

Asymmetrical interactions in the perception of face identity and emotional expression are not unique to the primate visual system

Fabian A. Soto

Department of Psychology, University of Iowa,
Iowa City, IA, USA



Edward A. Wasserman

Department of Psychology, University of Iowa,
Iowa City, IA, USA



The human visual system appears to process the identity of faces separately from their emotional expression, whereas the human visual system does not appear to process emotional expression separately from identity. All current explanations of this visual processing asymmetry implicitly assume that it arises because of the organization of a specialized human face perception system. A second possibility is that this finding reflects general principles of perceptual processing. Studying animals that are unlikely to have evolved a specialized face perception system may shed fresh light on this issue. We report two experiments that investigated the interaction of identity and emotional expression in pigeons' perception of human faces. **Experiment 1** found that pigeons perceive the similarity among faces sharing identity and emotion, and that these two dimensions are integral according to a spatial model of generalization. **Experiment 2** found that pigeons' discrimination of emotion was reliably affected by irrelevant variations in identity, whereas pigeons' discrimination of identity was not reliably affected by irrelevant variations in emotion. Thus, the asymmetry previously reported in human studies was reproduced in our pigeon study. These results challenge the view that a specialized human face perception system must underlie this effect.

Keywords: face recognition, categorization, object recognition

Citation: Soto, F. A., & Wasserman, E. A. (2011). Asymmetrical interactions in the perception of face identity and emotional expression are not unique to the primate visual system. *Journal of Vision*, 11(3):24, 1–18, <http://www.journalofvision.org/content/11/3/24>, doi:10.1167/11.3.24.

Introduction

Face perception is indubitably one of the most important achievements of the human visual system and it plays a vital role in our everyday social functioning. One interesting aspect of face processing is that faces can be simultaneously classified according to a number of different criteria. A single face can be categorized according to its identity, age, gender, emotional expression, gaze direction, and lip movements during speech.

Some researchers have proposed that the perceptual processes underlying some of these varied categorization abilities operate independently of each other. In their classic model of human face perception, Bruce and Young (1986) proposed that face perception starts with *structural encoding*, resulting in the representation of a face through a set of interconnected descriptions, some of them view-centered and others more abstract renderings of features and their spatial configuration. Later, the processing of emotional expression and identity progress through parallel processing routes. View-centered descriptions provide information for the analysis of expression, whereas abstract, expression-independent descriptions provide information for the analysis of identity.

This parallel processing model has generally been supported by the results of early neurobiological studies (for a critical view, see Tiberghien, Baudouin, Guilleme, & Montoute, 2003). Neuropsychological studies showed that some prosopagnosic patients with an impairment in face recognition also exhibited a relatively preserved ability to recognize emotional expressions, whereas other patients presented the reverse pattern of performance (e.g., Humphreys, Donnelly, & Riddoch, 1993). Neuroimaging studies have found that face identification is related to high activation of the fusiform gyri, whereas emotional expression categorization is related to high activation of several other brain areas, including the superior temporal sulcus (for a review, see Posamentier & Abdi, 2003). Finally, neurophysiological studies in monkeys have found that different cortical cells respond selectively during the processing of emotional expression and identity (e.g., Hasselmo, Rolls, & Baylis, 1989).

The results of many behavioral studies, however, do not support the hypothesis of strict independence of processing facial identity and emotional expression. Some of these studies have used the *Garner interference task* (Garner, 1974), in which participants are asked to classify a set of stimuli according to a relevant dimension (e.g., the identity of faces) in two different conditions: in the control

condition, the stimuli vary only in the relevant dimension, whereas in the orthogonal condition, the stimuli also vary orthogonally in a dimension that is irrelevant to the classification task (e.g., the emotional expression of faces). The rationale behind this design is that, if the two dimensions are processed separately—so that the participants can selectively attend to the relevant dimension and ignore the irrelevant dimension—then there should be no disparity in performance between the control and orthogonal conditions. If, on the other hand, the two dimensions are processed integrally—and participants cannot selectively attend to the relevant dimension without also processing the irrelevant dimension—then an impairment in performance should be observed in the orthogonal condition compared to the control condition.

Studies using the Garner interference task have found an asymmetrical pattern of interference between the processing of identity and emotional expression. Whereas variations in identity have an effect on the classification of emotional expression, variations in emotional expression do not have an effect on the classification of identity (Schweinberger, Burton, & Kelly, 1999; Schweinberger & Soukup, 1998).

A second set of studies has examined the issue of dimensional interaction using a face adaptation paradigm (Leopold, O'Toole, Vetter, & Blanz, 2001). Here, participants are shown morphed faces varying along a continuum that has two different target faces as endpoints. People who are pre-exposed to one of these target faces show a perceptual aftereffect in which an ambiguous face (that is, a face in the middle of the continuum) is more likely to be identified as belonging to the face at the opposite end of the continuum from the pre-exposed face. The same type of aftereffect has been shown for faces varying in gender, race, and emotional expression (Webster, Kaping, Mizokami, & Duhamel, 2004).

Experiments using the face adaptation paradigm have found that the identity adaptation aftereffect is relatively unchanged across changes in emotional expression (Fox, Oruç, & Barton, 2008); that is, the magnitude of this aftereffect remains relatively constant regardless of whether the adapted face shows the same emotional expression as the test face or a different emotional expression. By contrast, the magnitude of the adaptation aftereffect for emotional expression varies depending on whether the adapted face and the ambiguous face belong to the same individual or to a different individual (Fox & Barton, 2007); this effect cannot be explained by differences among individuals in their way of expressing emotion (Ellamil, Susskind, & Anderson, 2008).

Thus, behavioral evidence points toward an asymmetrical pattern of dependence in the processing of these two stimulus dimensions; although identity seems to be processed separately from emotional expression, emotional expression does not seem to be processed separately from identity. This conclusion is at odds with the classical parallel model of Bruce and Young (1986), but it is in line

with the more recent neurobiological model proposed by Haxby, Hoffman, and Gobbini (2000).

In this newer neurobiological model, the visual analysis of face information is carried out by a neural system comprising three regions in the visual cortex. The inferior occipital gyri are in charge of the early perception of facial features, providing input to the other two systems that process different aspects of faces in a relatively independent fashion. Invariant aspects of faces are processed by the lateral fusiform gyrus; thus, this area is in charge of the perception of face identity, gender, etc. Changeable aspects of faces are processed by the superior temporal sulcus; thus, this area is responsible for the perception of emotional expression, eye gaze, lip movements during speech, etc. Although the latter two systems are assumed to be anatomically separated, Haxby et al. emphasize that their degree of functional separation is unclear.

In line with Haxby et al.'s model, the same pattern of interaction discussed above for the perception of facial identity and emotional expression has been found using other invariant and changeable aspects of faces, such as gender and emotional expression (Atkinson, Tipples, Burt, & Young, 2005), as well as identity and facial speech gestures (Schweinberger & Soukup, 1998). Similarly, studies using the face adaptation paradigm have found that the emotional expression aftereffect is affected by changes in other permanent aspects of faces beyond identity, such as gender and race (Bestelmeyer, Jones, DeBruine, Little, & Welling, 2010).

Some results suggest that how identity and emotional expression are represented in the areas identified by Haxby et al. may account for the asymmetrical interaction between these dimensions that is observed in behavioral studies. For example, in a functional MRI adaptation study, Winston, Henson, Fine-Goulden, and Dolan (2004) found that the fusiform cortex may be exclusively involved in the processing of face identity, without any processing of emotional expression (but see Ganel, Valyear, Goshen-Gottstein, & Goodale, 2005). On the other hand, areas in the STS are involved in processing emotional expression without identity (mid-STS), plus other areas in the STS are involved in processing both facial properties (posterior STS); thus, the STS may not be a clearly separate anatomical substrate for processing emotion.

Several other researchers have proposed theoretical views that depart from the assumption of complete independence in processing identity and emotional expression, while being more explicit than Haxby et al. about any assumed interaction between these processes. For example, Schweinberger et al. (1999) proposed that the processing of identity and emotional expression occur in parallel, but there is also a unidirectional interaction in which the outcome of the identity system influences the perception of emotion. Fox and Barton (2007) proposed that the visual system computes an identity-dependent representation of emotional expression in addition to the identity-independent representation proposed by other authors.

Ganel and Goshen-Gottstein's (2004; see also Ganel et al., 2005) *structural reference hypothesis* proposes that emotional expression is coded as dynamic variations from the invariant structure of faces. This invariant structure influences how each individual expresses emotion. Therefore, the processing of invariant aspects of a face, such as identity, should influence the processing of its dynamic aspects. However, because only some very specific invariant structures can produce a particular emotional expression, this hypothesis predicts a symmetrical interaction between emotion and identity that most studies have not found (but see Ganel & Goshen-Gottstein, 2004; for a critical view of their results, see Martens, Leuthold, & Schweinberger, 2010).

Regardless of which of these (or other) models may ultimately prevail, there are two important ideas to extract from the behavioral data and theoretical developments in this interesting realm of research. First, an impressive body of evidence suggests that the pattern of interaction between identity and emotional expression in human face perception is asymmetrical. This evidence prompts researchers to conclude that the “invariant dimensions of a stimulus are more useful referents for computing information about changeable aspects of that stimulus than vice versa” (Atkinson et al., 2005, p. 1211). Second, all current theoretical explanations of this effect have implicitly assumed that a specialized *human* system for face processing exists and that the asymmetrical pattern of interference between identity and emotion results from the way in which this uniquely human system is organized.

This implicit assumption accords with an important and growing body of behavioral and neurobiological data that has been interpreted as evidence that a specialized face processing system does indeed exist for humans (and perhaps for some other primates; Farah, 1996; Kanwisher, 2000; Kanwisher & Yovel, 2006; Tsao & Livingstone, 2008). As is the case with other purportedly specialized systems, it has been suggested that the human system for face processing has been specialized for this task because of strong evolutionary pressures (Pascalis & Kelly, 2009) and that some of its mechanisms are innate and inheritable (Farah, Rabinowitz, Quinn, & Liu, 2000; Sugita, 2008; Wilmer et al., 2010).

