ORIGINAL ARTICLE

Variation in carotenoid-containing retinal oil droplets correlates with variation in perception of carotenoid coloration

Eleanor M. Caves [1,2] · Lorian E. Schweikert [2,3] · Patrick A. Green [1,2] · Matthew N. Zipple [2] · Carlos Taboada [2] · Susan Peters [2] · Stephen Nowicki [2,4] · Sönke Johnsen [2]

Abstract
In the context of mate choice, males may vary continuously in their expression of assessment signals, typically reflecting information about variation in mate quality. Similarly, females may exhibit variation in mate preference, which could be due to differences in how individual females perceive signals. The extent to which perception varies across individuals, however, and whether differences in sensory physiology underlie perceptual differences is poorly understood. Carotenoid pigments create the orange-red coloration of many assessment signals, and they also play a role in color discrimination in many vertebrates via their presence in retinal oil droplets. Here, we link variation in oil droplet carotenoid concentration with the ability of female zebra finches (*Taeniopygia guttata*) to discriminate an orange-red color continuum that parallels variation in male beak color, a mate assessment signal. We have shown previously that zebra finch females perceive this color range categorically, meaning they label color stimuli from this continuum as belonging to two categories and exhibit better discrimination between colors from different categories as compared with equally different colors from within a category. We quantified behavioral color discrimination and R-type (red) cone oil droplet spectral absorption, a proxy for carotenoid concentration. Oil droplet absorption was strongly predictive of variation in behavioral color discrimination ability. In particular, higher carotenoid concentration in oil droplets correlated with increased discrimination of colors from different sides of the previously identified category boundary. These data show that differences in the sensory periphery can correlate with individual variation in perception of a signal-relevant color range.

Significance statement
Signal receivers vary in their preferences for signaling traits, but whether this is due to variation in how different receivers perceive signals is not well-understood. We show that variation between individual zebra finch females in perception of an orange-red continuum range correlates with the carotenoid concentration of retinal oil droplets. These data provide the first direct evidence that individual variation in oil droplet carotenoid concentration can lead to variation in color discrimination ability. Linking variation in signal-relevant color discrimination ability with variation in retinal physiology suggests a potential mechanism contributing to individual variation in signal assessment.

Communicated by K. McGraw

✉ Stephen Nowicki
  snowicki@duke.edu

✉ Sönke Johnsen
  sjohnsen@duke.edu

1   Centre for Ecology and Conservation, University of Exeter, Penryn TR10 9FE, UK

2   Department of Biology, Duke University, Box 90338, Durham, NC 27708, USA

3   Department of Biology, Florida International University, Miami, FL 33199, USA

4   Department of Neurobiology, Duke University Medical School, Durham, NC, USA

Introduction

In a variety of behavioral contexts, such as mate choice or aggression, animals evaluate one another using assessment signals. The expression of these signaling traits often varies among individuals, reflecting (on average) reliable information about signaler quality (Maynard Smith and Harper 2003; Searcy and Nowicki 2005). In addition to variation in signaling traits, females sometimes vary in their preferences for male traits (Jennions and Petrie 1997; Ronald et al. 2012; Ah-King and Gowaty 2016), which could be due to variation in sensory perception across individuals (Ronald et al. 2012). However, the extent to which variation in receiver preferences results from variation in perception within a species, and whether those perceptual differences are due to variation in sensory physiology, remains poorly understood in non-human animals.

Carotenoid pigments, which underlie red, orange, and yellow coloration in many animals (Fox 1979), are an important class of compounds in signaling traits, particularly in mate choice (Searcy and Nowicki 2005). This is because carotenoids also play important roles in health-related physiological processes such as immune function and oxidant protection (Olson and Owens 1998; Hill 1999). Vertebrates cannot synthesize carotenoids de novo, and individuals vary in their ability to acquire (Hill et al. 1994; Toomey et al. 2010), metabolize (Borel 2012; Weaver et al. 2018), and allocate carotenoids to different functions (Blount 2004; Toomey and McGraw 2011), suggesting these pigments are a reliable indicator of the quality of a potential mate (Olson and Owens 1998; Hill et al. 2002; Searcy and Nowicki 2005; Casagrande et al. 2014; but see Koch and Hill 2018).

Carotenoids also play an important role in the color vision of certain vertebrates, including fish, turtles, birds, and diurnal lizards (Walls 1942; Toomey and Corbo 2017; Wilby and Roberts 2017), via their presence in retinal oil droplets. Specifically, oil droplet carotenoids act as long-pass cut-off filters. Light passes through the droplet, which selectively absorbs short-wavelength light (Wald and Zussman 1937; Meyer et al. 1971; Goldsmith et al. 1984) before it reaches the photoreceptor's light-absorbing visual pigment (Hart and Vorobyev 2005). Filtering by oil droplets narrows the spectral sensitivity of cones, thus reducing overlap between neighboring sensitivity curves, which is thought to enhance color discrimination ability (Vorobyev et al. 1998; Vorobyev 2003).

While across species it is clear that the characteristics of carotenoid-containing oil droplets result in differences in color discrimination, recent studies have yielded conflicting results for the hypothesis that variation in carotenoid levels within a species can impact the ability to discriminate certain color stimuli. For example, dietary carotenoid supplementation improved the ability of house finches (*Carpodacus mexicanus*) to extract red items from amid black and tan distractors (Toomey and McGraw 2011), and improved the ability of Japanese quail

(*Coturnix japonica*) to discriminate between red, orange, and yellow targets (Lim and Pike 2016). However, total retinal carotenoid content did not predict female preference for red over orange and yellow males in house finches (Toomey and McGraw 2012). One potential reason for these mixed results is that the effects of dietary manipulation on retinal carotenoids are small (e.g., Knott et al. 2010; Toomey and McGraw 2010), having negligible impacts on predicted spectral sensitivity (Knott et al. 2010). More recent work, however, has shown that standing variation in the concentration of retinal carotenoids among wild-caught cowbirds (*Molothrus ater*) is potentially large enough to contribute to differences in perception of signaling traits across individuals (Ronald et al. 2017), though no behavioral tests were performed in that study.

In this study, we quantify oil droplet absorbance, a measure of spectral filtering and a proxy for carotenoid concentration (e.g., Lipetz 1984), in female zebra finches. We then relate this measure to behavioral performance on a signal-relevant color discrimination task. Zebra finches are an especially interesting species for such a test, because Caves et al. (2018) previously demonstrated that females categorically perceive an orange-red color continuum that parallels color variation found in male beaks, an assessment signal known to influence female mate choice (e.g., Burley and Coopersmith 1987; Vos 1995; Collins and ten Cate 1996; de Kogel and Prijs 1996). Specifically, when discriminating between pairs of colors across this orange-red continuum, female zebra finches labeled this continuum as lying in two categories with a perceptual boundary between them. Furthermore, the females showed increased discrimination between colors that occur on opposite sides of the perceptual boundary as compared with color pairs that were equally distant in color space but that occurred on the same side of the boundary. This enhanced discrimination between colors that cross a perceptual boundary is referred to as categorical perception (Harnad 1987). Caves et al. (2018) also observed substantial variation among individuals in the strength of the category boundary; that is, how much color discrimination ability improved as a result of discriminating between colors from across the boundary as compared with colors from the same category. Thus, here we test the hypothesis that variation in the carotenoid concentration of retinal oil droplets contributes to variation in perception of signal-relevant colors across individuals. In particular, we link variation among females in color discrimination ability with variation in the spectral filtering of retinal oil droplets.

Methods

Experimental subjects

Subjects were sexually mature female zebra finches obtained from a colony maintained by Richard Mooney at Duke

University (IACUC A258-14-10). (Although females ranged in age from 17 to 64 months at the start of the experiment, age was not a significant factor in models describing our results; see "Results"). Outside of behavioral trials, lighting was kept on a 15-h:9-h light to dark cycle with overhead lighting produced by fluorescent bulbs (Ecolux with Starcoat SP 35/41, color temperature 3500–4100 K, General Electric) with electronic ballasts (Hi-Lume 3D/Eco-10, Lutron Electronics) to match the lighting and light cycle in their original colony. Electronic ballasts produce little or no flicker; operating at 60 Hz, any flickering that might be associated with this lighting would occur at 120 Hz, well above the critical flicker fusion rate of 55 Hz measured for zebra finches (Crozier and Wolf 1941). Rooms were maintained at 25–27 °C. Prior to the start of the dietary manipulation (below), all birds were maintained on an ad libitum diet of zebra finch food (Kaytee Forti-Diet Pro Health Finch diet). All methods were approved under Duke University IACUC protocol A004-17-01.

