

Experimental trait mismatches uncover specificity of evolutionary links between multiple signaling traits and their interactions in hummingbirds

Richard K. Simpson^{1,2,3}  and Kevin J. McGraw¹

¹*School of Life Sciences, Arizona State University, Tempe, Arizona 85287*

²*Department of Biological Sciences, University of Windsor, Windsor, Ontario N9B 3P4, Canada*

³*E-mail: rksimpson9@gmail.com*

Received September 16, 2018

Accepted November 19, 2018

Many animal signals co-occur, and these signals may coevolve due to their interactive properties. Previous work has demonstrated ecological drivers of evolutionary relationships between signals and the environment, which leads to questions about why specific signal pairs evolved among species that possess multiple signals. We asked whether the coloration of different species was optimized for presentation with its natural behavioral display. We investigated this in “bee” hummingbirds, where males exhibit angle-dependent structurally-colored plumage and a stereotyped courtship (shuttle) display, by experimentally creating mismatches between the behavior and plumage of five species and quantifying how these mismatches influenced male color appearance during a display. Specifically, we photographed the plumage from a given species as we moved its feathers through the position-and-orientation-specific courtship display path of other species and quantified the resulting color appearance during the display in order to compare the mismatched color appearance to each species’ natural color appearance. We found that mismatches significantly altered display flashiness (% change in coloration during displays) compared to the natural plumage-behavior pairings, and that such departures in flashiness were predicted by differences in shuttle behaviors alone. These results illustrate a tight evolutionary relationship between shuttle displays and color flashiness in these hummingbirds. Further, we found that interspecific variation in male plumage, behavior, and natural color appearance predicted deviations between natural and mismatched flashy color appearance. Altogether, our work provides a new method for testing signal coevolution and highlights the complex evolutionary relationships between multiple signals and their interactions.

KEY WORDS: Angle-dependent structural plumage, bee hummingbirds, courtship, coloration, shuttle display.

Many animals possess multiple signals that they use in communication (e.g., peacock jumping spider dances and colorful ornaments; Girard et al. 2011, 2015). There are two sets of hypotheses that are typically invoked to explain diversity in multiple signals. The first set are hypotheses focused on the information contained within the signals (Hebets and Papaj 2005), which can shed light how multiple signals coevolved with other, nonsignaling traits, such as the relationship between northern cardinal (*Cardinalis cardinalis*) bill color and body condition and breast color and offspring feeding rate (e.g., Jawor and Breitwisch 2004; Jawor

et al. 2004). The second set of hypotheses focuses on signal efficacy and how signals are optimized for transmission and perception in their particular environments (Endler 1992; Hebets and Papaj 2005). For example, various neotropical bird species tend to display or hold territories in environments that make them more detectable or contrast better against the background (Heindl and Winkler 2003; Gomez et al. 2004; Endler et al. 2014; Simpson and McGraw 2018b). Although these studies have revealed important evolutionary links between environmental/nonsignaling traits and animal signals, questions remain about how and why certain

signals coevolve together (or antagonistically). Multiple signals are often used together, and this can lead to signal interactions, such as one signal making the other more detectable/discriminable or two signals interacting to produce a new, emergent signal (Hebets and Papaj 2005). By studying the driving mechanisms that lead to variation in these signal interactions, we can better understand the specific coevolutionary coupling between signals and their interactions.

Signaling traits can interact with each other and the environment to produce certain, quantifiable, interactive properties (e.g., emergent or composite signals; Hutton et al. 2015). For example, a butterfly's wing beats, angle-dependent colorful scales, and position relative to the sun can produce a strobe-like, flashy color-display that is not possible without trait interactions (White et al. 2015; also see peacocks *Pavo cristatus*; Dakin and Montgomerie 2013). Other animals, like some hummingbirds, position and orient their colorful ornaments during a display relative to the sun in a way that produces a consistent color-display (Simpson and McGraw 2018c), and the consistent color appearances are not solely the result of the ornament being a specific brightness/coloration but are the result of the interactions between colorful plumage, behaviors, and the environment (Simpson and McGraw 2018a). Signal interactions do not just vary among distantly related species (e.g., butterflies vs. hummingbirds) but can also vary among closely related species, and these signal interactions can coevolve with the original signals themselves. For example, hummingbird flashiness during a display positively covaries with courtship display properties (e.g., display width, velocity), whereas how bright and colorful a species appears during a display positively covaries with plumage properties (i.e., feather brightness; Simpson 2018). These results lead to questions about how optimized a hummingbird species' plumage is for its particular behavioral display (or vice versa) and how optimized one or both signals are to produce specific signal interactions. There is evidence that selection does favor specific trait pairings (e.g., evolution of hummingbird bill shape to flower shape; Snow and Snow 1980; Stiles 1981; Smith et al. 1996), including for specific forms of communication (e.g., the evolution of birdsong to match a species' habitat; Tobias et al. 2010). If we can disrupt the pairing between two different signals in a species (i.e., have species A perform a behavioral display with the plumage of a species B), we can test how tightly or diffusely these signal pairings have coevolved, based on if/how the interaction properties change relative to how they appear naturally in each species.

The evolutionary linkage between two coexpressed signaling traits can be disrupted, both naturally and experimentally. One example of a natural disruption between two signals is found within *Sceloporus* lizards, where most species within the genus possess a colorful underbelly and exhibit push-up/head-bob displays, which they use to communicate to rival males (Martins et al. 2015).

However, some species have lost their underbelly coloration, and this has led to changes in their behaviors, such that males of the species without underbelly coloration exhibit more behavioral displays (Martins et al. 2015). While *Sceloporus* lizards provide an example of a natural disruption of the links between multiple signals, experimental disruptions of signal linkages, especially signals with known and quantifiable signal interaction properties, may also provide key insights into the coevolved links and interactions between multiple signals.

Our aim in this study was to experimentally disrupt the evolved linkage between courtship displays and angle-dependent structural coloration in a monophyletic clade of North American "bee" hummingbirds (Tribe Mellisugini, McGuire et al. 2014), by creating signal mismatches (i.e., plumage/behavior mismatches: running the feathers of one species through the courtship display of different species; Fig. 1; see more below) and quantifying if/how these mismatches alter the signal interaction properties previously studied in this system (Simpson 2018). Previous work on *Schizocosa* wolf spiders tested the effects of signal alterations, or mismatches, on female choice, and found that the addition of leg ornaments on males did not affect female choice for species that do not naturally have visual components to their courtship displays (Hebets and Uetz 2000). On the other hand adding or removing leg ornaments increased or decreased, respectively, female receptivity in species that did have visual components to their displays (Hebets and Uetz 2000). However, in this study we aim to test the effect of mismatches on signal interaction properties, and we are not completely removing or adding signal properties, as in the spider work, but expanding or reducing male signals based on natural, interspecific variation within the clade. These mismatches allow us to quantify the robustness of each evolved signal combination (plumage-behavior-appearance), providing a tool to measure the strength of the evolutionary pairing between those signals.

Hummingbirds from the "bee" clade vary in their striking angle-dependent head/throat coloration and a stereotyped courtship display: the shuttle display (Simpson 2018). Shuttles are characterized by a male flying rapidly back and forth in front of a female while erecting his colorful plumage (Feo and Clark 2010; Clark et al. 2012, 2013, Simpson and McGraw 2018a, c). In our previous work, we tested the correlational relationships between properties of male shuttles, plumage, display environment, and their interactions to begin to understand how these signals coevolved (Simpson 2018). To better understand the evolution of multiple signals and their interactions, we aimed to disrupt the evolved links (i.e., species-specific signal forms) between these signals and test how this disruption altered the signal interaction properties (i.e., male color appearance during displays). Previously, we conducted shuttle-display reconstructions by moving plucked feathers from a given species through that species'