Thus, one prominent possibility is that identity can be processed separately from emotion by means of an independent, modular perceptual system that is highly specialized in processing invariant facial information (Schweinberger & Soukup, 1998). Nevertheless, a second, less prominent possibility also exists: perhaps the earlier reviewed pattern of performance is due to the operation of general principles of perceptual processing instead of an adaptive specialization.

For example, work with principal components analysis of face images has shown that functional dissociations between facial characteristics can be explained as the result of compact coding based solely on statistical regularities across faces (Calder, Burton, Miller, Young, &

Akamatsu, 2001). Most of the components that carry important information for the classification of identity are different from those that carry information for the classification of emotion, although there are also components that are common to both tasks. Furthermore, a single-layer neural network, using the representations generated from principal components analysis, can account for some of the behavioral dissociations between identity and emotion reported in the literature (Cottrell, Branson, & Calder, 2002).

One strategy for determining whether a particular behavior is the result of a specialized adaptation or a general perceptual system is to compare the performance of distantly related species. Distantly related species that share a remote common ancestor are likely to share quite general mechanisms of information processing that were functional in that ancestor. These general mechanisms ought to be used to solve tasks that are important for survival across diverse environments (e.g., object recognition, in general, rather than face recognition, in particular); thus, such general mechanisms should be conserved across evolution (Bitterman, 2000; Papini, 2002). Adaptive specializations, on the other hand, should be present in only one of two distantly related species being compared, unless similar evolutionary pressures prompted convergent evolution of the same processes (Papini, 2002).

Birds are particularly useful to compare with primates in vision science, because both taxonomic groups possess the most advanced biological visual systems, they share a common evolutionary ancestor in the early amniotes, and the visual systems of all amniotes share basic organizational properties (Husband & Shimizu, 2001; Shimizu, 2009; Shimizu & Bowers, 1999; for reviews of behavioral studies on high-level visual cognition in birds, see Cook, 2001; Wasserman & Zentall, 2006). However, it is highly unlikely that birds have evolved any specialized perceptual system to process human faces. Any similarities between birds and primates in the perception of human faces is therefore likely to be the result of mechanisms that have been conserved across evolution or processes that are similar because of convergent evolution under similar environmental exigencies (for example, those resulting from similar structures of the visual environment; for an example of this approach using an invertebrate species, see Avargues-Weber, Portelli, Bernard, Dyer, & Giurfa, 2009).

Several previous studies have assessed birds' ability to categorize human faces. From these studies, we know that pigeons and crows learn to categorize human faces according to gender (Bogale, Aoyama, & Sugita, 2011; Huber, Troje, Loidolt, Aust, & Grass, 2000; Troje, Huber, Loidolt, Aust, & Fieder, 1999) and emotion (Jitsumori & Yoshihara, 1997) and that they transfer this learning to novel visual images. Pigeons' performance in face categorization tasks comes under the control of several facial features or other stimulus properties, with variations in responding being explained quite well as a linear function of the presence or absence of such features (Huber & Lenz, 1993; Jitsumori & Yoshihara, 1997).

The actual stimulus properties controlling performance vary across studies, probably depending on factors such as categorization task and characteristics of the stimulus set. In categorization of faces by gender, color and shading have been shown to be particularly important, whereas sharp edges are less relevant (Huber et al., 2000; Troje et al., 1999). Pigeons use information near the eyes and chin to discriminate male versus female faces and they use information near the mouth to discriminate happy faces versus neutral faces (Gibson, Wasserman, Gosselin, & Schyns, 2005).

All of these previous studies have used human faces as a means to understand how birds process natural object categories in general. The approach taken here is different; we used pigeons to test for the possibility that a feature of *human* face recognition could be explained as the result of general visual processes. To the best of our knowledge, only one previous study has used this strategy in the study of face perception. In this study (Phelps & Roberts, 1994), both humans and squirrel monkeys showed an advantage in discriminating upright human and ape faces over their inverted versions, suggesting that a specialized primate system for processing of upright faces exists. To control for the possibility that this effect was due to an advantage of upright faces for processing by *any* biological visual system, these authors also trained pigeons with the same discrimination task and found that these animals did *not* show the same face inversion effect as primates.

Here, we explored pigeons' perception of human face identity and expression, in the hope of determining to what extent the pattern of interaction between these dimensions in humans results from principles of visual processing that are also functional in non-primates.

Experiment 1 was a preparatory study that sought to determine whether pigeons can perceive the similarity structure of stimuli that vary along the dimensions of human face identity and emotional expression. To do so, we used a task that has proven to be extremely effective in previous studies (Kirkpatrick-Steger & Wasserman, 1996). A second goal of this experiment was to determine whether the two dimensions combined integrally or separately according to a spatial model of multidimensional generalization (Soto & Wasserman, 2010a).

Experiment 2 explored in greater depth and detail the separability of each dimension with respect to the other using a behavioral test proposed in the context of General Recognition Theory (GRT; Ashby & Townsend, 1986); this test follows a very similar logic to the Garner interference task that was used in several of the previously reviewed human experiments.

Experiment 1

Experiment 1 was aimed at determining whether pigeons perceive the similarity structure of stimuli that

vary along the dimensions of face identity and emotional expression. To do so, we used a multidimensional generalization task involving 16 photographs of human faces, resulting from the factorial combination of 4 different identities and 4 different emotional expressions. For each pigeon, responses to only one of these stimuli, the target stimulus, were reinforced with food delivery, whereas responses to any of the other stimuli were not reinforced. **Figure 1** shows the 16 images used in the present experiment. To provide an example, the image in the top left corner of **Figure 1** has been marked as the target stimulus and all of the other images have been grouped together as a function of the properties that they share with the target. The level of generalized responding to these non-reinforced stimuli should reflect their similarity to the target stimulus; this generalization measure was available for stimuli differing from the target in a single dimension (either identity or emotion, grouped in the blue and green rectangles in **Figure 1**, respectively) or in two dimensions (both identity and emotion, grouped in the orange rectangle in **Figure 1**). If pigeons perceive similarity across stimuli sharing the same identity, then they should exhibit higher responding to those stimuli that share the same identity with the target stimulus, but have a different emotion, than to those stimuli that differ from the target stimulus in both emotion and identity. In the example provided in **Figure 1**, there should be greater generalization of responding to the stimuli enclosed in the green rectangle than to those stimuli enclosed in the orange rectangle. The same logic applies for testing whether pigeons perceive similarity across stimuli sharing the same emotion.

As well, if one of the dimensions were more difficult to discriminate than the other, then stimuli sharing the same value as the target in the less discriminable dimension should show significantly higher generalization scores than stimuli sharing the same value as the target in the other, more discriminable dimension. In terms of the example provided in **Figure 1**, if identity were more difficult to discriminate than emotion (relative to the specific target stimulus chosen), then there should be higher generalization of responding to the stimuli enclosed in the green rectangle than to those enclosed in the blue rectangle. The opposite pattern would be true if emotion were more difficult to discriminate than identity. It was highly desirable to have evidence indicating whether there were any important disparities in discriminability for stimuli that were changed across the two dimensions under study.

Importantly, the generalization data obtained from this experiment also proved to be useful for determining how these two dimensions combine according to a spatial model of multidimensional stimulus generalization (Soto & Wasserman, 2010a). A longstanding theoretical approach to the representation of similarities among stimuli involves positioning each stimulus in a multidimensional space, where the dimensions represent variable stimulus properties and the similarity between the stimuli is inversely

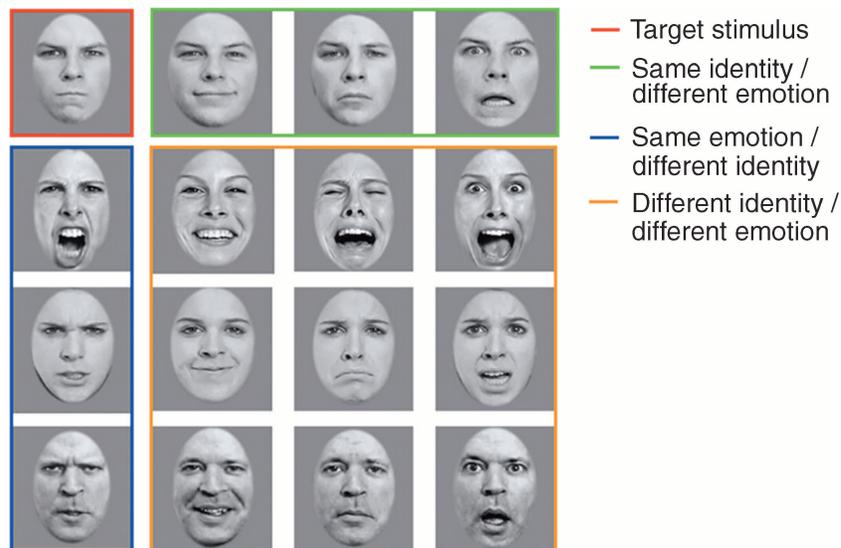


Figure 1. Stimuli used in the two experiments.

proportional to the distance between them in that space (for a review, see Nosofsky, 1992). The distance between stimuli can be computed from their coordinates in space following one of many metrics, but two of them have received particular attention from cognitive scientists. According to the City-Block metric, the distance between points is the sum of their distances along each dimension. According to the Euclidean metric, the distance between points is the length of the straight line that directly connects them and thus is computed using the Pythagorean formula.

Spatial models using the City-Block metric offer a good representation for stimuli varying along *separable* dimensions—those that can be attended to and processed independently of each other. On the other hand, spatial models using the Euclidean metric offer a good representation for stimuli varying along *integral* dimensions—those that cannot be attended to and processed independently of each other (Garner, 1974; Shepard, 1991).