Dietary carotenoid levels

Based on previous work that linked dietary carotenoid supplementation or depletion with color discrimination ability (e.g., Toomey and McGraw 2011; Lim and Pike 2016), we first attempted to increase the variation in oil droplet carotenoid concentration within our population by manipulating birds' diets. Individuals were either maintained on ad libitum zebra finch food as described above ("control"; $n = 9$), or a carotenoid-restricted diet, with ("carotenoid replaced"; $n = 12$) or without ("carotenoid minus"; $n = 10$) access to carotenoids in their drinking water. Birds were assigned to groups both to make sample sizes as equal as possible and based on social cohesion within groups, since they were housed in group cages (custom, $46 \times 30 \times 30$ cm) based on treatment group during the dietary manipulation. Two birds died in the course of the study resulting in a total of 31 birds in the diet manipulation.

The "carotenoid replaced" and "carotenoid minus" birds were, over the course of a 5-week wash-in period, introduced to a diet of 30% seed and 70% carotenoid-free pellets (diet no. 5C7V, TestDiet, St. Louis, MO, USA) by volume, following Knott et al. (2010). Although the seed content in this restricted diet was below the daily intake of the birds on an ad libitum diet, the overall volume of food provided was the same in all treatments, so birds were not food-restricted. We visually confirmed using video footage that birds in the diet-manipulated groups were consuming carotenoid-free pellets.

Birds in the "carotenoid minus" group were provided with tap water. Birds in the "carotenoid replaced" group were provided with tap water that was supplemented with a lutein-zeaxanthin carotenoid mixture (Oro Glo-11, Kemin Agrifoods Europe, Herentals, Belgium) at a concentration of 50 µg/ml (following Blount et al. 2003; Knott et al. 2010).

Lutein and zeaxanthin are two of the primary carotenoids that act as metabolic precursors to the carotenoids found in retinal oil droplets (Goldsmith et al. 1984; Schiedt et al. 1991; Bhosale et al. 2007). At the end of the 5-week wash-in period, birds were maintained on their respective diets for an additional 5 weeks before trials began. The 5-week maintenance period was selected because Knott et al. (2010) report that this is sufficient time to observe depletion of retinal carotenoids, although the effect sizes they report were small (~ 1-nm difference in photoreceptors containing R-type (i.e., red) oil droplets between control and carotenoid-depleted birds). Birds were weighed weekly starting at day 1 of the wash-in period, to ensure that the carotenoid-limited diets did not cause significant weight loss (Fig. S1).

Selection of color stimuli

To assess color discrimination ability, we followed the methods described in Caves et al. (2018). We began with a set of 40 Munsell color swatches (Pantone LLC, Carlstadt NJ, USA), ranging from orange to red, selected because they had previously been used to describe the gamut of beak color variation in male zebra finches (Burley and Coopersmith 1987; Collins et al. 1994; Birkhead et al. 1998). We then measured reflectance spectra from each color swatch using an integrating sphere with a built-in tungsten-halogen light source (ISP-REF; Ocean Optics). Measurements were taken in reference to a Spectralon 99% white reflectance standard (Labsphere). For each of the measured Munsell colors, we calculated normalized photon catch (the relative stimulation of each photoreceptor type in the eye) for zebra finch short-, medium-, and long-wavelength photoreceptors. Photon catches were calculated from 400 to 700 nm using zebra finch spectral sensitivity curves (Bowmaker et al. 1997; Lind 2016), an ambient light spectrum (as described below), and the spectral reflectance of each color, using the following formula:

$$Q_{r,c}(\lambda) \; \alpha \int_{400}^{700} S_r(\lambda) \times R_C(\lambda) \times I(\lambda) d\lambda$$

in which Q is the photon catch for photoreceptor type r in response to color c, S_r is the sensitivity of photoreceptor type r, R_c is the reflectance of color c, λ denotes wavelength, and I is the irradiance of the illuminant. The zebra finch spectral sensitivity curves that we used incorporate information about the transmittance of each oil droplet type and ocular media transmittance specific to the zebra finch; a detailed description can be found in Lind (2016). As an ambient light spectrum, we used the CIE Illuminant A standard tungsten bulb illuminance spectrum (color temperature 2856 K) which is nearly identical to the actual ambient light in our experimental room (Fig. S2). Using experimental light as the measure of ambient light rather than a standard had no effect on predicted discriminability.

We then used the photon catch values to calculate chromatic distance (ΔS, a measure of the predicted discriminability between two colors) using the receptor noise-limited (RNL) model of color discrimination (Vorobyev and Osorio 1998). We visualized ΔS using a perceptually uniform, two-dimensional space based on both hue and saturation/chroma, in which the Euclidean distance between two colors is equivalent to the RNL model-derived chromatic distance (equations describing the chromaticity space can be found in Hempel de Ibarra et al. 2001). Although the RNL-based chromaticity space was developed for trichromatic vision, it is appropriate to use here for three reasons. First, the reflected ultraviolet radiance from our Munsell chip stimuli under experimental lighting conditions is essentially zero (Fig. S3). Second, the quantum catch for the UV cone was on average (± standard deviation) only $0.26 \pm 0.10\%$ (range 0.14–0.42%) of total single-cone quantum catch (see Table S1). Third, recalculating ΔS using a tetrachromatic visual system (and thus including the UV cone catch) had minimal impact on predicted discriminability (Table S2), changing ΔS values by a mean (± standard deviation) of 0.26 ± 0.41 (range −0.18–0.99). Therefore, we expect that the impact of the UV cone on color perception was minimal, and thus, we did not include UV cone's photon catch in our calculations. Additionally, assuming a trichromatic visual system allowed us to visualize the relative positions of stimulus colors in the chromaticity space described above, in which Euclidean distance is equivalent to ΔS (Hempel de Ibarra et al. 2001). Visualizing the stimulus colors in this way allowed us to select eight stimulus colors that were roughly equally spaced in the chromaticity space and thus predicted to be equally discriminable from one another to a zebra finch visual system (Fig. S4), and which spanned the full range of previously described beak colors from the darkest red (color 1) to the brightest orange (color 8).

One factor included in the RNL model that can impact predicted ΔS is the relative density of each cone type (see Bitton et al. 2017). Because measures of photoreceptor noise and proportion of each cone type for the zebra finch are lacking, in our calculations of ΔS, we assumed equal cone-type proportions and a Weber fraction of 0.05 for the long-wavelength cone. Previous studies of zebra finch vision have used cone-type proportions of 1.5S:2M:3L (as reported in Lind 2016), which are based on numbers of each cone type measured using microspectrophotometry in Bowmaker et al. (1997). Notably, although recalculating ΔS using the cone-type proportions reported in Lind (2016) increased absolute ΔS (Table S2), it had minimal impact on the relative distance between color stimuli (see Table S2). Specifically, the three adjacent pairs of colors predicted to be most discriminable remained the same, though slight changes in order occurred, while the four color pairs predicted to be the least discriminable did not shift in order. Thus, we report both values of ΔS in

Table S2 and in the main text use ΔS based on the assumption of equal cone-type proportions, which is consistent with our previous work.

To investigate the effects of perceived brightness on color discrimination, we also calculated the relative photon catch of the zebra finch double cone (Lind 2016) for each Munsell color, since it is thought that in birds, the double cones encode brightness information (Osorio et al. 1999; Osorio and Vorobyev 2005; reviewed in Martin and Osorio 2008).

Behavioral tests of color discrimination

At the beginning of behavioral testing, birds were moved to individual cages ($12 \times 18 \times 13$ cm, Prevue Pet) outfitted with two wooden perches, a cuttlebone, water ad libitum, and either a normal seed diet (for control birds) or the carotenoid-limited diets described above (for "carotenoid replaced" and "carotenoid minus" birds). Birds in the "carotenoid minus" group continued to receive untreated tap water, while "carotenoid replaced" birds were provided with carotenoid-supplemented water as described above. On days during which behavioral trials were run, food was removed from each cage at 0900, to ensure that birds were motivated to perform the task. Trials began at 1400 each day and lasted for 20 days in total. During behavioral trials, lighting was provided by halogen bulbs (color temperature 2900 K, model number H&PC-61361, Philips Lighting) hung approximately 80 cm above the cage and filtered through vellum paper to provide diffuse, even lighting. Birds were allowed at least 5 min to acclimate to the experimental lighting conditions before trials began.

The eight selected colors (above) were used to create disc stimuli that were made by gluing two semi-circular halves together to form a circle. The two halves of the discs were either the same color ("solid") or different colors ("bicolor"). Discs were covered with an epoxy cover and fitted underneath with a rubber bumper that ensured they fit snugly into the wells.