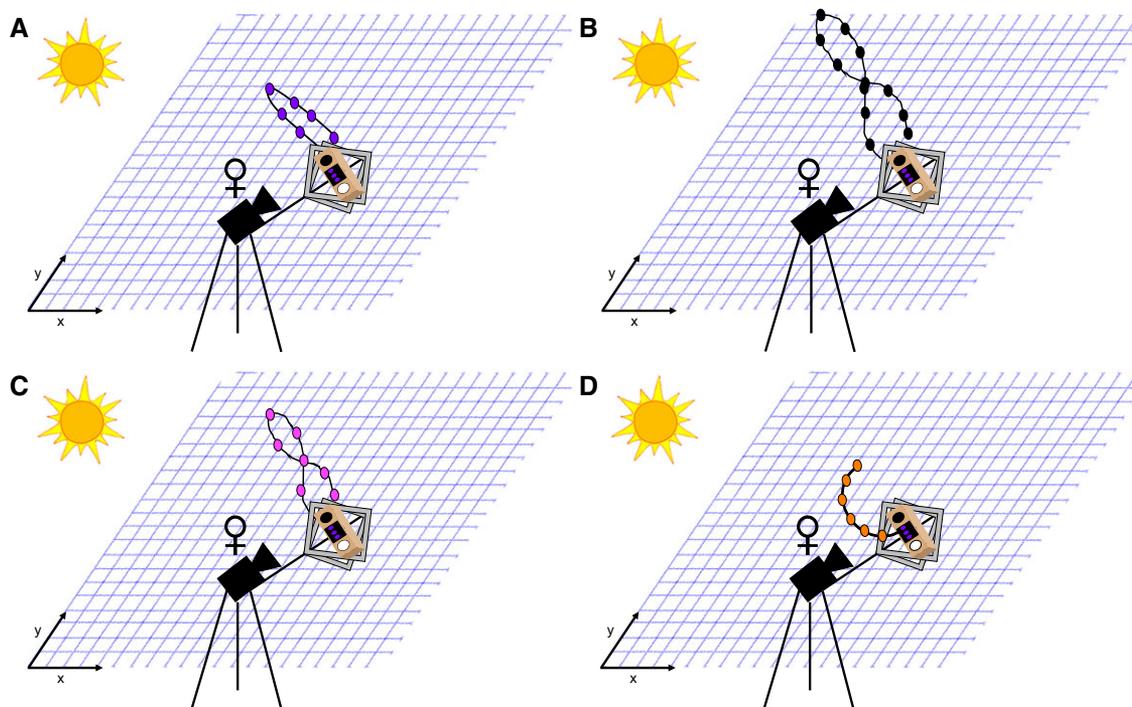


Figure 1. A visual representation of an original, species-specific shuttle re-creation, in this instance a Costa's hummingbird (A), and three plumage-behavior mismatches (B–D). In each figure, the camera on the tripod represents the modified DSLR camera we used to photograph male feathers from the female's point of view (also indicated by the female symbol). The wood block on gray squares represents our lazy-Susan apparatus, where we could move the feathers around the camera to simulate male display movements, and also orient the feathers to simulate male angles of orientation relative to the female. The grid represents the horizontal plane in which our measurements and re-creations are done (see Simpson 2018 for further details). In each panel, the lines represent the shuttle display path and the points indicate the representative points where we photographed the feathers. For each re-creation, we start the display based on the starting location of this male Costa's display, keeping the starting position of each display relative to the sun constant. (A) Represents a re-creation of a Costa's hummingbird shuttle, with that Costa's male's feathers. (B) Represents a re-creation of a black-chinned hummingbird shuttle with the same Costa's male's feathers. (C) Represents a re-creation of a broad-tailed hummingbird shuttle with the Costa's feathers. (D) Represents a re-creation of an Allen's hummingbird shuttle with the Costa's feathers. Note: display paths not accurately scaled by size, and the Calliope hummingbird path was not included for an even number of panels.

display, to quantify male *color appearance while shuttling*, or how a male's throat feathers appear to females as they display (Table 1; Simpson 2018; Simpson and McGraw 2018a). However, here we took colorful feathers from each of five “bee” hummingbird species and also moved them through the shuttle-display paths of the other four species, to quantify male color appearance during these mismatched color-displays (i.e., *mismatched color appearance*; Table 1; Fig. 1). With these plumage/behavior mismatches, we can test if/how species-specific, or *natural, color appearances* (Table 1; see Fig. S1 for examples) during a display are different from the mismatched color appearances, such that the magnitude of the deviations between the two color appearances represents the degree of specificity of the evolved signal pairing.

In general, we predicted that plumage-behavior mismatches would cause large departures in each species' natural color appearance—both in terms of color flashiness (changes in color throughout the display) and average brightness/coloration during

the display. Specifically, we predicted that the greater the difference between properties of the plumage patch (e.g., patch size, feather reflectance), courtship-display (e.g., shuttle width, speed), and/or display orientation to sun of the focal and heterospecific mismatched species, the greater the deviation in color appearance (i.e., the difference between a species' natural color appearance and mismatched color appearance; Table 1). Based on our current understanding of the coevolution between properties of signals and their signal interactions in these hummingbirds (Simpson 2018), we also sought to determine the specific properties of the signals themselves that might be driving interspecific deviations in color appearance. For example, because plumage patch size and feather reflectance positively covary with how bright and colorful a male appears on average during a display (Simpson 2018), we predicted that species with larger plumage patches and/or more colorful feathers would have greater deviations in average color appearance during mismatches. We also predicted that species

Table 1. A list of color variables and mismatch methods used throughout the text.

Term	Definition
Feather reflectance	The inherent reflectance properties (i.e. brightness, hue, chroma) of male feathers when measured under controlled laboratory conditions (i.e. outside of the display context).
Color appearance during a display	The appearance of male feathers from the perspective of a female as he performs his shuttle display in his natural environment. Also referred to as the <i>signal interaction property</i> .
Natural color appearance	The color appearance of a given male's feathers when run through his own, species-specific shuttle display.
Mis-matched color appearance	The color appearance of a given male's feathers when run through a different species' shuttle display OR the color appearance during a given male's shuttle display when using a different species' feathers.
% change in color appearance	The summed, absolute percent change in coloration (i.e. brightness, hue, chroma) between the color variables for each position in a re-created shuttle display. Also referred to as <i>flashiness</i> outside of the methods.
Average color appearance	The average coloration (i.e. brightness, hue, chroma) among the color variables for each position in a re-created shuttle display.
Plumage-behavior mis-match	A re-created shuttle display using feathers of the focal species but displays of every other species. For example, taking Costa's hummingbird feathers and running them through an Allen's, black-chinned, broad-tailed, and Calliope hummingbird shuttle display.
Behavior-plumage mis-match	A re-created shuttle display using shuttle display of the focal species but feathers from every other species. For example, taking feathers from an Allen's, black-chinned, board-tailed, and Calliope hummingbird and running them through a Costa's hummingbird shuttle display.
Deviation in color appearance	The absolute difference between a natural color appearance and a mismatched color appearance. For example, the difference in % color appearance between an Allen's feathers run through an Allen's display and an Allen's feathers run through a Costa's display.

When the terms are first presented in the text, they are italicized.

with more exaggerated display movements (e.g., wider shuttles) should exhibit greater deviations in color appearance during mismatches, because of the positive covariance between shuttle behaviors and flashiness (Simpson 2018). Finally, we predicted that species that normally appeared flashier and/or more colorful during their displays would exhibit greater deviations in their color appearance during mismatches.

Material and Methods

FIELD SHUTTLE AND PLUMAGE DATA COLLECTION

For this study, we used previously recorded shuttle displays (from 2014–2017) and plucked feathers for each hummingbird species (Simpson 2018; Simpson and McGraw 2018a, c): broad-tailed (*Selasphorus platycercus*) and black-chinned (*Archilochus alexandri*) hummingbirds, which were studied in Arizona, and Costa's (*Calypte costae*), Calliope (*Selasphorus calliope*), and Allen's (*Selasphorus sasin*) hummingbirds, which were studied in California (see Simpson 2018 for location and permit details). Anna's hummingbirds, which were included in our previous work (Simpson 2018), were excluded from this study due to the fact that they do not move during their courtship display, making it difficult to mismatch their displays with other species. Captured female hummingbirds from each species were used to elicit male

shuttle displays from their own species in the field (Simpson 2018; Simpson and McGraw 2018a, c). We presented a female in a wire-mesh cylindrical cage with a clear plexiglass bottom (see Simpson 2018 for full description of cage set up) to males on their territories, and then we placed a high-definition video camera (Sony HDR-CX330) underneath the caged female, pointing up, which enabled us to film male shuttle orientations and movements in the horizontal plane (Simpson and McGraw 2018a). After we filmed male shuttles, we captured the males that displayed on their territories and plucked feathers from their colorful throat patch (gorget), specifically from the area underneath their bill (Simpson and McGraw 2018a). We also quantified male plumage patch size by photographing each side of the male's head (before we plucked feathers, area in mm²; Simpson and McGraw 2018a).

QUANTIFYING MALE SHUTTLE DISPLAYS

We used the quantified shuttle display data from our previous work (Simpson 2018) to re-create male displays in the field. Briefly, we measured the *x-y* coordinates of each male throughout his display path, tracking the position of his gorget relative to the female. We then used these coordinates to calculate an average shuttle cycle, from which we calculated shuttle width (cm, distance between the turn-around point and starting point) and average translational

velocity (cm/s). We also quantified how males oriented themselves toward the female during shuttles by measuring the angle between the center of a male's gorget and the female's head at representative points (see Simpson and McGraw 2018a for additional details). We then calculated an average angle of orientation during a display, and the variance (standard deviation) in male angle of orientation relative to the female during a display. Finally, based on the direction of compass north, the female's location, and the time/date of each display, we calculated the solar azimuth and elevation for each display (Hoffmann 2017). We then quantified each male's display location relative to the solar azimuth and female (see Simpson and McGraw 2018a for additional details).

PLUMAGE REFLECTANCE AND ANGLE-DEPENDENCE MEASUREMENTS

To determine the color properties of feathers outside of the context of the display (in other words, the *feathers' inherent reflectance* when measured under controlled conditions; Table 1; Fig. S2), we used UV-Vis spectrometry to quantify the reflectance and angle-dependence of each male's plucked gorget feathers (Simpson and McGraw 2018a). These reflectance measurements were taken from individual feathers following the methods of Meadows et al. (2011). From these reflectance measurements, we used an ultraviolet (UV) sensitive avian visual model (Vorobyev and Osorio 1998; Herrera et al. 2008) to calculate standard tetrachromatic color variables (Stoddard and Prum 2008) in R using the *pavo* package (Maia et al. 2013). Specifically, we calculated hue theta (i.e., is the hue blue, yellow, etc.; hereafter, "red-green-blue" or "RGB hue"), hue phi (i.e., the degree of UV reflectance; hereafter, "UV hue"), chroma (i.e., the purity of the color; r-achieved in *pavo*; Stoddard and Prum 2008; Maia et al. 2013), and luminance (i.e., brightness). We also tilted the feathers from 0° to 20° to the side and measured the feather reflectance in 5° increments. From these tilted reflectance measures, we measured the slope between all angles for each tetrachromatic color variable, as our measures of feather angle dependence (Fig. S2; Simpson and McGraw 2018a).