The obtained stimulus generalization data were used to find the metric in a spatial model that would best describe the way in which these two dimensions interact with each other. A City-Block metric would provide evidence of dimensional separability, whereas a Euclidean metric would provide evidence of dimensional integrality.

Methods

Subjects

Four feral pigeons (*Columba livia*) were kept at 85% of their free-feeding weights by controlled daily feeding. The birds had previously participated in unrelated research.

Apparatus

The experiment used four $36 \times 36 \times 41$ cm operant conditioning chambers (a detailed description can be

found in Gibson, Wasserman, Frei, & Miller, 2004), located in a dark room with continuous white noise. The stimuli were presented on a 15-in LCD monitor located behind an AccuTouch resistive touch screen (Elo Touch-Systems, Fremont, CA) that was covered by a thin sheet of mylar for durability. A food cup was centered on the rear wall of the chamber. A food dispenser delivered 45-mg food pellets through a vinyl tube into the cup. A house light on the rear wall of the chamber provided illumination during sessions. Each chamber was controlled by an Apple eMac computer and the experimental procedure was programmed using Matlab Version 7.0.4 (The MathWorks) with the Psychophysics Toolbox extensions (Brainard, 1997). Images were displayed in a 7.5-cm square in the middle of the screen; the rest of the screen was black at all times.

Stimuli

The stimuli were 16 grayscale photographs of 4 individuals (2 males and 2 females, see Figure 1), each showing 4 different emotional expressions (Happiness, Anger, Sadness, Fear). The photographs were saved with a definition of 256×256 pixels. The main inner facial features (eyes, nose, and mouth) were aligned across all of the stimuli. The faces were shown through an elliptical aperture in a homogeneous gray screen; this presentation revealed only inner facial features and hid non-facial information, such as hairstyle and color.

Procedure

A different target stimulus was randomly selected for each pigeon from a set of four possible images with the following combinations of identity and emotion: {male 1, anger}, {female 1, happiness}, {female 2, sadness}, and

{male 1, fear} (see Figure 1). Responses to the target stimulus alone were consistently reinforced across the experiment.

All trials began with the presentation of a white square in the center display area of the screen. A single peck anywhere within the square led to the presentation of the trial stimulus. On a reinforced trial, the stimulus was presented and remained on for 15 s; the first response after 15 s turned the display area black and delivered food. On a non-reinforced trial, the stimulus was presented and remained on for 15 s, after which the display area automatically darkened and the intertrial interval began. On both reinforced and non-reinforced trials, scored responses were recorded only during the first 15 s that the stimulus was displayed on the screen. The intertrial interval randomly ranged from 6 to 10 s. Reinforcement consisted of 1 to 3 food pellets scheduled randomly from trial to trial.

The experiment began with a Baseline Training phase during which all of the trials were reinforced. Each daily session of Baseline Training was composed of 8 blocks of 24 trials each, for a total of 192 trials. In each block, there were 8 unscored presentations of the target stimulus and 1 scored presentation of each of the 16 visual stimuli, including the future target stimulus and the 15 future non-target stimuli. This trial organization was implemented to equilibrate the block structure of the Baseline phase and the following Discrimination Training phase. The order of trial presentation was randomized within each block.

Pigeons were kept on Baseline Training until the mean rate of responding to each of the 16 stimuli fell within 80% and 120% of the overall mean rate of responding. After the pigeons met criterion, they started Discrimination Training.

Each daily session of Discrimination Training was composed of 8 blocks of 24 trials, for a total of 192 trials. In each block, there were 8 reinforced and unscored presentations of the target stimulus and 1 non-reinforced and scored presentation of each stimulus, including the target stimulus and the 15 non-target stimuli. This procedure of reinforcing 33.3% of all of the trials assured sustained responding across Discrimination Training. Reinforcement of 88.9% of the target stimulus trials fostered high rates of responding to this stimulus, while equilibrating the total number of non-reinforced and scored presentations of the target stimulus and the non-target stimuli for data analysis. The order of trial presentation was randomized within each block.

A global measure of discrimination performance (Overall Discrimination Ratio, ODR) was computed by taking the mean response rate to all 15 of the non-target stimuli and dividing it by the mean response rate to the 1 target stimulus. We analyzed the data from all sessions from the inception of Discrimination Training until an ODR less than 0.30 was obtained across 3 sessions. Because at this point the discrimination between the target stimulus and all of the non-target stimuli was very strong, including data from later sessions added no useful information.

Model fit

The generalization data were fitted to a spatial model of multidimensional stimulus generalization, described in detail by Soto and Wasserman (2010a). This model takes as input measures of generalization to stimuli that differ from the target stimulus in only one property: either identity ($g_{id,0}$) or emotion ($g_{0,em}$). The model gives as output predictions of the generalization values that should be observed to a stimulus that differs from the target in both identity and emotion ($g_{id,em}$), according to the following equation:

$$g_{id,em} = \exp\left(-\sqrt[2]{(-\ln(g_{id,0}))^r + (-\ln(g_{0,em}))^r}\right), \quad (1)$$

where $g_{id,em}$ represents the generalization measure to a stimulus with identity “id” and emotion “em”, and where the number 0 in a subscript is reserved to label the identity or emotion of the target stimulus. Following the example provided in Figure 1, what this model does is to provide predictions of the generalization values that should be observed to the stimuli enclosed in the orange rectangle, which have a different identity and emotion than the target stimulus, using a combination of the generalization values observed to the stimuli enclosed in the green and blue rectangles, which share either the same identity or the same emotion with the target stimulus. Thus, suppose that we want to use the previous equation to predict responding to the image in the top left of the orange rectangle of Figure 1 (female 1, happy). We would have to take the observed generalization data to the images directly on top (male 1, happy) and to the left (female 1, angry) and replace them in the equation for the terms $g_{0,em}$ and $g_{id,0}$, respectively. Because this can be repeated for all of the stimuli that have a different identity and emotion from the target (within the orange rectangle in Figure 1), nine generalization values can be predicted for each pigeon and be compared to the empirical data.

The value of the parameter r in our model determines the specific metric that is used to compute the distances between stimuli in space from the distances along each dimension. The City-Block metric results from $r = 1$, whereas the Euclidean metric results from $r = 2$ (many other values are possible, each resulting in a different metric and a different combination rule).

To fit the behavioral data to this spatial model, the mean rate of responding to each stimulus across all of the scored training sessions was used as the behavioral measure of stimulus generalization. These data were transformed to a scale ranging from 0 to 1, by dividing them by the mean rate of responding to the target stimulus. In most behavioral studies, the lowest possible value of the generalization measures tends to be higher than 0, because such measures include data gathered during an initial learning phase when subjects respond indiscriminately. For this reason, it is common practice to rescale the generalization data to bring

its lowest values closer to 0 (Blough, 1988; Shepard, 1957; Soto & Wasserman, 2010a). Here, the obtained stimulus generalization data were separately rescaled for each pigeon via this linear transformation:

$$g_{ij} = (1 + \min(G_{ij}))G_{ij} - \min(G_{ij}). \quad (2)$$

We used the resulting stimulus generalization measures to find the best-fitting value of r in our model, using a least-squares estimation procedure. Lack-of-fit measures were also computed for the model, using values of r equal to 1 and 2, corresponding to the City-Block and Euclidean metrics, respectively.

Results and discussion

It took between 1 and 31 sessions ($M = 8.75$) for the pigeons to finish Baseline Training and between 5 and 35 sessions ($M = 13.5$) for them to finish Discrimination Training.

To determine whether each stimulus dimension separately controlled pigeons' discriminative behavior, we took the generalization scores for the stimuli that shared the same value along one dimension and we rank ordered them from the lowest to the highest. If the dimension along which these stimuli varied were controlling the pigeons' behavior, then the highest rankings should correspond to the stimulus that shared the same value along that relevant dimension with the target stimulus. For example, we could take all of the stimuli in the second column of Figure 1 that share the emotional expression of "happiness" and rank order them in terms of their generalization values. If the pigeon could perceive the similarity between the target stimulus (red rectangle) and the only stimulus in this column sharing the same identity with the target (the top stimulus, enclosed within the green rectangle), then this stimulus should show the highest ranking among those in the selected column. If this process is repeated for all of the columns that do not contain the target face, then there are three stimuli that share the same identity with the target stimulus, whose scores can be used to estimate the level of stimulus control by face identity. If the pigeons were unable to perceive the similarity of stimuli sharing the same identity, then the ranking of these stimuli would be equally distributed from 1 to 4 and the expected mean ranking would be 2.5. On the other hand, if the pigeons were able to perceive the similarity of stimuli sharing the same identity, then the ranking of these stimuli would be relatively high and the average ranking should be higher than 2.5. The same ranking analysis can be carried out for all of the rows of stimuli that do not contain the target face, to determine the control of generalized responding by emotional expression.

We found that the mean rank of the target identity was 3.59 ($SD = 0.12$) and the mean rank of the target emotion

was 3.22 ($SD = 0.54$). Both values were far above the value of 2.50 expected by chance; one-sample t -tests indicated that this result was statistically significant for both identity, $t(3) = 18.28$, $p < 0.001$, and emotion, $t(3) = 2.64$, $p < 0.05$. In sum, the results showed that the pigeons reliably generalized responding both across face identity and emotion, indicating that they did perceive the similarity among stimuli sharing the same value along each of these dimensions.

In order to compare the degree of stimulus generalization between the relevant dimensions, we averaged the generalization scores for the set of stimuli sharing the same identity as the target and for the set of stimuli sharing the same emotional expression as the target, which resulted in two marginal scores for each pigeon. The mean marginal score for identity ($M = 0.45$; $SD = 0.11$) was higher than the mean marginal score for emotional expression ($M = 0.34$; $SD = 0.09$), but this disparity did not reach statistical significance, $t(3) = 2.44$, $p > 0.05$. Thus, although variations in face identity tended to be more discriminable for pigeons than variations in emotional expression, this tendency was not statistically reliable.

