Received: 14 May 2021 | Revised: 12 October 2021 | Accepted: 13 October 2021

DOI: 10.1111/eth.13247

RESEARCH ARTICLE

ethology | WILEY

Testosterone, signal coloration, and signal color perception in male zebra finch contests

P. A. Green[1,2] | E. M. George[3,4] | K. A. Rosvall[3,4] | S. Johnsen[1] | S. Nowicki[1,5]

[1]Department of Biology, Duke University, Durham, North Carolina, USA

[2]Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn, UK

[3]Department of Biology, Indiana University Bloomington, Bloomington, Indiana, USA

[4]Center for the Integrative Study of Animal Behavior, Indiana University Bloomington, Bloomington, Indiana, USA

[5]Department of Neurobiology, Duke University, Durham, North Carolina, USA

Correspondence
P. A. Green, Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California, USA.
Email: pagreen@ucsb.edu

Present address
P. A. Green, Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California, USA

Funding information
This work was supported by funding provided by the Duke University Office of the Provost to SN. PAG is currently supported by a Human Frontiers Life Science Program Fellowship (LT000460/2019-L). EMG is supported by a predoctoral fellowship on Common Themes in Reproductive Diversity (NICHD-T32HD049336)

Abstract

Many animals use assessment signals to resolve contests over limited resources while minimizing the costs of those contests. The carotenoid-based orange to red bills of male zebra finches (*Taeniopygia guttata*) are thought to function as assessment signals in male–male contests, but behavioral analyses relating contest behaviors and outcomes to bill coloration have yielded mixed results. We examined the relationship between bill color and contests while incorporating measurements of color perception and testosterone (T) production, for an integrative view of aggressive signal behavior, production, and perception. We assayed the T production capabilities of 12 males in response to a gonadotropin-releasing hormone (GnRH) challenge. We then quantified the initiation, escalation, and outcome of over 400 contests in the group, and measured bill color using calibrated photography. Finally, because signal perception can influence signal function, we tested how males perceive variation in bill coloration, asking if males exhibit categorical perception of bill color, as has been shown recently in female zebra finches. The data suggest that males with greater T production capabilities than their rivals were more likely to initiate contests against those rivals, while males with redder bills than their rivals were more likely to win contests. Males exhibited categorical color perception, but individual variation in the effect of categorical perception on color discrimination abilities did not predict any aspects of contest behavior or outcomes. Our results are consistent with the hypotheses that T plays a role in zebra finch contests and that bill coloration functions as an aggressive signal. We suggest future approaches, based on animal contest theory, for how links among signals, perception, and assessment can be tested.

KEYWORDS
assessment signaling, bill color, categorical perception, dominance signaling, GnRH challenge

1 | INTRODUCTION

Contests allow animals to monopolize indivisible resources such as mates and food (Briffa & Hardy, 2013). Contests also come with costs, however, including energy expenditure and risk of injury or death (e.g., Briffa & Sneddon, 2007; Riechert, 1988). To resolve these important contests while minimizing costs, many animals use signals that communicate their quality or ability to win. Understanding the evolution of signaling systems requires testing how variation in the form of signals used in contests (also known as "aggressive signals") predicts contest behaviors and outcomes (Searcy & Beecher, 2009; Searcy & Nowicki, 2005).

Red-shouldered widowbirds (*Euplectes axillaris*), for example, use red-colored epaulets as signals of competitive ability. Males with larger and redder epaulets are more likely to hold territories (Pryke & Andersson, 2003a), and males with epaulets manipulated

to be larger and redder are more likely to win territorial contests in which males with smaller, less red epaulets often retreat without physical fights (Pryke & Andersson, 2003b). Orange to red signals such as widowbird epaulets are thought to be reliable (i.e., "honest") signals of quality because this coloration is carotenoid-based, and the biochemistry of carotenoid metabolism is thought to enforce a reliable correlation between color and individual quality (Blount & McGraw, 2008; Koch et al., 2018). Specifically, because vertebrates cannot produce carotenoids *de novo*, these compounds must be taken in from the diet. Once ingested, carotenoids are metabolized into forms used for display and forms used to support immune system function or parasite resistance, resulting in a potential tradeoff between the two (Olson & Owens, 1998; Weaver et al., 2018). Individuals that can better withstand this tradeoff should be able to produce higher intensity (e.g., larger or redder) carotenoid-based signals, thereby communicating their quality or ability (reviewed in Searcy & Nowicki, 2005).

Here, we test whether the carotenoid-based bill color of male zebra finches (*Taeniopygia guttata*) functions as an aggressive signal. Previous studies have linked the orange to red color continuum of male zebra finch bills with immune function (Blount et al., 2003; deKogel & Prijs, 1996; George et al., 2017; McGraw & Ardia, 2003; but see Birkhead et al., 1998), suggesting that bill color may reflect individual quality and, by extension, competitive ability. However, although one behavioral study has supported the role of carotenoid-based colors in aggressive signaling (Ardia et al., 2010), others have found no evidence of a signaling function (Bolund et al., 2007; Burley & Coopersmith, 1987; Etman et al., 2001).

These contrasting results may be due, in part, to methodological differences, such as how bill color is measured and how contest behavior is assayed. Studies have quantified bill color using various techniques, including subjectively matching colors to a color standard (Burley & Coopersmith, 1987), spectrophotometry (Ardia et al., 2010), and a combination of these approaches (Bolund et al., 2007). Researchers also have varied in how they stage contests and quantify contest behavior. Wild zebra finches live in groups of up to hundreds of individuals (Zann, 1996), yet prior studies finding no evidence of a competitive signaling function have measured the time single males spent near other single males in an adjacent cage (Burley & Coopersmith, 1987; Etman et al., 2001) or the amount of time males in staged dyads spent acting aggressively (Bolund et al., 2007). The only study that supported a competitive signaling function for bill coloration staged contests among triads of males and measured the number of times each male initiated a contest (Ardia et al., 2010). Experimental studies of assessment during contests often test how relative trait expression affects contest behaviors and outcomes; for example, testing how focal individuals' likelihoods of winning contests are correlated with relative (e.g., focal - rival) signal expression (Arnott & Elwood, 2009; Briffa et al., 2013). However, to our knowledge, no studies of aggressive signaling in zebra finches have used this approach.

Interacting physiological variables, such as testosterone (T) levels, may also influence if and how carotenoid-based bill coloration functions as an assessment signal in zebra finches. Treatment with exogenous T has been shown to enhance the size and/or color of carotenoid signals in several bird species (Ardia et al., 2010; Blas et al., 2005; Casagrande et al., 2011; Lindsay et al., 2011; Martínez-Padilla et al., 2014; Khalil et al., 2020, but see Stoehr & Hill, 2001). While these experimental manipulations demonstrate that T can influence the expression of carotenoid-based colors, fewer studies have addressed whether naturally occurring individual differences in T are likewise associated with signal variation and function (Ardia et al., 2010; Khalil et al., 2020). Because T levels can fluctuate greatly within individuals (Williams, 2008), injection with gonadotropin-releasing hormone (GnRH) has been used to maximally stimulate the hypothalamic-pituitary-gonadal axis, providing a repeatable and standardized metric of an individual's T production capability (Jawor et al., 2006). These GnRH challenge-induced T levels have been connected to signal variation and aggression (Bradley & Stoddart, 1997; Cain & Pryke, 2017; McGlothlin et al., 2008); however, little work to date has explicitly explored the integration between GnRH challenge-related T production abilities, carotenoid-based signals, and contests.

Finally, another key factor that could influence the function of carotenoid-based signals is the ability of receivers to perceive relevant signals. Receiver perception is known to impact signal function (Guilford & Dawkins, 1991; Miller & Bee, 2012; Rowe, 2013). In zebra finches, recent work has shown that females perceive the range of orange to red coloration that is typical of male bills in a categorical, not continuous, fashion (Caves et al., 2018). Specifically, females labeled the orange to red continuum as belonging to two categories (i.e., "orange" or "red") and were better able to discriminate among color pairs when the two colors came from opposite sides of the boundary between these two categories as compared to within-category color pairs, even for pairs that were equally distant in color space. Females varied in their color discrimination abilities, with some females better able to discriminate cross-boundary colors than others (Caves et al., 2018). Male zebra finches might also perceive carotenoid-based coloration categorically and, like females, males may differ in the degree to which this categorical perception influences color discrimination. If so, then variation among males in color discrimination abilities may result in males perceiving bill colors differently from one another, thus influencing the efficacy of bill color as an aggressive signal.

