the final block, the three pigeons had an average discrimination index of 0.88 ($SD = .03$), comparable to the performance of the actor-conditional birds from Experiment 1.

The bottom panel of Fig. 3 shows the peck distributions for the three pigeons in this experiment. Given that the only model was the bystander, the right half of this standardized graph represents the area of the colored stimulus. This panel showed that all three pigeons favored the bystander's side of the display during their successful conditional discrimination. Two birds showed a clear switch from Experiment 1 (mean proportion of pecks toward bystander, 84% and 99%), while the third, more variable, bird shifted its behavior towards the bystander side of the display, $M = 54\%$, $ts(19) > 8.6$, $ps < .001$.

Discussion

The results of these two experiments suggest that pigeons can more readily process actions when the critical conditional information resides within the actor than in a bystander. All four pigeons in the actor-conditional group learned their conditional discrimination, while the bystander-conditional group failed, with only one bird learning that discrimination in a comparable period. In Experiment 2, the same fundamental spatial and perceptual factors for the bystander model remained in place as in Experiment 1, but all three pigeons were able to learn when a large color cue replaced the action and actor. The successful acquisition with this modification in the second experiment indicates that their prior learning difficulty was not related to the physical and spatial arrangement of the display or due to the discriminability of the different bystander models. Instead, these data suggest that the key limitation was related to connecting an actor's behavior with an adjacent bystander's identity, though further tests are needed to rule out other alternatives. Taken together, these results tentatively suggest for the first time that for pigeons, the actions of a digital model are intrinsically bound, configured, or processed integrally with the acting model.

The unification of the important properties of the display at the same spatial location in the actor-conditional discrimination quite likely contributes to making this condition easier. Acquiring both the necessary action information and the necessary model identity information in a single moment from a single location must be advantageous. Pigeons' difficulty in processing spatially separated visual information is well established in a variety of contexts (Lamb & Riley, 1981; Qadri & Cook, 2015). The distributions of pecking across the display by the pigeons are certainly consistent with the benefit of unification. Three of the four actor-conditional pigeons clearly pecked at the actor's different locations across trials. Similar tracking behavior has been found to indicate attentional focus (Brown, Cook, Lamb, & Riley, 1984; L. Dittrich et al., 2010; Jenkins & Sainsbury, 1970), suggesting that these birds found that the actor contained the critical information for the discrimination.

If unification or colocalization was particularly critical, the difficulty faced by three of the four bystander-conditional pigeons in learning their discrimination is perhaps not surprising. The spatial separation between the relevant bystander model and to-be-discriminated actions in the actor likely made "tracking" both at the same time harder. Perhaps a form of attentional capture might have also contributed to disrupting learning, given that three of these bystander-conditional pigeons (including the one bird who learned) pecked towards the actor in 54%–60% of their pecks. Motion is a perceptual feature known to be a dominant factor in attracting attention in humans and other animals (Abrams & Christ, 2003; W. H. Dittrich & Lea, 1993; Pratt, Radulescu, Guo, & Abrams, 2010). Consequently, these pigeons may have spent their time processing the moving model to the detriment of evaluating or detecting the identity of the critical bystander. However, since up to 60% of pecking was directed toward the actor, 40% of pecks remained directed toward the bystander's portion of the display. This reduces the likelihood that attentional capture can account for all of their difficulties, since a considerable amount of pecking/attention seems distributed across both panels. Furthermore, the 8-cm distance between the display screen and the touchscreen reduced the total visual angle of the stimuli, which should also have promoted the processing of the entire display. Thus, while physical separation and attentional limitations may have contributed to their difficulties, the bystander-conditional birds still had the opportunity to attend to both models and still did not learn the discrimination. In examining the behavior of the lone bird that learned in the bystander-conditional group, we could not identify any behavior or strategy that accounted for his relative success.