We tested color discrimination using a food-reward protocol in which birds were presented with a foraging grid containing 12 wells. All birds used in this study had been previously trained on this protocol and used in Caves et al. (2018). Six of the wells were covered by the disc stimuli described above, two by bicolor discs and the remaining four by solid discs (two of each color in the bicolor discs; see Fig. S4 inset). Using discs made of the two endpoint colors, 1 and 8 ("1|8"), we trained the birds to search for food rewards placed beneath bicolor discs. Birds passed a trial if they flipped over both bicolor discs before flipping any solid discs within 2 min. Birds that passed six out of seven consecutive training trials began experimental trials. Six birds (one control, three "carotenoid replaced," and two "carotenoid minus") did not meet this criterion to move from training to experimental trials, resulting in a total of 25 birds that participated in behavioral

trials ($n = 8$ control, $n = 9$ "carotenoid replaced," and $n = 8$ "carotenoid minus").

In experimental trials, the makeup of the discs on the grid was the same as in the training trials, but we varied the two colors comprising the discs. Experimental trials involved color combinations that were either one (i.e., 1|2, 2|3, 3|4, etc.), two (i.e., 1|3, 2|4, 3|5, etc.), or three (i.e., 1|4, 2|5, 3|6, etc.) color steps apart (where a color step refers to two colors that are adjacent on the continuum from 1 to 8). Each day, experimental trials began with a 1|8 refresher task. Trials ended with a motivation check, in which we recorded the amount of time it took birds to begin to eat out of their regular seed dish once it was returned to the cage to ensure that birds had remained hungry and motivated throughout the task (see Table S3).

We randomized the location of discs on the grid for each trial using the *sample* function in R (R Development Core Team 2018). Although not all birds saw the same color combination on a given day, all birds performed one-apart tasks on the same day, two-apart tasks on the same day, etc. For each bird, we performed a total of seven trials for each color combination and calculated the proportion of trials that they passed for a given color combination (which we term "pass frequency"). If an individual did not flip at least two discs in at least three of the seven trials for a given color combination (which occurred in 5% of trials), we excluded that data point from analyses on the basis that we did not have enough data to assess whether that individual could discriminate that particular color combination. It was not possible to record behavioral data blind because our study involved watching focal animals perform cognitive tasks in the lab.

Retina extraction and microspectrophotometry of oil droplets

At the conclusion of behavioral trials, we used microspectrophotometry (MSP) to measure transmission spectra through individual photoreceptors containing red (R-type) oil droplets. We measured transmittance spectra from R-type oil droplets for three reasons. First, as reported in Knott et al. (2010), R-type oil droplet transmittance differed by only 1 nm between ventral and dorsal retinal regions (the smallest difference of any oil droplet type). Therefore, measuring R-type oil droplets minimized the noise that would be present in our data as a result of measuring oil droplets from across the retina (see below). Second, based on data in Knott et al. (2010), the predicted effect size of the diet manipulation on oil droplet carotenoids is smallest in the R-type droplet, as they are the most densely pigmented. Thus, measuring R-type droplets likely provided a conservative measure of change in carotenoid concentration given that a greater absolute change would be required to detect an effect that is likely also present in other oil droplet types. Third, we were able to unambiguously identify R-type droplets by color and size, and to obtain

reliable measurements with minimal noise (Goldsmith et al. 1984; Bowmaker et al. 1997). During all measurements, the MSP operator (L.S.) was blind to the identity of the sample.

Birds were euthanized by decapitation for retinal analysis the day following completion of their behavioral trials (control birds on August 22, 2018, and "carotenoid replaced" and "carotenoid minus" birds in September 9 and 10, 2018, respectively). The left eye of each bird was removed, and each retina was whole-mounted photoreceptor-side-up on a no. 1 ½, 22 × 30-mm glass coverslip (Electron Microscopy Sciences, Hatfield, PA), covered with a drop of 100% glycerol (Millipore Sigma, Merck, Darmstadt, Germany). To reduce scattering during MSP recordings, oil droplets were then isolated by using a razor blade to lightly dissociate them from the retinal tissue. Under a dissecting microscope, isolated oil droplets were swept from macerated retina to the sample edge using a paintbrush, and positioned circumferentially around the preparation. Thus, in the final preparation, we measured a haphazard sample of oil droplets from all retinal regions, by selecting oil droplets to measure from along the entire circular boundary of the retinal preparation. This allowed us to obtain an average of the variation across the entire retina for a given individual. Finally, a second cover slip was placed on top of the preparation and pressed gently (allowing the oil droplets to retain their shape) into a ring of silicone grease that had been placed around the tissue.

We performed MSP using a Nikon Diaphot-TMD inverted compound microscope (Melville, NY). A 20-W quartz tungsten-halogen lamp (Optometrics LLC, San Francisco, CA) provided white light, which was passed through a 50-μm-diameter fiber (Ocean Optics, Dunedin, FL) and a Zeiss ×32 Ultrafluar microscope condensing objective before passing through the sample. Light from the condensing objective was focused through a single oil droplet, collected by a Zeiss ×16/0.40 PH2 Neofluar microscope objective, which then passed the light through a UV-transparent beam splitter, a portion of which then entered a 1-mm-diameter fiber (Ocean Optics) connected to a USB2000 spectrometer (Ocean Optics) with a detector range of 200–1100 nm. Reference scans were taken through glycerol only, in the space between isolated oil droplets. Transmittance spectra of 15 R-type oil droplets were measured for each bird using OceanView (version 1.6.7) Software (Ocean Optics) (Fig. S5).

Calculation of λ_{mid}

We calculated λ_{mid}, defined as the wavelength halfway between the maximum and minimum recorded transmittance (Lipetz 1984; Hart and Vorobyev 2005), as a measure of oil droplet spectral filtering. λ_{mid} is a standard metric for describing oil droplets and is highly correlated with other commonly used metrics that describe oil droplet transmittance, such as λ_{cut} (for example, Hart and Vorobyev (2005) report that

Spearman's rank correlation coefficient between λ_{mid} and λ_{cut} equals 0.99 across all oil droplet types across a wide range of bird species). Because of its relationship with oil droplet spectral filtering, therefore, λ_{mid} is likely a useful proxy for carotenoid concentration, though we should note that the exact relationship between given λ_{mid} values and the precise concentrations of particular carotenoids is unknown. Additionally, given that oil droplet filtering is the functionally important aspect of carotenoid pigmentation, λ_{mid} also provides a direct link to the function of the oil droplet.

To calculate λ_{mid}, we restricted the data to include points between 520 and 680 nm. This isolated the area of interest in the transmittance spectrum and also excluded the noise in the spectrum that occurs at the shortest and longest wavelengths. We then fit the logistic function to the spectrum:

$$\frac{\text{Asym}}{1 + e^{\left(\frac{\lambda_{mid} - \text{input}}{\text{scal}}\right)}}$$

where Asym is a numeric parameter representing the asymptote, λ_{mid} is the x value at the inflection point of the curve, input is a vector at which to evaluate the function (i.e., 520–680 nm), and scal is a numeric scale parameter on the input axis.

The fitting was performed using a custom code in R version 3.4.3 (R Development Core Team 2018) that used nonlinear least squares with the self-starting logistic function "SSlogis" (see Supporting Data: MSP analysis code and sample oil droplet transmission file). From the model fit, we then calculated the midpoint (λ_{mid}). To assess the fit of the sigmoidal model, we calculated R^2 values. The mean ± standard deviation in R^2 value across all measured oil droplets was 0.98 ± 0.01, indicating that the data were well-described by the logistic model. Individual oil droplet spectra with R^2 fits of less than 0.95 were excluded from the dataset. As an additional check, we visually inspected all of the oil droplet spectra and confirmed that each of the spectra with an R^2 of less than 0.95 contained significant noise likely indicative of the oil droplet moving off-center while the transmission spectrum was captured. In total, only 10 oil droplets were excluded from the total of 450 that were measured (final $n = 440$ oil droplets from 31 individuals).

Statistical analysis

To analyze our behavioral color discrimination data ($n = 25$ birds), we built linear mixed-effects models using the R package *lme4* (Bates et al. 2015). Models included behavioral discrimination data from one-, two-, and three-apart trials. To calculate p values for the fixed-effects models, we used the package *afex* (Singmann et al. 2015), which utilizes *lmerTest* (Kuznetsova et al. 2017) to estimate degrees of freedom and calculate p values via Satterthwaite's method.