We tested whether variation in bill color, T production capabilities, and categorical perception predict the likelihood of a male initiating, escalating, or winning a contest. We first measured T levels from 12 male zebra finches in response to a GnRH challenge. We then observed contests within the group of 12 males across several days, quantifying the number of times each male initiated a contest with another male, initiated an escalated contest, and won a contest. We also quantified bill color from these males using calibrated photography. We assayed whether males exhibited categorical perception of the same orange to red colors as that shown for females by Caves et al. (2018), and extracted a metric of individual-level variation in the degree to which categorical perception affected color

discrimination. Finally, we correlated these metrics of bill color, T production, and categorical perception with the likelihood that a focal individual from each contest initiated, escalated, and won the contest.

We hypothesized that bill color functions as an aggressive signal and that aggression is mediated by T. Therefore, we established several predictions regarding the effects of beak color and T on contest behaviors and outcomes. We first predicted that males with redder bills, and with higher GnRH-induced T, would be more likely to initiate and win contests against rivals. Because signals are thought to decrease contest costs, we predicted that males with redder bills would be less likely to engage in escalated contests. Because T can enhance aggression, we predicted that males with higher GnRH-induced T would be more likely to engage in escalated contests. We also hypothesized that males that discriminated bill coloration in a more categorical fashion would resolve their contests with lower costs, in a manner that interacted with bill redness (the signal being perceived). Thus, we predicted that males with stronger categorical perception responses would more frequently initiate contests against males with less red bills than their own, and resolve these contests without escalation. As a result, males with stronger categorical perception responses were predicted to be more likely to win contests against rivals with less red bills. In comparison, males with weaker categorical perception responses would make more perceptual errors, initiating and escalating more contests with rivals having redder bills than their own, and losing more contests against rivals with less red bills.

2 | METHODS

2.1 | Subjects and animal maintenance

Subjects were 12 male zebra finches, seven obtained from Magnolia Bird Farm (Riverside, California, USA) and five obtained from Dr. Richard Mooney's breeding colony at Duke University. Birds were maintained in individual cages (46 × 23 × 23 cm, Prevue Pet, Chicago, IL), fed *ad libitum* with Kaytee Forti-Diet Pro Health Finch diet, and kept on a 15h:9h light schedule under fluorescent lighting (Ecolux with Starcoat SP 35/41, correlated color temperature 3500–4100K, General Electric), except when perception trials were being conducted (see below). All housing and experimental procedures were approved by the Duke University IACUC (protocol A004-17–10).

2.2 | Testosterone measurements

We followed methods in George and Rosvall (2018) to measure each bird's plasma testosterone level at baseline and in response to a GnRH challenge. At the beginning of the study, when animals were housed individually, we sampled blood for T between 10:46 AM and 12:42PM. We took a small (~70 μl) blood sample from the brachial vein using a heparinized microcapillary tube within ~3 min of removing each bird from its cage. We then injected a solution of 1.25 μg chicken GnRH-I (Bachem H-3106, Torrence, CA, USA) in 50 μl phosphate-buffered saline using a 50 μl Hamilton syringe into the left pectoral muscle and placed each bird individually in an opaque paper bag. Thirty minutes after the injection, we again sampled blood to assess each bird's ability to produce T when maximally stimulated, after which we returned the bird to its individual cage. Samples were centrifuged and plasma was stored at −20°C. Hormone extraction and ELISA were carried out by EMG at Indiana University. Briefly, for each sample we extracted hormones from 10μl of plasma using three rounds of liquid-liquid ether extraction, and reconstituted the extracted sample in 500 μl assay buffer. We used competitive-binding commercial ELISA kits (Enzo ADI# 900-176) to measure the T concentration of reconstituted plasma extract. T concentration was calculated by comparing a sample's absorbance to the absorbance of the assay's 9-point standard curve (Gen5 curve-fitting software, BioTek EPOCH plate reader, Winooski, VT, USA). We chose assay volumes based on predicted T concentrations from existing zebra finch literature (Prior et al., 2017); however, we found that 11 out of 12 baseline samples fell outside the linear portion of the standard curve (i.e., max binding ≥80%). We conservatively estimated a maximum possible concentration of 0.98 ng/ml plasma in those samples, based on the assay absorbance of wells with 80% max binding, adjusted for plasma and buffer volumes. We focused downstream analyses on GnRH-induced T levels, which fell in the most accurate and sensitive part of the standard curve. All samples were run on a single plate, with intra-plate variation of 4.7%. We used a two-tailed *t*-test to test whether GnRH-induced T levels ($\log_{10}$-transformed to meet assumptions of normality) were different than the adjusted baseline levels.

2.3 | Quantifying competitive interactions

Four days after blood sampling, we placed all 12 birds in one large cage (91 × 61 × 53 cm). This timing is long enough that any effects of the GnRH challenge will have subsided (Bergeon Burns et al., 2014; Rosvall et al., 2016). In this group cage, birds had access to a feeder from which at least four birds could feed at one time. Birds also had access to two single-bird watering tubes. One day after birds were introduced to the cage, we conducted two three-day sets of contest trials with a three-day gap in between, for a total of six days of contest data.

We assayed competitive interactions in a feeding context following Etman et al. (2001) and Ardia et al. (2010). Before each trial, we removed the large feeder and left birds without food for ~1 hr. We then placed a single-access feeder on the side of the cage and recorded all interactions occurring at or near the feeder for the next 30 min using a GoPro Hero 3+ camera (San Mateo, CA, USA). After 30 min, we removed the single feeder and replaced the original multi-bird feeder. We left birds in the large group cage between each daily trial and over the three-day gap between trial sets.

We coded all contests that occurred during trials at or near the single-access feeder, identifying birds by their unique color bands. We first noted which bird initiated each contest; that is, which bird made a clear, directed motion toward another bird, after which a contest began. We coded each contest as being one of three types: "displacements," "pecks," and "bill fencing." Displacements occurred when one bird approached another and sat next to it, usually making body contact, after which one or the other bird was displaced. A peck was recorded when one bird pecked the other with its bill. When bill fencing, both birds would assume an erect posture, knocking their bills together laterally, usually for several seconds (Morris, 1954; Zann, 1996). Morris (1954) stated that bill fencing occurs most frequently between birds that are closest in dominance; based on this reference and our own observations that bill fencing interactions took longer (and may, therefore, be costlier) than displacements and pecks, we considered bill fencing interactions to be escalated interactions. Finally, we recorded which bird won the contest and which lost: the loser was the bird that left the area of the interaction, while the winner was the bird that stayed. We only coded interactions in which we could unambiguously identify the initiator, the winner, and the loser. Over six trial days, we recorded a total of 403 contests. Each male participated in an average of 67 contests (SD = 46, min = 6, max = 147).

2.4 | Bill color measurements

Four days after the contest trials had ended, we followed the methods of Tedore and Johnsen (2012) and Johnsen (2016) to quantify bill color using calibrated photography. While gently holding a bird, we placed the ventral side of its lower bill on a black foam cube near a reflectance standard made of eight gray paint swatches of varying reflectance. We took three photographs of the upper bill (RAW file type) from approximately a 45-degree angle (to minimize specular, i.e., glossy, reflectance from the bill) using a Nikon D700 camera. The area was lit from above with a single halogen light (color temperature 2900K, model number H&PC-61361, Philips Lighting, Somerset, NJ, USA) shone through a layer of vellum paper placed 8 cm from the lamp to ensure diffuse illumination.