Table 1 summarizes the main results of fitting the spatial model to the generalization data for each pigeon and for the pooled data of all of the pigeons. The first row shows the best-fitting value of the parameter r in the model. The three lower rows show the Root Mean Squared Error (RMSE) values that were obtained for the best-fitting value of r and for the predictions generated by the model using the City-Block ($r = 1$) and Euclidean ($r = 2$) metrics. For each pigeon and for their pooled data, the Euclidean metric better fit the data than the City-Block metric.

In sum, the obtained stimulus generalization results suggest that pigeons do perceive similarity across human faces sharing the same identity and emotional expression. Although the birds tended to perceive faces sharing the same identity as more similar to one another than faces sharing the same emotional expression, this tendency was not statistically significant. Finally, the generalization data from each of the birds better fitted a spatial model with a Euclidean metric than a spatial model with a City-Block metric. Because the results of prior studies have related the Euclidean metric to dimensional integrality (Garner, 1974; Shepard, 1991), this fact suggests that our pigeons did not process identity and emotional expression as separable dimensions.

However, the present experiment has one important shortcoming: namely, fitting the stimulus generalization data to a spatial model provides only a global measure of separability. If we accept the conclusion that the two dimensions combine integrally, then it is not clear whether emotional expression is not separable from identity, identity is not separable from emotional expression, or both. To provide a more precise answer to this question, we need to conduct a study that directly evaluates both forms

Bird	24W	27W	79W	34R	Group
Best-fitting r	390.85	1.55	1.81	1.92	2.11
RMSE					
Best fit	0.01014	0.00162	0.05080	0.00278	0.01861
City-Block	0.03408	0.00476	0.05783	0.01285	0.02738
Euclidean	0.01844	0.00237	0.05092	0.00281	0.01863

Table 1. Best-fitting values of parameter r and RMSE values, both for the individual pigeon data and for the pooled data in [Experiment 1](#).

of separability, just as the human studies reviewed earlier have done. For this reason, [Experiment 2](#) was conducted to more precisely evaluate the separability of each stimulus dimension with respect to the other, using a test proposed within the framework of GRT (Ashby & Townsend, 1986; Kadlec & Townsend, 1992).

Experiment 2

GRT (Ashby & Perrin, 1988; Ashby & Townsend, 1986) is a multidimensional generalization of Signal Detection Theory. GRT proposes that a given physical stimulus does not produce the same perceptual effect each time it is presented; rather, it produces a number of perceptual effects that follow a probability distribution. Each perceptual effect is usually described as a point in a continuous multidimensional space; a common assumption is that the distribution of such points is multivariate Gaussian. The observer sets up decision criteria, thereby dividing the perceptual space into different regions that are associated with particular responses. Each time a perceptual effect lands in one of these regions, it is classified using the response that is associated with it.

GRT has the advantage of offering a formal framework within which relatively ambiguous concepts, such as perceptual separability and independence, can be more rigorously defined. If A and B are components of a multidimensional stimulus, then Component A is perceptually separable from Component B if the distribution of perceptual effects of A is the same across all levels of B (Ashby & Townsend, 1986; Kadlec & Townsend, 1992). A consequence is that if Component A is perceptually separable from Component B, then the discriminability of stimuli varying across A should be equal across all levels of B, where discriminability is measured via the traditional value d' from Signal Detection Theory (Green & Swets, 1966). Although more criteria must be met to lend strong support to the conclusion that two dimensions *are separable* (see Kadlec & Townsend, 1992), finding that the discriminability of Component A varies for different levels of Component B provides strong support for the conclusion that Component A *is not separable* from Component B.

In [Experiment 2](#), we again trained pigeons in a discrimination task involving the same 16 photographs

of human faces that we used in the prior experiment (see [Figure 1](#)). We randomly assigned the pigeons to two groups. In Group Emotion, the birds had to report the emotional expression of the presented faces, regardless of variations in the identity of each particular face. In Group Identity, the birds had to report the identity of the presented faces, regardless of variations in the emotional expression of each particular face. For both groups, the proportion of correct and incorrect responses made by the pigeons provides sufficient information to compute discriminability measures for each value of the relevant dimension across all levels of the irrelevant dimension.

Given that the generalization data from [Experiment 1](#) best fit a spatial model using the Euclidean metric—suggesting perceptual integrality of identity and emotional expression—we suspected that at least one of the two dimensions might not pass the GRT test for separability from the other dimension. Specifically, if the results of the human studies reviewed earlier derive from the general principles of visual perception that are shared by primates and birds, then we might expect our pigeons to produce an asymmetrical pattern of interaction between identity and emotional expression: their processing of identity should be separable from emotional expression, but their processing of emotional expression should not be separable from identity.

To assess whether the behavior of the pigeons in this experiment could be predicted by the physical similarity among training images, we also trained an ideal observer model (Tjan, Braje, Legge, & Kersten, 1995) in the categorization tasks given to Group Emotion and Group Identity. The behavior of this ideal observer represents the optimal performance in these tasks for an agent relying on the pixel-by-pixel similarities among the images to solve them. The comparison between the pattern of results shown by these ideal observers and the pattern of results shown by the animal subjects would indicate what aspects of the experimental results can and cannot be explained by low-level stimulus similarities.

Methods

Subjects

Eight other feral pigeons (*C. livia*) were kept at 85% of their free-feeding weights by controlled daily feeding. The birds had previously participated in unrelated research.

Stimuli and apparatus

The stimuli and apparatus were the same as those described in [Experiment 1](#). The stimuli were presented in a central display screen with the same dimensions described in [Experiment 1](#). Additionally, four different 2.7×2.7 cm black-and-white icons were used as response keys on each trial. The icons represented abstract shapes that were completely unrelated to the visual categorization tasks presented to each pigeon (e.g., a white triangle inside a black square). A different response key was displayed close to each of the four corners of the display screen, with the center of these keys 3.5 cm from the corners of the display screen.

Procedure

The birds were randomly assigned to two experimental groups, Group Identity and Group Emotion, each trained on a different categorization task using a four-alternative forced-choice procedure. The same 16 photographs of faces were to be sorted according to different categories in these two tasks. Pigeons in Group Identity had to classify the photographs according to face identity to obtain food; each of the four keys was arranged to be correct for each of the four different people. Pigeons in Group Emotion had to classify the photographs according to emotional expression to obtain food; each of the four keys was arranged to be correct for each of the four different emotions. The assignment of categories to responses was partially counterbalanced within each group by means of a Latin square design.

A trial began with the presentation of a black cross in the center of the white display screen. Following one peck anywhere on the display, a training photograph appeared. The bird had to peck the stimulus a number of times (gradually increasing from 5 to 55 pecks, with the final number depending on individual performance on the task), a condition arranged to promote learning of the discrimination task. After this observing response requirement was met, the four response keys appeared and the pigeon had to peck one of them to advance through the trial.

If the pigeon pecked the correct response key, then food was delivered and an intertrial interval of 5 s ensued. If the pigeon pecked an incorrect response key, then the house light and the monitor screen darkened for a Time-Out period and a correction trial was given. The Time-Out period lasted between 5 and 35 s (depending on individual performance on the task). Correction trials continued to be given until the correct response was made. All responses were recorded, but only the first report response of each trial was scored in data analysis.

Each daily training session consisted of 10 blocks of 16 trials, for a total of 160 trials. Each block consisted of a single presentation of each of the 16 photographs in the training set.

The pigeons were trained until they reached a criterion of 74% correct responses overall with at least 55% correct responses on each response key in a single session. All of the data from the inception of training to the point at which each pigeon met criterion were pooled to compute the proportion of responses to each response key for each training stimulus. From these proportions, d' measures were calculated using Algorithm 2 in Smith (1982). This process yielded 16 d' measures for each pigeon—one for each training photograph.

Ideal observer

A different ideal observer simulation was run for each of the eight birds in the experiment, in order to match the pigeon's overall performance level and training period. This matching ensured that any disparities between the animals and the ideal observers would not be due to these factors and that enough data would be provided to run comparable analyses in the two observer groups. Both the task and the stimuli presented to the ideal observers were the same as those presented to the animal subjects. However, noise was added to the images shown to the ideal observers in order to bring their performance to the level shown by the pigeons. On each trial, one of the images was selected and Gaussian white noise with zero mean and variable standard deviation was added to it. The standard deviation of the noise was changed across trials using the QUEST algorithm (Watson & Pelli, 1983), in order to keep the overall performance of the ideal observer equal to that of the corresponding pigeon. The noise statistics were available to the ideal observer, which compared the resulting image against all of the images included in the task, giving as output the category label with the highest posterior probability computed according to Equation 4 in Tjan et al. (1995). This method corresponds to Bayesian *a posteriori* maximization assuming a generative model in which, on each trial, an object category generates a particular two-dimensional image that is perturbed by Gaussian noise.

Results and discussion

Animal subjects

It took a mean of 37.0 sessions for birds in Group Identity to reach criterion, with individual values ranging from 14 to 74 sessions. It took a mean of 88.8 sessions for birds in Group Emotion to reach criterion, with individual values ranging from 34 to 189 sessions. The disparity between the groups did not attain statistical significance, $t(6) = 1.40$, $p > 0.10$.