To examine how differences in color related to pass frequency, we built models that isolated the contribution of each "color step" (e.g., 1–2, 2–3, and 3–4) to pass frequency. For example, a large contribution of the 5–6 color step to pass frequency would indicate that birds perform well on discrimination tasks that include that color step (for example, 3|6, 4|6, and 5|6), regardless of whether the task is between colors that are closer (as in a one-apart, e.g., 5|6) or farther (as in a three-apart, e.g., 3|6) from one another in color space. We coded these color steps as binary measures of each comparison. For example, comparison 1|3 includes color steps 1–2 and 2–3 but not steps 3–4, 4–5, 5–6, et cetera.

Because our color steps are not perfectly equally spaced in chromaticity space (see Fig. S4 and "Selection of color stimuli," above), we first ensured that describing our color discrimination data using color steps as opposed to chromatic distance was appropriate. To do so, we built a linear mixed-effects model of pass frequency with dietary treatment group and binary measures of each color step (i.e., whether or not a given comparison included a given color step) as fixed-effects as well as bird ID as random intercepts. This color step model performed far better than an equivalent model that included chromatic distance rather than color steps (chromatic distance model Akaike information criterion (AIC) = 1.08, color steps model AIC = −121.3, ΔAIC = 122.3, see Table S4). Additionally, because real zebra finch beaks vary in brightness, our color stimuli also varied in brightness, so we built a model that included Michelson Contrast (ratio of the sum to the difference of double cone photon catches) of each color pair rather than color steps. The contrast model performed significantly worse than the color step model (contrast model AIC = −58.6, color steps model AIC = −121.3, ΔAIC = 62.7), indicating that considering color steps better described our data than using differences in brightness.

We next built a series of linear mixed-effects models to examine the contributions of each color step and λ_{mid} to pass frequency, with each model including additional parameters to a baseline ("null") model. We then took an information criterion approach (using AIC scores) to determine which of the models was best, and we report that model as our final model in the main text. Our simplest model (the "null" model) was identical to the color step model referenced in the previous paragraph, and included pass frequency for a given comparison as the response variable, with each color step and dietary treatment group as fixed-effects, and bird ID as a random intercept. A version of this "null" model that included no random intercept of bird ID performed much worse than the "null" model (ΔAIC = 36), verifying that there was a significant variation between individuals in color discrimination ability.

The second model (the "lambda model") additionally included λ_{mid} as a predictor variable to examine the independent effect of oil droplet absorbance on pass frequency. The third model (the "5–6 interaction model") included all terms from

the "lambda model" as well as an interaction term between λ_{mid} and the 5–6 color step. The fourth model (the "5–6 interaction + random slope model") included all terms from the "5–6 interaction model" as well as random slopes of the effect of the 5–6 color step for each bird's color discrimination ability. Finally, the fifth model (the "all interactions model") included all terms from the "5–6 interaction model" as well as interaction terms between λ_{mid} and all color steps. Based on previous work demonstrating a categorical boundary between colors 5 and 6 and observed between-individual variation in the strength of that boundary (Caves et al. 2018), we predicted that increases in carotenoid concentration would specifically affect discrimination of comparisons that included the 5–6 color step, and thus, the "5–6 interaction + random slopes model" would be the best-fit.

To assess model fit between these models, we used the Akaike information criterion (AIC; Akaike 1974; Burnham and Anderson 2002), and then assigned ΔAIC values by calculating the difference between the AIC value of a given model and the AIC value of the best-fit model (i.e., that with the lowest AIC). We considered ΔAIC values of 3–7 to indicate models that were considered possibly worse in fit to the null model, while models with ΔAIC > 7 were considered definitely worse in fit than the best-fit model (Burnham et al. 2011; Symonds and Moussalli 2011).

As predicted, of the five models we built to describe our behavioral color discrimination data, the "5–6 interaction + random slopes" model had the lowest AIC value (Table 1). This model performed much better than simpler models, with ΔAIC > 21 for the "null," "lambda," and the "5–6 interaction models" relative to the best-fit model. Continuing to add additional interactions between λ_{mid} and additional color steps (i.e., the "all interactions model") made the model fit worse, not better (ΔAIC = 3.1). We therefore report the output of the "5–6 interaction + random slopes" model as our final model in the main text. The raw data and the R script used to generate each model and its output can be found in the supplementary material and on the Duke University Data Repository (https://doi.org/10.7924/r4jw8dj9h), to allow readers to reproduce each of these models.

Table 1 AIC values and ΔAIC values relative to the best-fit model (the "5–6 interaction + random slopes model"), for each of the five models used to describe our behavioral discrimination data

Model	AIC	ΔAIC
"5–6 interaction + random slopes"	− 163.9	0
"All interactions"	− 160.8	3.1
"5–6 interaction"	− 141.2	22.7
"Lambda"	− 130.2	33.7
"Null"	− 121.3	42.6

To visualize the relationship between inter-individual variation in λ_{mid} and the effect of a comparison crossing the 5–6 boundary on color discrimination ability, we extracted the random slopes from a version of the "null" model that additionally included random slopes of the effect of the 5–6 color step on individual discrimination. (The reason to use the "null" model here as the base model, rather than the "5–6 interaction + random slopes model" is that we wanted to visualize the relationship between λ_{mid} and the 5–6 effect. Had we taken the coefficients from the "5-6 interaction + random slopes model," the coefficients from that model would already be controlling for the effect of λ_{mid}, given that λ_{mid} is a fixed effect in that model.) We then plotted these values against our measures of individuals' λ_{mid} values, allowing us to directly examine the relationship between λ_{mid} and the effect of the 5–6 boundary on individuals' discrimination performance. Finally, we performed a simple linear regression in which the effect of the 5–6 boundary for each bird was the response variable and λ_{mid} was the predictor variable. Such an approach treats the individual bird as the unit of analysis and uses the estimates derived from the mixed model as a single measure of a trait specific to each bird, allowing us to estimate the proportion of variance in the effect of the 5–6 boundary for each bird that was explained by between-individual variation in λ_{mid}.

Results

Dietary manipulation

We found no significant differences in λ_{mid} among dietary treatment groups (ANOVA; $F_{2,28} = 2.78$, $p = 0.08$; Fig. 1a), although the trends were in the expected direction. The lowest (mean ± standard deviation) λ_{mid} occurred in the "carotenoid minus" group (596 ± 2.4 nm) and the highest in the control group (599 ± 2.6 nm), with the "carotenoid replaced" group being intermediate (598 ± 2.9 nm). The effect size of the manipulation was small (as expected based on Knott et al. 2010), with mean λ_{mid} differing by only 3 nm between the control and "carotenoid minus" groups.

ANOVAs indicated that there were no differences between treatment groups in overall participation rate (the mean rate at which birds flipped at least two discs in a trial, as opposed to one or none; $F_{2,23} = 0.87$, $p = 0.43$), pass frequency on a daily refresher task using colors 1 and 8 ($F_{2,23} = 0.94$, $p = 0.41$), or motivation to feed, as indicated by the mean amount of time it took birds to start eating from their regular seed dish at the end of each set of trials ($F_{2,23} = 0.27$, $p = 0.77$). Thus, the birds in all three treatment groups were equally motivated to perform the task, retained the initial training task equally well over the course of the experiment, and participated daily at equal rates (Table S3).
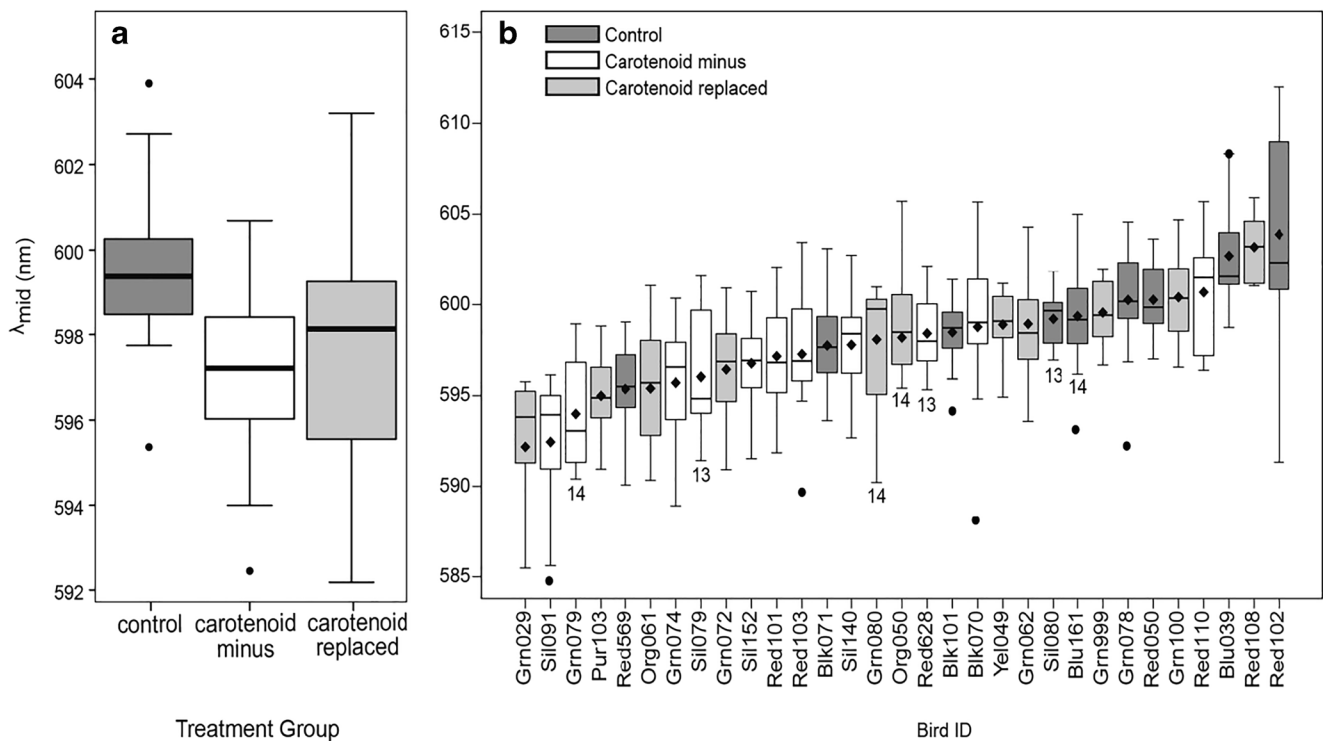
Fig. 1 Microspectrophotometric measures of λ_{mid} in R-type oil droplets, **a** within dietary treatment groups and **b** across individuals. In **b**, sample size was 15 R-type oil droplets per individual, except where noted by a number under the boxplot; diamonds depict the mean value for each individual. In both plots, box plots depict the median (horizontal line), 25th and 75th percentiles (box), 25th and 75th percentiles $\pm 1.5\times$ inter-quartile range (whiskers), and outliers (circles)