We took reflectance measurements from each square on the gray standard using an integrating sphere with a tungsten-halogen light source (ISP-REF; Ocean Optics, Largo, FL, USA). Then, with Adobe Photoshop CC 2018 (using the Adobe Standard image profile), we measured RGB values from an approximately 50 × 50 pixel region of each swatch of the gray standard. In Microsoft Excel 2016, we plotted the RGB values of the gray standards against their known reflectance values (from integrating sphere measurements) and fit exponential equations to these values. This process created a calibration equation for each photograph from which RGB values on the bill could be converted to reflectance averaged over each color channel. In Photoshop, we measured RGB values from 30 × 30 pixel regions on both the left and ride side of the bill for each of the three photos for each bird. We intentionally avoided measuring areas of high specular reflectance and areas that appeared to be in shade. We converted these bill RGB values to reflectance using the calibration equations generated from the gray standard. Note that we used a gray standard instead of another (e.g., red to orange) standard because gray colors produce a simple, monotonic curve for calibrating RGB values to reflectance. While this precluded direct comparison of the subject males' bill colors to the colors used in color perception tests (see below), it allowed for the most accurate quantification of male bill color. Moreover, the colors used in color perception tests have previously been identified as spanning the range of male bill color (Caves et al., 2018).

To create a single metric of bill color for each bird, we first took the average of the R and, separately, G channel reflectance from the left and right bill region, generating an average R and an average G reflectance for the whole bill (R_{avg} and G_{avg}, respectively). To approximate bill redness while accounting for differences in brightness, we used the formula $[(R_{avg} - G_{avg})/(R_{avg} + G_{avg})]$ (Tedore & Johnsen, 2012). This metric spans from −1 to 1 and describes the amount of light captured by the red channel of the camera, relative to the green channel (which approximates the brightness channel in avian vision; e.g., Spottiswoode & Stevens, 2010). Greater values represent redder bills. We refer to this metric as "bill redness."

2.5 | Color perception tests

Starting the day after taking bill color photographs, we tested how male zebra finches perceive variation in orange to red coloration. We tested perception of the same colors used in Caves et al. (2018), as these colors have previously been used to describe the color range of male zebra finch bills. We briefly describe the methods used to choose these colors below and refer the reader to Caves et al. (2018) for further details. We first identified 40 colors from the Munsell color system (Pantone Corporation, Grand Rapids, MI, USA) that had previously been matched to the continuum of male zebra finch bill color (Birkhead et al., 1998; Burley & Coopersmith, 1987; Collins et al., 1994). To map how these colors are predicted to differ from each other according to an avian viewer, we calculated the quantum catch (i.e., how stimulated each photoreceptor type is when viewing a given color) of each color for zebra finch short-, medium-, and long-wavelength photoreceptors (SW, MW, LW, respectively). Using the SW, MW, and LW quantum catch values and an ambient light spectrum (CIE Illuminant A), we calculated the chromatic distance between all 40 color stimuli as ΔS using the receptor noise-limited (RNL) model of color discrimination (Vorobyev & Osorio, 1998). We plotted each color in a two-dimensional chromaticity space based on hue and saturation/chroma, in which the Euclidean distance between points is equal to ΔS (Hempel de Ibarra et al., 2002). Because the chromaticity space is built for trichromatic vision, we did not include UV or double cone quantum catch. Modeling colors in trichromatic space is appropriate here because (1) our color stimuli had near-zero UV reflectance, due to both low reflectance and low illumination from our light source in this portion of the spectrum (Caves

et al., 2018); (2) the double cone is typically assumed to mediate perceived brightness, the effects of which we tested using a different metric (see below and Supporting Information); and 3) prior work has shown that calculating quantum catch for these colors using a tetrachromatic visual system has minimal impact on perceived discriminability (Caves et al., 2020). Using the chromaticity space, we chose eight colors that (1) were relatively equally distant from each other (and, therefore, predicted to be equally discriminable) and (2) spanned the full range of previously described bill colors in male zebra finches (Figure S1).

We created color stimuli using these eight colors. Stimuli were 1-inch disks comprised of two semi-circular halves of Munsell paper covered with a clear epoxy cover and with a small bumper on the bottom. Disks were either "bicolor," where each disk half was a different one of the eight colors selected above, or "solid," where each disk half was the same color. For example, a disk with one half as color 2 and one half as color 5 was "bicolor 2|5," whereas a disk with both halves color 2 was "solid 2".

The color stimuli were used in both training and testing trials. All training and testing protocols, including housing protocols, were identical to those in Caves et al. (2018). In training, we presented birds with two foraging trays, each with six wells (12 total wells). Six of the wells were covered by color disks: two bicolor 1|8, two solid 1, and two solid 8. Disk location on the wells was randomized using the sample function in R (R Core Team, 2020). By placing a food reward under the bicolor disks, we trained birds to flip over the bicolor disks, but not the solid disks. A bird passed a trial if it flipped both bicolor disks before any solid disk. We moved a bird from training to testing once it passed six of seven consecutive training trials.

In testing trials, we varied the colors on the bicolor and solid disks. For example, in 3|8 trials, we presented birds with two 3|8 bicolor disks, two solid 3 disks, and two solid 8 disks. If a bird flipped both bicolor disks before flipping any solid disks, it passed the trial. A bird failed a trial if it flipped only one solid disk. If a bird flipped no disks or only one bicolor disk, we removed that trial from the dataset. Each day, testing trials began with a 1|8 refresher trial and ended with a motivation check, in which we returned each bird's seed dish and ensured the bird ate within 2 min (showing that birds remained motivated throughout the trials). We allowed each bird 10 trials with each color combination and only analyzed data from birds that participated in five or more trials of a given color combination. To ensure that birds did not pass trials simply by smelling the food reward, we also gave birds five trials each of 1|1 and 8|8 comparisons, where all disks were either solid 1 or solid 8. We recorded all trials using a Logitech Webcam Pro 9000 (Newark, CA, USA).

Our protocol allowed us to test both hallmarks of categorical perception: labeling and discrimination (Green et al., 2020). To test labeling, we presented birds with bicolor disks showing color 1 against all other colors (1|8, 1|7, 1|6…1|2) and color 8 against all other colors (1|8, 2|8, 3|8…7|8). For each bird and each color pair, we calculated pass frequency: the number of trials each bird passed divided by the number of trials in which it participated. The expected pass frequency if birds flipped disks by chance was 1/15, or 0.07. Potential perceptual boundaries are indicated where there is a sharp and reciprocal discontinuity in pass frequency along the continuum, suggesting that stimuli on either side of this boundary are labeled differently. For female zebra finches, this discontinuity occurred between colors 5 and 6 (Caves et al., 2018).

While labeling identifies the location of hypothesized perceptual boundaries, to demonstrate categorical perception, birds should be more accurate at discriminating between two colors that come from opposite sides of the hypothesized boundary as compared with two equally different colors that come from the same size of the boundary (i.e., that are within a category). To test discrimination, we presented birds with color pairs in which the distance between colors was constant at either one color step apart ("one aparts," 1|2, 2|3…7|8), two steps apart ("two aparts," 1|3, 2|4…6|8) or three steps apart ("three aparts," 1|4, 2|5…5|8). In female zebra finches, discrimination was significantly better for colors crossing the 5|6 hypothesized boundary than for colors within each category (Caves et al., 2018).

2.6 | Statistical analyses

2.6.1 | Categorical perception

To test if male zebra finches exhibit the same categorical perception as females (i.e., with a category boundary between colors 5 and 6), we combined our dataset of 12 males with data on female zebra finch color perception from Caves et al. (2018). Following methods in Caves et al. (2018), we built linear mixed models (LMMs) from the combined labeling and discrimination data using the lme4 package (Bates et al., 2015) in R. These models represented the hypotheses that zebra finches discriminate colors based on a combination of chromatic distance and categorical perception, or on chromatic distance alone.

The first model (model 1) had pass frequency as the response variable and, as predictor variables, sex, chromatic distance (ΔS) and a binary term for whether or not compared colors crossed a hypothesized perceptual boundary between colors 5 and 6. We also included interaction terms between sex and chromatic distance and sex and the cross-boundary term. The model had a random intercept of bird ID and random slope of the cross-boundary term—these terms quantify inter-bird variation in pass frequency for trials that crossed the 5|6 color boundary, as compared to trials that did not.