Whatever the source of the limitation, the substitution of the two colors for the two actions relieved the difficulty. After this exchange, the pigeons in Experiment 2 were able to use the information from both halves of the display to conditionalize their pecking. In addition to the research on the primacy of motion as a salient visual cue (e.g., W. H. Dittrich & Lea, 1993), different lines of evidence suggest that pigeons may process stimuli holistically and are unable to fully suppress processing of irrelevant features (Honig, 1993; Lea et al., 2018; Smith et al., 2011). Therefore, beyond the removal of motion cues, perhaps the reduction of available dimensions to two or the reduction of available form features may have been critical. Additionally, the color panels in Experiment 2 spanned the entire video frame, potentially reducing the amount of attention given to that panel and facilitating attention towards the neutrally posed "bystander" model. Indeed, two of the three birds exhibited shifts to pecking at the "bystander" in this condition, in one case directing 99% of pecks to the digital model, suggesting that the pigeons were directing their attention to the bystander. A yet more skeptical explanation would suggest that these pigeons learned more slowly overall, and the change in stimulus conditions was

simply coincidental with their learning rate. Further research evaluating the different aspects of the conditionalizing cue with different sets of pigeons can evaluate the contribution each of these aspects makes on the pigeons' learning. Presently, we tentatively take their success in this experiment to suggest the prior difficulty had been in connecting the actors' actions to the bystander's identity rather than spatial or model discriminability factors. Again, this is consistent with the idea that in Experiment 1, the action features were bound closely to the actor's identity, preventing it from being easily processed in conjunction with the bystander's identity.

The initial conditional discrimination of simultaneously recognizing the behavior of the actor and the identity of the actor would seem to be a form of configural processing. The representation underlying the behavior would have a single unified object (e.g., white shirt + Indian dance or black shirt + martial arts) that gains association with reward. Models of learning and discrimination using these configural representations have had success in accounting for a diversity of results in animal cognition (George & Pearce, 2012; Pearce, 2002; Pearce & Bouton, 2001). Conceivably, the color patch could have generated a similar configural display-wide representation (e.g., purple shirt + red square or green shirt + green square). Alternatively, the large size of color patches could instead have created novel contexts in which the pigeons learned to search for the reinforced bystander identity. If so, then these color cue acts might have acted as occasion setter, modulating how the pigeon processes the identity of the bystander (Bonardi, Bartle, & Jennings, 2012). In this explanation, the pigeon first recognizes the occasion setter, the large color panel, and this recognition changes how or what is processed in the bystander portion of the display. This sort of hierarchical analysis could pose an easier problem to solve than the original two-model requirement, which might require simultaneous processing of the features in order to generate the configural representation. This potential change in the strategy could also account for the success of the pigeons in Experiment 2.

Our previous research using a similar conditional go/no-go discrimination showed that pigeons can process both action and pose information from the same digital model when discriminating complex action sequences (Qadri & Cook, 2017). The current results suggest that along with actions and poses, the model's identity or appearance is also readily available to the pigeon. This facility seems to allow the conjunction of identity and behavior to be processed in a way that is not possible when the same information is distributed across two models. The actions of an actor or agent are seemingly more easily configured together by pigeons when they are portrayed within a single model. Perhaps these are features that are integral, such that processing behavior and identity are linked, to the point that even if one were irrelevant, it could cause

interference in processing (Garner & Felfoldy, 1970). Recent research is beginning to examine how humans process and determine who is acting in a scene given brief glimpses of information (Hafri et al., 2013; Hafri et al., 2018). Given the spontaneous, rapid, and accurate processing of this information in humans, it seems likely that the neural processes underlying action recognition in humans also configures actions as properties of actors. If both humans and pigeons perceive actions and actors as unified features of a configuration, an important question is raised about whether this is the result of convergent evolution or conservation of visual mechanisms across evolutionary time (Cook, Qadri, & Keller, 2015).

