Female and male plumage colour signals aggression in a dichromatic tropical songbird

Ana V. Leitão [a, *], Michelle L. Hall [a], Kaspar Delhey [b], Raoul A. Mulder [a]

[a] School of BioSciences, University of Melbourne, Parkville, Melbourne, VIC, Australia
[b] School of Biological Sciences, Monash University, Clayton, Melbourne, VIC, Australia

Animal signals of competitive ability allow contests for limited resources to be settled without costly physical fights. Traits indicating competitive ability are diverse and span visual, acoustic or chemical modalities. Although animal signalling has been intensively studied, research has focused mainly on male traits. Little is known about the extent and functionality of competitive signals in females and whether there are sex differences in signal function. We studied whether plumage colour signals competitive ability in female and male lovely fairy-wrens, *Malurus amabilis*. In this species, both sexes sport elaborate but sexually dichromatic ornamental plumage. Using a mirror image stimulation test, we first assessed the relationship between male and female colour and agonistic behaviour, controlling for other physical, social and ecological variables. We then tested whether colourful plumage influenced aggressive response in both sexes by experimentally manipulating plumage colour and measuring individual responses to their mirror image. Females and males were more aggressive towards naturally less colourful reflections of the cheek patch in the mirror. However, when we manipulated plumage colour, both females and males responded more aggressively to experimentally increased cheek colour reflection in the mirror. Our findings suggest that plumage colour signals competitive ability in an aggressive context in both sexes and raises the possibility that signal reliability may be maintained by social interactions where individuals police and punish dishonest signals.

© 2019 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Competition for limited resources is a vital component of survival and reproduction and a major selective force. Historically, most studies have focused on male–male competition (Andersson, 1994), although females also compete for sexual and nonsexual resources (e.g. food, territories), in a range of taxa including reptiles (While, Sinn, & Wapstra, 2009), fish (Forsgren, Amundsen, Borg, & Bjelvenmark, 2004), insects (Papadopoulos, Carey, Liedo, Mulle, & Sentuerk, 2009), birds (Cain & Langmore, 2015; Pryke, 2007) and mammals (Holekamp, Smale, & Szykman, 1996), including humans (Vaillancourt, 2013). In females, aggressive interactions can help improve conditions for breeding (Rosvall, 2008), access to high-quality reproductive resources (Sandell & Smith, 1997) or generally increase reproductive success (Sinn, While, & Wapstra, 2008).

Traits such as colourful plumage or elaborate vocalizations (hereafter 'ornamental traits') can be signals of social status that help settle contests by honestly informing rivals about the costs of escalating aggression (status signalling theory; Rohwer, 1975; Senar, 2006). Males and females of many species may display such competitive trait signals (Tobias, Montgomerie, & Lyon, 2012) that are driven by selection acting differentially between sexes (Dale, Dey, Delhey, Kempenaers, & Valcu, 2015), but while these are generally well studied in males, our understanding of female signals is still limited (Rosvall, 2011). It has been hypothesized that female ornamental traits are maintained as inherited nonfunctional by-products of selection on male ornaments (genetic correlation: Lande, 1980) or shaped by natural selection (Bortolotti, 2006; Wallace, 1891). Alternatively, female ornamental traits may be maintained by direct selection, as sexual signals (Amundsen, 2000) or generally as social signals for competition over sexual and nonsexual resources (Tobias et al., 2012; West-Eberhard, 1979). Female ornamental traits may or may not be shaped by the same selective processes as in males (Clutton-Brock, 2007; LeBas, 2006). Thus, to understand and address the different hypotheses for female ornament evolution, we must evaluate the possible function(s) of female ornaments and directly compare them to those of males. So far, a few studies have contrasted female and male ornaments in same-sex competition and intersexual differences in

* Correspondence: A.V. Leitão, School of BioSciences, University of Melbourne, Building 147, Parkville, Melbourne, 3010, Victoria, Australia.
E-mail address: anamvleitao@gmail.com (A. V. Leitão).

signal function (Crowhurst, Zanollo, Griggio, Robertson, & Kleindorfer, 2012; Mónus, Liker, Pénzes, & Barta, 2016; Murphy et al., 2014; Reedy et al., 2017; van Dongen & Mulder, 2008; Young, Cain, Svedin, Backwell, & Pryke, 2017).

Generalized insights about ornamental trait function, such as plumage coloration, have been drawn primarily from studies of temperate species. This is problematic because tropical regions have the greatest species diversity (Stutchbury & Morton, 2001), and in the tropics males and especially females are more colourful than in temperate environments (Dale et al., 2015). Therefore, selective pressures on colourful plumages might vary between temperate and tropical regions.

The objective of this study was to investigate whether plumage colour functions as a competitive signal in both sexes of the lovely fairy-wren, *Malurus amabilis*, a resident tropical passerine with year-round territories and reproduction. The lovely fairy-wren provides an opportunity to investigate possible functional aspects of coloration, since males and females are both colourful but sexually dichromatic, a good criterion for studying competitive traits (Amundsen & Pärn, 2006). Furthermore, within the Australian fairy-wrens, female lovely fairy-wrens express the most colourful plumage, and males, in contrast to most Australian fairy-wrens, do not lose their colourful plumage outside the breeding season (Leitão, Hall, Venables, & Mulder, 2019). Additionally, lovely fairy-wrens have tropical life history traits, such as year-round breeding and territorial defence (Leitão et al., 2019), which may offer a contrasting view on the evolution and function of ornaments compared to well-studied temperate species.

We used mirror image stimulation to experimentally assess same-sex competition. An individual's reaction to its own reflection in a mirror was interpreted as its response to a perceived conspecific or rival, since birds typically do not recognize their mirror image (e.g. New Caledonian crows, *Corvus macrorhynchos*: Medina, Taylor, Hunt, & Gray, 2011; jackdaws, *Corvus monedula*: Soler, Pérez-Contreras, & Peralta-Sánchez, 2014; great tits, *Parus major*: Kraft, Forštová, Urhan, Exnerová, & Brodin, 2017; the sole exception is the Eurasian magpie, *Pica*: Prior, Schwarz, & Güntürkün, 2008). Mirrors have been used to study aggression in a range of species, such as crayfish (May & Mercier, 2007), fish (Oliveira, Carneiro, & Canário, 2005; Wilson, de Boer, Arnott, & Grimmer, 2011), birds (Branch, Kozlovsky, & Pravosudov, 2015; Hirschenhauser, Wittek, Johnston, & Möstl, 2008) and mammals (Svendsen & Armitage, 1973). Mirror image stimulation provides an opportunity to understand how individuals perceive physical traits such as plumage coloration since it has the benefit of standardizing the opponent's behaviour (to 'mirror' the subject), remove confounding effects by testing in social isolation and excluding all stimuli (acoustic, tactile and chemical) except for visual ones (Gallup Jr, 1968; Ruzzante, 1992).

We aimed to investigate whether male and female plumage colours in lovely fairy-wrens (1) are a signal and (2) function in a competitive context. We tested whether: (1) the degree of aggression birds expressed when exposed to mirror image stimulation was related to plumage colour; (2) experimental manipulations of plumage colour changed the aggressive response of birds subjected to mirror image stimulation; and (3) males and females behaved differently in response to mirror image stimulation. We predicted first that if plumage coloration signals competitive ability, birds would behave more aggressively to unmanipulated 'rivals' perceived to be less colourful (and thus less intimidating). If increased coloration reflects the degree of threat or challenge, the reverse expectation (higher aggression towards a more threatening signal) might be expected (see Searcy, Anderson, & Nowicki, 2008). Second, if aggression is an intrinsic state related to the individual's plumage state or colour, manipulation should have no effect, and

individuals' responses to a mirror image will be similar to those in the unmanipulated condition. Alternatively, if elaborate plumage colour signals status, individuals with enhanced colour should be less aggressive towards the more colourful 'rival' reflected image. Finally, if females and males have similar roles in competition, both should show similar patterns between plumage coloration and aggression.

METHODS

### Study Species and Field Data Collection

The lovely fairy-wren is a small nonmigratory bird, endemic to the wet tropics of Australia that breeds throughout the year, but primarily in the dry season (Leitão et al., 2019). It is a facultative cooperative breeder that forms long-term pair bonds and maintains territories year round. Resident males, females and subordinates engage in coordinated territorial disputes that can escalate to physical aggression between same-sex opponents (Leitão et al., 2019). Adult males and females are colourful but dichromatic (Appendix Fig. A1); fledglings of both sexes have a duller version of female-like plumage. Males and females may take up to several months to moult to their adult plumage, both maintaining bright plumage year round thereafter (no seasonal change in plumage colour: Leitão et al., 2019).

We studied colour-banded lovely fairy-wrens in the Cairns region, North Queensland, Australia in 2013–2017. Adults and juveniles were captured in mist-nets (lured with playback calls) for banding and collecting morphometric measurements as well as for colour measurements and the mirror image stimulation test (described below). Birds were marked with a numbered metal band (Australian Bird and Bat Banding Scheme) and a unique colour combination. We measured morphological traits, including tarsus, wing and head–bill lengths (vernier calliper, ±0.05 mm).

After release, birds were monitored weekly for several months a year, and social status, nesting behaviour, bird density and territory size were recorded (details in Leitão et al., 2019).