We compared the fit of this model to a second model (model 2) that did not include terms for crossing the 5|6 boundary and, therefore, did not account for categorical perception. This model had a random intercept of bird ID and random slope of chromatic distance.

Categorical perception would be supported if model 1 was a better fit to the data than model 2, as measured by a decrease in AIC score (ΔAIC) of more than two, following Burnham et al. (2010). To test whether the sexes differed in categorical perception responses, we examined the significance of the ΔS:sex and the across:sex interaction in model 1 by using the drop1 function in R to compare the likelihood ratio of the full model to a model without each interaction

term. We then built a reduced model without interaction terms (none of which were significant) and tested for the significance of the sex fixed effect by using the drop1 function to compare the likelihood ratio of a model with the sex effect to a model without the sex effect.

We also built models to test whether brightness differences between colors better predicted pass frequency than a cross-boundary term, and whether males sourced from different suppliers (Magnolia Farms vs. Mooney Lab) showed differences in categorical perception. We discuss these models and their results in the Supporting Information.

To quantify individual variation in the effect of categorical perception on color discrimination in males, we re-ran model 1 using only data from males and used the coef function to extract the random slope estimates of the 5|6 cross-boundary effect from the model. We called this metric "cross-boundary perception score." Birds with higher cross-boundary perception scores showed greater increases in pass frequency for trials that crossed the 5|6 boundary as compared to trials that did not, suggesting a stronger effect of categorical perception on color discrimination abilities (see also Caves et al., 2018).

2.6.2 | Predicting contest initiation, escalation, and outcomes

We first used two-tailed *t*-tests to test whether males from different source populations (i.e., Magnolia Farm vs. Mooney Lab) differed in either bill redness or post-GnRH challenge T levels ($\log_{10}$-transformed). We then built generalized linear mixed models (GLMMs) using the lme4 package to test whether males from different source populations differed in their likelihood of initiating, escalating, or winning contests (focal bird ID was a random effect, as below). Finally, we used two-tailed Pearson product-moment correlation tests (cor.test function in R) to test for correlations between bill redness, post-GnRH challenge T levels, and cross-boundary perception scores. The distributions of all variables used in these tests matched assumptions of normality.

To test how bill redness, GnRH-induced T, and categorical perception affected contest initiation, escalation, and outcomes, we built GLMMs using the lme4 package. We first randomly selected a focal individual from each contest using the sample function. The other individual in the contest was the "rival." We built three GLMMs predicting the binary (1 if yes, 0 if no) response variables: likelihood that the focal male initiated the contest, likelihood that the focal male initiated a contest that involved bill fencing (as opposed to displacement or pecking), and likelihood that the focal male won the contest. Each GLMM had the predictor variables: relative (focal - rival) bill redness, relative (focal - rival) T level 30 min post-GnRH challenge, and focal cross-boundary perception score. We also included the two-way interaction between relative bill redness and focal cross-boundary perception score. This interaction term reflects the prediction that the effect of relative bill redness on contest behaviors or outcomes could be modified according to the focal individual's cross-boundary perception. Finally, we included random effects of each individual in the interaction (focal and rival) and the source population (Magnolia Farms or Mooney Lab) of each individual. All predictor variables were centered with unit variance using the scale function, and all models had a binomial error with a logit link function. After fitting the full model, we then used the summary function to extract the estimates for each predictor.

The estimates for each predictor will vary according to which bird is randomly assigned focal and rival status. For example, the estimate of the effect of focal cross-boundary perception on the likelihood of focal contest initiation will vary depending on which males are assigned as focal males. To ensure that our statistical results were robust to this random focal/rival assignment, we conducted a resampling analysis (Adams & Anthony, 1996; Crowley, 1992). We repeated the random focal/rival assignment for each contest 10,000 times. In each iteration, we repeated the model fitting approach described above. This process generated, for each predictor in each model, 10,000 observed estimate values.

To statistically test our predictions (see Introduction), we asked whether the distribution of observed values for the model estimates was greater or less than that of a null dataset. In each null dataset, for each male we randomly sampled (without replacement) the value of each predictor from the values of all males. We then followed the same modeling approach as above and generated 10,000 null estimate values. We calculated two-tailed *p*-values for each effect in our models as two times the proportion of estimates from the null data that were greater (or less) than the median estimate value of the observed dataset (i.e., $p = 0.05$ occurs when 250/10,000 null estimates are greater than the median observed estimate). In the Results, we report *p*-values alongside the distribution of estimate values and measures of effect size, to facilitate interpretation of the strength of model effects. Tables S1–S6 show the output of all models. Figures S4-S9 show the distributions of null data estimate values plotted with the median estimate from the observed data, as well as the number of null data estimates that were greater than and less than the median value of the observed data estimates.

3 | RESULTS

3.1 | Testosterone levels

Baseline T levels (after adjusting for samples with max binding ≥80%) averaged 0.99 ng/ml (standard deviation = 0.02 ng/ml), while post-GnRH challenge T levels averaged 3.69 ng/ml (standard deviation =2.50 ng/ml). The GnRH challenge increased T levels (*t*-test post- vs. adjusted baseline $\log_{10}$(T); $t = 5.19$, df =11.02, $p < 0.001$).

3.2 | Categorical color perception

A plot of the labeling data showed that pass frequency increased as chromatic distance increased, and that the greatest change in pass

frequency occurred between colors 5 and 6, irrespective of whether comparisons were against color 1 or color 8 (Figure 1a). Plots of discrimination data for two-apart (Figure 1b) and three-apart (Figure S2) trials revealed higher pass frequencies for color pairs that came from opposite sides of the hypothesized 5|6 boundary as compared to within-category comparisons. Discrimination plots for one-apart (Figure S3) data showed invariant pass frequencies across the hypothesized 5|6 boundary. Thus, patterns of labeling and discrimination for males closely matched those found for females by Caves et al. (2018).

The statistical analysis of combined labeling and discrimination data confirmed that males exhibit the same categorical perception previously found for females. Model 1, which included a term for crossing the 5|6 boundary, was a substantially better fit than model 2, which did not include a boundary term (model 2 - model 1 ΔAIC = 206.24). Model 1 did not include a significant interaction between either chromatic distance and sex (β = 0.00, SE = 0.00, $\chi^2_1 = 0.35$, $p = 0.55$) or crossing the 5|6 boundary and sex (β = 0.05, SE = 0.05, $\chi^2_1 = 1.06$, $p = 0.30$). The fixed effect of sex was not significant in the reduced model (β = 0.00, SE =0.02, $\chi^2_1 = 0.09$, $p = 0.77$).

The models predicting pass frequency from brightness differences or including the effect of male source did not fit the data better than model 1 (Supporting Information).

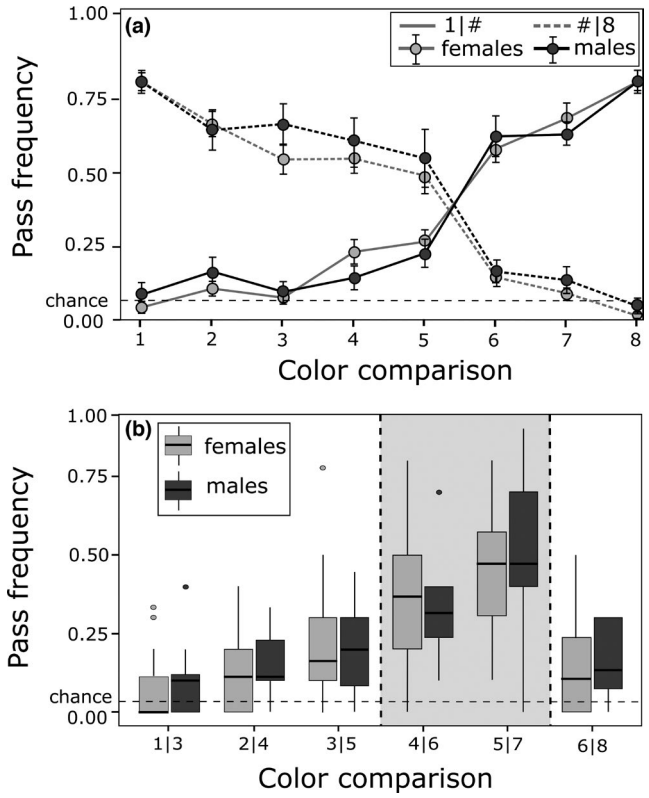

FIGURE 1 Categorical perception of carotenoid-based color in male (dark gray) and female (light gray) zebra finches. Female data are from Caves et al. (2018). Results from (a) labeling and (b) two-apart discrimination data. In both (a) and (b), the dashed horizontal line on *y*-axis shows chance pass frequency of 0.07. In (a), solid lines show comparisons against color 1 while dashed lines show comparisons against color 8; points show mean values and vertical bars show standard error. In (b), comparisons that cross the hypothesized 5|6 boundary are within the light gray-shaded region in the vertical dashed lines; boxes show median (dark line), 25th and 75th percentiles (boxes), 25th/75th percentiles × interquartile range (whiskers), and outliers (circles)