The configuration of identity and action also raise a type of the traditional and challenging binding problem (Treisman, 1996). In early vision, motion, color, and shape features are extracted from the visual scene in part along separate pathways (Butler & Hodos, 1996; Farah, 2000; Shimizu & Karten, 1993; Wang, Jiang, & Frost, 1993). However, humans and other animals behave as if the world has objects that contain each of those features. How does the brain reconnect the information from those separate pathways into a single representation for further processing? While traditionally structured as a problem for low-level feature maps such as color and orientation, the problem of action recognition is also distributed across numerous areas in the human brain and faces the same the difficulty (Rizzolatti, Fogassi, & Gallese, 2001). If action recognition in pigeons is also distributed in the same way, then the displays tested here would have challenged those brain regions to coordinate their processing for the final configural representation of the acting model. In both humans and nonhuman animals, determining the nature of this configural representation might advance our understanding of these disconnected brain regions' relationships.

We think that stating that the pigeons perceived our digital human models as being "agents" would be an over interpretation of our current results. With that said, the use of similar digital, two-body stimuli have revealed deficits by high-functioning individuals with autism compared with typical controls in understanding communicative behavior (Lühe et al., 2016). It has been proposed, for example, that individuals with autism have deficits in visual processing, especially as related to agency detection and action recognition (Blake, Turner, Smoski, Pozdol, & Stone, 2003; Farran & Brosnan, 2011; Simmons et al., 2009). Understanding how actors (mates/conspicuous/predators) and their different behaviors are encoded and represented by the observing animal raises interesting questions related to theory of mind, the role of action and actor representations in social processing, and the attribution of intentionality. Our results start by suggesting that spatial colocalization and processing as a single object is likely critical. Variations on the conditional

discrimination involving multiple figures explored here are worthy of further study.

Author notes This research and its preparation were supported by a grant from the National Eye Institute (#RO1EY022655). Home page (www.pigeon.psy.tufts.edu). The motion data used in this project was obtained from mocap.cs.cmu.edu. The database was created with funding from NSF EIA-0196217. We thank Suzanne Gray for her insightful comments on an earlier version of this manuscript. Data associated with this manuscript can be found at (<https://doi.org/10.6084/m9.figshare.9642929>). Neither of these experiments were preregistered.