[Figure 2](#) shows the mean d' values computed from the training data, with the scores from Group Identity plotted in the top panel and the scores from Group Emotion

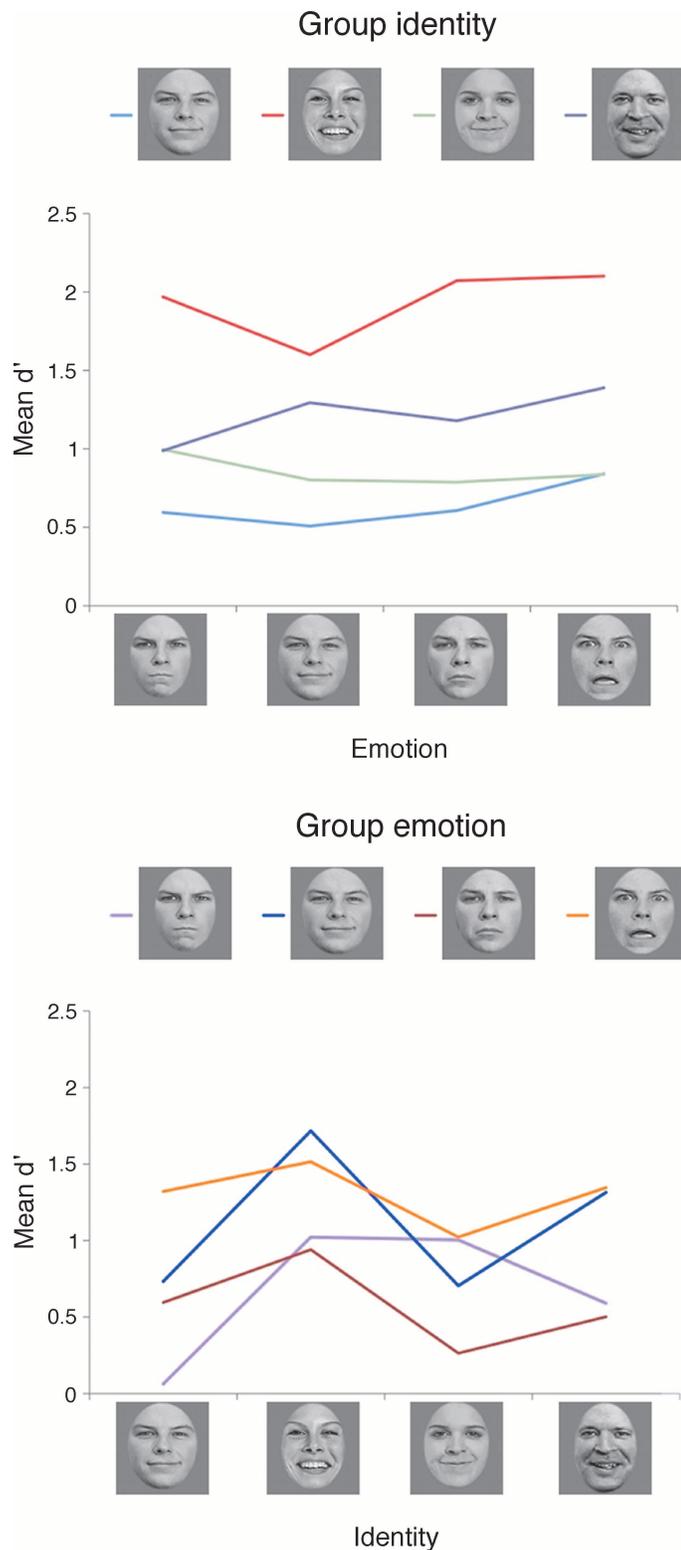


Figure 2. Mean d' values computed from the data of Group Identity and Group Emotion in Experiment 2.

plotted in the bottom panel. In each graph, the y -axis represents d' , the x -axis represents variations in the irrelevant dimension, and lines of different colors represent different values along the dimension that was relevant

for the discrimination task of each group. If the relevant dimension were separable from the irrelevant dimension, then each line in the graph should have been parallel to the x -axis; that is, in Group Identity, the discriminability of each identity should be more or less the same across variations in emotion, whereas in Group Emotion, the discriminability of each emotion should be more or less the same across variations in identity. Visual comparison of the top and bottom panels in Figure 2 clearly reveals that, although there were deviations from separability in the averaged data of both groups, those deviations were much more marked in Group Emotion than in Group Identity.

To determine whether any of the variations in mean d' observed for Group Identity (top panel of Figure 2) were statistically significant, a 4 (Identity) \times 4 (Emotion) ANOVA with d' as the dependent variable was performed. The analysis yielded a significant main effect of Identity, $F(3, 9) = 4.13$, $p < 0.05$, indicating that some identities were more difficult to discriminate for the pigeons than others. On the other hand, the main effect of Emotion was not significant, $F(3, 9) = 0.86$, $p > 0.1$, nor was the interaction between Identity and Emotion, $F(9, 27) = 1.56$, $p > 0.1$. Thus, variations in emotional expression did not affect the average discriminability of identity to an extent that would contradict the separability of identity from emotion.

To more precisely evaluate whether any of the differences between means for each individual identity were statistically significant, 4 one-way ANOVAs were conducted, each testing the effect of variations of emotional expression over mean d' for a particular identity. In none of these analyses was the effect of Emotion significant (male 1: $F(3, 9) = 1.45$, $p > 0.1$; female 1: $F(3, 9) = 2.10$, $p > 0.1$; female 2: $F(3, 9) = 0.32$, $p > 0.1$; male 2: $F(3, 9) = 1.25$, $p > 0.1$).

To determine whether any of the variations in mean d' observed for Group Emotion (bottom panel of Figure 2) were statistically significant, a 4 (Emotion) \times 4 (Identity) ANOVA with d' as the dependent measure was performed. The main effect of Emotion was not significant, $F(3, 9) = 3.33$, $p > 0.05$, although there was a tendency for different emotional expressions to be more difficult for pigeons to discriminate than others. On the other hand, the ANOVA indicated that there was a significant main effect of Identity, $F(3, 9) = 7.89$, $p < 0.01$, as well as a significant interaction between Emotion and Identity, $F(9, 27) = 5.97$, $p < 0.001$. Given that irrelevant variations of identity have a strong impact on the discriminability of different emotions, these results support the conclusion that emotional expression is not separable from identity in pigeons' perception of human faces.

To determine exactly which emotions were affected by variations in identity, 4 one-way ANOVAs were conducted, each testing the effect of variations in identity over mean d' for a particular emotion. There were significant effects of identity variations in the discrimination of

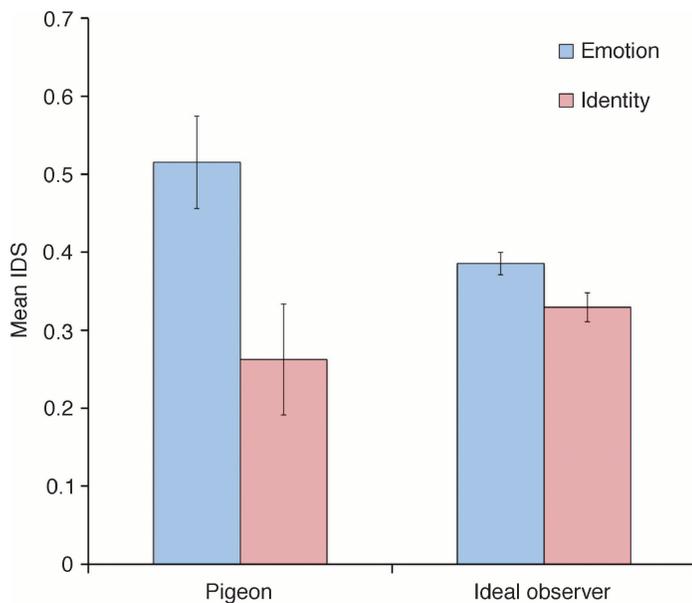


Figure 3. (Left) Mean Index of Deviations from Separability (IDS) for Group Identity and Group Emotion in Experiment 2 and (right) their corresponding ideal observers. The IDS measures deviations from complete separability in the individual pigeon data (see text for full explanation).

anger, $F(3, 9) = 5.35, p < 0.05$, happiness, $F(3, 9) = 10.19, p < 0.01$, sadness, $F(3, 9) = 7.44, p < 0.01$, and fear, $F(3, 9) = 5.39, p < 0.05$. In other words, variations in identity affected the discrimination of all four of the emotional expressions under study.

Therefore, the analysis of mean d' scores in both groups gives strong support to the conclusion that emotional expression is not perceived by pigeons as separable from identity and weak evidence (see Kadlec & Townsend, 1992) that identity is perceived by pigeons as separable from emotional expression.

The analysis of group data, however, only suggests that there were *systematic* deviations from separability in Group Emotion that were not found in Group Identity, that is, deviations that were similar across birds and were thus reflected in the mean d' . The relative lack of departure from dimensional separability that was observed in Group Identity (depicted in the top panel of Figure 2) may have, in part, been due to the averaging of d' scores across pigeons. Inspection of the individual data from pigeons in Group Identity disclosed what seemed to be substantial deviations from separability in all but one bird. Given that this observation was not reflected in the pooled data, it is possible that the variability in individual data is due to *random* error and not to replicable, *systematic* effects of one dimension over the other.

It thus seemed necessary to find a way to measure deviations from separability in the data of each individual pigeon and then to compare these deviations between the experimental groups. With this goal, an Index of Deviations from Separability (IDS) was computed for each

pigeon. The IDS was computed by first taking the absolute difference between all pairs of d' scores sharing the same value along the relevant dimension (emotion in Group Emotion; identity in Group Identity). The mean of these values was then divided by the mean d' to standardize the measure in order to take into account overall individual performance in the task. The final value was the individual IDS for each bird.

As shown in the left panel of Figure 3, the mean IDS for Group Emotion ($M = 0.52, SE = 0.06$) was much higher than the mean IDS for Group Identity ($M = 0.26, SE = 0.07$), a difference that proved to be statistically significant according to an independent samples t -test, $t(6) = 2.71, p < 0.05$. In addition, the mean IDS was significantly higher than 0 in both Group Emotion, $t(3) = 8.70, p < 0.01$, and in Group Identity, $t(3) = 3.68, p < 0.05$. Thus, although the individual data from the pigeons in both experimental groups showed statistically significant deviations from separability, these deviations were much more marked in Group Emotion than in Group Identity, thereby confirming an asymmetrical relation between identity and emotional expression in pigeons' perception of human faces.