3.3 | Contest initiation, escalation, and outcomes

Males from the two sources did not differ in either bill redness (*t*-test $t = 0.72$, df = 9.39, $p = 0.49$) or GnRH-induced T levels ($\log_{10}$-transformed data; *t*-test $t = 1.26$, df =6.9, $p = 0.25$). As described above, males from the two sources also did not differ in color discrimination (see also Supporting Information). Males from the two sources also did not differ in the likelihood of initiating (Mooney Lab β = −0.48, $\chi^2_1 = 0.63$, $p = 0.44$), escalating (β = 0.53, $\chi^2_1 = 3.36$, $p = 0.07$), or winning (β = −0.55, $\chi^2_1 = 0.83$, $p = 0.36$) contests. Bill redness, post-GnRH challenge T, and cross-boundary coefficient were not significantly correlated with each other (redness and cross-boundary coefficient correlation = 0.05, $t_{10} = 0.17$, $p = 0.87$; redness and T correlation = −0.20, $t_{10} = -0.64$, $p = 0.54$; T and cross-boundary coefficient correlation = −0.11, $t_{10} = -0.36$, $p = 0.73$).

The likelihood that a focal male initiated a contest was positively correlated with relative post-GnRH challenge T (null data median β [5%, 95%] = 0.0081 [−0.7968, 0.8443]; observed data median β [5%, 95%] = 0.9101 [0.7852, 1.0225]; Figure 2a). Only 364 of 10,000 null data estimates were greater than the median estimate of the observed data; however, this result did not have a significant two-tailed *p*-value ($p = 0.0728$). From the observed estimate distributions, the median probability of initiating a contest for a male with a GnRH-induced T value of one standard deviation below the mean was 32.6% (5%, 95% probabilities = 27.1%, 40.1%). The median probability of initiating a contest for a male with a GnRH-induced T value of one standard deviation above the mean was 71.8% (5%, 95% probabilities = 56.1%, 84.4%).

No predictor variables were strongly correlated with the likelihood that a focal male initiated a bill fencing contest.

The likelihood that a focal male won a contest was positively correlated with relative bill redness (null data median β [5%, 95%] = 0.0013 [−0.7299, 0.7227]; observed data median β [5%, 95%] = 0.7504 [0.6641, 0.8777]; Figure 2b). Of 10,000 null data estimates, 437 were greater than the median observed data estimate; however, this result did not have a significant two-tailed *p*-value ($p = 0.0874$). The median probability of winning a contest for a male with a beak redness value one standard deviation below the mean was 37.6% (5%, 95% probabilities = 31.4%, 43.4%). The median probability of winning for a male with a beak redness value one standard deviation above the mean was 65.0% (5%, 95% probabilities = 52.5%, 78.2%). Finally, the likelihood a focal male won a contest was correlated with the likelihood the focal male initiated that contest (Pearson correlation = 0.75).

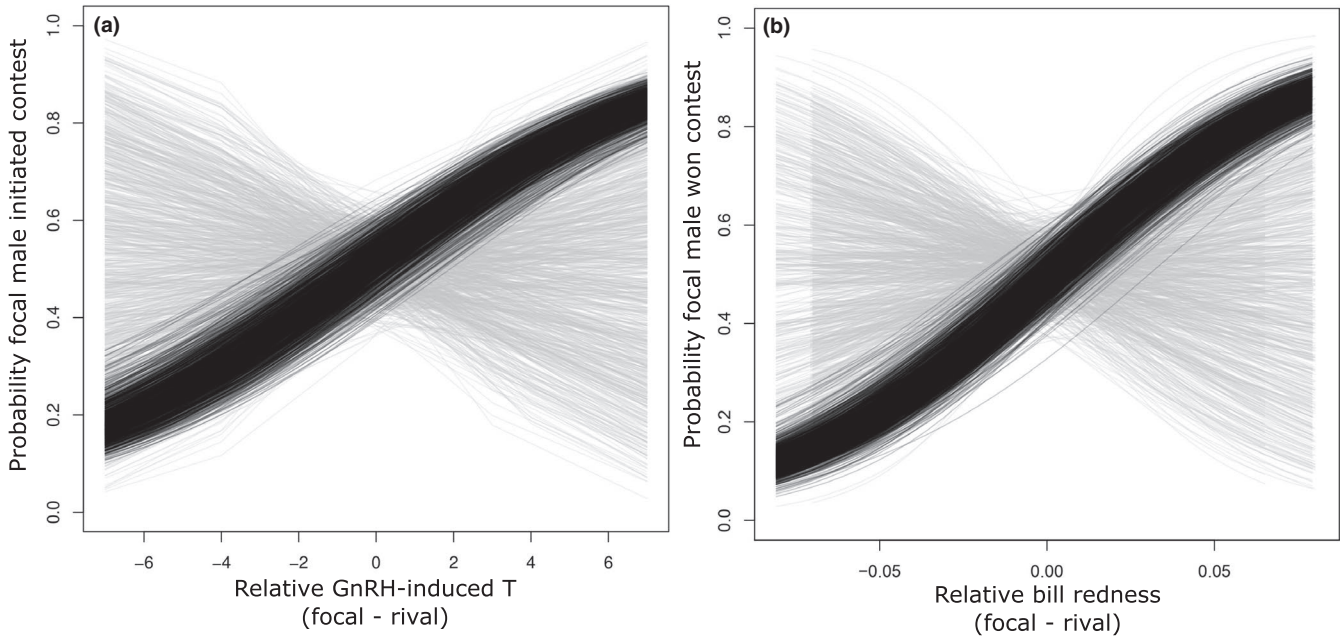

FIGURE 2 (a) Relative GnRH-induced T (ng/ml) plotted against the probability a focal male initiated a contest, and (b) relative bill redness plotted against the probability a focal male won a contest. Lines represent 1,000 randomly sampled estimates (from 10,000 GLMMs, see Methods) for both observed (black lines) and null (gray lines) data

4 | DISCUSSION

This study integrates carotenoid-based bill coloration, hormonal traits, and receiver perception to understand how they impact the initiation, escalation, and outcomes of competitive interactions. The data suggest that male zebra finches having more pronounced T production capabilities than their rivals were more likely to initiate contests (Figure 2a), while males with redder bills than their rivals were more likely to win contests (Figure 2b). Males exhibited the same categorical perception of carotenoid-based coloration as previously shown for female zebra finches (Figure 1). However, individual variation in categorical perception did not predict the initiation, escalation, or outcomes of contests. Our results are consistent with the hypotheses that T plays an important role in aggressive interactions and that carotenoid-based bill color functions as an aggressive signal in contests within zebra finch groups. Below, we discuss the implications of these results and suggest future studies to better understand competitive interactions, signaling, and signal perception.