References

- Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, 14(5), 427–432. doi:<https://doi.org/10.1111/1467-9280.01458>
- Asen, Y., & Cook, R. G. (2012). Discrimination and categorization of actions by pigeons. *Psychological Science*, 23(6), 617–624. doi:<https://doi.org/10.1177/0956797611433333>
- Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology*, 37(2), 213–230. doi:[https://doi.org/10.1016/0022-0965\(84\)90001-8](https://doi.org/10.1016/0022-0965(84)90001-8)
- Blake, R., & Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology*, 58(1), 47–73. doi:<https://doi.org/10.1146/annurev.psych.57.102904.190152>
- Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. *Psychological Science*, 14(2), 151–157. doi:<https://doi.org/10.1111/1467-9280.01434>
- Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2(8), 561–567. doi:<https://doi.org/10.1038/35086023>
- Bonardi, C., Bartle, C., & Jennings, D. (2012). US specificity of occasion setting: Hierarchical or configural learning? *Behavioural Processes*, 90(3), 311–322. doi:<https://doi.org/10.1016/j.beproc.2012.03.005>
- Brown, M. F., Cook, R. G., Lamb, M. R., & Riley, D. A. (1984). The relation between response and attentional shifts in pigeon compound matching-to-sample performance. *Animal Learning & Behavior*, 12(1), 41–49. doi:<https://doi.org/10.3758/BF03199811>
- Butler, A. B., & Hodos, W. (1996). *Comparative vertebrate neuroanatomy: Evolution and adaptation*. New York, NY: Wiley-Liss.
- Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage*, 12(3), 314–325. doi:<https://doi.org/10.1006/nimg.2000.0612>
- Cook, R. G., & Katz, J. S. (1999). Dynamic object perception by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 25(2), 194–210. doi:<https://doi.org/10.1037/0098-7403.25.2.194>
- Cook, R. G., Qadri, M. A. J., & Keller, A. M. (2015). The analysis of visual cognition in birds: Implications for evolution, mechanism, and representation. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 63, pp. 173–210). Elsevier. doi:<https://doi.org/10.1016/bs.plm.2015.03.002>
- D'Eath, R. B. (1998). Can video images imitate real stimuli in animal behaviour experiments? *Biological Reviews of the Cambridge Philosophical Society*, 73(3), 267–292. doi:<https://doi.org/10.1111/j.1469-185X.1998.tb00031.x>
- Dittrich, L., Rose, J., Buschmann, J. U. F., Bourdonnais, M., & Güntürkün, O. (2010). Peck tracking: A method for localizing critical features within complex pictures for pigeons. *Animal Cognition*, 13(1), 133–143. doi:<https://doi.org/10.1007/s10071-009-0252-x>
- Dittrich, W. H., & Lea, S. E. G. (1993). Motion as a natural category for pigeons: Generalization and a feature-positive effect. *Journal of the Experimental Analysis of Behavior*, 59(1), 115–129. doi:<https://doi.org/10.1901/jeab.1993.59-115>
- Dittrich, W. H., & Lea, S. E. G. (1994). Visual perception of intentional motion. *Perception*, 23(3), 253–268. doi:<https://doi.org/10.1068/p230253>
- Dittrich, W. H., & Lea, S. E. G. (2001). Motion discrimination and recognition. In R. G. Cook (Ed.), *Avian visual cognition* [Online]. Available at: pigeon.psy.tufts.edu/avc/dittrich/
- 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. doi:<https://doi.org/10.1901/jeab.1998.70-281>
- Farah, M. J. (2000). *The cognitive neuroscience of vision*. New York, NY: Blackwell Publishing.
- Farran, E. K., & Brosnan, M. J. (2011). Perceptual grouping abilities in individuals with autism spectrum disorder: Exploring patterns of ability in relation to grouping type and levels of development. *Autism Research*, 4(4), 283–292. doi:<https://doi.org/10.1002/aur.202>
- Fleishman, L., & Endler, J. (2000). Some comments on visual perception and the use of video playback in animal behavior studies. *Acta Ethologica*, 3(1), 15–27. doi:<https://doi.org/10.1007/s102110000025>
- Frohnwieser, A., Murray, J. C., Pike, T. W., & Wilkinson, A. (2016). Using robots to understand animal cognition. *Journal of the Experimental Analysis of Behavior*, 105(1), 14–22. doi:<https://doi.org/10.1002/jeab.193>
- Gallup, A. C., Swartwood, L., Militello, J., & Sackett, S. (2015). Experimental evidence of contagious yawning in budgerigars (*Melopsittacus undulatus*). *Animal Cognition*, 18(5), 1051–1058. doi:<https://doi.org/10.1007/s10071-015-0873-1>
- Galoch, Z., & Bischof, H.-J. (2007). Behavioural responses to video playbacks by zebra finch males. *Behavioural Processes*, 74(1), 21–26. doi:<https://doi.org/10.1016/j.beproc.2006.09.002>
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1(3), 225–241. doi:[https://doi.org/10.1016/0010-0285\(70\)90016-2](https://doi.org/10.1016/0010-0285(70)90016-2)
- George, D. N., & Pearce, J. M. (2012). A configural theory of attention and associative learning. *Learning & Behavior*, 40(3), 241–254. doi:<https://doi.org/10.3758/s13420-012-0078-2>
- Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology: General*, 142(3), 880–905. doi:<https://doi.org/10.1037/a0030045>
- Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, 175, 36–52. doi:<https://doi.org/10.1016/j.cognition.2018.02.011>
- Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243–259. doi:<https://doi.org/10.2307/1416950>
- Hermstein, R. J., Loveland, D. H., & Cable, C. (1976). Natural concepts in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 2(4), 285–302. doi:<https://doi.org/10.1037/0097-7403.2.4.285>