Ideal observer

Figure 4 shows the mean d' values computed from the data of the ideal observers, with scores from the identity categorization task plotted in the top panel and scores from the emotion categorization task plotted in the bottom panel. As with the pigeon results shown in Figure 2, the y -axis represents d' , the x -axis represents variations in the irrelevant dimension, and lines of different colors represent different values along the dimension that was relevant for the discrimination task of each group. Visual comparison of the top and bottom panels in Figure 4 reveals that the ideal observers showed marked deviations from separability in *both* experimental tasks.

The data provided by the ideal observers in Group Identity (top panel of Figure 4) were analyzed using a 4 (Identity) \times 4 (Emotion) ANOVA with d' as the dependent variable. Just as reported for the pigeons before, this analysis resulted in a significant main effect of Identity, $F(3, 9) = 34.33, p < 0.001$, indicating that some identities were more difficult to discriminate than others. However, in this case, there was also a significant main effect of Emotion, $F(3, 9) = 10.28, p < 0.01$, and, more importantly, there was a significant interaction between Identity and Emotion, $F(9, 27) = 64.477, p < 0.001$.

To evaluate more precisely which individual identities showed a statistically significant difference between means, 4 one-way ANOVAs were conducted. These analyses indicated a significant effect of Emotion for all identities [male 1: $F(3, 9) = 4.11, p < 0.05$; female 1: $F(3, 9) = 62.736, p < 0.001$; female 2: $F(3, 9) = 17.50, p < 0.01$; male 2: $F(3, 9) = 74.83, p < 0.01$].

These results suggest that observers relying only on the physical similarity among the stimuli in the identity

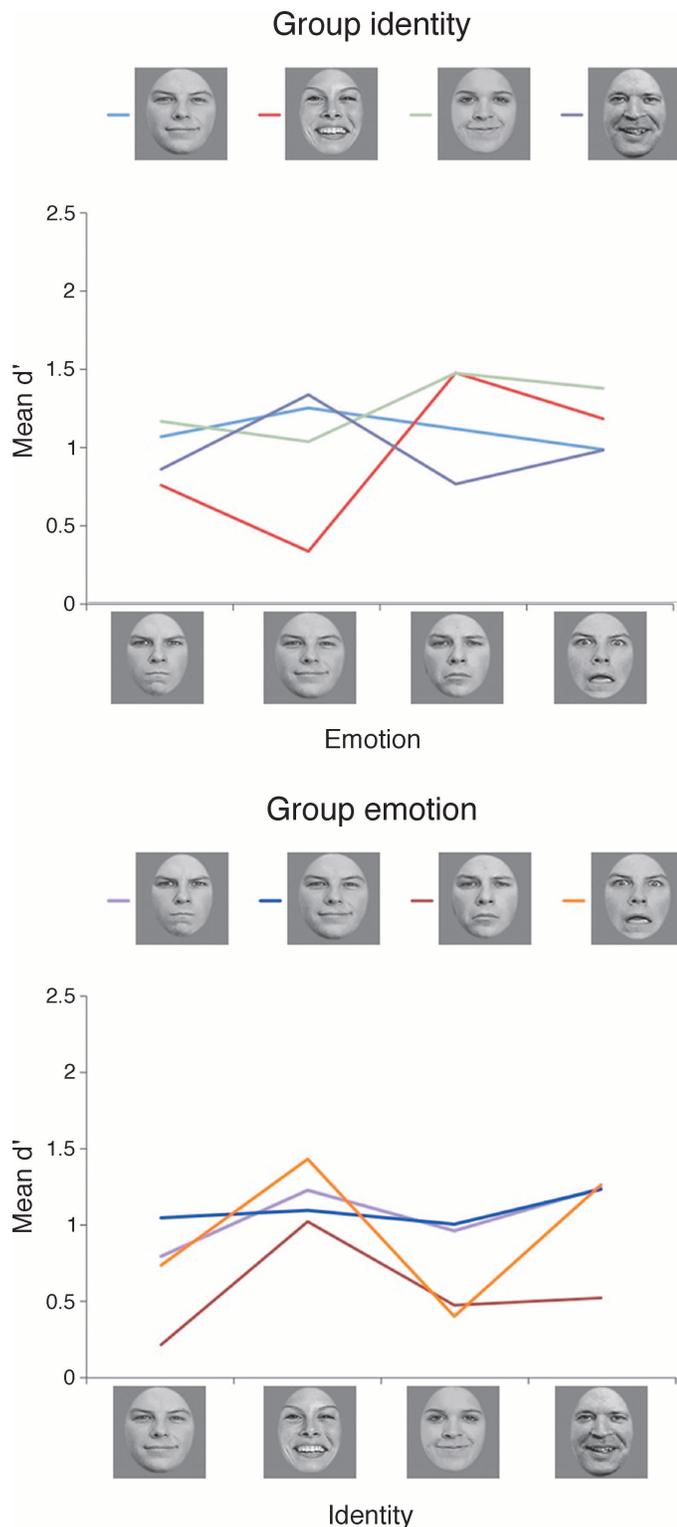


Figure 4. Mean d' values computed from the data of ideal observers trained on the same tasks as pigeons in Group Identity and Group Emotion in Experiment 2.

categorization task would be expected to show important deviations in the separability of identity from emotion. This result is in contrast to what was found for pigeons in Group Identity, who showed no statistically

significant deviations in the separability of identity from emotion.

The data provided by the ideal observers in Group Emotion (bottom panel of Figure 4) were analyzed using a 4 (Emotion) \times 4 (Identity) ANOVA with d' as the dependent variable. Unlike the results observed for pigeons, here the main effect of Emotion did reach significance, $F(3, 9) = 147.03, p < 0.001$, indicating that some emotions were more difficult to discriminate than others. As observed for the pigeons, there was also a significant main effect of Identity, $F(3, 9) = 59.59, p < 0.001$, and a significant interaction between Identity and Emotion, $F(9, 27) = 37.93, p < 0.001$.

To evaluate more precisely which individual emotions showed a statistically significant difference between means, 4 one-way ANOVAs were conducted. These analyses indicated a significant effect of Identity for all emotions [Anger: $F(3, 9) = 26.46, p < 0.001$; Happiness: $F(3, 9) = 11.77, p < 0.01$; Sadness: $F(3, 9) = 59.47, p < 0.001$; Fear: $F(3, 9) = 64.62, p < 0.001$].

Just as in the identity categorization task, observers relying only on the physical similarity among stimuli would be expected to show important deviations from separability in the emotion categorization task. However, in this case, the results observed for pigeons could be predicted on the basis of the mere physical similarity among the stimuli.

To evaluate more precisely whether the pattern of results observed for each pigeon could be predicted from the performance of the ideal observers, we correlated the d' values obtained from each pigeon with the corresponding d' values obtained from the ideal observer matched to that pigeon in overall performance. To compute average correlation coefficients for each group, we converted the individual correlation coefficients to Fisher's z values, which were then averaged, and the result converted back to a correlation coefficient, denoted by r_z (Corey, Dunlap, & Burke, 1998). The average correlation with ideal observer performance was negative for Group Identity ($r_z = -0.21$), whereas it was positive for Group Emotion ($r_z = 0.39$), a difference that was statistically significant, $z = 3.44, p < 0.01$. The results of this analysis thus confirmed the prior ANOVA results, suggesting that physical stimulus similarity could be driving the performance of pigeons in Group Emotion but not the performance of pigeons in Group Identity.

One possible problem with the previous comparison is that lower correlations between the pigeon and the ideal observer data could be due to higher random error in the data from pigeons in Group Identity than in the data from pigeons in Group Emotion. Although this possibility would not explain why the correlations in Group Identity were negative, we evaluated this possibility by computing the correlations between the data sets of different pigeons in the same group. If the data from a particular pigeon did not correlate positively with the ideal observer data because it was too "noisy," then it should also show low correlations with the data from other pigeons in the same group. The

data patterns for pigeons in Group Identity were quite similar to each other, with all correlations being positive ($r_z = 0.53$) and some showing values around 0.70. The same was true for pigeons in Group Emotion, with an average correlation ($r_z = 0.63$) that did not differ significantly from the mean correlation in Group Identity, $z = 0.67$, $p > 0.1$.

In sum, pigeons in Group Identity exhibited a consistent pattern of behavior that did not accord with what would be predicted if they were basing their responding on the physical similarity among the stimuli shown in this task. Pigeons in Group Emotion also exhibited a consistent pattern of behavior that did accord with what would be predicted if they were basing their responding on the physical similarity among the stimuli shown in this task.

IDS values were computed to measure deviations from separability in the data of each ideal observer, just as was done before for the pigeon data. As shown in the right panel of Figure 3, the mean IDS for Group Emotion ($M = 0.39$, $SE = 0.01$) was only slightly above the mean IDS for Group Identity ($M = 0.33$, $SE = 0.02$), a disparity that did not reach statistical significance according to an independent samples t -test, $t(6) = 2.35$, $p > 0.05$. Thus, when tested using the same procedure as was deployed for the pigeons, the ideal observer data did not support the conclusion that the pattern of physical similarity among stimuli would predict significant differences between groups in their degree of separability.

The outcome of this IDS analysis is also important for interpreting the previous ANOVA results. The data gathered from the ideal observers were much less variable than the pigeon data. Thus, a possible explanation for why we obtained significant deviations from separability in Group Identity for the ideal observer data and not for the pigeon data is simply that the former test had more statistical power than the latter. In the IDS test, however, the statistical power gained by low error variance was not enough to support the conclusion of a difference between groups. It appears that the most parsimonious interpretation of this pattern of results is that the effect size observed in the pigeon data (Cohen's $d = 0.70$) cannot be explained as arising from the mere physical similarity among the stimuli used in the present study (Cohen's $d = 0.31$).