4.1 | Contests, bill redness, and testosterone

Our findings that males with relatively higher GnRH-induced T levels than their rivals were more likely to initiate contests, and that males with relatively redder bills were more likely to win contests, are consistent with results across a variety of vertebrate taxa showing that testosterone production capabilities can influence contest behavior and that carotenoid-based coloration can function as a signal (Hamilton et al., 2013, reviewed in Blount & McGraw, 2008). The evidence is not universal; however. McGraw and Hill (2000), for example, found no competitive signaling function of carotenoid colors in house finches. Earlier studies with zebra finches suggested no competitive signaling function for carotenoid bill coloration (Bolund et al., 2007; Burley & Coopersmith, 1987; Etman et al., 2001). However, Ardia et al. (2010) did find that males with redder bills and higher T levels were more likely to initiate contests, similar to our findings.

Our results suggest that different techniques for measuring bill color may impact the findings of studies of aggressive signaling in zebra finches. Prior studies have used matching to Munsell chips (Burley & Coopersmith, 1987), spectrophotometric measurements (Ardia et al., 2010), or both techniques (Bolund et al., 2007) to measure bill color. We used a calibrated photography technique that allows for averaging across large regions of the bill, circumventing the issues that (1) spectophotometric measurements over large regions can result in high variation (Johnsen, 2016) and that (2) human observers perceive color in different ways than birds do, making subjective matching to standards such as Munsell chips problematic (Bennett et al., 1994; Caves et al., 2019). Our approach also has limitations that are important to consider. For example, we had to avoid regions with high levels of specular reflectance in the mid-bill region. Calibrated photography also requires summarizing color into three bands (R, G, B) that cannot be directly connected to avian vision, as can be done with a full-spectrum approach like spectrophotometry. However, we note that some spectrophotometry studies use approaches such as principal components analysis to summarize spectrophotometry data (e.g., Bolund et al., 2007). Comparisons of bill color measurements using multiple techniques,

especially comparing spectrophotometry, multispectral imaging, and calibrated photography, could reveal the best approaches for future studies of carotenoid-based coloration.

Our findings also suggest that the ways in which aggressive interactions are studied might influence tests of aggressive signaling in zebra finches. A first consideration is the number of individuals allowed to interact. Results supporting an aggressive signaling function of bill color have staged interactions among at least three males (Ardia et al., 2010, the present study), while those using dyads or choice experiments with single individuals found negative results (Bolund et al., 2007; Burley & Coopersmith, 1987; Etman et al., 2001). Because zebra finches typically live in large groups in the wild (Zann, 1996), tests using groups may be most relevant for understanding aggressive signaling in this species. It is also important to consider which aspects of aggressive behavior are measured. Studies supporting a signaling function for bill color have tested how bill color predicts whether individuals initiated (Ardia et al., 2010) and/or won contests (present study). Studies showing no relationship have measured the time males spent near other males (Burley & Coopersmith, 1987; Etman et al., 2001) or the time males spent competing (Bolund et al., 2007). In a contest, a focal individual should use a signal to gather information about a competitor's ability, using that information to make decisions such as whether to initiate, escalate, or leave a contest (Arnott & Elwood, 2009; Briffa & Hardy, 2013). Therefore, metrics of these decisions might be most applicable to understanding the signal function of bill color in zebra finches. A final consideration is that contests must involve (at least) two competitors. While this seems obvious, ours is the first study in zebra finches, to our knowledge, to use relative (focal - rival) metrics of bill color as predictors of contest behaviors and outcomes. Some theoretical models suggest that competitors assess relative traits, like fighting ability or signals of ability, when making competitive decisions (Arnott & Elwood, 2009). Therefore, tests using relative metrics of signal expression (e.g., bill redness) may be best suited for understanding how signals function in contests.

Adding to earlier findings that testosterone influences aggression in zebra finches, we found a positive relationship between relative GnRH-induced T levels and the likelihood that focal individuals initiated contests. The relationship between aggression and T in zebra finches has been known at least since Arnold (1975), who showed that castrating males reduces aggression, while later injection of testosterone propionate (TP, a non-aromatizable androgen) recovers aggression. More recently, Ardia et al. (2010) found that experimentally injecting low-ranking males with TP led to increases in the number of contests those males initiated. These remove-and-replace hormonal experiments clearly show that T is associated with aggression-related traits in this system. However, further studies using individual-level measures of endogenous T, such as were used by Ardia et al. (2010), are needed to understand whether variation in T secretion regulates aggressive interactions and/or signals. Baseline T levels are notoriously variable, leading to the suggestion that individual differences simply reflect noise (summarized in Williams, 2008). Measuring T following a GnRH challenge, as we did here, bypasses these concerns by providing an individually repeatable and standardized metric of what each individual is capable of when their hypothalamic-pituitary-gonadal axis is maximally stimulated. Post-GnRH challenge T levels have been correlated with ornament size (McGlothlin et al., 2008) and aggression (Bradley & Stoddart, 1997) but, to our knowledge, ours is the first study to relate this measure of T to contests and a putative signal of competitive ability in the same cohort of study animals.

For many years, researchers have sought to understand how T integrates multiple components of the phenotype, including coloration and behavior (e.g., Ketterson et al., 2009; Lipshutz et al., 2019). In addition to our results showing a correlation between T production and contest initiation, we found that post-GnRH challenge T levels were negatively correlated, albeit weakly, with both our bill redness metric and the strength of perception across the 5|6 boundary (Results). In contrast to these results, Ardia et al. (2010) found a strong, *positive* correlation between zebra finch bill redness (red chroma) and T. Our results may be different because we measured T after a GnRH challenge, which is thought to reflect maximal T responses (Jawor et al., 2006), and because we measured T from zebra finches that were caged individually. Ardia et al. (2010), by comparison, measured T without a GnRH challenge from birds that had been in cages of three individuals for at least two days before measurement. Further research into if and how zebra finch carotenoid-based color is regulated by T, and by variation in T receptors, may clarify the role T plays in reliable signaling during contests.

4.2 | Perception and aggression

Male zebra finches exhibited categorical perception of a carotenoid-based color continuum mirroring that of male bills (Figure 1). The location and strength of the category boundary were virtually identical to those shown previously for female zebra finches (Caves et al., 2018), and did not differ between males raised under different lighting environments (Supporting Information).

We found that the cross-boundary perception score, our metric of inter-individual variation in categorical perception, did not predict whether focal individuals initiated, escalated, or won contests. The cross-boundary perception score reflects each male's change in color discrimination abilities for colors that crossed the 5|6 category boundary as compared to colors within categories; that is, it does not test whether broader aspects of color discrimination influence contests. To address this, we replaced the cross-boundary score predictors in our models (including interaction terms) with the average pass rate of each individual across all trials (excluding 1|1 and 8|8 comparisons). Pass rates reflect overall color discrimination abilities: males with higher pass rates showed generally better color discrimination abilities, irrespective of whether compared colors crossed the 5|6 category boundary. Pass rate was strongly correlated with cross-boundary coefficient scores (Pearson correlation = 0.96), and models using pass rate showed similar results to models using cross-boundary coefficients (Tables S4–S6, Figures S7–S9). In the model

predicting contest outcomes, the interaction term between relative bill redness and focal pass rate had a negative estimate ($p = 0.0640$; Figure S9, Table S6). Though in the opposite direction of our predictions, this suggests a potential effect of perception on contest outcomes.

One reason we may have found little effect of color perception on contest behavior is that we tested perception in a different context than staged contests. In perception tests, we tested how well males could discriminate between two colors with a shared border in the context of accessing a food reward. In contrast, during contests males are likely viewing a single color stimulus (a rival male's bill) and, presumably, using that single color stimulus to determine their own contest behavior. Variation in behavioral and ecological context, as well as in perceptual task, can influence findings related to color vision (Menzel & Backhaus, 1989; Olsson et al., 2018). Therefore, future studies may develop perceptual assays that are more closely linked to the behavior of interest. For instance, future work may test how males respond to videos of rival males in which bill color is digitally manipulated.

Our resampling approach allowed us to account for the random sampling of focal and rival males, an important consideration in our analyses of variables related to perception. For example, we predicted that focal males' perception of bill redness would impact their behavior. Therefore, we used raw (not relative) metrics of perception for focal males. Focal and rival status was randomly assigned and, by chance, some of these random assignments resulted in significant effects of perception on contest behavior and outcomes. However, using our resampling procedure, we found that most of these random assignments resulted in models with non-significant effects of perception on contest behaviors and outcomes (Tables S1–S6, Figures S4–S9). When studies of contests use focal and rival assignment, and especially when predictor variables are not relative (e.g., focal - rival values), we suggest the use of resampling analyses to ensure that effects are robust to random focal and rival assignments. Finally, to avoid pseudoreplication in these analyses, it is crucial to randomize data such that individuals maintain their (randomized) phenotype across all contests within a given iteration, but phenotypes may change between iterations.