DISPLAY RE-CREATIONS, PLUMAGE-DISPLAY MISMATCHES, AND QUANTIFYING MALE COLOR APPEARANCE

In our previous work, we determined a male's natural color appearance during his display by taking a male's plucked feathers into the field, positioning them relative to the sun as that male had positioned himself while shuttling, moving the feathers through that male's re-created shuttle display, and photographing the feathers to acquire our color-appearance metrics (Fig. S1; Simpson 2018; Simpson and McGraw 2018a). For this study, we re-created *plumage-display mismatches* (Table 1) in the field by moving feathers from each of the hummingbird species through

the average shuttle display paths for the other species (Fig. 1). For example, we moved the six gorget feathers that we plucked from a Costa's hummingbird through the quantified average shuttle cycles of an individual male Allen's, black-chinned, broad-tailed, and Calliope hummingbird (Fig. 1). Because we were focused on how plumage-behavior mismatches would affect male color appearance, we re-created each mismatch in similar environmental (lighting) conditions of the nonmismatched species (in the example above – all Costa's display mismatches occurred under Costa's environmental conditions). Further, we kept the mismatched displays positioned relative to the sun based on the original display (Fig. 1). By re-creating the mismatches under the same environmental conditions of the original species, we could eliminate any added effects of species/environment mismatches to better determine the evolutionary link between the plumage and behavioral signals themselves and their interactions. We randomly selected (using a randomly generated sequence from random.org, which used an atmospheric noise model without replacements) feathers/behaviors from five individuals per species for these mismatches. We were able to pluck feathers from only two male Allen's and four male black-chinned hummingbirds, but we had multiple displays from each individual, so we were able to use a unique feather/behavior combination for our mismatches (i.e., a different Allen's/black-chinned display path per mismatch replicate). Unfortunately, we only filmed three displays from two Calliope male hummingbirds, so we randomly (see above) selected which two of the three shuttles would be used twice in order to achieve five mismatches with this species. Overall, this resulted in 20 plumage/behavior mismatch sets and a total of 100 re-created mismatches ($n = 5$ species, 4 mismatches/species, 5 replicates/mismatch).

During each of these re-created mismatches, we followed the exact same methods as we did previously to determine color appearance of male feathers during his shuttle display (i.e., the within-species matches), by moving plucked feathers through re-created shuttles and then photographing the feathers from the female's point of view at each representative position, using a full-spectrum DSLR camera (Canon 7D with a quartz sensor) equipped with an El Nikkor 80 mm enlarging lens and two Bradaar light filters (Stevens et al. 2007; Troscianko and Stevens 2015) attached to a lazy-Susan apparatus (described in Simpson 2018; Simpson and McGraw 2018a, c). For each position, we calculated RGB hue, UV hue, and chroma using the relative cone stimulation values from each multispectral photo through *pavo* (Maia et al. 2013), and calculated luminance from the double-cone stimulation values from each photo using the Multispectral Image Calibration and Analysis Toolbox (Troscianko and Stevens 2015) in ImageJ (Schneider et al. 2012). Finally, as our measures of trait interactions (i.e., degree of match/mismatch between plumage and behavior), we used the tetrachromatic color

variables taken from the multispectral photos to calculate the *average color appearance* (i.e., how males appeared on average during their display; Table 1) and *% change in color appearance* during a display (i.e., how male coloration changes as he displays; Table 1) for each mismatch (Simpson and McGraw 2018c).

STATISTICAL METHODS

Testing the effects of plumage/behavior mismatches on color appearance

All statistical analyses were conducted in R (R Development Core Team 2017). To initially test if the plumage-display mismatches caused divergences in male color appearance from the species-appropriate matches, we compared the average natural color appearance to the mismatched color appearances for each species. First, using principle components analysis (PCA), we collapsed the % change in color appearance variables (luminance, chroma, RGB hue, and UV hue; $n = 20/\text{variable}$) into a single principle component (PC): “% change in color appearance PC,” with higher values indicating males that had higher percent changes in luminance, chroma, RGB hue, and UV hue color appearance during a display (Table S1). We also collapsed the variables for how males appeared on average during a display into a single PC: “average color appearance PC,” with higher values indicating males that appeared brighter, more chromatic, more red-shifted, and exhibited lower UV reflectance during a display (Table S1).

We used analysis of variance (ANOVA) to test whether the color appearance from each mismatch resulted in a departure from the natural (i.e., species-specific) color appearance within each species (see example below; and see Simpson 2018 for samples sizes per species for natural color appearances). We then used Tukey-Kramer post-hoc tests to evaluate the significant differences in male color appearance due to mismatches with different species. For example, we tested whether the color appearance of an Allen’s hummingbird was different than the color appearance from mismatches of: (1) Allen’s feathers with black-chinned shuttle, (2) Allen’s feathers with broad-tailed shuttle, (3) Allen’s feathers with Calliope shuttle, and (4) Allen’s feathers with Costa’s shuttle (Fig. 1). Then, we reanalyzed those data to simulate *behavior-plumage mismatches* (where the focal species’ shuttle is used with other species’ feathers; Table 1) and to test whether the color appearance of, for example, an Allen’s hummingbird was different than the color appearance from mismatches of: (1) Allen’s shuttle with black-chinned feathers, (2) Allen’s shuttle with broad-tailed feathers, (3) Allen’s shuttle with Calliope feathers, and (4) Allen’s shuttle with Costa’s feathers. This dual analysis allowed us to assess whether mismatching the display relative to the plumage or the plumage relative to the display (or both) can cause departures in color appearance.

For the purposes of this study, we ignored all post-hoc results between different mismatch groups (i.e., Allen’s:Costa’s

vs. Allen’s:black-chinned) and only focused on whether each mismatch differed from the one species-appropriate match. Through these analyses, we found that plumage-behavior mismatches (i.e., using plumage of focal species for mismatch; Table 1) had little to no effect on average color appearance (see Results for more details). Further, we found that behavior-plumage mismatches (i.e., using shuttle of focal species for mismatch; Table 1) had no effect on male % change in color appearance and only affected average color appearance during a display as predicted by interspecific differences in feather color (e.g., orange vs. purple feathers; see Results for more details). Thus, for the remainder of our analyses, we focused only on further understanding what about mismatches of plumage relative to the display (plumage-display mismatches) might be driving the deviations in % change in color appearance during a display.

Testing if/how differences in signal properties between focal and mismatched species predict deviations in color appearance

Next, we aimed to understand what about these mismatches might be driving the deviations in % change in color during a display, such as differences in shuttle display behavior, plumage, or male display position relative to the sun between the focal species and mismatched species. We first averaged the % change in color appearance variables from the five individual re-creations per mismatch pair per species. For example, we calculated single average % change in luminance during a display for the (1) Allen’s-black-chinned mismatch, (2) Allen’s-broad-tailed mismatch, (3) Allen’s-Calliope mismatch, and (4) Allen’s-Costa’s mismatch (thus $n = 4$ per species). We then took each of these average % change in color appearance mismatch variables (luminance, chroma, RGB hue, and UV hue) for a species and found the absolute difference between them and the natural % change in color appearance of that species to create a *deviation in % color change* (Table 1) for each color variable. We also calculated the absolute difference in shuttle-behavior, plumage, and solar-environment properties for each mismatch (e.g., |Allen’s shuttle width—black-chinned shuttle width|).

We collapsed (separate PCA from above) the deviation in male color change variables into a single PC: “deviation in % color change PC,” with higher values indicating larger differences between the focal species’ and mismatched species’ variables for % change in luminance, chroma, RGB hue, and UV hue appearance during a display (Table S2). We also collapsed differences in the three shuttle display properties between the focal and mismatch species (width, velocity, and variation in angles of orientation) into a single PC: “differences in shuttle behavior PC,” with higher values indicating larger differences between the focal species’ and mismatched species’ shuttle behaviors (Table S2). Additionally, we collapsed differences in feather reflectance between the focal

and mismatched species into two PCs: (1) “differences in feather hue PC,” with higher values indicating larger differences between the focal species’ and mismatched species’ feather RGB/UV hue (Table S2); and (2) “differences in feather luminance/chroma PC,” with higher values indicating larger differences between the focal species’ and mismatched species’ feather luminance/chroma (Table S2). Finally, we collapsed differences in feather angle-dependence between the focal and mismatched species into two PCs: (1) “differences in hue angle-dependence PC,” with higher values indicating larger differences between the focal species’ and mismatched species’ angle-dependence in RGB/UV hue (Table S2); and (2) “differences in luminance/chroma angle-dependence,” with higher values indicating larger differences between the focal species’ and mismatched species’ angle-dependence in luminance/chroma (Table S2).

We used an information-theoretic, model-averaging approach to test whether differences between the focal species’ and mismatched species’ signal properties explained variation in the deviations in % color change. We built a global-mixed linear model, with deviations in % color change PC as our response variable, species as a random effect (as these mismatch data are not necessarily species-specific data points but data on differences between species), and differences in plumage patch size, feather hue PC, feather luminance/chroma PC, hue angle-dependence PC, luminance/chroma angle-dependence PC, shuttle behavior PC, average male orientation angle toward the female, male display position relative to the sun, and solar elevation between the focal and mismatched species as our fixed effects. Then we used Akaike weights from the global and all subsequent models to calculate the summed weight for each fixed effect, or relative importance (RI), while also calculating the average beta value for each fixed effect. While RI values are informative, they can be misleading (Galipaud et al. 2014), so we created a final-mixed linear model, with deviations in % color change PC as our dependent variable, using fixed effects that had a RI greater than 0.5 (Simpson and McGraw 2018a). We only interpreted fixed effects that were significant in this final model.