In summary, the pattern of results in the present experiment is essentially the same as that found in prior human studies using different tasks. This concordance suggests that the asymmetrical interaction between identity and emotional expression is the result of the general principles of visual perception that are shared across evolutionarily distant species rather than the result of a specialized face processing mechanism that is unique to humans and other primates. Furthermore, this asymmetrical pattern cannot be explained as arising from the mere physical similarity among the stimuli in this experiment, as revealed by our analysis of ideal observers' performance in the relevant tasks. Pigeons in Group Emotion performed their task as would be expected if they were basing their

responding on mere physical similarity: they showed significant deviations from separability in the ANOVA and their performance correlated with the corresponding ideal observer performance. On the other hand, pigeons in Group Identity did not perform their task as would be expected if they were basing their responding on mere physical similarity: they did not show significant deviations from separability in the ANOVA and their performance correlated negatively with the corresponding ideal observer performance. In addition, the disparity between groups in deviations from separability was high and significant for the pigeons; this would not be expected from the performance of the ideal observers, which showed a small and non-significant difference between groups.

General discussion

The goal of the present study was to determine whether pigeons show the asymmetrical interaction between identity and emotional expression that has previously been found in humans' perception of faces. Evidence for an asymmetrical interaction in a non-primate species would suggest that this pattern of results can arise from principles of visual processing that are not specialized for face processing.

In Experiment 1, we found that pigeons perceive similarity across different face stimuli that varied in identity and emotional expression; the birds reliably generalized their responding on the basis of shared identity and emotional expression. We also found that there was no statistically reliable disparity in the discriminability of identity and emotion in our stimulus set. Finally, fitting the pigeons' data to a spatial model of multidimensional stimulus generalization indicated that emotional expression and identity combined according to a Euclidean metric, which has been associated with integral processing of dimensions, rather than to a City-Block metric, which has been associated with separable processing of dimensions.

Experiment 2 was a more direct test of the separability of each stimulus dimension. A group-level analysis of our pigeons' behavior yielded a statistically reliable effect of variations in identity over the discrimination of emotion, but the effect of variations in emotion over the discrimination of identity was not statistically reliable. An individual-level analysis of deviations from separability showed that there were reliable deviations both when pigeons discriminated emotional expression across variations in identity and when pigeons discriminated identity across variations in emotional expression; critically, the deviations in the case of the emotion discrimination were reliably higher than the deviations in the case of the identity discrimination. Overall, this pattern of results confirms an asymmetric interaction between identity and emotional expression in pigeons' perception of human faces.

The fact that the individual pigeons in Group Identity of [Experiment 2](#) showed deviations from complete separability merits further discussion. This result suggests that, although pigeons exhibited an asymmetrical relation between both face dimensions, we cannot conclude that identity is completely separable from emotion. Importantly, complete separability of identity from emotional expression has not been uniformly found in human experiments either.

For example, Ganel and Goshen-Gottstein (2004, Experiment 1) found that, if the discriminability of identity and emotion in the images is equilibrated, then it is possible to find interference from emotional expression to identity in a Garner interference task with adult humans. However, the size of the interference from identity to emotional expression was higher than from emotional expression to identity, confirming an asymmetrical pattern of interaction between the dimensions. A similar pattern of results was found by Atkinson et al. (2005, Experiment 2) in their study of the interaction between sex and emotion. Children from 9 to 11 also exhibited the same pattern of results when they were trained in a Garner interference task with stimuli that were not controlled for discriminability (Baudouin, Durand, & Gallay, 2008).

Furthermore, experiments reporting no Garner interference from emotion to identity have reached this conclusion by analyzing the *pooled* data of all subjects. Our study also failed to find an effect of variations in emotion over identity discrimination in the group-level analysis of the pooled data from [Experiment 2](#); only when the individual-level data were examined was there a reliable effect of variations in emotion over identity discrimination. These analyses therefore suggest that any effects of variations in emotion over identity discrimination were not systematic across pigeons, thereby canceling each other when the individual data were averaged. It is unclear whether the same averaging artifact occurred in human studies reporting complete separability of identity from emotion.

One possible problem with interpreting the results of the present study concerns the task used to test dimensional separability. As reviewed in the [Introduction](#) section, previous human studies have assessed separability using either the Garner interference task (Schweinberger et al., 1999; Schweinberger & Soukup, 1998) or the face adaptation paradigm (Ellamil et al., 2008; Fox & Barton, 2007; Fox et al., 2008). On the other hand, we used a categorization task in which the visual stimuli varied both along a dimension that was relevant for the task and a dimension that was irrelevant for the task. Although this method is similar to the orthogonal condition in the Garner interference task, the relevant comparison here entails the discriminability of stimuli in the relevant dimension across changes in the irrelevant dimension; the critical comparison in the Garner interference task concerns overall performance in a condition involving stimulus changes along the

irrelevant dimension versus a condition that does not include such changes.

Our reasons for using the GRT test for separability were twofold. First, it allowed us to test dimensional separability in a non-verbal species. The Garner interference task cannot be used in pigeons because it depends critically on verbal instructions. In addition, previous face perception studies with the Garner interference task used response times as the dependent measure, which can be difficult to measure with precision using pigeons as experimental subjects. Second, GRT provides a precise definition of perceptual separability within a theoretical framework that is widely accepted and applied in psychophysics. This definition stands in contrast with the more informal and ambiguous definition of perceptual separability/independence on which some of the prior human studies rest (for a more in-depth discussion, see Ashby & Townsend, 1986).

Admittedly, our results can be interpreted as comparable with those of prior human studies only if we assume that all of these tasks, as superficially different as they might be, measure a common psychological construct: namely, dimensional separability. This is a common assumption underlying studies of dimensional separability in people (Garner, 1974; Maddox, 1992; Shepard, 1991), including those investigations that have evaluated the perceptual interaction between identity and emotional expression in human face processing.

Work with hierarchical models of object recognition has found that simple principles of shape processing can explain both neural and behavioral aspects of facial processing (Jiang et al., 2006; Riesenhuber, Jarudi, Gilad, & Sinha, 2004). These models extend the processing hierarchy that was first proposed by Hubel and Wiesel (1968), which hypothesized that shape processing in the ventral visual system depends on a bottom-up hierarchy of feature detectors, tuned to increasingly complex features and becoming increasingly tolerant to image variations as we move up the processing hierarchy. The fact that such models can predict the data patterns that are reported in face recognition experiments serves as “a computational counterexample to theories which posit that human face discrimination necessarily relies on face-specific processes” (Jiang et al., 2006, p. 167).

Electrophysiological evidence shows that the early stages of visual shape processing that were proposed by Hubel and Wiesel (1968) are at work in the pigeon tectofugal visual pathway (Li, Xiao, & Wang, 2007). In addition, we have recently found computational evidence indicating that a hierarchical model of object recognition, extended by an error-driven learning stage (see Soto & Wasserman, 2010b), can also explain many aspects of pigeons’ object recognition (Soto & Wasserman, 2011). If the simple principles of shape processing that are embodied in hierarchical models are indeed shared among avian and mammalian species, then they are likely

candidates to explain the finding that both pigeons and humans show asymmetrical interference between emotion and identity in the perception of human faces.

It is important to emphasize that our results should not be taken to suggest that either pigeons process faces as people do or that people lack specialized processes for face perception. Current research in comparative cognition is based on the idea that any form of complex cognition must arise from a number of subprocesses, with some of them being specialized in a particular species and others being more widespread across species (de Waal & Ferrari, 2010; Shettleworth, 2000, 2010). It is very likely that both specialized and general processes are involved in human face recognition; it is thus an empirical matter to determine the contributions of each kind of process to this complex cognitive feat. The current study reports evidence suggesting that a particular feature of human face perception might be the result of processes that are shared across species. However, mechanisms that are specialized to process biologically relevant visual stimuli are likely to exist in both the pigeon and human visual systems.

The current study further emphasizes the importance of avoiding two misleading assumptions that many researchers make when thinking about the evolution of a specialized face processing system. First, evidence for specialized processes should not lead to the conclusion that such processes are somehow “encapsulated,” or free from the influence of more general processes, and that human faces engage only the specialized face perception system. It is obvious that human faces are not biologically relevant for pigeons, and hence, they ought to be processed in the same way as any other stimulus. The fact that an asymmetry in the processing of identity and expression is found in the pigeon suggests that some aspects of human face perception might not arise as the result of specialized processes only. Second, the evolution of a face recognition system did not solely involve the specialization of perceptual processes but also the specialization of the human face as an efficient transmitter of facial signals (Schyns, Petro, & Smith, 2009; Smith, Cottrell, Gosselin, & Schyns, 2005). The human face could have been specialized through evolution to transmit signals that would be easily decoded by existing visual processes. If those visual processes were present also in other species, then that could explain why pigeons tend to process identity separately from emotional expression as well as why monkeys show an inversion effect with human faces and not with monkey faces (Phelps & Roberts, 1994; Wright & Roberts, 1996).

We would like to close by commenting on the wider impact of comparative studies such as ours. Currently, it is a popular practice among researchers in perception and cognition to speculate about “specialized mechanisms” to explain their experimental data. In many cases, the underlying assumption is that a particular neural system has been specialized through evolution for the task under study. These speculations are often not supported by any empirical data.

We believe that results of comparative studies are a key ingredient to move from mere speculation about the evolutionary roots of human cognition and perception to a real empirical understanding of these processes. Comparative studies involving closely related species, such as non-human primates, have the potential to enrich our understanding of the origin and mechanisms of specialized systems (e.g., Sugita, 2008). Primate studies, however, leave open the question as to whether the system under study is indeed unique to a small group of species that are closely related to humans.