To summarize, our data suggest that (1) T production capabilities influence contest behavior, (2) bill color functions in aggressive signaling, (3) male zebra finches show the same categorical perception previously found in females, but (4) individual variation in perception, as we tested it, poorly predicts contest behaviors and outcomes. These findings contribute to our understanding of the signaling function of carotenoid-based coloration, the role of hormonal states in aggression, and how perception impacts animal signaling systems.

ACKNOWLEDGMENTS

We thank Drs. E.M. Caves, M.N. Zipple, and C.R. Campbell; J. Granger and C. Howell; and two anonymous reviewers for feedback on the manuscript and analysis. We also thank D. Bharath, A. Crabill, R. Ginjupalli, L. Kinsella, M. Knauer, E. Olszewski, A. Rowe, and N. Tarn for valuable help in carrying out the experiments, and S. Peters for management of the project.

CONFLICT OF INTERESTS

There are no conflicts of interest to declare.

AUTHORS' CONTRIBUTIONS

Conceptualization: PAG, KAR, SJ, SN; methodology: PAG, EMG, KAR, SJ, SN; data collection: PAG, EMG, KAR, SN; data analysis: PAG, EMG; writing—original draft: PAG; writing—review and editing: PAG, EMG, KAR, SJ, SN; funding acquisition: PAG, EMG, SN; resources: KAR, SJ, SN; supervision: KAR, SJ, SN.

CODE AVAILABILITY

All code are available on FigShare: https://doi.org/10.6084/m9.figshare.14595987.

DATA AVAILABILITY STATEMENT

All data are available on FigShare: https://doi.org/10.6084/m9.figshare.14595987.

ORCID

P. A. Green https://orcid.org/0000-0002-2434-8795
E. M. George https://orcid.org/0000-0002-1223-5857
K. A. Rosvall https://orcid.org/0000-0003-3766-9624
S. Johnsen https://orcid.org/0000-0002-3943-8320
S. Nowicki https://orcid.org/0000-0002-6564-905X

REFERENCES

Adams, D. C., & Anthony, C. D. (1996). Using randomization techniques to analyse behavioural data. *Animal Behaviour*, *51*, 733–738.

Ardia, D. R., Broughton, D. R., & Gleicher, M. J. (2010). Short-term exposure to testosterone propionate leads to rapid bill color and dominance changes in zebra finches. *Hormones and Behavior*, *58*(3), 526–532.

Arnold, A. P. (1975). The effects of castration and androgen replacement on song, courtship, and aggression in zebra finches (*Poephila guttata*). *Journal of Experimental Zoology*, *191*, 309–326.

Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests. *Animal Behaviour*, *77*(5), 991–1004.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bennett, A. T., Cuthill, I. C., & Norris, K. J. (1994). Sexual selection and the mismeasure of color. *American Naturalist*, *144*(5), 848–860.

Bergeon Burns, C. M., Rosvall, K. A., Hahn, T. P., Demas, G. E., & Ketterson, E. D. (2014). Examining sources of variation in HPG axis function among individuals and populations of the dark-eyed junco. *Hormones and Behavior*, *65*(2), 179–187.

Birkhead, T. R., Fletcher, F., & Pellatt, E. J. (1998). Sexual selection in the zebra finch *Taeniopygia guttata*: Condition, sex traits and immune capacity. *Behavioral Ecology and Sociobiology*, *44*(3), 179–191.

Blas, J., Pérez-Rodríguez, L., Bortolotti, G. R., Viñuela, G., & Marchant, T. A. (2005). Testosterone increases bioavailability of carotenoids: Insights into the honesty of sexual signaling. *Proceedings of the National Academy of Sciences of the United State of America*, *103*(49), 18633–18637.

Blount, J. D., & McGraw, K. J. (2008). Signal functions of carotenoid colouration. In G. Britton, & S. Liaaen-Jensen (Eds.), *Carotenoids* (Vol. *4*, 213–236). Birkhauser.

Blount, J. D., Metcalfe, N. B., Birkhead, T. R., & Surai, P. F. (2003). Carotenoid modulation of immune function and attractiveness in zebra finches. *Science*, *300*, 125–127.

Bolund, E., Schielzeth, H., & Forstmeier, W. (2007). Intrasexual competition in zebra finches, the role of beak colour and body size. *Animal Behaviour*, *74*(4), 715–724.

Bradley, A. J., & Stoddart, D. M. (1997). Plasma androgen, social position, and response to GnRH in the marsupial sugar glider *Petaurus breviceps* (Marsupialia: Petauridae). *Journal of the Zoological Society of London*, *241*, 579–587.

Briffa, M., & Hardy, I. C. W. (2013). Introduction to animal contests. In I. C. W. Hardy, & M. Briffa (Eds.), *Animal contests* (pp. 1–4). Cambridge University Press.

Briffa, M., Hardy, I. C. W., Gammell, M. P., Jennings, D. J., Clarke, D. D., & Goubault, M. (2013). Analysis of animal contest data. In I. C. W. Hardy, & M. Briffa (Eds.), *Animal contests* (pp. 47–85). Cambridge University Press.

Briffa, M., & Sneddon, L. U. (2007). Physiological constraints on contest behaviour. *Functional Ecology*, *21*, 627–637.

Burley, N. T., & Coopersmith, C. B. (1987). Bill color preferences of zebra finches. *Ethology*, *76*, 133–151.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2010). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35.

Cain, K. E., & Pryke, S. R. (2017). Testosterone production ability predicts breeding success and tracks breeding stage in male finches. *Journal of Evolutionary Biology*, *30*(2), 430–436.

Casagrande, S., Dijkstra, C., Tagliavini, J., Goerlich, V. C., & Groothuis, T. G. (2011). Differential effects of testosterone, dihydrotestosterone and estradiol on carotenoid deposition in an avian sexually selected signal. *Journal of Comparative Physiology A*, *197*(1), 1–13.

Caves, E. M., Green, P. A., Zipple, M. N., Peters, S., Johnsen, S., & Nowicki, S. (2018). Categorical perception of colour signals in a songbird. *Nature*, *560*(7718), 365–367.

Caves, E. M., Nowicki, S., & Johnsen, S. (2019). Von Uexküll revisited: Addressing human biases in the study of animal perception. *Integrative and Comparative Biology*, *59*(6), 1451–1462.

Caves, E. M., Schweikert, L. E., Green, P. A., Zipple, M. N., Taboada, C., Peters, S., Nowicki, S., & Johnsen, S. (2020). Variation in carotenoid-containing retinal oil droplets correlates with variation in perception of carotenoid coloration. *Behavioral Ecology and Sociobiology*, *74*(7). https://doi.org/10.1007/s00265-020-02874-5.

Collins, S. A., Hubbard, C., & Houtman, A. M. (1994). Mate choice in the zebra finch: The effect of male beak colour and male song. *Behavioral Ecology and Sociobiology*, *35*(1), 21–25.

Crowley, P. H. (1992). Resampling methods for computation-intensive data analysis in ecology and evolution. *Annual Review of Ecology and Systematics*, *23*, 405–447.

deKogel, C. H., & Prijs, H. J. (1996). Effects of brood size manipulations on sexual attractiveness of offspring in the zebra finch. *Animal Behaviour*, *51*(3), 699–708. https://doi.org/10.1006/anbe.1996.0073

Etman, E., Lelieveld, H. M., & ten Cate, C. (2001). Male bill colour and competition in zebra finches. *Behavioural Processes*, *55*, 119–124.

George, D. B., Schneider, B. C., McGraw, K. J., & Ardia, D. R. (2017). Carotenoids buffer the acute phase response on fever, sickness behavior and rapid bill color change in zebra finches. *Journal of Experimental Biology*, *220*(Pt 16), 2957–2964.