Testing if/how natural signal variation predicts deviations in color appearance

Lastly, we tested the prediction that interspecific variation in signal properties (i.e., shuttle behavior, plumage size, natural color appearance) predicts variation in deviations in % color change due to mismatches. For example, we predicted that species with wider shuttles will exhibit greater deviations between their natural and mismatched % change in color appearance. Thus, we calculated a species-average for each courtship trait and for their deviations in % color change due to mismatches. However, we only tested whether plumage patch properties, shuttle properties, and natural color appearances predicted interspecific variation in deviations

in % color change, as these variables were previously found to strongly, evolutionarily covary (Simpson 2018).

We were unable to neatly collapse species-average deviations in % color change variables into PCs, so we retained them as individual variables in analyses. We did, however, collapse species-average natural % change in color appearance and average male color appearance during a display into two single PCs (in a separate PCA than above): (1) “% change in color appearance PC” and (2) “average color appearance PC,” both of which had the same relationships as in *Statistical methods: section 1*. (Table S3). We collapsed shuttle velocity, shuttle width, and variation in male orientation toward the female into a single PC: “shuttle behavior PC,” with higher values indicating species with faster and wider shuttles and more variation in how they oriented toward the female during a display (Table S3). We collapsed species-average feather reflectance variables into a single PC: “feather reflectance PC” with higher values indicating species with brighter, more chromatic, more red-shifted, and more UV reflecting feathers (Table S3). Finally, we collapsed species-average feather angle-dependence variables into two PCs: (1) “angle-dependence PC1,” with higher values indicating species with more angle-dependence in chroma, RGB hue, and UV hue; and (2) “angle-dependence PC2,” with higher values indicating species with more angle-dependence in chroma and luminance (Table S3).

We then created univariate models for each independent variable explaining each of the four species-average deviations in % color change variables separately ($n = 20$ models). We calculated and evaluated the AIC values and weights for each model, per color variable (i.e., five for luminance, five for chroma, etc.), to test which (if any) variables best explained variation in each deviation in % color change variable.

Results

EFFECTS OF PLUMAGE-SHUTTLE MIS-MATCHES ON COLOR APPEARANCE

We found significant differences in the % change in color appearance PC between species-specific and plumage-behavior mismatched pairings (i.e., different shuttle for the same feathers) for all five species: Allen’s (ANOVA: $F_{4,21} = 7.8$, $P < 0.001$), black-chinned (ANOVA: $F_{4,24} = 17.9$, $P < 0.001$), broad-tailed (ANOVA: $F_{4,29} = 6.1$, $P = 0.001$), Calliope (ANOVA: $F_{4,18} = 11.6$, $P < 0.001$), and Costa’s (ANOVA: $F_{4,48} = 14.4$, $P < 0.001$). Through Tukey-Kramer post-hoc tests on black-chinned hummingbird mismatch comparisons, we found that the natural flashiness (i.e., % change in color appearance PC) of displaying male black-chinned hummingbirds was significantly higher than that for all the mismatch groups (Table 2, Fig. 2B). In the post-hoc tests on Allen’s hummingbird mismatches, we only found that flashiness during a display was significantly

Table 2. Results from Tukey-Kramer post-hoc tests, testing whether species-specific % change in color appearance was significantly different during a plumage-behavior mismatch (i.e., mismatched behavior relative to plumage).

Focal species (feathers that were used)	Mismatched species' display	Difference between means	<i>P</i> -value
Allen's	Black-chinned	3.91	<0.01
Allen's	Broad-tailed	2.14	0.09
Allen's	Calliope	0.66	0.92
Allen's	Costa's	0.35	0.99
Black-chinned	Allen's	3.16	<0.01
Black-chinned	Broad-tailed	2.05	0.01
Black-chinned	Calliope	2.87	<0.01
Black-chinned	Costa's	2.83	<0.01
Broad-tailed	Allen's	2.10	0.05
Broad-tailed	Black-chinned	1.51	0.24
Broad-tailed	Calliope	1.24	0.43
Broad-tailed	Costa's	1.65	0.17
Calliope	Allen's	0.70	0.80
Calliope	Black-chinned	2.81	<0.01
Calliope	Broad-tailed	1.54	0.15
Calliope	Costa's	0.39	0.97
Costa's	Allen's	0.83	0.60
Costa's	Black-chinned	3.55	<0.01
Costa's	Broad-tailed	1.33	0.16
Costa's	Calliope	0.50	0.91

The names of the species of the feathers used for the mismatch (and the natural % change in color) are in the first column, followed by the names of the species of the display used in the mismatch, and the *P*-value in the final column. Significant effects are in bold.

higher during a black-chinned shuttle, compared to the natural flashiness (Table 2, Fig. 2A). From post-hoc tests with broad-tailed hummingbird mismatches, we found significantly greater natural flashiness compared to the flashiness of a broad-tailed's feathers during an Allen's shuttle, but no significant differences for other mismatches (Table 2, Fig. 2C). Lastly, from Costa's and Calliope hummingbird mismatch post-hoc tests, we found that the natural flashiness was significantly lower than the flashiness of Costa's/Calliope feathers during a black-chinned shuttle (Table 2, Fig. 2D, E). Thus, we found species that naturally tended to have flashier color appearances during displays (black-chinned and broad-tailed) had their flashiness reduced by plumage-behavior mismatches, while species that naturally tended to have more consistent color appearances (Allen's, Costa's, and Calliope), had increased flashiness during plumage-behavior mismatches. We found no significant differences between species-specific and plumage-behavior mismatched groups in average color appearance PC for any species (ANOVAs; Table S4).

We found significant differences in average color appearance PC between species-specific and behavior-plumage mismatched pairings (i.e., different feathers for the same shuttle) for all five species: Allen's (ANOVA: $F_{4,21} = 28.9$, $P < 0.001$), black-chinned (ANOVA: $F_{4,24} = 31.4$, $P < 0.001$), broad-tailed (ANOVA: $F_{4,29} = 26.9$, $P = 0.001$), Calliope (ANOVA:

$F_{4,18} = 27.6$, $P < 0.001$), and Costa's (ANOVA: $F_{4,48} = 23.8$, $P < 0.001$). Using Tukey-Kramer post-hoc tests on Allen's hummingbird mismatch comparisons we found that the natural average color appearance of displaying Allen's hummingbird males was significantly different than that for all the mismatch groups (Table 3, Fig. 2F). In post-hoc tests on black-chinned hummingbird mismatches, we found that average color appearance during a display was significantly different during an Allen's, broad-tailed, and Calliope hummingbird's shuttle, compared to the natural average color appearance (Table 3, Fig. 2G). With the broad-tailed hummingbird mismatch comparisons, we found significantly different natural average color appearance compared to the average appearance of an Allen's, black-chinned, and Costa's hummingbirds' feathers during a broad-tailed hummingbird shuttle (Table 3, Fig. 2H). For Calliope hummingbird mismatch comparison tests, their natural average appearance was significantly different than Allen's, black-chinned, and Costa's hummingbird's feathers during a Calliope hummingbird shuttle (Table 3, Fig. 2I). Lastly, for the post-hoc tests on Costa's hummingbird mismatches, their natural average appearance was significantly different than Allen's and broad-tailed hummingbird's feathers during a Costa's hummingbird shuttle (Table 3, Fig. 2J). We found no significant differences between species-specific and behavior-plumage mismatched groups in % change in color PC for any species