Mirror Image Stimulation Test

During the dry seasons of 2015 and 2016, when birds were captured, we performed a mirror image stimulation test, close to the capture location in a modified bird cage (60 × 30 cm and 30 cm
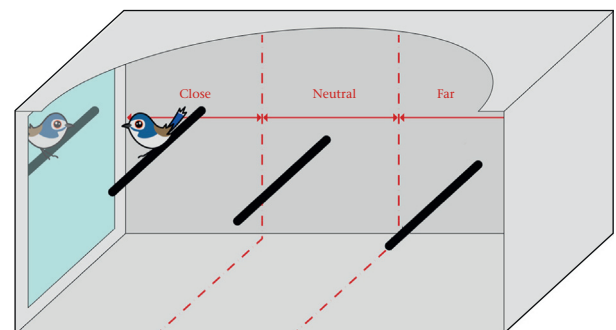

Figure 1. Diagrammatic front view of the mirror image test cage. The cage had three perches at different heights. One of the ends of the cage was covered by a removable plastic board covering a mirror, and on the opposite side birds were introduced through a small hole that was then covered. The mirror was uncovered (from a distance) approximately 5 min after the bird entered the cage. The walls of the cage were white, and the front was covered with wire to allow illumination and video recording. To minimize external influence, a white cloth wall covered the camera and the wired part of the cage. We analysed time spent in each area of the cage relative to the mirror (close, neutral or far). Image reproduced, with permission, from Osvaldo Branquinho.

high; Fig. 1), with a mirror (30 × 30 cm) on the side wall of the cage concealed by a removable plastic cover.

We conducted two experiments. In experiment 1, (28 May to 17 November 2015), we assessed aggressive behaviour of females and males in response to mirror image stimulation, using birds with unmanipulated plumage colour ($N = 23$ females and 19 males). We excluded tests of subordinates ($N = 4$), to control for social status. In experiment 2 (25 July to 23 November 2016), we experimentally manipulated the colour of the cheek patch in both sexes (the only variable significantly related to aggression in experiment 1, see Results), and again recorded aggressive responses to mirror image stimulation (total for all treatments $N = 35$ females and 29 males, see below). We excluded from the analysis subordinates ($N = 5$), juveniles ($N = 2$), males that were moulting ($N = 5$) and individuals whose spectral files were lost ($N = 2$).

Cheek colour was manipulated in experiment 2 using three different treatments of water-based nontoxic colour marker pens (Ecoline brush pens, Royal Talens, Apeldoorn, Netherlands). We tested different colour pens on feather samples and measured their reflectance. The resultant reflectance spectra were then compared with the natural plumage spectra of the individuals, to select the best match for each treatment. In treatment 1 we used a 'blender' (uncoloured) pen intended as a control; in treatment 2 we used 'warm grey' pen colour (hereafter 'grey') intended to decrease coloration; and in treatment 3 we used 'ultramarine blue deep' pen colour (hereafter 'blue') intended to enhance blue coloration (treatment effect is shown in Fig. 2). Overall, treatment 1 'blender' maintained coloration as intended, but both treatment 2 'grey' and treatment 3 'blue' increased coloration (Appendix Table A1). Most treatments fell within the natural variation (Appendix Tables A1
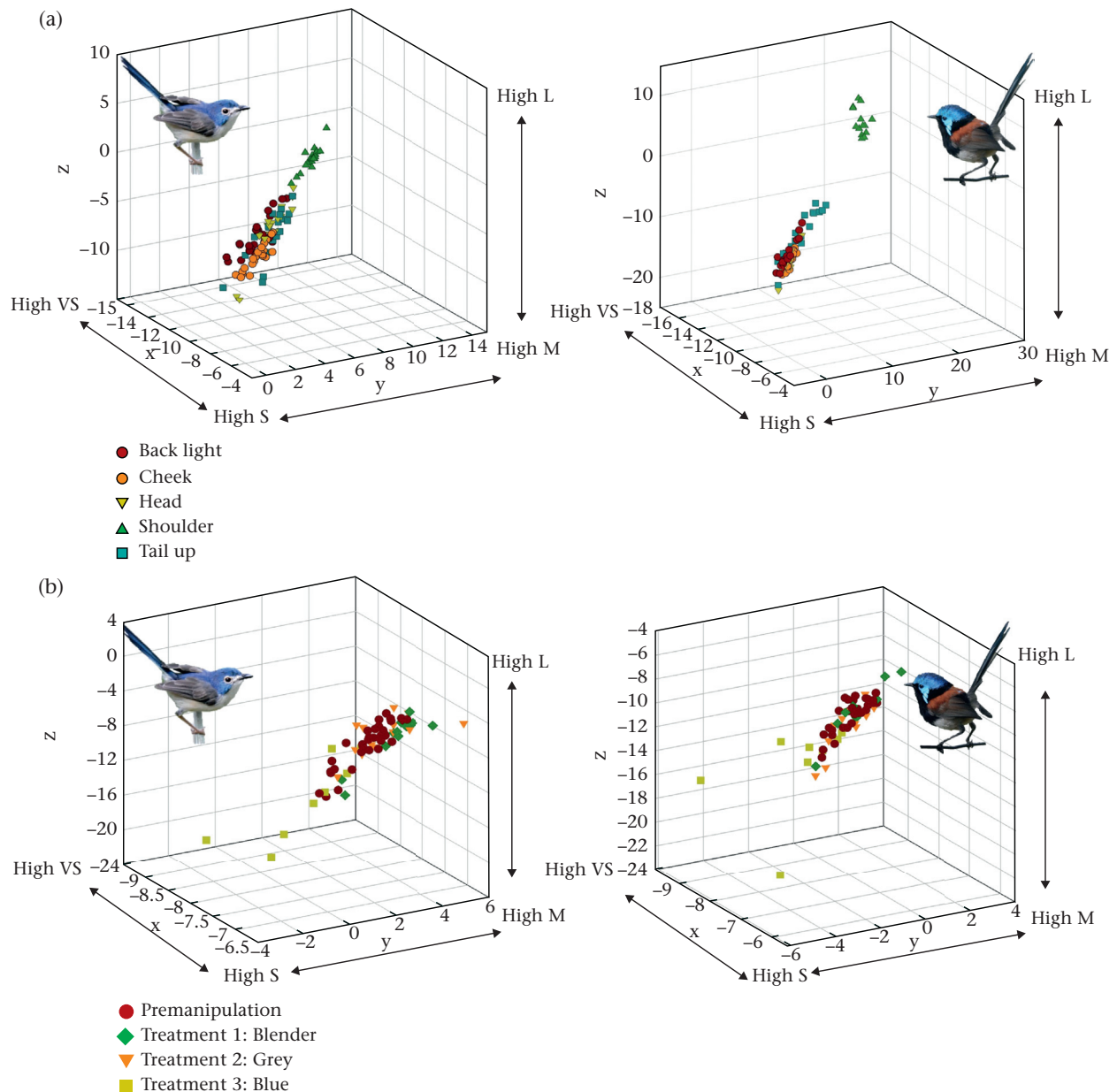
Figure 2. Graphic representation of chromatic coordinates (xyz) plotted in the visual space of U-type (ultraviolet-sensitive) birds, for plumage patches of female and male lovely fairy-wrens. (a) Experiment 1: natural variability of each plumage patch ($N = 23$ females and 19 males). (b) Experiment 2: pre- and postmanipulation chromatic coordinates of the cheek patch ($N = 35$ females and 29 males). The x axis represents the relative stimulation of the S cone in relation to the VS cone; y represents relative stimulation of the M cone in relation to the VS and S cones; and z represents the relative stimulation of the L cone in relation to the VS, S and M cones.

and A2), although in some cases the blue treatment increased coloration outside the observed natural range. For this reason, we consider treatment 1 'blender' as maintaining colour, treatment 2 'grey' as increasing colour within the natural variation and treatment 3 'blue' as increasing colour within and beyond the natural variation (for more details see Appendix, Experiment 2: Treatment effect). In total, we tested 10 'blender', eight 'blue', 11 'grey' males and 12 'blender', 10 'blue', 13 'grey' females in adult plumage ($N = 35$ females and 29 males). Treatment reduced the existing correlation between cheek colour, shoulder colour and body size, but cheek colour was correlated in a similar manner with other blue patches (Appendix Table A3).

Behaviour in the mirror test was recorded with a video camera (GoPro HD Hero2, GoPro Inc., San Mateo, CA, U.S.A.) for 10–20 min. In the first half of the test the mirror was covered; and in second half it was exposed by pulling the plastic cover from a distance. Most trials were conducted between 0600 and 1300 hours, local time (GMT +10), although some captures were also made in the afternoon (1500–1800 hours). From the video recordings, we analysed time spent in each area of the cage relative to the mirror (close to the mirror, neutral or far; Fig. 1; Supplementary Video S1) and aggressive behaviours, that is, the total number of displays (feathers erected, wings extended, usually performed at neutral distance), pecks and swoops (physical contact) towards the mirror (Supplementary Video S1). Since aggressive behaviours were all correlated ($r_\rho > 0.22$, $P < 0.02$), we combined them in a single score (sum of all aggressive displays). All behaviours were quantified separately for the periods with the mirror covered and the mirror uncovered. Aggressive behaviours were only observed after the mirror was exposed and were similar to aggressive behaviours observed in the field (Leitão et al., 2019).

We obtained high-quality sound recordings from the mirror test experiments (in 2015 recorder Marantz PMD660 with a microphone Sennheiser ME67, in 2016 recorder ZOOM H5 XYH-5 with a microphone Olympus ME-15). We used spectrograms generated in Avisoft-SASLab Pro v5.2.09 (Avisoft Bioacoustics, Berlin, Germany) to count the songs ('reels': Rowley & Russell, 1997) from females and males and analysed song rate (songs/min) from before and after the mirror was exposed.

Colour Measurement

Plumage reflectance of captured birds was measured on five patches: head, cheek, shoulder, back and tail; we obtained five readings per patch from each bird. For experiment 2, we measured the reflectance of the cheek patch immediately before and after the colour manipulation. Reflectance measurements were taken in the bird-visible wavelength range (300–700 nm), using an Ocean Optics JAZ spectrometer and inbuilt PX-3 Pulsed Xenon light source, with a probe and a machined 45° angle end (UV-VIS fibre-optic reflectance, Ocean Optics, Largo, FL, U.S.A.). The probe end excluded all ambient light and maintained a fixed distance to the feather surface. Reflectance was calibrated relative to a white standard (Ocean Optics WS-2).