Carotenoid concentration in oil droplets

Despite minimal differences in λ_{mid} between dietary treatment groups, we found large variation across individuals in λ_{mid}, which ranged from 592 to 603 nm ($n = 31$ individuals; Fig. 1b). We also found relatively large within-individual variation in λ_{mid}, which was expected given that oil droplets, which are known to vary in carotenoid concentration across different retinal regions, were sampled across the entirety of the retina to generate a representative mean for each individual. However, λ_{mid} was strongly and significantly repeatable within individuals ($p < 0.0001$, $R = 0.42$, 95% confidence interval [0.28–0.54]), with the magnitude of the R-statistic indicating that approximately 42% of total variance in λ_{mid} is explained by individual ID. Thus, we feel confident that we captured real between-individual variation in λ_{mid}.

Behavioral color discrimination

We observed substantial inter-individual variation in how much better at discriminating colors birds became when crossing the 5–6 boundary relative to within-category comparisons (range = 0.07–0.55, median = 0.27, mean = 0.30, coefficient of variation = 46%, Fig. 2). The best-fit model that described color discrimination ability (Table 2) showed that several color steps, including 3–4, 4–5, 5–6, and 6–7 each contributed significantly

to pass frequency ($p < 0.001$) as indicated by the fact that the 95% confidence interval around the coefficient did not overlap zero (Fig. 3a). However, as indicated by the model coefficients, crossing the 5–6 color step had by far the largest contribution to pass frequency, resulting in an average 30 percentage point increase in pass frequency (95% confidence interval = 25–35 percentage points), as compared with the second largest increase of (12 points) that results from crossing the 6–7 step (95% CI = 7–17 percentage points) (Fig. 3a). The confidence interval for the 5–6 step also did not overlap with that of any other color step. The disproportionate impact of the 5–6 color step on pass frequency and the lack of overlap between confidence intervals confirmed our previous finding that a category boundary exists between colors 5 and 6 (Caves et al. 2018). Of note is that while the RNL model (Vorobyev and Osorio 1998) did predict slight differences in discriminability of the chosen colors as indicated by the slightly unequal ΔS values between color pairs (Fig. S4), a model including chromatic distance (ΔS) rather than color steps did not describe the data well (see "Methods" and Table S4), ΔS of each color step did not correlate with the model coefficients (Fig. 3b), and the disproportionate effect of the 5–6 step on pass frequency was not predicted by the RNL model (Table S2).

λ_{mid} was strongly associated with birds' average pass frequency across trials (the "lambda model," $t = 3.38$, $df = 21$, $p = 0.003$, see Table S5). Thus, individuals with
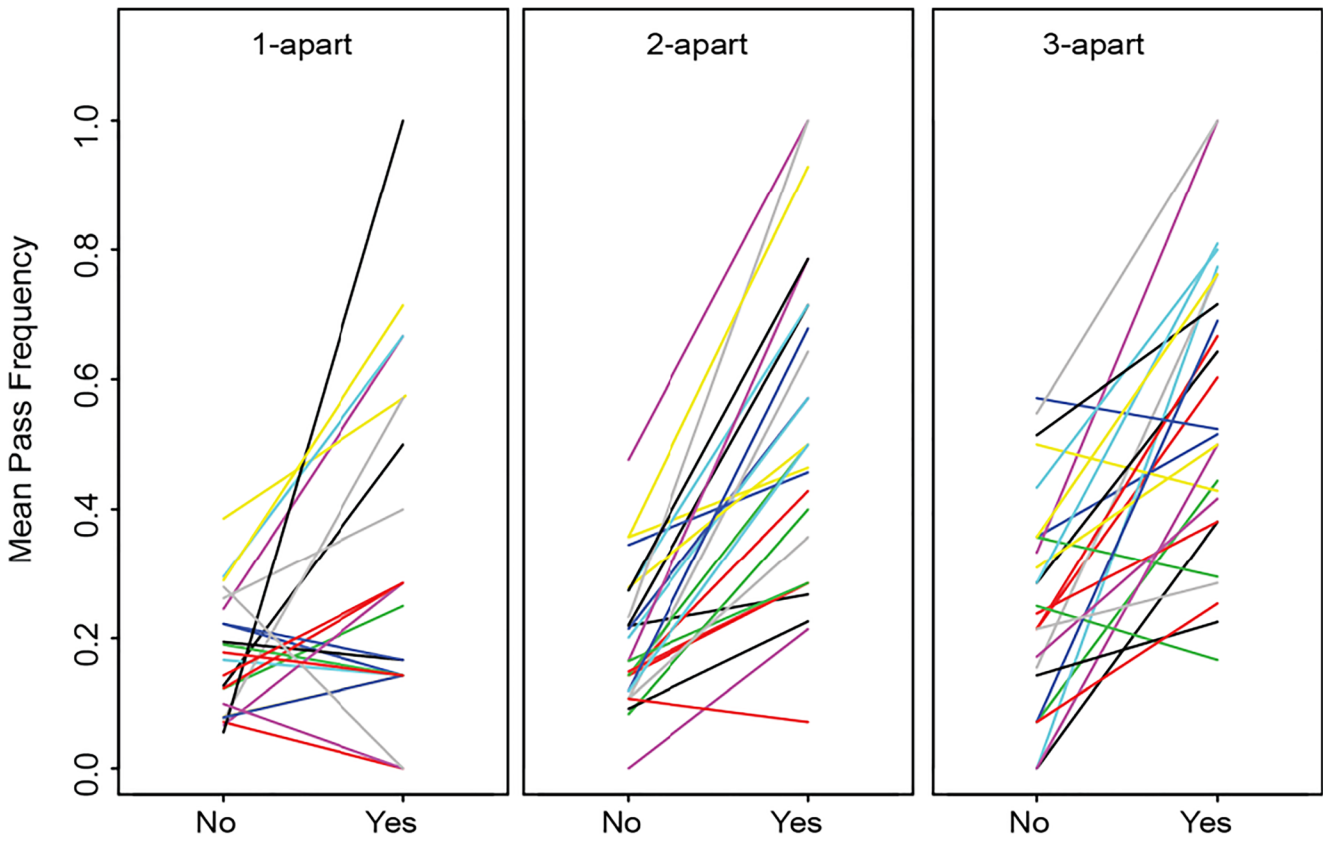
Fig. 2 The strength of the category boundary, i.e., the increase in mean pass frequency between comparisons that did and did not cross the 5–6 category boundary, varied across individuals for 1-, 2-, and 3-apart comparisons. Each line indicates an individual

higher λ_{mid} values also had higher overall pass frequencies on color discrimination tasks (Fig. S6). The results of the best-fit model (Table 2), however, revealed that the bulk of the increase in pass frequency for birds with higher λ_{mid} was specifically due to those birds being better at discriminating between colors that came from