George, E. M., & Rosvall, K. A. (2018). Testosterone production and social environment vary with breeding stage in a competitive female songbird. *Hormones and Behavior*, *103*, 28–35.

Green, P. A., Brandley, N. C., & Nowicki, S. (2020). Categorical perception in animal communication and decision-making. *Behavioral Ecology*, *31*(4), 859–867.

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, *42*(1), 1–14.

Hamilton, D. G., Whiting, M. J., & Pryke, S. R. (2013). Fiery frills: Carotenoid-based coloration predicts contest success in frillneck lizards. *Behavioral Ecology*, *24*(5), 1138–1149.

Hempel de Ibarra, N., Giurfa, M., & Vorobyev, M. (2002). Discrimination of coloured patterns by honeybees through chromatic and achromatic cues. *Journal of Comparative Physiology, A*, *188*(7), 503–512.

Jawor, J. M., McGlothlin, J. W., Casto, J. M., Greives, T. J., Snajdr, E. A., Bentley, G. E., & Ketterson, E. D. (2006). Seasonal and individual variation in response to GnRH challenge in male dark-eyed juncos (*Junco hyemalis*). *General and Comparative Endocrinology*, *149*(2), 182–189.

Johnsen, S. (2016). How to measure color using spectrometers and calibrated photographs. *Journal of Experimental Biology*, *219*(Pt 6), 772–778.

Ketterson, E. D., Atwell, J. W., & McGlothlin, J. W. (2009). Phenotypic integration and independence: Hormones, performance, and response to environmental change. *Integrative and Comparative Biology*, *49*(4), 365–379.

Khalil, S., Welklin, J. F., McGraw, K. J., Boersma, J., Schwabl, H., Webster, M. S., & Karubian, J. (2020). Testosterone regulates CYP2J19-linked carotenoid signal expression in male red-backed fairywrens (*Malurus melanocephalus*). *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *287*(1935), 20201687.

Koch, R. E., Hill, G. E., & Sandercock, B. (2018). Do carotenoid-based ornaments entail resource trade-offs? An evaluation of theory and data. *Functional Ecology*, *32*(8), 1908–1920.

Lindsay, W. R., Webster, M. S., & Schwabl, H. (2011). Sexually selected male plumage color is testosterone dependent in a tropical passerine bird, the red-backed fairy-wren (*Malurus melanocephalus*). *PLoS One*, *6*(10), e26067.

Lipshutz, S. E., George, E. M., Bentz, A. B., & Rosvall, K. A. (2019). Evaluating testosterone as a phenotypic integrator: From tissues to individuals to species. *Molecular and Cellular Endocrinology*, *496*, 110531.

Martínez-Padilla, J., Perez-Rodriguez, L., Mougeot, F., Ludwig, S., & Redpath, S. M. (2014). Intra-sexual competition alters the relationship between testosterone and ornament expression in a wild territorial bird. *Hormones and Behavior*, *65*(5), 435–444.

McGlothlin, J. W., Jawor, J. M., Greives, T. J., Casto, J. M., Phillips, J. L., & Ketterson, E. D. (2008). Hormones and honest signals: Males with larger ornaments elevate testosterone more when challenged. *Journal of Evolutionary Biology*, *21*(1), 39–48.

McGraw, K. J., & Ardia, D. R. (2003). Carotenoids, immunocompetence, and the information content of sexual colors: An experimental test. *American Naturalist*, *162*(6), 704–712.

McGraw, K. J., & Hill, G. E. (2000). Carotenoid-based ornamentation and status signaling in the house finch. *Behavioral Ecology*, *11*(5), 520–527.

Menzel, R., & Backhaus, W. (1989). Color vision in honey bees: Phenomena and physiological mechanisms. In D. G. Stavenga, & R. C. Hardie (Eds.), *Facets of Vision* (pp. 281–297). Springer.

Miller, C. T., & Bee, M. A. (2012). Receiver psychology turns 20: Is it time for a broader approach? *Animal Behaviour*, *83*(2), 331–343. https://doi.org/10.1016/j.anbehav.2011.11.025.

Morris, D. (1954). The reproductive behaviour of the zebra finch (*Poephila guttata*), with special reference to pseudofemale behaviour and displacement activities. *Behaviour*, *6*(4), 271–322.

Olson, V. A., & Owens, I. P. F. (1998). Costly sexual signals: Are carotenoids rare, risky or required? *Trends in Ecology & Evolution*, *13*(12), 510–514. https://doi.org/10.1016/S0169-5347(98)01484-0

Olsson, P., Lind, O., Kelber, A., & Simmons, L. (2018). Chromatic and achromatic vision: Parameter choice and limitations for reliable model predictions. *Behavioral Ecology*, *29*(2), 273–282.

Prior, N. H., Yap, K. N., Mainwaring, M. C., Adomat, H. H., Crino, O. L., Ma, C., Guns, E. S., Griffith, S. C., Buchanan, K. L., & Soma, K. K. (2017). Sex steroid profiles in zebra finches: Effects of reproductive state and domestication. *General and Comparative Endocrinology*, *244*, 108–117.

Pryke, S. R., & Andersson, S. (2003a). Carotenoid-based epaulettes reveal male competitive ability: Experiments with resident and floater red-shouldered widowbirds. *Animal Behaviour*, *66*(2), 217–224. https://doi.org/10.1006/anbe.2003.2193

Pryke, S. R., & Andersson, S. (2003b). Carotenoid-based status signalling in red-shouldered widowbirds (*Euplectes axillaris*): Epaulet size and redness affect captive and territorial competition. *Behavioral Ecology and Sociobiology*, *53*(6), 393–401. https://doi.org/10.1007/s00265-003-0587-2

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Riechert, S. E. (1988). The energetic costs of fighting. *American Zoologist*, *28*, 877–884.

Rosvall, K. A., Bergeon Burns, C. M., Jayaratna, S. P., & Ketterson, E. D. (2016). Divergence along the gonadal steroidogenic pathway: Implications for hormone-mediated phenotypic evolution. *Hormones and Behavior*, *84*, 1–8.

Rowe, C. (2013). Receiver psychology: A receiver's perspective. *Animal Behaviour*, *85*(3), 517–523.

Searcy, W. A., & Beecher, M. D. (2009). Song as an aggressive signal in songbirds. *Animal Behaviour, 78*(6), 1281–1292.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication: Reliability and deception in signaling systems*. Princeton University Press.

Spottiswoode, C. N., & Stevens, M. (2010). Visual modeling shows that avian host parents use multiple visual cues in rejecting parasitic eggs. *Proceedings of the National Academy of Sciences*, *USA*, *107*(19), 8672–8676.

Stoehr, A. M., & Hill, G. E. (2001). The effects of elevated testosterone on plumage hue in male House Finches. *Journal of Avian Biology*, *32*, 153–158.

Tedore, C., & Johnsen, S. (2012). Weaponry, color, and contest success in the jumping spider *Lyssomanes viridis*. *Behavioural Processes*, *89*(3), 203–211.

Vorobyev, M., & Osorio, D. (1998). Receptor noise as a determinant of colour thresholds. *Proceedings of the Royal Society of London*, *Series B: Biological Sciences*, *265*, 351–358.

Weaver, R. J., Santos, E. S. A., Tucker, A. M., Wilson, A. E., & Hill, G. E. (2018). Carotenoid metabolism strengthens the link between feather coloration and individual quality. *Nature Communications*, *9*(1), 73.

Williams, T. D. (2008). Individual variation in endocrine systems: Moving beyond the 'tyranny of the Golden Mean'. *Philosophical Transactions of the Royal Society of London*, *Series B: Biological Sciences*, *363*(1497), 1687–1698.

Zann, R. A. (1996). *The Zebra Finch: A synthesis of field and laboratory studies*. Oxford University Press.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Green, P. A., George, E. M., Rosvall, K. A., Johnsen, S., & Nowicki, S. (2021). Testosterone, signal coloration, and signal color perception in male zebra finch contests. *Ethology*, 00, 1–12. https://doi.org/10.1111/eth.13247