Reflectance spectra were down-sampled to 5 nm steps and summarized using psychophysical models of avian vision to describe chromatic variation (Vorobyev & Osorio, 1998; Vorobyev, Osorio, Bennett, Marshall, & Cuthill, 1998), as implemented by Delhey, Delhey, Kempenaers, and Peters (2015) using formulas in Cassey et al. (2008). In birds, colour vision is mediated by four types of single cones that are sensitive to very short (VS), short (S), medium (M) and long (L) wavelengths. Variation in visual sensitivity between bird species is mainly in the VS and S cones and species can generally be ultraviolet sensitive (U-type) and violet sensitive

(V-type). Since lovely fairy-wrens are U-type birds (Ödeen, Pruett-Jones, Driskell, Armenta, & Håstad, 2012), we used average U-type peak sensitivity: 372, 456, 544, 609 nm (Endler & Mielke, 2005). The noise-to-signal ratio of each cone type was calculated based on formula 10 in Vorobyev et al. (1998) from the average cone proportions of U-type birds from Hart (2001; VS = 0.35, S = 0.6, M = 1, L = 1), combined with a Weber fraction of 0.1 for the L cone (Olsson, Lind, Kelber, & Simmons, 2017). As irradiance we used the spectrum of standard daylight (D65; Vorobyev et al., 1998).

Visual models computed using the approach described above reduce each spectrum to a set of three (xyz) chromatic coordinates that define its position in avian visual space where the x axis represents the relative stimulation of the S cone in relation to the VS cone, the y axis represents relative stimulation of the M cone in relation to the VS and S cones and the z axis represents the relative stimulation of the L cone in relation to the VS, S and M cones. Chromatic distances are expressed in JNDs (just noticeable differences), whereby values below 1 are thought not to be discriminable (Kelber, Vorobyev, & Osorio, 2003). After carrying out the visual models, we averaged xyz coordinates of the five readings for each colour patch to have a single set of values per plumage patch (chromatic colour variation for each patch is shown in Fig. 2).

We summarized chromatic variation (xyz coordinates) with a principal component analysis (PCA) using a covariance matrix to maintain the JND units of the original data (Delhey et al., 2015). A separate PCA analysis was conducted for each colour patch, and in experiment 2 for the cheek patch before and after treatment (Appendix Table A4). All PCA analyses resulted in one factor ($PC1_{chroma}$) that explained more than 90% of chromatic variance, with low positive loadings for the x axis, intermediate negative values for y and high negative loadings for z (Appendix Table A4). Thus, spectra with high $PC1_{chroma}$ values provide low stimulation of the L cone relative to VS + S + M cones, and low stimulation of the M cone relative to S + VS. In other words, birds with high $PC1_{chroma}$ values had colours richer in shorter wavelengths (UV/blue) and poorer in longer wavelengths (red).

Within each sex, the colours of different (blue) patches were correlated (head, tail and back; Appendix Table A5); the back was excluded from further analysis because it was highly correlated with other blue patches (Appendix Table A5), and had the lowest visual impact compared to other plumage patches (hidden by wings and head). Cheek and shoulder (blue and red/brown, respectively) were less strongly correlated with one another and with other patches (Appendix Table A5). Cheek patch was the most colourful patch in both sexes (Fig. 2, reflectance spectra in Appendix Fig. A2). Blue plumage patches were more colourful in males than females (more negative values indicate shortwave-rich colours) as was the red/brown shoulder patch (more positive values indicate longwave-rich colours) (Appendix Table A6, Fig. A2). Despite these differences, female and male homologous blue colour patches followed the same main axis of variation, which is different from the shoulder patch (Fig. 2).

Ethical Note

Our study complied with the ASAB/ABS Guidelines for the Use of Animals in Research and was approved by the University of Melbourne animal ethics committee (register 1613868.1). Fieldwork was carried out under licence from Queensland Parks and Wildlife Service (WISP13237913), and birds were banded under Australian Bird and Bat Banding Scheme banding permits. All personnel involved were trained in the appropriate techniques to capture and handle birds. Birds were captured in mist-nets which were

monitored continuously, and individuals were removed immediately upon capture. Birds were held for banding, measurements and experiments for up to 30 min. Colour manipulation was applied in less than 2 min and was done with water-pen markers (instead of typical alcohol-based markers) with the purpose of being easily removable with water and self-maintenance behaviour. Individual appearance was not visibly changed after treatment and could only be confirmed by spectrophotometry. Birds were released back at their capture location, and were monitored weekly, through censuses and 1 h focal observations. No adverse effects on the birds' behaviour or survival were detected, as all birds were resighted on their territory for at least 1 month after the experiments, by which time the treatment is expected to have vanished completely (Delhey, Peters, Johnsen, & Kempenaers, 2007).

Statistical Analysis

Statistical analyses were performed with R v3.4.3 (The R Foundation for Statistical Computing, Vienna, Austria). We analysed behavioural responses to the mirror image stimulation using independent-samples Mann–Whitney tests to compare sex differences and paired-sample Wilcoxon tests to compare behavioural changes before and after the mirror was uncovered; we used the behaviours that changed (aggression, time spent close to and far from the mirror) as response variables in subsequent models. All these variables were correlated ($r_\rho > 0.25$, $P < 0.01$). To reduce the dimensionality of the behavioural response to the mirror, we ran a PCA based on the behavioural data collected after the mirror was exposed (Table 1). Variables aggression and time spent distant were log transformed to improve normality before running the PCA. We did a PCA separately for experiments 1 and 2 and a third PCA with both experiments (to facilitate comparing them; Table 1). Each PCA resulted in a single factor ($PC1_{aggression}$) that explained more than 50% of the variation, where positive values indicated higher aggressiveness and proximity to the mirror and less time spent far from it (Table 1). $PC1_{aggression}$ was used to compute the scores of behavioural responses to the mirror, hereafter 'aggression score'.

For experiment 1, we used a general linear mixed model (GLMM) to assess which factors best explained variation in aggression score ($PC1_{aggression1}$) in the mirror test (assuming normal distribution of the error term for the dependent variable). As fixed effects in the model we included chromatic variation of the different colour patches ($PC1_{chroma1}$ for head, cheek, shoulder, tail), and other physical, social and ecological variables that might be related to aggressive behaviour: sex, body size, breeding stage (no nest, nesting, fledglings or unknown), number of neighbours (proxy for number of direct competitors in the area), Julian day (to account for differences within season) and time of day when tested. Group identity was included as a random effect to account for nonindependence of data collected from multiple birds of the same social group. Body size was estimated from the first factor of a PCA using tarsus, wing and head–bill length, which explained 63% of variance, with high positive loadings for all variables (tarsus 0.5, wing 0.6, head–bill 0.7). We first report the results from the full model including all variables and complement this with multimodel inference using the Akaike information criterion corrected for small sample size (AICc) to select best fitting models based on AIC weights (Appendix Table A7). The AIC was calculated for models containing all possible combinations of fixed effects, using package 'nlme' (Pinheiro, Bates, DebRoy, & Sarkar, 2011) and the function 'dredge' from package 'MuMIn' (Barton, 2017).

From experiment 2, having found a relationship between cheek patch chromatic variation and aggression score (see Results), we tested whether aggression score ($PC1_{aggression2}$) was influenced by natural plumage (before manipulation) or by treatment type ('blender', 'grey', 'blue'). In all models, we also included as fixed effects sex and its interaction with colour variable/treatment, and group identity as a random effect. We performed the same analysis excluding individuals that had colours outside of the natural range and that were detectable (see Appendix Tables A1 and A2 for ranges). We also tested for chromatic variation of the cheek patch postmanipulation, and to control for individuals' natural plumage colour we tested the interaction between their natural colour and the treatment applied (see Appendix, chromatic variation of the cheek patch postmanipulation). To understand whether there were any differences between experiments, we tested whether aggression score or natural colour of the cheek varied between experiments 1 and 2, and we did not find any differences (see Appendix, differences between experiments).

In the main text, we report analyses from both females and males, testing for sex differences by using colour PCAs with the sexes combined, since colours follow the same axis of variation (Fig. 2). In Appendix Tables A8–A12, we report results using separate PCA colours for each sex and models for each sex that did not differ qualitatively from those reported in the main text.

In all the analyses, models did not include variables that were highly correlated (Pearson correlations, $r_\rho < 0.43$ for all variables included) to avoid problems of collinearity ($|r| > 0.7$; Dormann et al., 2013). GLMMs were performed using package lme4 (Bates & Maechler, 2010) and the significance of factors and degrees of freedom (Satterthwaite's method for approximating degrees of freedom) were assessed using the 'lmerTest' package (Kuznetsova, Brockhoff, & Christensen, 2017).

RESULTS

Behavioural Responses to Mirror Image Stimulation

Individuals changed their behaviour when the mirror was exposed, spending less time at intermediate distances and more time close to the mirror (experiments 1 and 2, paired-sample Wilcoxon test: $Z = -2.68$, $N = 106$, $P < 0.01$) as well as far from it ($Z = -3.09$, $N = 106$, $P < 0.01$). Two per cent of females and 17% of males performed displays towards the mirror (feathers erected, wings extended) and 40% of females and 60% of males pecked or swooped at the mirror (physical contact with the mirror; Appendix Fig. A3 and Supplementary Video S1). Females and males sang at higher rates before than after the mirror was exposed (mean songs/min ± SD: 2.12 ± 2.44 before, 1.42 ± 2.29 after; $Z = -2.84$, $N = 102$, $P < 0.01$). Both sexes sang at similar rates (songs/min) when the mirror was covered (independent-samples Mann–Whitney test: $Z = -1.04$, $P = 0.15$; female: 3.55 ± 2.48, $N = 58$; male: 2.41 ± 2.63, $N = 48$), but females sang at higher rates than males when the mirror was exposed ($Z = -2.81$, $P < 0.01$; female: 2.09 ± 2.69; male: 0.33 ± 0.59). Females that sang more were also more aggressive

Table 1
Loadings for the principal component analysis (PCA) of behaviours (aggression, time spent close to the mirror, time spent far from the mirror) in the mirror image stimulation test, after uncovering the mirror

	Experiment 1	Experiment 2	Experiments 1 and 2
	$PC1_{aggression1}$	$PC1_{aggression2}$	$PC1_{aggression1+2}$
Aggression (pecks+swoops+displays)	0.51	0.56	0.55
Time spent close to the mirror	0.66	0.62	0.63
Time spent far from the mirror	−0.55	−0.54	−0.55
Eigenvalues (% of variance)	1.23 (50)	1.42 (67)	1.35 (61)

The PCA is for experiment 1, experiment 2 and both experiments combined.