Table 2   Results from the "5-6 interaction + random slopes model." Interactions are indicated by an asterisk (*). Note that each color step was included as a separate explanatory factor. Significant model effects are shown in italics

| Fixed-effects	Coefficient	Standard error	T statistic	P	Interpretation
Treatment: Car. replaced	*0.14*	*0.04*	*3.53*	*0.002*	*↑Pass frequency for birds in the carotenoid replaced group as compared with control*
Treatment: Car. minus	*0.10*	*0.04*	*2.09*	*0.028*	*↑Pass frequency for birds in the carotenoid minus group as compared with control*
λ_{mid}	*0.02*	*0.01*	*2.21*	*0.039*	*↑Pass frequency with ↑λ_{mid}*
1–2 color step	0.02	0.03	0.54	0.59	—
2–3 color step	0.05	0.03	1.80	0.07	—
3–4 color step	*0.08*	*0.02*	*3.72*	*0.0002*	*↑Pass frequency for comparisons that cross the 3–4 color step*
4–5 color step	*0.10*	*0.02*	*4.06*	*< 0.0001*	*↑Pass frequency for comparisons that cross the 4–5 color step*
5–6 color step	*0.30*	*0.04*	*8.2*	*< 0.0001*	*↑Pass frequency for comparisons that cross the 5–6 color step*
6–7 color step	*0.12*	*0.03*	*4.88*	*< 0.0001*	*↑Pass frequency for comparisons that cross the 6–7 color step*
7–8 color step	− 0.01	0.03	− 0.20	0.84	—
*λ_{mid} *5–6 color step*	*0.03*	*0.01*	*2.09*	*0.048*	*↑Effect of crossing 5–6 step with ↑λ_{mid}*
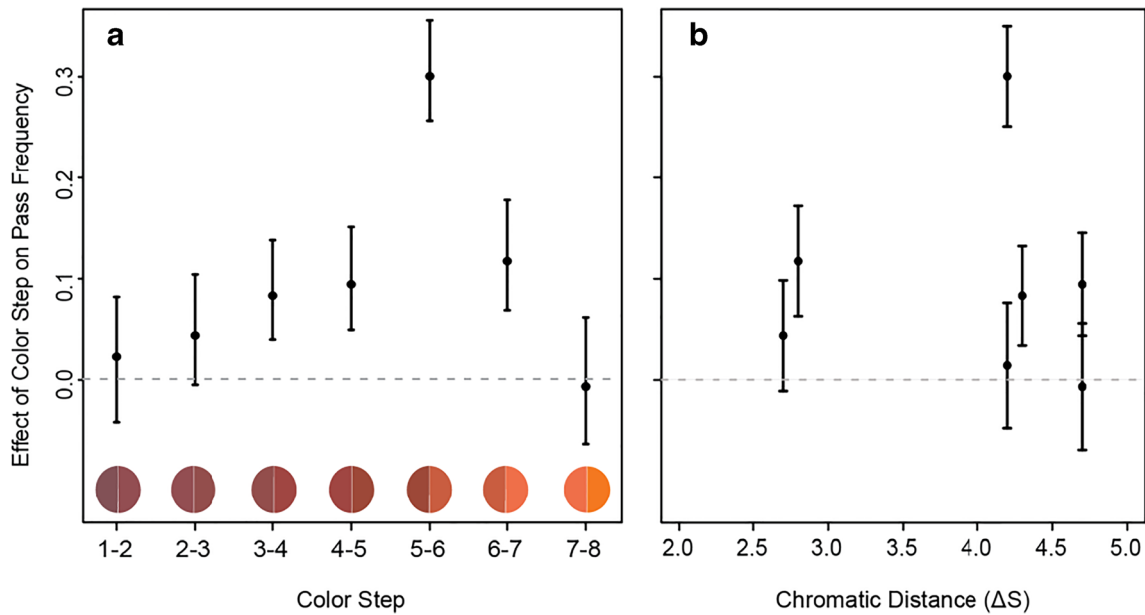
Fig. 3 **a** The effects of each color step on pass frequency. Points represent estimates (model coefficients) and bars represent 95% confidence intervals corresponding to the contribution of each color step to birds' pass frequency. The colored circles indicate the two Munsell colors that fall on either end of a given step. **b** The color steps with the greatest effect on pass frequency are not those that are the farthest apart in terms of chromatic distance (ΔS), indicated on the x-axis

different sides of the 5–6 category boundary. Specifically, the best-fit model (Table 2) showed that both λ_{mid} ($p = 0.039$, Table 2) and the interaction term between λ_{mid} and the 5–6 color step ($p = 0.048$ Table 2) were significant, and adding additional interaction terms between λ_{mid} and other color steps did not further improve the model fit (see Table 1). The value of the coefficient for the interaction term (0.03; Table 2) indicates that higher values of λ_{mid} are associated with increased ability to discriminate between cross-boundary pairs of colors, but not those from within a category. In addition, the model showed a significant effect of treatment group that was independent of λ_{mid}, in that both "carotenoid replaced" and "carotenoid minus" birds had higher overall pass rates than control birds (Table 2).

Consistent with the model results, a linear regression showed that the effect of crossing the 5–6 boundary for each bird was significantly and positively correlated with λ_{mid} ($b = 0.022$, $t = 2.3$, $p = 0.03$) and that variation in λ_{mid} explained 18% of the between-individual variation in the effect of crossing the 5–6 boundary on discrimination ($R^2 = 0.18$, Fig. 4). Thus, higher λ_{mid} values are associated with a greater increase in pass frequency when comparing colors from across the 5–6 boundary as opposed to colors from within the same category.

Finally, because various aspects of retinal physiology can vary with age, we performed a post hoc analysis in which age was added to the best-fit model; however, this analysis did not identify a significant effect of bird age on pass frequency ($p = 0.22$).

Discussion

Our data indicate that λ_{mid} in R-type oil droplets, a proxy for the concentration of carotenoids, correlates with variation in behavioral color discrimination of a carotenoid color continuum. Given the many physiological functions of carotenoids and the costs associated with obtaining and metabolizing them into useful forms, their expression as an ornamental color is
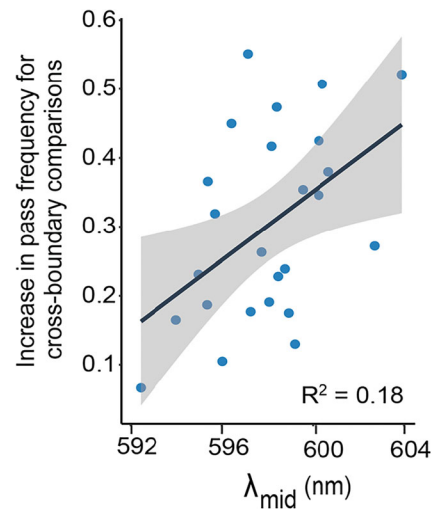

Fig. 4 Mean λ_{mid} and the effect of crossing the 5–6 color step on discrimination ability across individuals. Coefficients (y-axis) show the strength of the category boundary for each individual. These coefficients were derived from a version of the "null model" built with random slopes of crossing the 5–6 boundary. The gray-shaded area represents the 95% confidence interval surrounding a best-fit line drawn through these points

thought to be a signal of a potential mate's quality (Olson and Owens 1998; Hill et al. 2002; Searcy and Nowicki 2005; Casagrande et al. 2014; but see Koch and Hill 2018 for a review of the debate regarding the indicator function of carotenoids). The link shown here between carotenoid levels and color discrimination suggests that carotenoid availability may influence not only just how carotenoid-based color signals are expressed but also how they are perceived as well.

Across the 31 subjects in this study, the λ_{mid} of R-type oil droplets ranged from 592 to 603 nm, a range that overlaps with a previously published value of 597 nm for the mean R-type λ_{mid} in zebra finches (Bowmaker et al. 1997; Hart and Vorobyev 2005). The range of variation, 11 nm, is also similar to reported variation in mean R-type oil droplet λ_{cut} in wild cowbirds *Molothrus ater* (Ronald et al. 2017), adding to our understanding of intra-individual variation in oil droplet absorbance. We also found that individuals varied in their ability to discriminate colors spanning the range of male zebra finch beak coloration, a signal involved in mate choice (e.g., Burley and Coopersmith 1987; Vos 1995; Collins and ten Cate 1996; de Kogel and Prijs 1996). In particular, this effect was observed most strongly with respect to discrimination of color pairs that crossed the color category boundary we had identified previously (i.e., 5–6 color boundary; Caves et al. 2018). When color pairs crossed this boundary, pass frequency increased by a minimum of 7 percentage points to as much as 55 percentage points. This variation correlated positively with variation in λ_{mid} ($R^2 = 0.18$), showing that at least some of the observed variation in the ability of different females to discriminate between signal-relevant colors is explained by variation in retinal carotenoids.