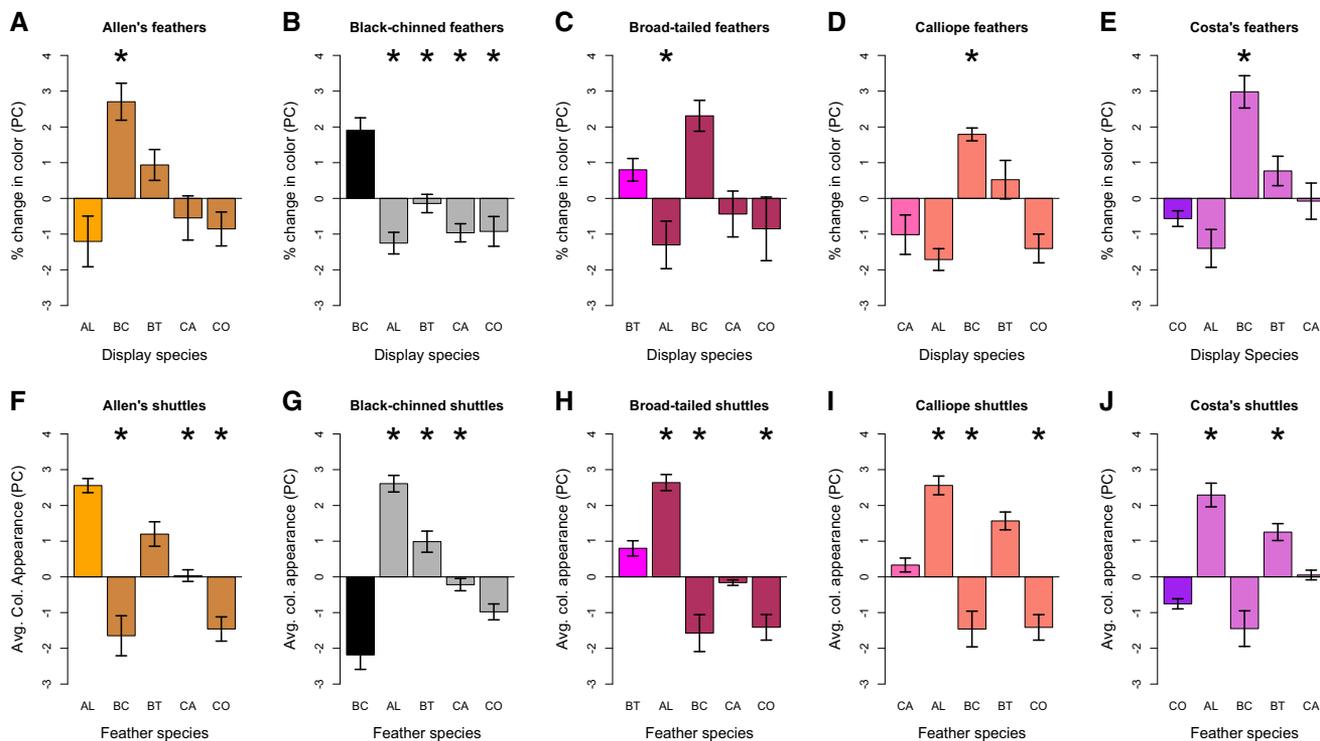


Figure 2. (A–E) Plumage-shuttle mismatches (i.e., mismatched behavior relative to plumage) significantly affected natural male color appearance during a display—specifically % change in male coloration, or flashiness. (F–J) Shuttle-plumage mismatches (i.e., mismatched plumage relative to behavior) significantly affected natural male color appearance during a display—specifically how males appeared on average. In each barplot, the appearance during a natural plumage-shuttle combination is the first column, while the other columns represent the mismatches. The species whose feathers/display were used is indicated in the title, and the species whose display/feather was used for each re-creation is along the x-axis (AL = Allen's, BC = black-chinned, BT = broad-tailed, CA = Calliope, and CO = Costa's). Asterisk above a bar indicate a significant difference between the natural and mismatched color appearances. Error bars represent standard errors.

(ANOVAs; Table S5). In sum, we found that average color appearance during a display varied during behavior-plumage mismatches as predicted by each species' feather reflectance properties. For example, a species with naturally purple feathers (e.g., Costa's hummingbird) had a different color appearance during its shuttle when a species with naturally orange feathers (e.g., Allen's hummingbird) were used. However, when another species with naturally purple feathers (e.g., black-chinned hummingbird) was used in a Costa's hummingbird display, the mismatched average color appearance was similar to the natural average color appearance.

SIGNAL DIFFERENCES PREDICTING COLOR APPEARANCE DEVIATIONS

We found that the best predictor of deviations in % color change PC was differences in shuttle behavior PC between focal and mismatched species (RI = 0.99; Fig. 3), and differences in shuttle behavior PC was significant in the final model ($mR^2 = 0.85$; Table 4). Thus, we supported our prediction that greater differences between species-specific and mismatched shuttles (i.e., width and velocity and variation in male orientation angle toward

the female during a display) led to greater deviations between natural and plumage-behavior mismatched flashiness (Table 4). However, we did not find support for our other predictions about how differences in plumage or environmental properties between focal and mismatched species predict deviations in % color change PC with an RI greater than 0.5 (Fig. 3).

SPECIES-AVERAGE SIGNALS PREDICTING VARIATION IN COLOR APPEARANCE DEVIATIONS

We found that the best univariate model explaining variation in species-average deviation in % change in luminance was plumage patch size (weight = 0.66; Table 5), such that species with larger plumage patches had greater deviations between their natural and mismatched flashiness in luminance (Fig. 4A). We found that the best model explaining deviation in % change in chroma was flashiness during a display (% change in color PC; weight = 0.86; Table 5). Specifically, species that naturally appeared flashier during their displays had greater deviations between their natural and mismatched flashiness in chroma (Fig. 4B). Additionally, we found that the best model explaining variation in deviation

Table 3. Results from Tukey-Kramer post-hoc tests, testing whether species-specific average color appearance was significantly different during a behavior-plumage mismatch (i.e., mismatched plumage relative to behavior).

Focal species (display that was used)	Mismatched species' feathers	Difference between means	P-value
Allen's	Black-chinned	4.20	<0.01
Allen's	Broad-tailed	1.36	0.06
Allen's	Calliope	2.52	<0.01
Allen's	Costa's	4.02	<0.01
Black-chinned	Allen's	4.80	<0.01
Black-chinned	Broad-tailed	3.17	<0.01
Black-chinned	Calliope	1.97	<0.01
Black-chinned	Costa's	1.21	0.10
Broad-tailed	Allen's	1.84	<0.01
Broad-tailed	Black-chinned	2.37	<0.01
Broad-tailed	Calliope	0.96	0.15
Broad-tailed	Costa's	2.21	<0.01
Calliope	Allen's	2.23	<0.01
Calliope	Black-chinned	1.79	0.03
Calliope	Broad-tailed	1.23	0.22
Calliope	Costa's	1.75	0.04
Costa's	Allen's	3.05	<0.01
Costa's	Black-chinned	0.69	0.37
Costa's	Broad-tailed	2.01	<0.01
Costa's	Calliope	0.81	0.23

The names of the species of the displays used for the mismatch (and the natural average color appearance) are in the first column, followed by the names of the species of the feather used in the mismatch, and the *P*-value in the final column. Significant effects are in bold.

Table 4. Results from final linear-mixed model, after conducting an information-theoretic, model-averaging approach, containing fixed effects with a relative importance (RI) of 0.5 or greater, in which we tested the influence of differences in male shuttling behaviors between focal and mismatched species on deviations natural % change in male color appearance during courtship displays.

Response Variable	Fixed Effects (RI)	Estimate	Std. Err.	<i>t</i> -value	<i>P</i> -value
Dev. In % Change in Color PC	Intercept	<0.01	0.14	<0.01	1.00
$R^2_m = 0.85$	Shuttle behavior PC (0.99)	0.89	0.08	10.55	<0.01

Marginal R^2 value are listed below the response variable, which measures the variation explained by the fixed effects in each model. The relative importance of the fixed effect, calculated from model-averaging (see methods), is given next to the effect. Significant effects are in bold.

in % change in RGB hue was shuttle behavior PC (weight = 0.54; Table 5), with species that have more exaggerated shuttles having greater deviations between their natural and mismatched flashiness in RGB hue (Fig. 4C). Finally, we found that the best model explaining variation in deviation in % change in UV hue was feather reflectance PC (weight = 0.91; Table 5). Species with more exaggerated feather reflectance had greater deviations between their natural and mismatched flashiness in UV hue (Fig. 4D). Overall, we supported our predictions that interspecific variation in courtship signals predict how much plumage-behavior mismatches affect species-specific color appearance during a display.

Discussion

We employed a novel experimental signal-mismatch procedure in attempt to understand the coevolved links between two signaling traits (ornamental plumage and courtship behavior) and their interactions in several hummingbird species from the "bee" hummingbird tribe. We have previously shown that hummingbird plumage, shuttle behaviors, and display orientation relative to the sun can interact in complex ways to produce different male color appearances during courtship (Simpson and McGraw 2018a, c). Further, we have evidence that these signal interaction properties do coevolve with the signals that interact to produce them (Simpson 2018), as opposed to evolving independently, as some

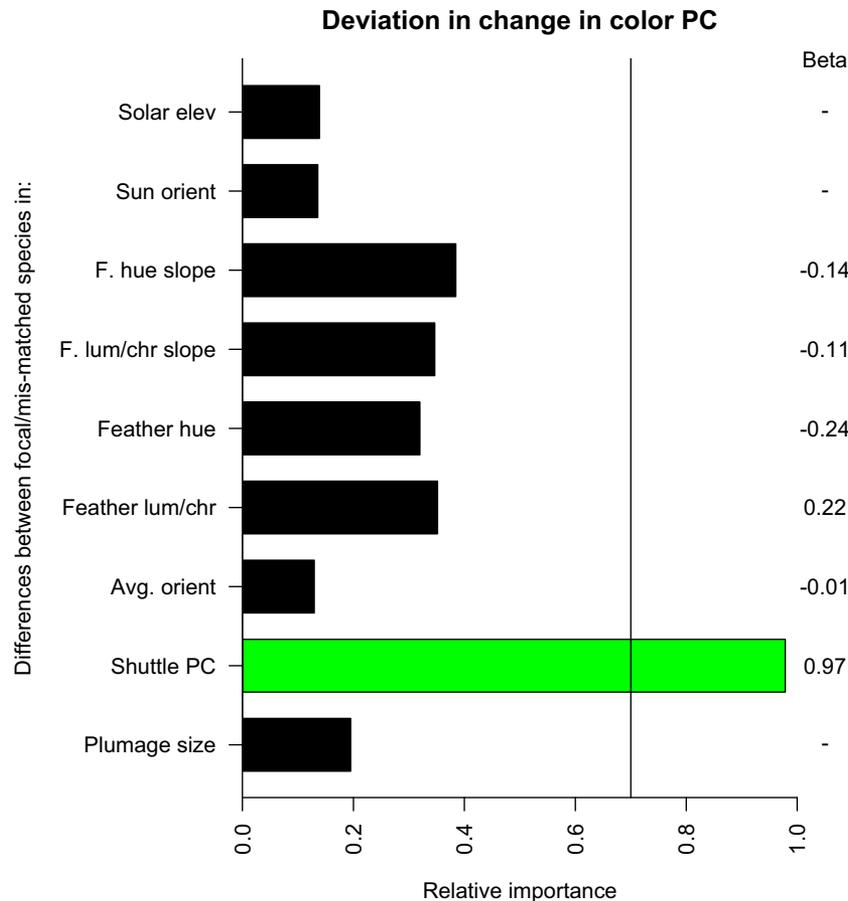


Figure 3. Results from information-theoretic model-averaging analyses, illustrating the relative importance (RI) of each fixed effect (left-hand side) on deviations from natural % change in color appearance PC. Fixed effects with an RI greater than 0.7 are indicated with green bars, while fixed effects with an RI below 0.7 are in black. The average beta for each effect is on the right-hand side of the plot, with betas lower than 0.01 are indicated by a dash.

signal pairs have done (e.g., *Trogonidae* song and plumage; Ornelas et al. 2009). By tracking color changes of feathers from the different species as they were moved through the paths of other species' shuttle displays, we found that one of the two color appearance variables (color appearance flashiness) differed significantly between the natural species condition and the heterospecific mismatches. Additionally, we found that the differences between the natural and mismatched flashiness was positively related to deviations between focal and mismatched shuttle parameters (e.g., shuttle width). Altogether, these results illustrate the species-specific pairings between shuttle displays and color appearance but not plumage and color appearance, because as differences between species-specific and mismatched shuttles increased, there was an increased disruption in the natural flashiness of a males and this was not the case with differences in plumage patch sizes.