(correlation with female aggression score: $r_\rho = 0.32$, $P = 0.01$), but this was not the case in males (correlation with male aggression score: $r_\rho = -0.13$, $P = 0.38$).

Experiment 1: Aggressive Behaviour towards the Mirror

Individuals with less blue cheek colour were significantly more aggressive towards their mirror image than those with bluer cheeks (Fig. 3). Other factors such as body size, breeding stage, density or season did not explain variability in aggressiveness in the full model (Table 2) and were also not present in the best fitted model (Appendix Table A7). The best fitting model explaining aggression score contained the fixed effects colour of the cheek (values for reduced model: $t_{35.60} = -2.76$, $P < 0.01$) and sex, with males being more aggressive towards the mirror than females (values for reduced model: $t_{38.99} = 3.58$, $P < 0.001$).

Experiment 2: Colour as an Aggressive Signal

Treatment type was related to aggression towards the mirror (Table 3), and males tended to be more aggressive than females although this difference was not significant (Table 3). Post hoc analysis revealed that individuals with colour expression increased by treatment 3 'blue' (i.e. bluer colours) were significantly more aggressive than individuals from treatment 1 'blender' (Fig. 4, Table 3); individuals in treatment 2 'grey' group also tended to be more aggressive than treatment 1, although this difference was not significant (Table 3). Excluding individuals that had detectable postmanipulation colours outside the natural range (see Appendix Tables A1 and A2 for ranges) gave results following the same trend (see Appendix, Results excluding individuals whose colour was outside the natural range). Results were also similar when we tested the effect of the continuous chromatic variation of the cheek

Table 2
The effect of predictor variables on aggression score in experiment 1, with group identity included as a random factor

| Parameter	β	SE	df	t	P
Fixed effects					
Intercept	−30.91	72.15	26.35	−0.43	0.67
Sex[a]	1.78	1.16	2.77	1.53	0.14
Body size	−2.03e-02	0.20	21.64	−0.10	0.92
Cheek colour	−0.43	0.19	28.99	−2.27	**0.03**
Head colour	6.90e-02	0.11	27.79	0.65	0.52
Shoulder colour	−1.81e-02	2.62e-02	29.0	−0.69	0.49
Tail colour	−2.26e-02	5.06e-02	28.46	−0.45	0.66
Breeding stage[b]					
Nest	0.45	0.70	28.93	0.64	0.53
Fledglings	−0.60	0.95	28.39	−0.64	0.53
Unknown	0.48	0.64	26.41	0.75	0.46
No. of neighbours	1.54e-02	0.17	23.70	0.09	0.93
Julian day	1.92e-03	4.73e-03	26.29	0.41	0.69
Time of day	4.86e-04	7.94e-04	28.73	0.61	0.55
Random effect	σ^2				
Group identity	0.57				
Residual	0.83				

Significant predictors are in bold. Effects and associated statistics are based on the full model with all predictors.
[a] Sex is a categorical term and the reference is female.
[b] Breeding stage is a categorical term with four levels: no nest, unknown, nest and fledglings; the reference is no nest.

patch postmanipulation (see Appendix, Chromatic variation of the cheek patch postmanipulation).

The subject's own colour (natural plumage colour before the manipulation) was not significantly related to aggression score ($t_{59.51} = 1.40$, $P = 0.16$; Fig. 4), and neither was sex, either alone ($t_{56.70} = 0.29$, $P = 0.77$) or in interaction with natural plumage (cheek PC1 ∗ sex: $t_{55.85} = -0.65$, $P = 0.52$).
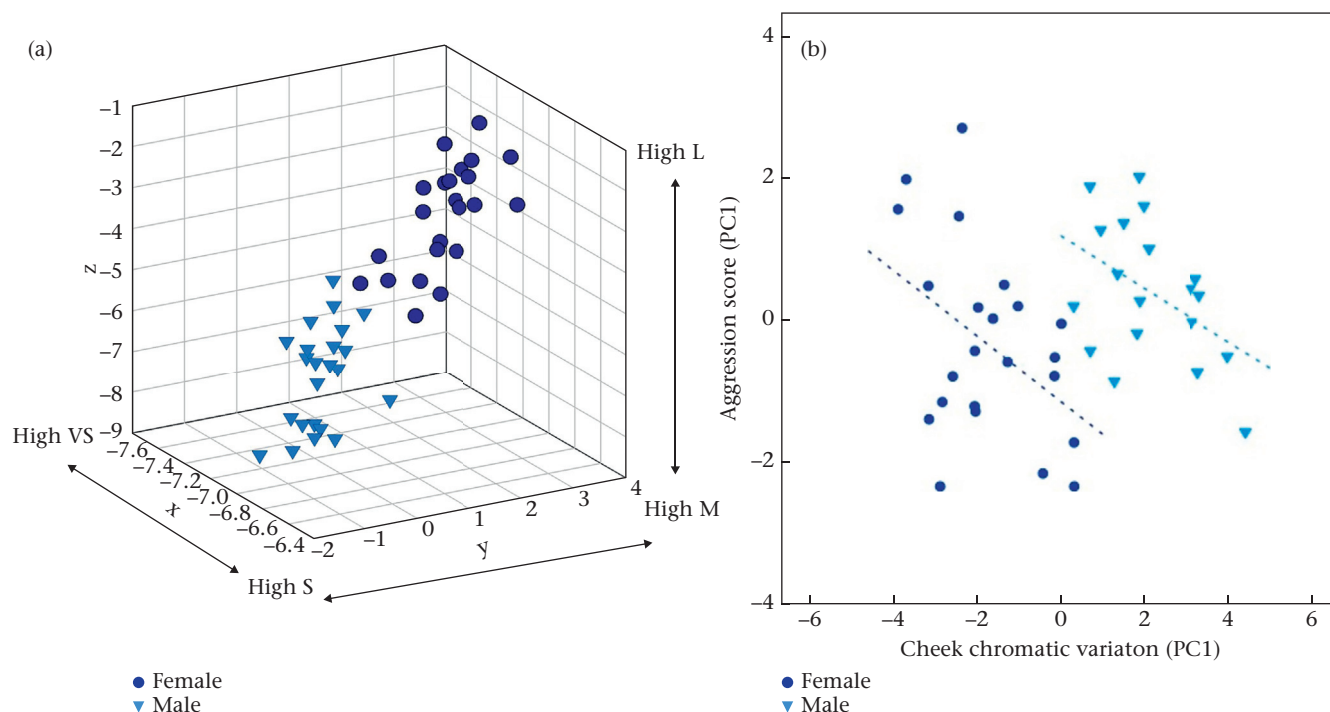
Figure 3. Experiment 1. (a) Chromatic variation of female and male cheek plumage colour represented within the tetrahedral colour space of birds. Most variation of the cheek colour is in the z axis, which represents long relative to shorter wavelength reflectance. Male and female cheek colours follow the same axis, although males have higher negative values towards UV and blue than females. Variation along the x axis represents stimulation of the VS cone relative to the S cone, the y axis represents stimulation of the M cone relative to the VS and S cones and the z axis represents stimulation of the L cone relative to the M, S and VS cones. Units of xyz are in JNDs (just noticeable differences). (b) Chromatic colour of the cheek patch (PC1$_{chroma1}$ of the cheek) in relation to aggression score (PC1$_{aggression1}$). Dashed lines depict the linear regression lines.

Table 3
Effect of fixed predictors on aggression score ($PC1_{aggression2}$) in experiment 2 for treatments and post hoc comparisons between treatments

Parameter	β	SE	df	t	P
Intercept	−1.23	0.39	56.03	−2.89	**0.005**
Treatment[a]					
Blue	1.30	0.58	55.52	2.26	**0.027**
Grey	1.44	0.53	50.29	2.71	**0.009**
Sex[b]	1.09	0.54	33.99	2.02	0.051*
Treatment[a]∗Sex[b]					
Blue∗Sex	−0.57	0.85	50.02	−0.67	0.50
Grey∗Sex	−1.16	0.76	39.66	−1.51	0.13
Treatment pairwise comparison					
Blue−grey	0.14	0.43	59.91	0.34	0.93
Blender−blue	−1.07	0.43	59.62	−2.47	**0.04**
Blender−grey	−0.92	0.40	60.00	−2.29	0.06*

Results are from the Tukey method for post hoc comparisons among the least-squares means treatments (treatment 1 = 'blender', treatment 2 = 'grey', treatment 3 = 'blue'; $N = 70$ comparisons), using package 'lsmeans' (Lenth, 2016). Significant predictors are in bold; an asterisk indicates marginal significance. Female $N = 35$, male $N = 29$.
[a] Treatment is a categorical term with three levels: 'blender', 'grey' and 'blue'; the reference is 'blender'.
[b] Sex is a categorical term and the reference is female.

DISCUSSION

Using mirror image stimulation to experimentally assess aggressive behaviour in close-range same-sex interactions, we showed that female and male lovely fairy-wrens both reacted more aggressively to their mirror image reflection when they were less ornamented themselves (i.e. less blue or shortwave-rich colours). However, experimental manipulations of the cheek plumage revealed the opposite pattern: individuals with enhanced (richer in shortwaves) cheek colour behaved more aggressively towards the mirror image. Below we discuss these results and consider possible explanations for the difference between experiments.