- Honig, W. K. (1993). The stimulus revisited: My, how you've grown. In T. R. Zentall (Ed.), *Animal cognition: A tribute to Donald A. Riley* (pp. 19–33). Hillsdale, NJ: Erlbaum.
- Jenkins, H., & Sainsbury, R. S. (1970). Discrimination learning with the distinctive feature on positive or negative trials. In D. I. Mastofsky (Ed.), *Attention: Contemporary theory and analysis* (pp. 239–273). New York, NY: Appleton-Century-Crofts.
- Johansson, G. (1973). Visual perception of biological motion and a model of its analysis. *Perception & Psychophysics*, 14(2), 201–211. doi:<https://doi.org/10.3758/BF03212378>
- Klüver, M., Hecht, H., & Troje, N. F. (2016). Internal consistency predicts attractiveness in biological motion walkers. *Evolution and Human Behavior*, 37(1), 40–46. doi:<https://doi.org/10.1016/j.evolhumbehav.2015.07.001>
- Koban, A., & Cook, R. G. (2009). Rotational object discrimination by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 35(2), 250–265. doi:<https://doi.org/10.1037/a0013874>
- Lamb, M. R., & Riley, D. A. (1981). Effects of element arrangement on the processing of compound stimuli in pigeons (*Columba livia*). *Journal of Experimental Psychology: Animal Behavior Processes*, 7(1), 45–58. doi:<https://doi.org/10.1037/0097-7403.7.1.45>
- Lea, S. E. G. (1984). In what sense do pigeons learn concepts? In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition* (pp. 263–276). Hillsdale, NJ: Erlbaum.
- Lea, S. E. G. (2010). What's the use of picture discrimination experiments? *Comparative Cognition & Behavior Reviews*, 5, 143–147. doi:<https://doi.org/10.3819/ccbr.2010.50010>
- Lea, S. E. G., & Dittrich, W. H. (2000). What do birds see in moving video images? In J. Fagot (Ed.), *Picture perception in animals* (pp. 143–180). East Sussex, Brighton, UK: Psychology Press.
- Lea, S. E. G., & Harrison, S. N. (1978). Discrimination of polymorphous stimulus sets by pigeons. *The Quarterly Journal of Experimental Psychology*, 30(3), 521–537. doi:<https://doi.org/10.1080/0033557843000106>
- Lea, S. E. G., Pothos, E. M., Wills, A. J., Leaver, L. A., Ryan, C. M. E., & Meier, C. (2018). Multiple feature use in pigeons' category discrimination: The influence of stimulus set structure and the salience of stimulus differences. *Journal of Experimental Psychology: Animal Learning and Cognition*, 44(2), 114–127. doi:<https://doi.org/10.1037/xan0000169>
- Lea, S. E. G., Slater, A. M., & Ryan, C. M. E. (1996). Perception of object unity in chicks: A comparison with the human infant. *Infant Behavior and Development*, 19(4), 501–504. doi:[https://doi.org/10.1016/S0163-6383\(96\)90010-7](https://doi.org/10.1016/S0163-6383(96)90010-7)
- Lühe, T. V. D., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693), 20150373. doi:<https://doi.org/10.1098/rstb.2015.0373>
- Osthaus, B., Lea, S. E. G., & Slater, A. M. (2005). Dogs (*Canis lupus familiaris*) fail to show understanding of means-end connections in a string-pulling task. *Animal Cognition*, 8(1), 37–47. doi:<https://doi.org/10.1007/s10071-004-0230-2>
- Partan, S., Yelda, S., Price, V., & Shimizu, T. (2005). Female pigeons, *Columba livia*, respond to multisensory audio/video playbacks of male courtship behaviour. *Animal Behaviour*, 70(4), 957–966. doi:<https://doi.org/10.1016/j.anbehav.2005.03.002>
- Patterson-Kane, E., Nicol, C. J., Foster, T. M., & Temple, W. (1997). Limited perception of video images by domestic hens. *Animal Behavior*, 53(5), 951–963. doi:<https://doi.org/10.1006/anbe.1996.0385>
- Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, 30(2), 73–95. doi:<https://doi.org/10.3758/BF03192911>
- Pearce, J. M., & Bouton, M. E. (2001). Theories of associative learning in animals. *Annual Review of Psychology*, 52, 111–139. doi:<https://doi.org/10.1146/annurev.psych.52.1.111>