We found that differences in shuttle behavior between focal and heterospecific mismatched species predicted deviations in male flashiness, suggesting that the orientation-and-position-

specific movements that produce flashy color displays are finely tuned and specially paired within each "bee" hummingbird species. The importance of behavioral movements, orientations, and postures in creating a flashy color-display has also been demonstrated in two species of butterflies and in peacocks, albeit intraspecifically in each case (*Pavo cristatus*; Rutowski et al. 2007; Dakin and Montgomerie 2013; White et al. 2015). Further, some colorful ornaments are only observable due to specific behavioral manipulations, such as uncovering a colorful patch (e.g., lifting a wing and showing an underwing color patch; Hansen and Rohwer 1986; Zanollo et al. 2013), positioning the color patch toward the receiver (e.g., male *Harporhynchus* jumping spiders ensuring they directly face a female throughout their courtship display; Echeverri et al. 2017), or repositioning various body parts to create a color-display not possible in a natural body position (e.g., superb bird-of-paradise, *Lophorina superba*; Laman and Scholes 2012). Thus, for many animals, behaviors are important to optimizing color presentations during displays and/or coloration is important to accentuating behavioral

Table 5. Results from AIC model comparisons testing which species-average signal properties (color appearance, plumage, and shuttle behaviors) best predicted species-average deviations from natural % change in color variables (luminance, chroma, RGB hue, and UV hue, separately) due to mismatches.

Response variable	Fixed effect	AICc	Δ AIC	Akaike weight
Deviation in % Lum. Change	Plumage patch size	37.3	0.00	0.66
	Feather reflectance PC	41.2	3.90	0.10
	% Change in color PC	41.4	4.06	0.09
	Average color PC	41.6	4.30	0.08
	Shuttle PC	41.6	4.33	0.08
Deviation in % chroma change	% Change in color PC	25.3	0.00	0.86
	Shuttle PC	29.7	4.49	0.09
	Average color PC	31.9	6.60	0.03
	Feather reflectance PC	33.2	7.98	0.02
	Plumage patch size	35.5	10.22	0.01
Deviation in % RGB hue change	Shuttle PC	38.2	0.00	0.54
	% Change in color PC	39.8	1.63	0.24
	Plumage patch size	41.8	3.66	0.09
	Average color PC	42.2	4.07	0.07
	Feather reflectance PC	43.2	4.11	0.07
Deviation in % UV hue change	Feather reflectance PC	24.6	0.00	0.91
	Average color PC	29.7	5.10	0.07
	Shuttle PC	34.0	9.37	0.01
	% Change in color PC	34.8	10.18	0.01
	Plumage patch size	35.4	10.80	<0.01

The top model for each of the four comparisons is bolded.

performances (Byers et al. 2010), and our results here shed further light on this behavior/color appearance relationship by demonstrating how alterations in a species' natural display behavior reduce the efficacy of their natural color-display. Further, there are other nonvisual signals that exhibit directionality (Rosenthal 2007; Patricelli et al. 2016), such as the directional acoustic signals of blackbirds (Larsen and Dabelsteen 1990; Patricelli et al. 2007), treehoppers (Cocroft et al. 2000), and cicadas (Bennet-Clark and Young 1998) and chemical signals that are directionally transmitted through wind (Pokorny et al. 2017) or water flow (Atema 1995). Therefore, our results can be interpreted across multiple signal modalities, demonstrating that the need for signalers to behaviorally alter, position, and/or orient themselves or their signals relative to their intended receiver and environment, and these behaviors should closely coevolve with the directional aspects of their signals for optimal presentation and transmission through the environment.

Interestingly, we did not find that plumage-behavior mismatches affected how bright/colorful males appeared on average during a display. In our previous work (Simpson and McGraw 2018a, c), we found that within species plumage properties do not strongly predict variance in how colorful/bright males appear on average, suggesting that intraspecific variation in plumage properties may be relatively less important or play a secondary role to receivers during courtship displays. For example, colorful

ornaments have been suggested to amplify or enhance behavioral displays in *Anolis* lizards and bowerbirds (Fleishman 1988; Endler et al. 2014) instead of being the primary trait of interest. Additional work in *Schizocosa* wolf spiders has demonstrated lower female receptiveness to ornamental traits that are presented alone as opposed to when paired with behavioral displays or compared to behaviors presented alone (Hebets and Uetz 2000; Uetz et al. 2009). However, among species, we found that how species appear on average during a display did covary with species-specific plumage properties (i.e., feather reflectance; Simpson 2018), indicating that interspecific variation in how bright/colorful a species appears during a display is due to the color properties of male feathers and not behavior. Further, interspecific variation in average color appearance during a display did not covary with shuttle display properties (Simpson 2018). Thus, it is possible that average color appearance is more robust to changes in behavior among species due to the specific pairings between plumage patch properties and average color appearance, while within a species, average color appearance is mainly influenced by how displays are oriented relative to the sun and overall less sensitive to plumage variation (perhaps because there is little variation in plumage properties within most species). Altogether these results showcase the complex, multi-level relationships among behavioral and color signals and their interactions.

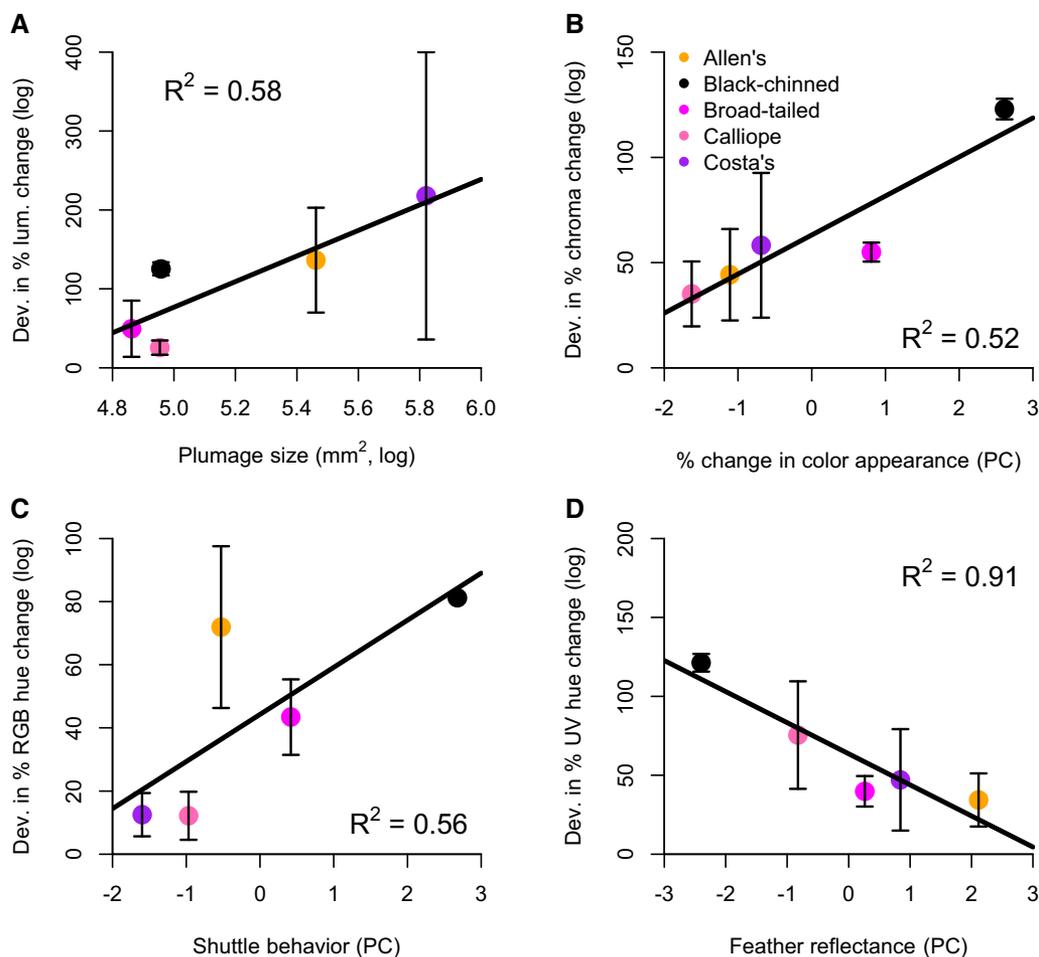


Figure 4. The relationship between species-average shuttle and plumage signals and species-average deviations from natural % change in color appearance due to mismatches. (A) A significant positive relationship between plumage patch size and deviations from natural % change in luminance. (B) A significant positive relationship between naturally occurring flashiness (% change in color appearance PC) and deviations from natural % change in chroma. (C) A significant positive relationship between shuttle behavior PC and deviations from natural % change in RGB hue. (D) A significant negative relationship between feather reflectance PC and deviations in from natural change in UV hue. Each point is color coded by species, with the color-species legend in panel B. R^2 for each relationship are also present in each panel, and error bars represent standard errors.