Same-Sex Interactions with Mirrored Rivals

Males and females reacted strongly to the mirror image stimulation, either responding aggressively with both attacks and displays to the mirrored 'rival' or avoiding it. Aggressive responses to the mirror (Appendix Fig. A3 and Supplementary Video S1) were similar to behaviour observed in the field during natural competitive interactions, where both sexes are aggressive towards same-sex conspecifics (Leitão et al., 2019). Although some studies have shown that mirror tests do not elicit the same hormonal responses (Oliveira et al., 2005) or the same aggressive intensity as live opponents (Balzarini, Taborsky, Wanner, Koch, & Frommen, 2014; Elwood, Stoilova, McDonnell, Earley, & Arnott, 2014), the same studies show correlations between responses to mirror and real opponents. In contrast to natural aggressive interactions, the 'rival' does not operate independently of the subject (and perhaps differently between subjects), that is, the mirror does not initiate a behaviour, so the pace and (de-)escalation of contests are driven by the subject and mirrored by the rival.

In natural settings, long-range acoustic signals usually precede short-range visual signals in escalating interactions in birds (Bradbury & Vehrencamp, 1998). The mirror image stimulation test simulated a short-range interaction, where visual signals are expected to be most important. In fact, when individuals were exposed to the mirror, they reduced their song rate, consistent with the view that birdsong is a long-range signal. This allowed us to test for the effects of plumage colour, eliminating other confounding factors (such as song).

To our knowledge, our study is the first to explore the role of plumage coloration with the mirror image stimulation test. This method has promise for future studies (particularly to test short-range signals) and indeed could be broadly employed in studies of visual signals, much as playback experiments are widely used in acoustic signal studies (Searcy & Nowicki, 2000).
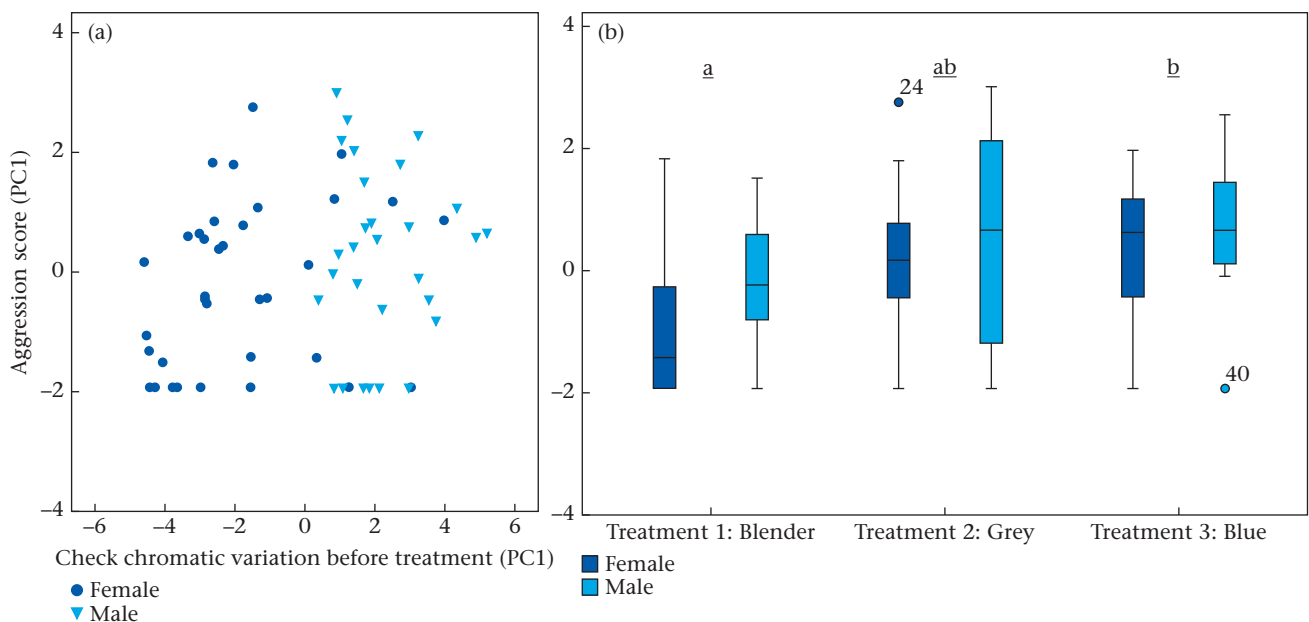
Figure 4. Experiment 2: relation between aggression score and (a) chromatic variation in cheek colour ($PC1_{chroma2}$) before treatment and (b) the three treatments (blender, grey, blue) for females and males. Box plots show the median and the interquartile range, whiskers indicate the values within 1.5 times the interquartile range and circles are outliers; different letters above bars indicate statistical significance ($P < 0.05$).

Plumage Colour as a Signal of Aggression

In the first experiment, natural colour of the cheek patch in both females and males predicted behaviour in the mirror test. We found that individuals with less blue plumage colours were more aggressive towards the mirror than those with richer blue plumages. There are at least two possible interpretations of this result. First, individuals might have an intrinsic aggressive state which is tied to their plumage colour (McGraw & Hill, 2000a; Pryke & Griffith, 2006), with less colourful birds being more aggressive, independently of the opponent's colour. Alternatively, individuals may use colour as a signal of social status (Griggio, Zanollo, & Hoi, 2010; Midamegbe, Grégoire, Perret, & Doutrelant, 2011), and be more likely to attack a less colourful 'rival' because it is perceived as weaker. The first explanation is unlikely, because if coloration was related to an intrinsic propensity for aggression, we should have observed a relationship between natural colour (before treatment) and aggression in experiment 2, which was not the case. Thus, we consider it more likely that colour is used as a signal in aggressive contexts.

Status signals have been shown to either increase or decrease aggression. For example, in red-winged blackbirds, *Agelaius phoeniceus*, male models with reddened epaulettes elicited greater responses from territorial owners and more displays from other males (other than the territory owner) compared to normal, dull or no epaulette models (Yasukawa et al., 2009). Also, in purple-crowned fairy-wrens, *Malurus coronatus*, 3D models in breeding colours received more aggression from resident breeder males than models in nonbreeding (dull) colours (Fan et al., 2018). In contrast, in house finches, *Carpodacus mexicanus*, duller males are dominant over bright males (McGraw & Hill, 2000a, 2000b) and similar examples are found with acoustic signals, where individuals either respond strongly to low-threat signals or avoid high-threat signals (Cramer & Price, 2007; Hardouin, Reby, Bavoux, Burneleau, & Bretagnolle, 2007 and see Searcy et al., 2008). It seems that status signals can either increase aggressive displays towards individuals that bear high-quality signals or deter attackers. In our experiment, we predicted that individuals would behave less aggressively to a more threatening signal because individuals were in a novel, unfamiliar environment (cage with mirror). We reasoned this to be more akin to a response to an intrusion onto the territory of a conspecific, rather than in defence of an individual's own territory, and in this context, individuals have more to lose if engaging in a fight due to owner–intruder asymmetry (Bradbury & Vehrencamp, 1998), where territorial ownership increases motivation to fight.

In the second experiment, manipulated colour of the cheek patch in both females and males predicted behaviour in the mirror test, but surprisingly, birds were more aggressive to increased than reduced coloration, in direct contrast to experiment 1. A few studies have shown similar results, where the experimental manipulation changed the expected result: in red bishops, *Euplectes orix*, less colourful individuals were more likely to win a contest, but when the colour was experimentally manipulated, more colourful individuals were more likely to win (Edler & Friedl, 2010); In northern cardinals, *Cardinalis cardinalis*, colour manipulation (either increasing or reducing) did not show any effect on dominance rank over food resources; however, naturally more colourful individuals were dominant over dull ones (Wolfenbarger, 1999); In the crimson finch, *Neochmia phaeton*, natural carotenoid plumage coloration was positively related to the probability of winning a contest, but when the colour was manipulated, this pattern reversed, with the underlying natural colour negatively related to the probability of winning a contest, and no effect of the experimental colour on contest outcome (Young, Cain, Svedin, Backwell,

& Pryke, 2016). We encourage other studies using mirror image stimulation to investigate visual signal function to assess responses to both unmanipulated and manipulated colour patches to test how widespread such contrasts are.

We considered several possible explanations for why birds reacted differently between experiments. First, our experimental manipulations of plumage colour may have resulted in 'unnatural' overall appearances of birds in the experimental treatments. If birds with the most altered cheek patches also deviated the most from 'normal' phenotypic appearance, this could have triggered higher levels of aggression. However, this was probably not the case, since most of the treated birds were within the natural variation, and even when excluding the individuals that were outside the natural range, the findings remained qualitatively similar.

Second, while plumage coloration may be an important signal in competition, other traits might be mediating aggression. Age, a factor that we could not control for, can affect aggressive behaviour (Edler & Friedl, 2010), as can previous experience (Collis & Borgia, 1992). Similarly, hormones such as testosterone can also mediate aggression and correlate positively with agonistic signals such as plumage (Tibbetts, 2014). In white-shouldered fairy-wrens, *Malurus alboscapulatus*, where females of different subspecies lack or possess male-like melanized black-and-white plumage (ornamentation), females of ornamented populations had higher levels of testosterone and were also more aggressive to a simulated territorial intrusion than those from populations with dull females, suggesting that testosterone may mediate aggression and female ornamentation (Enbody, Boersma, Schwabl, & Karubian, 2018).

A third nonmutually exclusive possibility might be that plumage colour is a signal that has costs imposed by receivers through social punishment (Hurd & Enquist, 2005; Tibbetts, 2013; Tibbetts & Dale, 2004), and that increased aggression in response to enhanced colour plumage reflects punishment via social policing of a perceived signalling 'cheat'. Some studies have shown that individuals with experimentally altered ornaments receive more aggression than 'honest' ones (Ligon & McGraw, 2016; Rohwer & Rohwer, 1978; Senar, 1999; Tibbetts & Dale, 2004; reviewed in Webster, Ligon, & Leighton, 2018). In a controlled experiment, Tibbetts and Izzo (2010) showed that female paper wasps, *Polistes dominulus*, with their status signal altered and mismatched with behaviour resulted in social punishment, but when both were experimentally altered to match, individuals did not suffer more aggression. If receivers can detect false signals due to incongruence between different cues, aggression can be used as a punishment and maintain signals by social costs. In the present study, the correlation between the manipulated cheek colour and shoulder patch colour and body size became weaker after the manipulation (Appendix Table A3). This discrepancy between the manipulated signal, more blue cheek, and other phenotypic traits, such as body size, may have triggered a punishment response of the 'rival' with increased aggression, although further experiments designed specifically to test this are needed to understand the mechanisms regulating signal honesty.