The results presented here explore the impact of one aspect of retinal physiology on the strength of the category boundary, but the precise mechanism underlying these results is unclear. One possibility is that predicted discriminability could shift with changes in the filtering effects of oil droplets. Specifically, the spectral sensitivity of a photoreceptor, which is used to calculate the predicted discriminability (ΔS) between two colors using the RNL model, depends in part upon filtering by the oil droplet. Thus, one plausible mechanism underlying our results may be that variation in oil droplet filtering leads to differences in ΔS in line with what we observed—i.e., that the predicted discriminability between colors 5 and 6 is much lower when an oil droplet is carotenoid-depleted than when it is carotenoid-enriched. We considered this possibility by modeling how variation in carotenoid concentration affects the shape of a photoreceptor's spectral sensitivity curve and thus predicted discriminability between different color pairs (see Modeling Supplement for details). The model showed that (1) although variation in oil droplet filtering may contribute to differences in predicted discriminability, the magnitude of variation we saw in predicted discriminability was too small to explain our results, and (2) the 5–6 color step was not predicted to be the most

discriminable under any filtering scenario (i.e., filters carotenoid-enriched or carotenoid-depleted) that we tested. These results are in line with a previous study which found that variation in λ_{cut} (an alternative metric used to quantify oil droplet transmittance, one that is highly correlated with λ_{mid}, see "Methods") of 2–10 nm had no impact on spectral sensitivity of photoreceptors and, thus, no effect on predicted color discriminability (Knott et al. 2010).

We currently lack a comprehensive understanding of how much variation in predicted color discriminability (ΔS) can lead to variation in behavior. In part, mismatches between predicted discriminability and observed discrimination behavior could be due to a variety of additional factors not explicitly included in the RNL model (for a review see Emery and Webster 2019). For example, Ronald et al. (2017) found that incorporating information regarding individual variation in cone-type proportion into the RNL model could shift the predicted discriminability of signaling coloration, a potential source of variation in perception that has not been tested yet in the context of categorical perception. Above, we have shown that using different cone-type proportions to calculate predicted discriminability has minimal impacts on which color steps we would predict to be most or least discriminable. However, more relevant to this study would be if we could create individually tailored predictions for discriminability based on information regarding cone-type proportions in each individual. This was beyond the scope of the present experiment. Using a modeling approach, Price et al. (2019) have suggested that at least some of the difference in within- versus across-boundary discrimination seen in categorical perception can be explained by incorporating information about opponent channels into the RNL model (Price et al. 2019), although the precise nature of avian opponent channels is still unknown. Taken together, these studies and our results suggest that the variation in behavioral color discrimination observed here could at least in part arise as a result of processes occurring at the level of photoreceptors and the retina, as well as at higher levels of neural processing (e.g., Kelber 2019).

Several experimental limitations may have influenced our results, although we attempted to minimize the potential impacts of these limitations. First, we examined only R-type oil droplets, but changes in carotenoid concentrations of other oil droplet types likely also impact color perception. Of note, however, is that in the modeling approach described above, we allowed λ_{mid} in all four single-cone oil droplet types to vary over 20 nm and found no appreciable impact of that variation on predicted color discrimination (see Modeling Supplement for details). Additionally, we measured relatively few oil droplets per individual compared with other studies. Given that we sampled oil droplets from across the entire retina, however, we suggest that we likely captured much of the variation in λ_{mid} that exists in each individual and so we do not expect that increasing sample size would result in large

shifts in each individual's mean λ_{mid}. Lastly, carotenoids play an important role in many physiological functions outside of retinal oil droplets, including serving as antioxidants and as enhancers of immune system function (von Schantz et al. 1999; Toomey et al. 2010; Borel 2012; Weaver et al. 2018). Thus, other factors linked to carotenoid deprivation could have influenced the behavior of birds in the carotenoid-limited treatments. However, we found no differences between carotenoid-limited and control birds in either mean body weight throughout the experiment (Fig. S1), or in several indicators of behavior and motivation, including overall participation rate and motivation to participate (Table S3).

Our data indicate that females with different levels of retinal carotenoids differ in their ability to discriminate among carotenoid-relevant colors. Further studies are needed, but these results suggest that females may also differ in their ability to discriminate between potential mates based on beak color. If such proves to be the case, this finding will have important implications for understanding mate choice and the dynamics of sexual selection in this species. For example, lower quality females having lower overall carotenoid levels (as the result of poor diet or an immune challenge) may be less able to deploy carotenoids in their retinal oil droplets and may thus discriminate differently among males based on beak coloration than would a high-quality female. Therefore, in a system in which both signal production and signal perception are influenced by carotenoids, the quality of both the sender and the receiver may influence the outcome of a mate choice interaction. The possibility that high-quality females are better able to discriminate high- from low-quality mates than are low-quality females suggests an intriguing potential link between visual physiology and assessment signaling that deserves further attention, and further highlights the importance of considering variation in the perceptual abilities of the signal receiver when studying the dynamics of a signaling system.

Compliance with ethical standards

References

Ah-King M, Gowaty PA (2016) A conceptual review of mate choice: stochastic demography, within-sex phenotypic plasticity, and individual flexibility. Ecol Evol 6:4607–4642. https://doi.org/10.1002/ece3.2197

Akaike H (1974) A new look at the statistical model identification. IEEE T Automat Contr 19:716–723. https://doi.org/10.1109/TAC.1974.1100705

Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48

Bhosale P, Serban B, Da YZ, Bernstein PS (2007) Identification and metabolic transformations of carotenoids in ocular tissues of the Japanese quail *Coturnix japonica*. Biochemistry 46:9050–9057. https://doi.org/10.1021/bi700558f

Birkhead TR, Fletcher F, Pellatt EJ (1998) Sexual selection in the zebra finch *Taeniopygia guttata*: condition, sex traits and immune capacity. Behav Ecol Sociobiol 44:179–191

Bitton PP, Janisse K, Doucet SM (2017) Assessing sexual dicromatism: the importance of proper parameterization in tetrachromatic visual models. PLoS One 12:e0169810. https://doi.org/10.1371/journal.pone.0169810

Blount JD (2004) Carotenoids and life-history evolution in animals. Arch Biochem Biophys 430:10–15. https://doi.org/10.1016/j.abb.2004.03.039

Blount JD, Metcalfe NB, Birkhead TR, Surai PF (2003) Carotenoid modulation of immune function and sexual attractiveness in zebra finches. Science 300:125–127. https://doi.org/10.1126/science.1082142

Borel P (2012) Genetic variations involved in interindividual variability in carotenoid status. Mol Nutr Food Res 56:228–240

Bowmaker JK, Heath LA, Wilkie SE, Hunt DM (1997) Visual pigments and oil droplets from six classes of photoreceptor in the retinas of

birds. Vis Res 37:2183–2194. https://doi.org/10.1016/S0042-6989(97)00026-6

Burley N, Coopersmith CB (1987) Bill color preferences of zebra finches. Ethology 76:133–151. https://doi.org/10.1111/j.1439-0310.1987.tb00679.x

Burnham KP, Anderson DR (2002) Model selection and multimodel interference: a practical information-theoretic approach. Springer, New York

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav Ecol Sociobiol 65:23–35

Casagrande S, Pinxten R, Zaid E, Eens M (2014) Carotenoids, birdsong and oxidative status: administration of dietary lutein is associated with an increase in song rate and circulating antioxidants (albumin and cholesterol) and a decrease in oxidative damage. PLoS One 9: e115899. https://doi.org/10.1371/journal.pone.0115899

Caves EM, Green PA, Zipple MN, Peters S, Johnsen S, Nowicki S (2018) Categorical perception of colour signals in a songbird. Nature 560: 365–367. https://doi.org/10.1038/s41586-018-0377-7

Collins SA, ten Cate C (1996) Does beak colour affect female preference in zebra finches? Anim Behav 52:105–112. https://doi.org/10.1006/anbe.1996.0156

Collins SA, Hubbard C, Houtman AM (1994) Female mate choice in the zebra finch - the effect of male beak colour and male song. Behav Ecol Sociobiol 35:21–25. https://doi.org/10.1007/BF00167055

Crozier BYWJ, Wolf E (1941) The simplex flicker threshold contour for the zebra finch. J Gen Physiol 24:625–633

de Kogel CH, Prijs HJ (1996) Effects of brood size manipulations on sexual attractiveness of offspring in the zebra finch. Anim Behav 51:699–708. https://doi.org/10.1006/anbe.1996.0073