When comparing deviations between natural and mismatched flashiness to male signal properties (e.g., shuttle behaviors, plumage size), we found that species with flashier and more exaggerated shuttle displays had greater deviations from their natural flashy color appearance due to mismatches, which further supports the aforementioned specific coevolutionary pairing between behavior and flashy color appearance. Additionally, the negative relationship between species-average feather reflectances and deviations from their natural flashy color appearance suggests that the natural flashiness of species with more exaggerated plumage (which often have more consistent color appearances throughout a shuttle; Simpson 2018) are less sensitive to plumage-behavior mismatches (also mentioned above). We currently have evidence from a separate comparative study on our focal hummingbird species that exaggeration in

feather reflectance and bright/colorful consistent appearances covary among species, whereas feather exaggeration and flashiness did not (Simpson 2018). Thus, it is possible that brighter and more colorful feathers are better designed to produce a consistent color-display, similar to more conventional nonangle-dependent plumage patches, such as carotenoid-based plumage (McGraw 2006), melanin-colored ornaments (D'Alba et al. 2014), or bright white color patches (Ferns and Hinsley 2004), especially when presented under relatively static light environments (i.e., male display orientation relative to the sun does not vary much).

The positive relationship between plumage patch size and deviations from species natural flashy color appearance is interesting considering the negative coevolutionary relationship between these two signals, which we previously uncovered in this clade (Simpson 2018). It seems that, despite this antagonistic

coevolution, the flashy color appearance of species with larger plumage patches (which is also typically associated with a consistent color-display) is more sensitive to plumage-behavior mismatches, suggesting that these plumage patches might be more prone to imprecise presentation if not oriented properly (i.e., as occurs in a nonspecies-specific shuttle) and that plumage patch sizes specifically coevolved with shuttle displays. Previous work on widowbirds (*Euplectes ardens*) and barn swallows (*Hirundo rustica*) have suggested that exaggerated plumage ornaments, like long tails, are unwieldy and difficult to fly with (Barbosa 1999; Pryke et al. 2001; Andersson et al. 2002), so because hummingbirds erect their plumage while shuttling, they may have compensated for their large plumage patches by evolving shuttles with less exaggerated movements to avoid this issue (Møller 2008) and/or more optimally show off their larger plumage patches (Simpson and McGraw 2018c). However, recent work on peacocks, hummingbirds, and rhinoceros beetles (*Trypoxylus dichotomus*) found that longer, and theoretically more unwieldy, tails/horns do not negatively impact various aspects of flight (Clark and Dudley 2009; McCullough et al. 2012; Askew 2014), potentially dismissing this idea. Instead, the evolved pairing between display flashiness and plumage patch size among species may have arisen because a bigger area of feathers may be more difficult to coordinate and orient to produce uniform/specific flash patterns. Future work manipulating gorget size is needed to better test these hypotheses.

Animals exhibit a great diversity of signaling traits and understanding the evolutionary patterns of this diversity has long interested biologists. In this study, we have provided a new and unique set of methods to more deeply probe the potential causes of specific signal pairings among species. Our results in this “bee” hummingbird clade show specific pairings between flashiness and behavior and also specific pairings between consistent color-displays and plumage reflectance, and combined these results further illustrate the complex evolutionary dynamics of multiple signals and their signal interactions. We hope that this study further demonstrates the importance of incorporating and understanding the role of signal interactions in the evolution of multiple signals, instead of solely focusing on the properties of the signaling traits in isolation.

AUTHOR CONTRIBUTIONS

Conceptualization, R.K.S. and K.J.M.; Methodology, R.K.S.; Investigation, R.K.S.; Writing – Original Draft, R.K.S.; Writing – Review & Editing, R.K.S. and K.J.M.; Funding Acquisition, R.K.S. and K.J.M.

ACKNOWLEDGMENTS

We thank, Meghan E. Duell, Russell A. Ligon, Ronald L. Rutowski, Stephen C. Pratt, Jimmy A. McGuire, Christopher J. Clark, Emilia P. Martins, and the McGraw lab for their support, helpful discussions with the background and methods of this study, and for providing feedback on this manuscript. We thank Jessica Givens, Alysia Apple, Avery Underwood,

Christina Piarowski, and Jade Gates for their help with video analyses. This work was supported by Arizona State University (ASU), T & E Inc. grants for Conservation Biology Research, and the National Science Foundation, Doctoral Dissertation Improvement Grant (IOS-1702016).

DATA ARCHIVING

The doi for our data is <https://doi.org/10.5061/dryad.4pv7r04>

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

LITERATURE CITED

- Andersson, S., S. R. Pryke, J. Ornborg, M. J. Lawes, and M. Andersson. 2002. Multiple receivers, multiple ornaments, and a trade-off between agonistic and epigamic signaling in a widowbird. *Am. Nat.* 160:683–691.
- Askew, G. N. 2014. The elaborate plumage in peacocks is not such a drag. *J. Exp. Biol.* 217:3237–3241.
- Atema, J. 1995. Chemical signals in the marine environment: dispersal, detection, and temporal signal analysis. *Proc. Natl. Acad. Sci.* 92:62–66.
- Barbosa, A. 1999. Aerodynamic costs of long tails in male barn swallows *Hirundo rustica* and the evolution of sexual size dimorphism. *Behav. Ecol.* 10:128–135.
- Bennet-Clark, H. C., and D. Young. 1998. Sound radiation by the bladder cicada *Cystosoma Saundersii*. *J. Exp. Biol.* 201:701–715.
- Clark, C. J., and R. Dudley. 2009. Flight costs of long, sexually selected tails in hummingbirds. *Proc. R Soc. B Biol. Sci.* 276:2109–2115.
- . T. J. Feo, and K. B. Bryan. 2012. Courtship displays and sonations of a hybrid male broad-tailed × black-chinned hummingbird. *Condor* 114:329–340.
- . T. J. Feo, and W. F. D. van Dongen. 2013. Sounds and courtship displays of the Peruvian sheartail, Chilean woodstar, oasis hummingbird, and a hybrid male Peruvian sheartail × Chilean woodstar. *Condor* 115:558–575.
- Cocroft, R. B., T. D. Tieu, R. R. Hoy, and R. N. Miles. 2000. Directionality in the mechanical response to substrate vibration in a treehopper (Hemiptera: Membracidae: *Umbronia crassicornis*). *J. Comp. Physiol. A Sensory, Neural, Behav. Physiol.* 186:695–705.
- D’Alba, L., C. Van Hemert, K. A. Spencer, B. J. Heidinger, L. Gill, N. P. Evans, P. Monaghan, C. M. Handel, and M. D. Shawkey. 2014. Melanin-based color of plumage: role of condition and of feathers’ microstructure. *Integr. Comp. Biol.* 54:633–644.
- Dakin, R., and R. Montgomerie. 2013. Eye for an eyespot: how iridescent plumage ocelli influence peacock mating success. *Behav. Ecol.* 24:1048–1057.
- Echeverri, S. A., N. I. Morehouse, and D. B. Zurek. 2017. Control of signaling alignment during the dynamic courtship display of a jumping spider. *Behav. Ecol.* 28:1445–1453.
- Endler, J. A. 1992. Signals, signal conditions, and the direction of evolution. *Am. Nat.* 139:S125–S153.
- Endler, J., J. Gaburro, and L. Kelley. 2014. Visual effects in great bowerbird sexual displays and their implications for signal design. *Proc. R Soc. B Biol. Sci.* 281:20140235.
- Feo, T. J., and C. J. Clark. 2010. The displays and sonations of the black-chinned hummingbird (Trochilidae: *Archilochus alexandri*). *Auk* 127:787–796.
- Ferns, P. N., and S. a. Hinsley. 2004. Immaculate tits: head plumage pattern as an indicator of quality in birds. *Anim. Behav.* 67:261–272.