Differences and Similarities in Females and Males

Males were somewhat more aggressive to the mirror than females (significant in the reduced model in experiment 1, marginally nonsignificant in experiment 2), and males showed more aggressive displays towards the mirror, but in both sexes about half of the individuals interacted aggressively and physically towards it.

Although this species is dichromatic, both male and female cheek colour (which is the bluest and most reflective colour patch;

Fig. 3, Appendix Fig. A2) was correlated with aggression across unmanipulated and manipulated colours (even if in opposite directions). It is likely that, although male and female colours are different, plumage is used to mediate social competition in both sexes. Indeed, the colours of both male and female cheek patches follow a similar pattern of variation in visual space (Fig. 3). Competitive signals are expected to evolve and benefit both sexes because of identical characteristics, ecological and social needs (sexual and nonsexual resources: West-Eberhard, 1979). In the tropics, females and males commonly have similar roles in territory defence (Stutchbury & Morton, 2001), and thus both sexes may benefit from competitive traits such as plumage coloration because of year-round territorial defence (Murphy, Hernández-Muciño, Osorio-Beristain, Montgomerie, & Omland, 2009). This is likely to apply to lovely fairy-wrens, since they breed intermittently throughout the year, and maintain year-long territory defence. As adult survival in both females and males is high (Leitão et al., 2019), territory vacancies may be scarce and competitive signals might be effective during disputes.

Females and males sang at higher and similar rates before than after the mirror was exposed, suggesting that song might be used as a long-range signal that is suppressed when individuals perceive a rival nearby. The function of song might differ between females and males, since we found that females sang at higher rates than males when the mirror was exposed, and females that sang more were also more aggressive. It is also possible that song has a similar signalling function, but males escalate more readily into physical aggression, whereas females spend more time signalling to reduce costs of fighting. Further controlled studies will be necessary to illuminate the function and sex differences in song in this species.

Conclusions

By analysing female and male lovely fairy-wren responses to natural and experimentally manipulated colour of their bright blue cheek patch, we showed that colour was related to aggressive behaviour, suggesting that plumage colour can function as a competitive signal in aggressive contexts and that this signal plays a role in intrasexual competition in both sexes.

In many tropical species both males and females are colourful (Dale et al., 2015) and it has been suggested that this might be linked to investment in territorial defence (Stutchbury & Morton, 2001). If selection favours more competitive individuals in these environments, then signals that mediate competition for resources may be important to reduce the probability of extreme consequences of frequent fighting (e.g. time, probability of injuries and death). However, the fact that correlational and experimental results showed opposite patterns indicates that more research is needed to establish the precise nature of the link between plumage colour and aggression. Particularly, future experiments should find a treatment that effectively reduces the colour to understand whether dull birds are still more aggressive (as in experiment 1). The results of this study do not rule out that plumage colour is also used in intersexual interactions such as mate choice, since birds breed year-round unlike most Australian fairy-wrens (reviewed in Leitão et al., 2019), and future work should address how plumage traits relate to reproductive and nonreproductive resources (e.g. territories), to infer other functions and selective pressures.

Supplementary material

Supplementary material associated with this article is available, in the online version, at https://doi.org/10.1016/j.anbehav.2019.01.025.

References

Amundsen, T. (2000). Why are female birds ornamented? *Trends in Ecology & Evolution, 15*(4), 149–155. https://doi.org/10.1016/S0169-5347(99)01800-5.

Amundsen, T., & Pärn, H. (2006). Female coloration: Review of functional and non-functional hypothesis. In G. Hill, & K. McGraw (Eds.), *Bird coloration: Function and evolution* (Vol. II, pp. 280–345). Cambridge, MA: Harvard University Press.

Andersson, M. B. (1994). *Sexual selection*. Princeton, NJ: Princeton University Press.

Balzarini, V., Taborsky, M., Wanner, S., Koch, F., & Frommen, J. (2014). Mirror, mirror on the wall: The predictive value of mirror tests for measuring aggression in fish. *Behavioral Ecology and Sociobiology, 68*(5), 871–878. https://doi.org/10.1007/s00265-014-1698-7.

Barton, K. (2017). *MuMIn: Multi-model inference*. R package, version 1.40.4 https://cran.r-project.org/web/packages/MuMIn/index.html.

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed-effects models using S4 classes*. R package version 1.1-17 https://cran.r-project.org/package=lme4.

Bortolotti, G. R. (2006). Natural selection and avian coloration: Protection, concealment, advertisement, or deception?. In G. E. Hill, & K. J. McGraw (Eds.), *Bird coloration: Function and evolution* (Vol. II, pp. 3–35) Cambrige, MA: Harvard University Press.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*. Sunderland, MA: Sinauer Associates.

Branch, C. L., Kozlovsky, D. Y., & Pravosudov, V. V. (2015). Elevation related variation in aggressive response to mirror image in mountain chickadees. *Behaviour, 152*(5), 667–676. https://doi.org/10.1163/1568539X-00003248.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer-Verlag.

Cain, K. E., & Langmore, N. E. (2015). Female and male song rates across breeding stage: Testing for sexual and nonsexual functions of female song. *Animal Behaviour, 109*, 65–71. https://doi.org/10.1016/j.anbehav.2015.07.034.

Cassey, P., Ewen, J. G., Blackburn, T. M., Hauber, M. E., Vorobyev, M., & Marshall, N. J. (2008). Eggshell colour does not predict measures of maternal investment in eggs of *Turdus* thrushes. *Naturwissenschaften, 95*. https://doi.org/10.1007/s00114-008-0376-x.

Clutton-Brock, T. (2007). Sexual selection in males and females. *Science, 318*, 1882–1885. https://doi.org/10.1126/science.1133311.

Collis, K., & Borgia, G. (1992). Age-related effects of testosterone, plumage, and experience on aggression and social dominance in juvenile male satin bowerbirds (*Ptilonorhynchus violceus*). *The Auk, 109*(3), 422–434.

Cramer, E. R. A., & Price, J. J. (2007). Red-winged blackbirds *Agelaius phoeniceus* respond differently to song types with different performance levels. *Journal of Avian Biology, 38*(1), 122–127. https://doi.org/10.1111/j.2006.0908-8857.03839.x.

Crowhurst, C. J., Zanollo, V., Griggio, M., Robertson, J., & Kleindorfer, S. (2012). White flank spots signal feeding dominance in female Diamond Firetails, *Stagonopleura guttata*. *Ethology, 118*(1), 63–75. https://doi.org/10.1111/j.1439-0310.2011.01986.x.

Dale, J., Dey, C. J., Delhey, K., Kempenaers, B., & Valcu, M. (2015). The effects of life history and sexual selection on male and female plumage colouration. *Nature, 527*(7578), 367–370. https://doi.org/10.1038/nature15509.

Delhey, K. (2015). The colour of an avifauna: A quantitative analysis of the colour of Australian birds. *Scientific Reports, 5*, 18514. https://doi.org/10.1038/srep18514.

Delhey, K., Delhey, V., Kempenaers, B., & Peters, A. (2015). A practical framework to analyze variation in animal colors using visual models. *Behavioral Ecology, 26*(2), 367–375. https://doi.org/10.1093/beheco/aru198.

Delhey, K., Peters, A., Johnsen, A., & Kempenaers, B. (2007). Fertilization success and UV ornamentation in blue tits *Cyanistes caeruleus*: Correlational and

experimental evidence. *Behavioral Ecology, 18*(2), 399–409. https://doi.org/10.1093/beheco/arl096.

van Dongen, W. F. D., & Mulder, R. A. (2008). Male and female golden whistlers respond differently to static and dynamic signals of male intruders. *Behavioral Ecology, 19*(5), 1025–1033. https://doi.org/10.1093/beheco/arn061.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Edler, A. U., & Friedl, T. W. P. (2010). Plumage colouration, age, testosterone and dominance in male red bishops (*Euplectes orix*): A laboratory experiment. *Ethology, 116*(9), 806–820. https://doi.org/10.1111/j.1439-0310.2010.01799.x.

Elwood, R. W., Stoilova, V., McDonnell, A., Earley, R. L., & Arnott, G. (2014). Do mirrors reflect reality in agonistic encounters? A test of mutual cooperation in displays. *Animal Behaviour, 97*(0), 63–67. https://doi.org/10.1016/j.anbehav.2014.07.028.

Enbody, E. D., Boersma, J., Schwabl, H., & Karubian, J. (2018). Female ornamentation is associated with elevated aggression and testosterone in a tropical songbird. *Behavioral Ecology, 29*(5), 1056–1066. https://doi.org/10.1093/beheco/ary079.

Endler, J. A., & Mielke, P. W. (2005). Comparing entire colour patterns as birds see them. *Biological Journal of the Linnean Society, 86*(4), 405–431. https://doi.org/10.1111/j.1095-8312.2005.00540.x.

Fan, M., Teunissen, N., Hall, M. L., Hidalgo, A. N., Kingma, S. A., Roast, M., et al. (2018). From ornament to armament or loss of function? Breeding plumage acquisition in a genetically monogamous bird. *Journal of Animal Ecology, 87*(5), 1274–1285. https://doi.org/10.1111/1365-2656.12855.

Forsgren, E., Amundsen, T., Borg, Å. A., & Bjelvenmark, J. (2004). Unusually dynamic sex roles in a fish. *Nature, 429*(6991), 551–554. https://doi.org/10.1038/nature02562.

Gallup, G. G., Jr. (1968). Mirror-image stimulation. *Psychological Bulletin, 70*(6, Pt. 1), 782–793.