Emery KJ, Webster MA (2019) Individual differences and their implications for color perception. Curr Opin Behav Sci 30:28–33. https://doi.org/10.1016/j.cobeha.2019.05.002

Fox DL (1979) Biochromy: natural coloration of living things. University of California Press, Berkeley

Goldsmith TH, Collins JS, Licht S (1984) The cone oil droplets of avian retinas. Vis Res 24:1661–1671. https://doi.org/10.1016/0042-6989(84)90324-9

Harnad SR (1987) Categorical perception: the groundwork of cognition. University of Cambridge Press, Cambridge

Hart NS, Vorobyev M (2005) Modelling oil droplet absorption spectra and spectral sensitivities of bird cone photoreceptors. J Comp Physiol A 191:381–392. https://doi.org/10.1007/s00359-004-0595-3

Hempel de Ibarra N, Giurfa M, Vorobyev M (2001) Discrimination of coloured patterns by honeybees through chromatic and achromatic cues. J Comp Physiol A 187:215–224. https://doi.org/10.1007/s00359-002-0322-x

Hill GE (1999) Is there an immunological cost to carotenoid-based ornamental coloration? Am Nat 154:589–595. https://doi.org/10.2307/2463839

Hill GE, Montgomerie R, Inouye CY, Dale J (1994) Influence of dietary carotenoids on plasma and plumage colour in the house finch: intra- and intersexual variation. Funct Ecol 8:343–350

Hill GE, Inouye CY, Montgomerie R (2002) Dietary carotenoids predict plumage coloration in wild house finches. Proc R Soc Lond B 269: 1119–1124. https://doi.org/10.1098/rspb.2002.1980

Jennions MD, Petrie M (1997) Variation in mate choice and mating preferences: a review of causes and consequences. Biol Rev 72: 283–327

Kelber A (2019) Bird colour vision–from cones to perception. Curr Opin Behav Sci 30:34–40. https://doi.org/10.1016/j.cobeha.2019.05.003

Knott B, Berg ML, Morgan ER, Buchanan KL, Bowmaker JK, Bennett ATD (2010) Avian retinal oil droplets: dietary manipulation of colour vision? Proc R Soc Lond B 277:953–962. https://doi.org/10.1098/rspb.2009.1805

Koch RE, Hill GE (2018) Do carotenoid-based ornaments entail resource trade-offs? An evaluation of theory and data. Funct Ecol 32:1908–1920. https://doi.org/10.1111/1365-2435.13122

Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest package: tests in linear mixed effects models. J Stat Softw 82:1–26. https://doi.org/10.18637/jss.v082.i13

Lim HH, Pike TW (2016) Dietary carotenoid availability affects avian color discrimination. Behav Ecol 27:1579–1584. https://doi.org/10.1093/beheco/arw116

Lind O (2016) Colour vision and background adaptation in a passerine bird, the zebra finch (Taeniopygia guttata). R Soc Open Sci 3: 160383. https://doi.org/10.1098/rsos.160383

Lipetz LE (1984) A new method for determining peak absorbance of dense pigment samples and its application to the cone oil droplets of Emydoidea blandingii. Vis Res 24:597–604. https://doi.org/10.1016/0042-6989(84)90114-7

Martin GR, Osorio D (2008) Vision in birds. In: Masland D, Albright RM (eds) The senses: a comprehensive reference, vol 1. Academic press, San Diego, pp 25–52

Maynard Smith J, Harper D (2003) Animal signals. Oxford University Press, Oxford

Meyer DB, Stuckey SR, Hudson RA (1971) Oil droplet carotenoids of avian cones-I. Dietary exclusion: models for biochemical and physiological studies. Comp Biochem Physiol B 40:61–64. https://doi.org/10.1016/0305-0491(71)90062-9

Olson VA, Owens IPF (1998) Costly sexual signals: are carotenoids rare, risky or required? Trends Ecol Evol 13:510–514. https://doi.org/10.1016/S0169-5347(98)01484-0

Osorio D, Vorobyev M (2005) Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and color vision. Proc R Soc Lond B 272:1745–1752

Osorio D, Jones C, Vorobyev M (1999) Accurate memory for colour but not pattern contrast in chicks. Curr Biol 9:199–202

Price TD, Stoddard MC, Shevell SK, Bloch NI (2019) Understanding how neural responses contribute to the diversity of avian colour vision. Anim Behav 155:297–305. https://doi.org/10.1016/j.anbehav.2019.05.009

R Development Core Team (2018) R: a language and environment for statistical computing (version 3.5.0). R Foundation for Statistical Computing, Vienna. https://www.R-project.org/

Ronald KL, Fernández-Juricic E, Lucas JR (2012) Taking the sensory approach: how individual differences in sensory perception can influence mate choice. Anim Behav 84:1283–1294. https://doi.org/10.1016/j.anbehav.2012.09.015

Ronald KL, Ensminger AL, Shawkey MD, Lucas JR, Fernández-Juricic E (2017) Testing a key assumption in animal communication: between-individual variation in female visual systems alters perception of male signals. Biol Open 6:1771–1783. https://doi.org/10.1242/bio.028282

Schiedt K, Bischof S, Glinzb E (1991) Recent progress on carotenoid metabolism in animals. Pure Appl Chem 63:89–100

Searcy WA, Nowicki S (2005) The evolution of animal communication. Princeton University Press, Princeton

Singmann H, Bolker B, Westfall J, Aust F (2015) afex: analysis of factorial experiments. R Package version 013–145, https://CRAN.R-project.org/package=afex

Symonds MRE, Moussalli A (2011) A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav Ecol Sociobiol 65:13–21. https://doi.org/10.1007/s00265-010-1037-6

Toomey MB, Corbo JC (2017) Evolution, development and function of vertebrate cone oil droplets. Front Neural Circuits 11:1–17. https://doi.org/10.3389/fncir.2017.00097

Toomey MB, McGraw KJ (2010) The effects of dietary carotenoid intake on carotenoid accumulation in the retina of a wild bird, the house

finch (*Carpodacus mexicanus*). Arch Biochem Biophys 504:161–168. https://doi.org/10.1016/j.abb.2010.06.033

Toomey MB, McGraw KJ (2011) The effects of dietary carotenoid supplementation and retinal carotenoid accumulation on vision-mediated foraging in the house finch. PLoS One 6:e21653. https://doi.org/10.1371/journal.pone.0021653

Toomey MB, McGraw KJ (2012) Mate choice for a male carotenoid-based ornament is linked to female dietary carotenoid intake and accumulation. BMC Evol Biol 12:3. https://doi.org/10.1186/1471-2148-12-3

Toomey MB, Butler MW, McGraw KJ (2010) Immune-system activation depletes retinal carotenoids in house finches (*Carpodacus mexicanus*). J Exp Biol 213:1709–1716. https://doi.org/10.1242/jeb.041004

von Schantz T, Bensch S, Grahn M, Hasselquist D, Wittzell H (1999) Good genes, oxidative stress and condition-dependent sexual signals. Proc R Soc Lond B 266:1–12. https://doi.org/10.1098/rspb.1999.0597

Vorobyev M (2003) Coloured oil droplets enhance colour discrimination. Proc R Soc Lond B 270:1255–1261. https://doi.org/10.1098/rspb.2003.2381

Vorobyev M, Osorio D (1998) Receptor noise as a determinant of colour thresholds. Proc R Soc Lond B 265:351–358. https://doi.org/10.1098/rspb.1998.0302

Vorobyev M, Osorio D, Bennett ATD, Marshall NJ, Cuthill IC (1998) Tetrachromacy, oil droplets and bird plumage colours. J Comp Physiol A 183:621–633

Vos DR (1995) The role of sexual imprinting for sex recognition in zebra finches: a difference between males and females. Anim Behav 50:645–653. https://doi.org/10.1016/0003-3472(95)80126-X

Wald G, Zussman H (1937) Carotenoids of the chicken retina. Nature 140:197. https://doi.org/10.1038/140197a0

Walls G (1942) The vertebrate eye and its adaptive radiation. The Cranbook Press, Bloomfield Hills

Weaver RJ, Santos ESA, Tucker AM, Wilson AE, Hill GE (2018) Carotenoid metabolism strengthens the link between feather coloration and individual quality. Nat Commun 9:73. https://doi.org/10.1038/s41467-017-02649-z

Wilby D, Roberts NW (2017) Optical influence of oil droplets on cone photoreceptor sensitivity. J Exp Biol 220:1997–2004. https://doi.org/10.1242/jeb.152918