- Fleishman, L. J. 1988. Sensory influences on physical design of a visual display. *Anim. Behav.* 36:1420–1424.
- Galipaud, M., M. A. F. Gillingham, M. David, and F. X. Dechaume-Moncharmont. 2014. Ecologists overestimate the importance of predictor variables in model averaging: a plea for cautious interpretations. *Methods Ecol. Evol.* 5:983–991.
- Girard, M. B., D. O. Elias, M. M. Kasumovic, and M. B. Girard. 2015. Female preference for multi-modal courtship: multiple signals are important for male mating success in peacock spiders. *Proc. R. Soc. B Biol. Sci.* 282:12–14.
- Girard, M. B., M. M. Kasumovic, and D. O. Elias. 2011. Multi-modal courtship in the peacock spider, *Maratus volans* (O.P.–Cambridge, 1874). *PLoS One* 6:e25390.
- Gomez, D., M. Théry, and M. Thery. 2004. Influence of ambient light on the evolution of colour signals: comparative analysis of a neotropical rainforest bird community. *Ecol. Lett.* 7:279–284.
- Hansen, A. J., and S. Rohwer. 1986. Coverable badges and resource defence in birds. *Anim. Behav.* 34:69–76.
- Hebets, E. A., and D. R. Papaj. 2005. Complex signal function: developing a framework of testable hypotheses. *Behav. Ecol. Sociobiol.* 57:197–214.
- Hebets, E., and G. Uetz. 2000. Leg ornamentation and the efficacy of courtship display in four species of wolf spider (Araneae: Lycosidae). *Behav. Ecol. Sociobiol.* 47:280–286.
- Heindl, M., and H. Winkler. 2003. Vertical lek placement of forest-dwelling manakin species (Aves, Pipridae) is associated with vertical gradients of ambient light. *Biol. J. Linn. Soc.* 80:647–658.
- Herrera, G., J. C. Zagal, M. Diaz, M. J. Fernández, A. Vielma, M. Cure, J. Martinez, F. Bozinovic, and A. G. Palacios. 2008. Spectral sensitivities of photoreceptors and their role in colour discrimination in the green-backed firecrown hummingbird (*Sephanoides sephaniodes*). *J. Comp. Physiol. A Neuroethol. Sensory Neural. Behav. Physiol.* 194:785–794.
- Hoffmann, T. 2017. SunCalc.
- Hutton, P., R. A. Ligon, K. J. McGraw, B. M. Seymoure, and R. K. Simpson. 2015. Dynamic color communication. *Curr. Opin. Behav. Sci.* 6:41–49.
- Jawor, J. M., and R. Breitwisch. 2004. Multiple ornaments in male northern cardinals, *Cardinalis cardinalis*, as indicators of condition. *Ethology* 110:113–126.
- Jawor, J. M., N. Gray, S. M. Beall, and R. Breitwisch. 2004. Multiple ornaments correlate with aspects of condition and behaviour in female northern cardinals, *Cardinalis cardinalis*. *Anim. Behav.* 67:875–882.
- Laman, T., and E. Scholes. 2012. Birds of paradise: Revealing the world's most extraordinary birds. National Geographic, Washington, DC.
- Larsen, O. N., and T. Dabelsteen. 1990. Directionality of blackbird vocalization. Implications for vocal communication and its further study. *Ornis Scand.* 21:36–45.
- Maia, R., C. M. Eliason, P. P. Bitton, S. M. Doucet, and M. D. Shawkey. 2013. pavo: an R package for the analysis, visualization and organization of spectral data. *Methods Ecol. Evol.* 4:906–913.
- Martins, E. P., A. G. Ossip-Klein, J. J. Zúñiga-Vega, C. Vital García, S. M. Campos, and D. K. Hews. 2015. Evolving from static to dynamic signals: evolutionary compensation between two communicative signals. *Anim. Behav.* 102:223–229.
- McCullough, E. L., P. R. Weingarden, and D. J. Emlen. 2012. Costs of elaborate weapons in a rhinoceros beetle: how difficult is it to fly with a big horn? *Behav. Ecol.* 23:1042–1048.
- McGraw, K. J. 2006. Mechanics of carotenoid-based coloration. Pp. 177–242 in G. E. Hill and K. J. McGraw, eds. *Bird coloration volume 1: Mechanisms and measurements*. Harvard Univ. Press, Cambridge.
- McGuire, J. A., C. C. Witt, J. V. V. Remsen, A. Corl, D. L. Rabosky, D. L. Altshuler, and R. Dudley. 2014. Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.* 24:910–916.
- Meadows, M. G., N. I. Morehouse, R. L. Rutowski, J. M. Douglas, and K. J. McGraw. 2011. Quantifying iridescent coloration in animals: a method for improving repeatability. *Behav. Ecol. Sociobiol.* 65:1317–1327.
- Møller, A. P. 2008. The cost of secondary sexual characters and the evolution of cost-reducing traits. *Ibis* 138:112–119.
- Ornelas, J. F., C. González, and A. E. de los Monteros. 2009. Uncorrelated evolution between vocal and plumage coloration traits in the trogons: a comparative study. *J. Evol. Biol.* 22:471–84.
- Patricelli, G. L., M. S. Dantzker, and J. W. Bradbury. 2007. Differences in acoustic directionality among vocalizations of the male red-winged blackbird (*Agelaius phoeniceus*) are related to function in communication. *Behav. Ecol. Sociobiol.* 61:1099–1110.
- Patricelli, G. L., A. H. Krakauer, and C. C. Taff. 2016. Variable signals in a complex world: shifting views of within-individual variability in sexual display traits. *Adv. Study Behav.* 48:319–386.
- Pokorny, T., I. Vogler, R. Losch, P. Schlütting, P. Juarez, N. Bissantz, S. R. Ramírez, and T. Eltz. 2017. Blown by the wind: the ecology of male courtship display behavior in orchid bees. *Ecology* 38:42–49.
- Pryke, S. R., S. Andersson, and M. J. Lawes. 2001. Sexual selection of multiple handicaps in the red-collared widowbird: female choice of tail length but not carotenoid display. *Evolution* 55:1452–1463.
- R Development Core Team. 2017. R: a language and environment for statistical computing. Version 3.4.1. R foundation for statistical computing, Vienna.
- Rosenthal, G. G. 2007. Spatiotemporal dimensions of visual signals in animal communication. *Annu. Rev. Ecol. Evol. Syst.* 38:155–178.
- Rutowski, R. L., J. M. Macedonia, J. W. Merry, N. I. Morehouse, K. Yturralde, L. Taylor-Taft, D. Gaalema, D. J. Kemp, and R. S. Papke. 2007. Iridescent ultraviolet signal in the orange sulphur butterfly (*Colias eurytheme*): spatial, temporal and spectral properties. *Biol. J. Linn. Soc.* 90:349–364.
- Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9:671–675.
- Simpson, R. K. 2018. The evolution of hummingbird coloration and courtship displays. Arizona State University, Arizona.
- Simpson, R. K., and K. J. McGraw. 2018a. It's not just what you have, but how you use it: solar-positional and behavioural effects on hummingbird colour appearance during courtship. *Ecol. Lett.* 21:1413–1422.
- . 2018b. Multiple signaling in a variable environment: expression of song and color traits as a function of ambient sound and light. *Biotropica* 50:531–540.
- . 2018c. Two ways to display: male hummingbirds show different color-display tactics based on sun orientation. *Behav. Ecol.* 29:637–648.
- Smith, C. E., J. T. Stevens, E. J. Temeles, P. W. Ewald, R. J. Hebert, and R. L. Bonkovsky. 1996. Effect of floral orifice width and shape on hummingbird-flower interactions. *Oecologia* 106:482–492.
- Snow, D. W., and B. Snow. 1980. Relationships between hummingbirds and flowers in the Andes of Colombia. *Bull. Br. Museum Nat. Hist.* 38:105–139.
- Stevens, M., C. A. Parraga, I. C. Cuthill, J. C. Partridge, and T. S. Troscianko. 2007. Using digital photography to study animal coloration. *Biol. J. Linn. Soc.* 90:211–237.
- Stiles, F. G. 1981. Geographical aspects of bird-flower coevolution, with particular reference to Central America. *Ann. Missouri Bot. Gard.* 68:323–351.
- Stoddard, M. C., and R. O. Prum. 2008. Evolution of avian plumage color in a tetrahedral color space: a phylogenetic analysis of new world buntings. *Am. Nat.* 171:755–76.
- Tobias, J. A., J. Aben, R. T. Brumfield, E. P. Derryberry, W. Halfwerk, H. Slabbekoorn, and N. Seddon. 2010. Song divergence by sensory drive in Amazonian birds. *Evolution* 64:2820–2839.

- Troschianko, J., and M. Stevens. 2015. Image calibration and analysis toolbox—a free software suite for objectively measuring reflectance, colour and pattern. *Methods Ecol. Evol.* 6:1320–1331.
- Uetz, G. W., J. A. Roberts, and P. W. Taylor. 2009. Multimodal communication and mate choice in wolf spiders: female response to multimodal versus unimodal signals. *Anim. Behav.* 78:299–305.
- Vorobyev, M., and D. Osorio. 1998. Receptor noise as a determinant of colour thresholds. *Proc. R Soc. B Biol. Sci.* 265:351–358.
- White, T. E., J. Zeil, and D. J. Kemp. 2015. Signal design and courtship presentation coincide for highly biased delivery of an iridescent butterfly mating signal. *Evolution* 69:14–25.
- Zanollo, V., M. Griggio, J. Robertson, and S. Kleindorfer. 2013. Males with a faster courtship display have more white spots and higher pairing success in the diamond firetail, *Stagonopleura guttata*. *Ethology* 119:344–352.

Associate Editor: G. Höbel;

Handling Editor: Mohamed A. F. Noor

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Results from principal components analysis on % change in color appearance variables (luminance, chroma, RGB hue, UV hue) and average color appearance variables (luminance, chroma, RGB hue, UV hue).

Table S2. Results from principal components analysis on data for deviations from natural % change in color appearances (luminance, chroma, RGB hue, UV hue), and differences in 1) feather reflectance variables (luminance, chroma, UV hue), 2) feather angle-dependence variables (angle dependence in luminance, chroma, and UV hue), and 3) shuttle behavior variables (shuttle width, shuttle velocity, and variation in male angle of orientation relative to the female) between focal and mis-matched species.

Table S3. Results from principal components analysis on species-level data for % change in color appearance variables (luminance, chroma, RGB hue, UV hue), average color appearance variables (luminance, chroma, RGB hue, UV hue), feather reflectance variables (luminance, chroma, UV hue), feather angle-dependence variables (angle dependence in luminance, chroma, and UV hue), and shuttle behavior variables (shuttle width, shuttle velocity, and variation in male angle of orientation relative to the female).

Table S4. ANOVA results comparing natural and plumage-display mis-matched average color appearance.

Table S5. ANOVA results comparing natural and display-plumage mis-matched % change in color appearance.

Figure S1. Feather spectral reflectance measurements of each species' feathers taken in a controlled laboratory setting (i.e. not during courtship), and these examples illustrate the diversity in natural plumage reflectance and angle-dependence among species.

Figure S2. Examples of natural male color appearance during a display for each species, depicting the variation in color appearance among species, with some species appearing flashier (e.g. black-chinned hummingbirds) and others appearing consistently-colored but brighter/more colorful overall (e.g. Allen's hummingbirds).