Griggio, M., Zanollo, V., & Hoi, H. (2010). Female ornamentation, parental quality, and competitive ability in the rock sparrow. *Journal of Ethology, 28*(3), 455–462. https://doi.org/10.1007/s10164-010-0205-5.

Hardouin, L. A., Reby, D., Bavoux, C., Burneleau, G., & Bretagnolle, V. (2007). Communication of male quality in owl hoots. *American Naturalist, 169*(4), 552–562. https://doi.org/10.1086/512136.

Hart, N. S. (2001). Variations in cone photoreceptor abundance and the visual ecology of birds. *Journal of Comparative Physiology A, 187*(9), 685–697.

Hirschenhauser, K., Wittek, M., Johnston, P., & Möstl, E. (2008). Social context rather than behavioral output or winning modulates post-conflict testosterone responses in Japanese quail (*Coturnix japonica*). *Physiology & Behavior, 95*(3), 457–463. https://doi.org/10.1016/j.physbeh.2008.07.013.

Holekamp, K. E., Smale, L., & Szykman, M. (1996). Rank and reproduction in the female spotted hyena. *Journal of Reproduction and Fertility, 108*(2), 229–237.

Hurd, P. L., & Enquist, M. (2005). A strategic taxonomy of biological communication. *Animal Behaviour, 70*(5), 1155–1170. https://doi.org/10.1016/j.anbehav.2005.02.014.

Kelber, A., Vorobyev, M., & Osorio, D. (2003). Animal colour vision–behavioural tests and physiological concepts. *Biological Reviews, 78*(1), 81–118.

Kraft, F.-L., Forštová, T., Urhan, A. U., Exnerová, A., & Brodin, A. (2017). No evidence for self-recognition in a small passerine, the great tit (*Parus major*) judged from the mark/mirror test. *Animal Cognition, 20*(6), 1049–1057. https://doi.org/10.1007/s10071-017-1121-7.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13.

Lande, R. (1980). Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution, 34*, 292–305. https://doi.org/10.1111/j.1558-5646.1980.tb04817.x.

LeBas, N. R. (2006). Female finery is not for males. *Trends in Ecology & Evolution, 21*(4), 170–173. https://doi.org/10.1016/j.tree.2006.01.007.

Leitão, A. V., Hall, M. L., Venables, B., & Mulder, R. A. (2019). Ecology and breeding biology of a tropical bird, the Lovely Fairy-Wren (*Malurus amabilis*). *Emu, 119*(1), 1–13. https://doi.org/10.1080/01584197.2018.1498745.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01.

Ligon, R. A., & McGraw, K. J. (2016). Social costs enforce honesty of a dynamic signal of motivation. *Proceedings of the Royal Society B: Biological Sciences, 283*(1841). https://doi.org/10.1098/rspb.2016.1873, 20161873.

May, H. Y., & Mercier, A. J. (2007). Duration of socialization influences responses to a mirror: Responses of dominant and subordinate crayfish diverge with time of pairing. *Journal of Experimental Biology, 210*(24), 4428–4436. https://doi.org/10.1242/jeb.011288.

McGraw, K., & Hill, G. E. (2000a). Plumage brightness and breeding-season dominance in the house finch: A negatively correlated handicap? *The Condor, 102*, 456–461. https://doi.org/10.1650/0010-5422(2000)102[0456:PBABSD]2.0.CO;2.

McGraw, K. J., & Hill, G. E. (2000b). Carotenoid-based ornamentation and status signaling in the house finch. *Behavioral Ecology, 11*(5), 520–527. https://doi.org/10.1093/beheco/11.5.520.

Medina, F. S., Taylor, A. H., Hunt, G. R., & Gray, R. D. (2011). New Caledonian crows' responses to mirrors. *Animal Behaviour, 82*(5), 981–993. https://doi.org/10.1016/j.anbehav.2011.07.033.

Midamegbe, A., Grégoire, A., Perret, P., & Doutrelant, C. (2011). Female–female aggressiveness is influenced by female coloration in blue tits. *Animal Behaviour, 82*(2), 245–253. https://doi.org/10.1016/j.anbehav.2011.04.020.

Mónus, F., Liker, A., Pénzes, Z., & Barta, Z. (2016). Status signalling in male but not in female Eurasian Tree Sparrows *Passer montanus*. *Ibis, 159*(1), 180–192. https://doi.org/10.1111/ibi.12425.

Murphy, T. G., Hernández-Muciño, D., Osorio-Beristain, M., Montgomerie, R., & Omland, K. E. (2009). Carotenoid-based status signaling by females in the tropical streak-backed oriole. *Behavioral Ecology, 20*(5), 1000–1006. https://doi.org/10.1093/beheco/arp089.

Murphy, T. G., West, J. A., Pham, T. T., Cevallos, L. M., Simpson, R. K., & Tarvin, K. A. (2014). Same trait, different receiver response: Unlike females, male American goldfinches do not signal status with bill colour. *Animal Behaviour, 93*(0), 121–127. https://doi.org/10.1016/j.anbehav.2014.04.034.

Ödeen, A., Pruett-Jones, S., Driskell, A. C., Armenta, J. K., & Håstad, O. (2012). Multiple shifts between violet and ultraviolet vision in a family of passerine birds with associated changes in plumage coloration. *Proceedings of the Royal Society B: Biological Sciences, 279*(1732), 1269–1276. https://doi.org/10.1098/rspb.2011.1777.

Oliveira, R. F., Carneiro, L. A., & Canário, A. V. (2005). Behavioural endocrinology: No hormonal response in tied fights. *Nature, 437*(7056), 207–208. https://doi.org/10.1038/437207a.

Olsson, P., Lind, O., Kelber, A., & Simmons, L. (2017). Chromatic and achromatic vision: Parameter choice and limitations for reliable model predictions. *Behavioral Ecology, 29*(2), 273–282. https://doi.org/10.1093/beheco/arx133.

Papadopoulos, N. T., Carey, J. R., Liedo, P., Mulle, H.-G., & Sentuerk, D. (2009). Virgin females compete for mates in the male lekking species *Ceratitis capitata*. *Physiological Entomology, 34*(3), 238–245. https://doi.org/10.1111/j.1365-3032.2009.00680.x.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2011). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-137 https://cran.r-project.org/web/packages/nlme.

Prior, H., Schwarz, A., & Güntürkün, O. (2008). Mirror-induced behavior in the magpie (*Pica pica*): Evidence of self-recognition. *PLoS Biology, 6*(8), e202. https://doi.org/10.1371/journal.pbio.0060202.

Prum, R. O. (2006). Anatomy, physics, and evolution of structural colors. In G. E. Hill, & K. J. McGraw (Eds.), *Bird coloration: Mechanisms and measurements* (Vol. I, pp. 295–355). Cambridge, MA: Harvard University Press.

Pryke, S. R. (2007). Fiery red heads: Female dominance among head color morphs in the Gouldian finch. *Behavioral Ecology, 18*(3), 621–627. https://doi.org/10.1093/beheco/arm020.

Pryke, S., & Griffith, S. C. (2006). Red dominates black: Agonistic signalling among head morphs in the colour polymorphic Gouldian finch. *Proceedings of the Royal Society B: Biological Sciences, 273*(1589), 949–957. https://doi.org/10.1098/rspb.2005.3362.

Reedy, A. M., Pope, B. D., Kiriazis, N. M., Giordano, C. L., Sams, C. L., Warner, D. A., et al. (2017). Female anoles display less but attack more quickly than males in response to territorial intrusions. *Behavioral Ecology, 28*(5), 1323–1328. https://doi.org/10.1093/beheco/arx095.

Rohwer, S. A. (1975). The social significance of avian winter plumage variability. *Evolution, 29*(4), 593–610. https://doi.org/10.1111/j.1558-5646.1975.tb00853.x.

Rohwer, S. A., & Rohwer, F. C. (1978). Status signalling in Harris sparrows: Experimental deceptions achieved. *Animal Behaviour, 26*, 1012–1022. https://doi.org/10.1016/0003-3472(78)90090-8.

Rosvall, K. A. (2008). Sexual selection on aggressiveness in females: Evidence from an experimental test with tree swallows. *Animal Behaviour, 75*(5), 1603–1610. https://doi.org/10.1016/j.anbehav.2007.09.038.

Rosvall, K. A. (2011). Intrasexual competition in females: Evidence for sexual selection? *Behavioral Ecology, 22*(6), 1131–1140. https://doi.org/10.1093/beheco/arr106.

Rowley, I. R., & Russell, E. (1997). *Fairy-wrens and grasswrens: Maluridae*. Oxford, U.K.: Oxford University Press.

Ruzzante, D. E. (1992). Mirror image stimulation, social hierarchies, and population differences in agonistic behaviour: A reappraisal. *Canadian Journal of Fisheries and Aquatic Sciences, 49*, 1966–1968. https://doi.org/10.1139/f92-218.

Sandell, M. I., & Smith, H. G. (1997). Female aggression in the European starling during the breeding season. *Animal Behaviour, 53*(1), 13–23. https://doi.org/10.1006/anbe.1996.0274.

Searcy, W. A., & Nowicki, S. (2000). Male–male competition and female choice in the evolution of vocal signaling. In Y. Espmark, T. Amundsen, & G. Rosenqvist (Eds.), *Animal signals: Signalling and signal design in animal comunication* (pp. 301–315). Trondheim, Norway: Tapir.

Searcy, W. A., Anderson, R. C., & Nowicki, S. (2008). Is bird song a reliable signal of aggressive intent? A reply. *Behavioral Ecology and Sociobiology, 62*(7), 1213–1216. https://doi.org/10.1007/s00265-008-0569-5.

Senar, J. C. (1999). Plumage coloration as a signal of social status. In N. J. Adams, & R. H. Slotow (Eds.), *Proceedings of the International Ornithological Congress* (Vol. 22, pp. 1669–1686). Johannesburg, South Africa: Birdlife South Africa.