- Poppe, R. (2010). A survey on vision-based human action recognition. *Image and Vision Computing*, 28(6), 976–990. doi:<https://doi.org/10.1016/j.imavis.2009.11.014>
- Pratt, J., Radulescu, P. V., Guo, R. M., & Abrams, R. A. (2010). It's alive!: Animate motion captures visual attention. *Psychological Science*, 21(11), 1724–1730. doi:<https://doi.org/10.1177/0956797610387440>
- Qadri, M. A. J., Asen, Y., & Cook, R. G. (2014a). Visual control of an action discrimination in pigeons. *Journal of Vision*, 14(5), 1–19. doi:<https://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. doi:<https://doi.org/10.3819/ccbr.2015.100004>
- Qadri, M. A. J., & Cook, R. G. (2017). Pigeons and humans use action and pose information to categorize complex human behaviors. *Vision Research*, 131, 16–25. doi:<https://doi.org/10.1016/j.visres.2016.09.011>
- Qadri, M. A. J., Sayde, J. M., & Cook, R. G. (2014b). Discrimination of complex human behavior by pigeons (*Columba livia*) and humans. *PLOS ONE*, 9(11), e112342. doi:<https://doi.org/10.1371/journal.pone.0112342>
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2(9), 661–670. doi:<https://doi.org/10.1038/35090060>
- Rowe, C., & Skelhorn, J. (2004). Avian psychology and communication. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271(1547), 1435–1442. doi:<https://doi.org/10.1098/rspb.2004.2753>
- Ryan, C. M. E., & Lea, S. E. G. (1994). Images of conspecifics as categories to be discriminated by pigeons and chickens: Slides, video tapes, stuffed birds and live birds. *Behavioural Processes*, 33(1/2), 155–175. doi:[https://doi.org/10.1016/0376-6357\(94\)90064-7](https://doi.org/10.1016/0376-6357(94)90064-7)
- Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299–309. doi:[https://doi.org/10.1016/S1364-6613\(00\)01506-0](https://doi.org/10.1016/S1364-6613(00)01506-0)
- Shimizu, T. (1998). Conspecific recognition in pigeons (*Columba livia*) using dynamic video images. *Behaviour*, 135(1), 43–53. doi:<https://doi.org/10.1163/156853998793066429>
- Shimizu, T., & Karten, H. J. (1993). The avian visual system and the evolution of the neocortex. In H. P. Zeigler & H. J. Bischof (Eds.), *Vision, brain, and behavior in birds*. Cambridge, MA: MIT Press.
- Simmons, D. R., Robertson, A. E., McKay, L. S., Toal, E., McAleer, P., & Pollick, F. E. (2009). Vision in autism spectrum disorders. *Vision Research*, 49(22), 2705–2739. doi:<https://doi.org/10.1016/j.visres.2009.08.005>
- Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., & Grace, R. C. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Bulletin & Review*, 18(2), 414–421. doi:<https://doi.org/10.3758/s13423-010-0047-8>
- Takahasi, M., Ikebuchi, M., & Okanoya, K. C. A. (2005). Spatiotemporal properties of visual stimuli for song induction in Bengalese finches. *NeuroReport*, 16(12), 1339–1343. doi:<https://doi.org/10.1097/01.wnr.0000175251.10288.5f>
- Takahasi, M., & Okanoya, K. (2013). An invisible sign stimulus: Completion of occluded visual images in the Bengalese finch in an ecological context. *Neuroreport*, 24(7), 370–374. doi:<https://doi.org/10.1097/WNR.0b013e328360ba32>
- Treisman, A. M. (1996). The binding problem. *Current Opinion in Neurobiology*, 6(2), 171–178. doi:[https://doi.org/10.1016/S0959-4388\(96\)80070-5](https://doi.org/10.1016/S0959-4388(96)80070-5)

- Venables, W., & Ripley, B. (2002). *Modern applied statistics using S*. New York, NY: Springer.
- Wang, Y. C., Jiang, S., & Frost, B. J. (1993). Visual processing in pigeon nucleus rotundus: Luminance, color, motion, and looming subdivisions. *Visual Neuroscience*, 10(1), 21–30. doi:<https://doi.org/10.1017/s0952523800003199>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.