In order to rule out the possibility that more general, evolutionarily conserved processes underlie a particular behavior, what is required is the study of distantly related species (Bitterman, 2000). Only when there is evidence that such general principles cannot explain the behavior of interest it is possible to more confidently conclude that specialized mechanisms have arisen as the result of specific evolutionary pressures. We hope that future speculations concerning the evolution of human cognition and perception will be more fully informed by the comparative study of these processes.

Acknowledgments

We thank Frédéric Gosselin for providing the stimuli used in this study as well as Ramesh Bhatt and Lou Tassinari for participating in a project that sowed the seed for the present investigations.

This research was supported by National Institute of Mental Health Grant MH47313 and by National Eye Institute Grant EY019781.

Commercial relationships: none.

Corresponding author: Fabian A. Soto.

Email: fabian-soto@uiowa.edu.

Address: Department of Psychology, University of Iowa, Iowa City, IA 52242, USA.

References

- Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, *95*, 124–150.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Atkinson, A. P., Tipples, J., Burt, D. M., & Young, A. W. (2005). Asymmetric interference between sex and emotion in face perception. *Perception & Psychophysics*, *67*, 1199–1213.
- Avargues-Weber, A., Portelli, G., Bernard, J., Dyer, A., & Giurfa, M. (2009). Configural processing enables

- discrimination and categorization of face-like stimuli in honeybees. *Journal of Experimental Biology*, *213*, 593–601.
- Baudouin, J. Y., Durand, K., & Gallay, M. (2008). Selective attention to facial identity and emotion in children. *Visual Cognition*, *16*, 933–952.
- Bestelmeyer, P. E. G., Jones, B. C., DeBruine, L. M., Little, A. C., & Welling, L. L. M. (2010). Face aftereffects suggest interdependent processing of expression and sex and of expression and race. *Visual Cognition*, *18*, 255.
- Bitterman, M. E. (2000). Cognitive evolution: A psychological perspective. In C. M. Heyes & L. Huber (Eds.), *The evolution of cognition* (pp. 61–79). Cambridge, MA: MIT Press.
- Blough, D. S. (1988). Quantitative relations between visual search speed and target–distractor similarity. *Perception & Psychophysics*, *43*, 57–71.
- Bogale, B. A., Aoyama, M., & Sugita, S. (2011). Categorical learning between “male” and “female” photographic human faces in jungle crows (*Corvus Macrorhynchos*). *Behavioural Processes*, *86*, 109–118.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327.
- Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision Research*, *41*, 1179–1208.
- Cook, R. G. (Ed.). (2001). Avian visual cognition [Online]. Available: www.pigeon.psy.tufts.edu/avc/.
- Corey, D. M., Dunlap, W. P., & Burke, M. J. (1998). Averaging correlations: Expected values and bias in combined Pearson r s and Fisher’s z transformations. *Journal of General Psychology*, *125*, 245–261.
- Cottrell, G. W., Branson, K. M., & Calder, A. J. (2002). Do expression and identity need separate representations. In W. D. Gray & C. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 238–243). Mahwah, NJ: Lawrence Erlbaum Associates.
- de Waal, F. B., & Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences*, *14*, 201–207.
- Ellamil, M., Susskind, J. M., & Anderson, A. K. (2008). Examinations of identity invariance in facial expression adaptation. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 273.
- Farah, M. J. (1996). Is face recognition “special”? Evidence from neuropsychology. *Behavioural Brain Research*, *76*, 181–189.
- Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, *17*, 117.
- Fox, C. J., & Barton, J. J. S. (2007). What is adapted in face adaptation? The neural representations of expression in the human visual system. *Brain Research*, *1127*, 80–89.
- Fox, C. J., Oruç, I., & Barton, J. J. S. (2008). It doesn’t matter how you feel. The facial identity aftereffect is invariant to changes in facial expression. *Journal of Vision*, *8*(3):11, 1–13, <http://www.journalofvision.org/content/8/3/11>, doi:10.1167/8.3.11. [PubMed] [Article]
- Ganel, T., & Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: The parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 583–596.
- Ganel, T., Valyear, K. F., Goshen-Gottstein, Y., & Goodale, M. A. (2005). The involvement of the “fusiform face area” in processing facial expression. *Neuropsychologia*, *43*, 1645–1654.
- Garner, W. R. (1974). *The processing of information and structure*. New York: Lawrence Erlbaum Associates.
- Gibson, B. M., Wasserman, E. A., Frei, L., & Miller, K. (2004). Recent advances in operant conditioning technology: A versatile and affordable computerized touchscreen system. *Behavior Research Methods, Instruments, & Computers*, *36*, 355–362.
- Gibson, B. M., Wasserman, E. A., Gosselin, F., & Schyns, P. G. (2005). Applying bubbles to localize features that control pigeons’ visual discrimination behavior. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*, 376–382.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*, 203–218.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–232.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*, 215–243.
- Huber, L., & Lenz, R. (1993). A test of the linear feature model of polymorphous concept discrimination with pigeons. *Quarterly Journal of Experimental Psychology*, *46B*, 1–18.
- Huber, L., Troje, N. F., Loidolt, M., Aust, U., & Grass, D. (2000). Natural categorization through multiple feature

- learning in pigeons. *Quarterly Journal of Experimental Psychology*, 53B, 341–357.
- Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neuropsychological evidence. *Neuropsychologia*, 31, 173–181.
- Husband, S., & Shimizu, T. (2001). Evolution of the avian visual system. In R. G. Cook (Ed.), *Avian visual cognition*. Medford, MA: Tufts University. E-book available from <http://www.pigeon.psy.tufts.edu/avc/husband>.
- Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using fMRI and behavioral techniques. *Neuron*, 50, 159–172.
- Jitsumori, M., & Yoshihara, M. (1997). Categorical discrimination of human facial expression by pigeons: A test of the linear feature model. *Quarterly Journal of Experimental Psychology*, 50B, 253–268.
- Kadlec, H., & Townsend, J. T. (1992). Signal detection analysis of multidimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–231). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3, 759–763.
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361, 2109–2128.
- Kirkpatrick-Steger, K., & Wasserman, E. A. (1996). The what and the where of the pigeon's processing of complex visual stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 60–67.
- Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, 4, 89–94.
- Li, D., Xiao, Q., & Wang, S. (2007). Feedforward construction of the receptive field and orientation selectivity of visual neurons in the pigeon. *Cerebral Cortex*, 17, 885–893.
- Maddox, W. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Martens, U., Leuthold, H., & Schweinberger, S. R. (2010). Parallel processing in face perception. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 103–121.
- Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43, 25–53.
- Papini, M. R. (2002). Pattern and process in the evolution of learning. *Psychological Review*, 109, 186–201.
- Pascalis, O., & Kelly, D. J. (2009). The origins of face processing in humans: Phylogeny and ontogeny. *Perspectives on Psychological Science*, 4, 200–209.
- Phelps, M. T., & Roberts, W. A. (1994). Memory for pictures of upright and inverted primate faces in humans (*Homo sapiens*), squirrel monkeys (*Saimiri sciureus*), and pigeons (*Columba livia*). *Journal of Comparative Psychology*, 108, 114–125.
- Posamentier, M. T., & Abdi, H. (2003). Processing faces and facial expressions. *Neuropsychology Review*, 13, 113–143.
- Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society B: Biological Sciences*, 271, 448–450.
- Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Perception & Psychophysics*, 61, 1102–1115.
- Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1748–1765.
- Schyns, P. G., Petro, L. S., & Smith, M. L. (2009). Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: Behavioral and brain evidence. *PLoS ONE*, 4, e5625.
- Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22, 325–345.
- Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In J. Pomerantz & G. Lockhead (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 53–71). Washington, DC: American Psychological Association.
- Shettleworth, S. J. (2000). Modularity and the evolution of cognition. In C. M. Heyes & L. Huber (Eds.), *The evolution of cognition* (pp. 43–60). Cambridge, MA: MIT Press.
- Shettleworth, S. J. (2010). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, 14, 477–481.

- Shimizu, T. (2009). Why can birds be so smart? Background, significance, and implications of the revised view of the avian brain. *Comparative Cognition & Behavior Reviews*, *4*, 103–115.
- Shimizu, T., & Bowers, A. N. (1999). Visual circuits of the avian telencephalon: Evolutionary implications. *Behavioural Brain Research*, *98*, 183–191.
- Smith, J. E. (1982). Simple algorithms for M-alternative forced-choice calculations. *Perception & Psychophysics*, *31*, 95–96.
- Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science*, *16*, 184–189.
- Soto, F. A., & Wasserman, E. A. (2011). Visual object categorization in birds and primates: Integrating behavioral, neurobiological, and computational evidence within a “general process” framework. Manuscript submitted for publication.
- Soto, F. A., & Wasserman, E. A. (2010a). Error-driven learning in visual categorization and object recognition: A common elements model. *Psychological Review*, *117*, 349–381.
- Soto, F. A., & Wasserman, E. A. (2010b). Integrality/separability of stimulus dimensions and multidimensional generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*, 194–205.
- Sugita, Y. (2008). Face perception in monkeys reared with no exposure to faces. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 394–398.
- Tiberghien, G., Baudouin, J. Y., Guillame, F., & Montoute, T. (2003). Should the temporal cortex be chopped in two? *Cortex*, *39*, 121–128.
- Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, *35*, 3053–3069.
- Troje, N. F., Huber, L., Loidolt, M., Aust, U., & Fieder, M. (1999). Categorical learning in pigeons: The role of texture and shape in complex static stimuli. *Vision Research*, *39*, 353–366.
- Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience*, *31*, 411–437.
- Wasserman, E. A., & Zentall, T. R. (Eds.) (2006). *Comparative cognition: Experimental explorations of animal intelligence*. Madison, NY: Oxford University Press.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.
- Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*, 557–561.
- Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*, 5238–5241.
- Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, *92*, 1830–1839.
- Wright, A. A., & Roberts, W. A. (1996). Monkey and human face perception: Inversion effects for human faces but not for monkey faces or scenes. *Journal of Cognitive Neuroscience*, *8*, 278–290.