Senar, J. C. (2006). Color displays as intrasexual signals of aggression and dominance. In G. E. Hill, & K. J. McGraw (Eds.), *Bird coloration: Function and evolution* (Vol. II, pp. 87–136). Cambridge, MA: Harvard University Press.

Sinn, D., While, G., & Wapstra, E. (2008). Maternal care in a social lizard: Links between female aggression and offspring fitness. *Animal Behaviour, 76*(4), 1249–1257. https://doi.org/10.1016/j.anbehav.2008.06.009.

Soler, M., Pérez-Contreras, T., & Peralta-Sánchez, J. M. (2014). Mirror-mark tests performed on jackdaws reveal potential methodological problems in the use of stickers in avian mark-test studies. *PLoS One, 9*(1), e86193. https://doi.org/10.1371/journal.pone.0086193.

Stutchbury, B. J., & Morton, E. S. (2001). *Behavioral ecology of tropical birds.* London, U.K.: Academic Press.

Svendsen, G. E., & Armitage, K. B. (1973). Mirror-image stimulation applied to field behavioral studies. *Ecology, 54*(3), 623–627. https://doi.org/10.2307/1935349.

Tibbetts, E. A. (2013). The function, development, and evolutionary stability of conventional signals of fighting ability. *Advances in the Study of Behavior, 45*, 49–80. https://doi.org/10.1016/B978-0-12-407186-5.00002-1.

Tibbetts, E. A. (2014). The evolution of honest communication: Integrating social and physiological costs of ornamentation. *Integrative and Comparative Biology, 54*(4), 578–590. https://doi.org/10.1093/icb/icu083.

Tibbetts, E. A., & Dale, J. (2004). A socially enforced signal of quality in a paper wasp. *Nature, 432*, 218–222. https://doi.org/10.1038/nature02949.

Tibbetts, E. A., & Izzo, A. (2010). Social punishment of dishonest signalers caused by mismatch between signal and behavior. *Current Biology, 20*(18), 1637–1640.

Tobias, J. A., Montgomerie, R., & Lyon, B. E. (2012). The evolution of female ornaments and weaponry: Social selection, sexual selection and ecological competition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1600), 2274–2293. https://doi.org/10.1098/rstb.2011.0280.

Vaillancourt, T. (2013). Do human females use indirect aggression as an intrasexual competition strategy? *Philosophical Transactions of the Royal Society B: Biological Sciences, 368*(1631), 20130080. https://doi.org/10.1098/rstb.2013.0080.

Vorobyev, M., & Osorio, D. (1998). Receptor noise as a determinant of colour thresholds. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265*(1394), 351–358. https://doi.org/10.1098/rspb.1998.0302.

Vorobyev, M., Osorio, D., Bennett, A. T. D., Marshall, N. J., & Cuthill, I. C. (1998). Tetrachromacy, oil droplets and bird plumage colours. *Journal of Comparative Physiology A: Neuroethology Sensory Neural and Behavioral Physiology, 183*(5), 621–633. https://doi.org/10.1007/s003590050286.

Wallace, A. R. (1891). *Natural selection and tropical nature: Essays on descriptive and theoretical biology.* London, U.K.: Macmillan.

Webster, M. S., Ligon, R. A., & Leighton, G. M. (2018). Social costs are an underappreciated force for honest signalling in animal aggregations. *Animal Behaviour, 143*, 167–176. https://doi.org/10.1016/j.anbehav.2017.12.006.

West-Eberhard, M. J. (1979). Sexual selection, social competition and evolution. *Proceedings of the American Philosophical Society, 123*(4), 222–234.

While, G. M., Sinn, D. L., & Wapstra, E. (2009). Female aggression predicts mode of paternity acquisition in a social lizard. *Proceedings of the Royal Society of London B: Biological Sciences, 276*(1664), 2021–2029. https://doi.org/10.1098/rspb.2008.1926.

Wilson, A. J., de Boer, M., Arnott, G., & Grimmer, A. (2011). Integrating personality research and animal contest theory: Aggressiveness in the green swordtail *Xiphophorus helleri. PLoS One, 6*(11), e28024. https://doi.org/10.1371/journal.pone.0028024.

Wolfenbarger, L. L. (1999). Is red coloration of male Northern Cardinals beneficial during the nonbreeding season?: A test of status signaling. *The Condor, 101*(3), 655–663.

Yasukawa, K., Butler, L. K., & Enstrom, D. A. (2009). Intersexual and intrasexual consequences of epaulet colour in male red-winged blackbirds: an experimental approach. *Animal Behaviour, 77*(2), 531–540. https://doi.org/10.1016/j.anbehav.2008.11.013.

Young, C. M., Cain, K. E., Svedin, N., Backwell, P. R. Y., & Pryke, S. R. (2016). The role of pigment based plumage traits in resolving conflicts. *Journal of Avian Biology, 47*(2), 167–175. https://doi.org/10.1111/jav.00742.

Young, C. M., Cain, K. E., Svedin, N., Backwell, P. R. Y., & Pryke, S. R. (2017). Predictors of aggressive response towards simulated intruders depend on context and sex in Crimson Finches (*Neochmia phaeton*). *Behavioural Processes, 138*, 41–48. https://doi.org/10.1016/j.beproc.2017.01.017.

Appendix

Experiment 2: Treatment effect

Descriptors of treatments are given in Table A1. Treatments mostly varied in the z axis and most or half of the individuals fall within the natural variation (Table A1).

To determine how close each manipulated colour was to the natural range, we analysed the differences between each treated colour and the nearest premanipulated one in colour space given by Euclidean distance and expressed in JND units (just noticeable differences), with values below 1 JND thought not to be discriminable (Kelber et al., 2003). In both females and males, the Euclidean distances average for 'blender' and 'grey' were close to or below 1 JND, meaning that painted colours resembled natural colours, but the 'blue' treatment average was higher than 2 JND, meaning that this treatment may be perceived as different from the natural colour (Table A2). Overall, treatment 1 'blender' maintained coloration, treatment 2 'grey' increased coloration within the natural range and treatment 3 'blue' increased coloration within and outside the natural range (Tables A1 and A2; see Methods).

Chromatic variation of the cheek patch postmanipulation

From experiment 2, having found a relationship between cheek patch chromatic variation and aggression score (see Results), we tested whether aggression score ($PC1_{aggression2}$) was influenced by chromatic variation of the cheek patch postmanipulation and included as fixed effects sex and its interaction with colour as a dependent variable and group identity as a random effect. To understand whether response to the mirror test was influenced by the interaction between the individual's natural colour and the treatment applied, we ran another model with the same terms but using as fixed factors the PCA of the chromatic difference between before and after treatment colour (see Table A4 for PCA loadings).

Results showed that chromatic variation of the cheek patch postmanipulation was positively correlated with aggression score (Table A14, Fig. A4), suggesting that enhanced plumage (colour of 'rival' in mirror) elicited a more aggressive response from test subjects. Males and females did not differ in their overall aggressive response, and the interaction between colour and sex was marginally nonsignificant (Table A14, Fig. A4).

Colour treatment controlling for individual colour (difference between natural colour and manipulated colour) showed again that individuals with increased plumage reflection were more aggressive towards their enhanced mirror reflection; sex and its interaction with colour difference was marginally related to aggression (Table A14, Fig. A4).

Differences between Experiments

To analyse differences between experiment type, we built two GLMMs to test whether aggression (PCA aggression computed on both experiments: $PC1_{aggression1+2}$; Table 1), or natural colour of the cheek patch (PCA chroma computed on both experiments: $PC1_{chroma1+2}$; Table A4), varied between experiments 1 and 2. We included bird and group identity as random factors to account for nonindependence in the data.

Aggression score did not vary between experiments ($t_{13.91} = -0.62$, $P = 0.55$), nor did the natural colour of the cheek patch ($t_{3.11} = 1.25$, $P = 0.29$).

Results excluding individuals whose colour was outside the natural range

If we exclude the individuals that are outside the natural range ($N = 9$ excluded, five from female 'blue' treatment, three from male 'blue' treatment and one from 'grey' treatment) results are similar to those reported in the main text (Table 3): treatment type was related to aggression towards the mirror and males tended to be more aggressive than females (Table A13). Individuals with colour expression increased in treatment 2 'grey' were more aggressive than those in treatment 1 'blender'; individuals in treatment 'blue' were more aggressive than individuals from treatment 1 'blender' although because of the smaller sample size this difference was marginally nonsignificant (Table A13).

Table A1
Descriptors of visual coordinates for experiment 2

		Female				Male			
		Mean	Minimum	Maximum	Within natural range (%)	Mean	Minimum	Maximum	Within natural range (%)
Premanipulation (N=35 F, 29 M)	x	−6.993	−7.57	−6.605		−6.479	−6.917	−5.981	
	y	1.697	−0.21	3.253		−1.995	−3.599	−0.482	
	z	−5.561	−12.94	−2.703		−7.559	−10.31	−6.139	
Blender (N=12 F, 10 M)	x	−6.985	−7.497	−6.615	100	−6.738	−7.129	−6.456	90
	Y	2.821	1.025	4.252	83	−1.379	−3.125	0.727	90
	z	−5.392	−11.513	−2.302	92	−7.939	−11.91	−5.452	80
Grey (N=13 F, 11 M)	x	−7.118	−7.437	−6.649	100	−6.805	−7.137	−7.137	73
	Y	2.607	0.884	5.162	69	−1.754	−2.783	−2.783	100
	z	−5.317	−9.592	−2.888	100	−9.199	−13.22	−13.219	82
Blue (N=10 F, 8 M)	x	−7.692	−8.387	−7.241	60	−7.137	−8.488	−6.721	50
	Y	−0.11	−2.899	1.882	60	−2.783	−5.914	−2.195	75
	z	−15.86	−27.613	−6.801	50	−13.22	−21.45	−9.931	63

F: females; M: males.

Table A2
Detectability of differences in colour treatment

	Female			Male		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Premanipulation−blender	0.893	0.254	1.495	0.729	0.104	1.746
Premanipulation−grey	0.765	0.296	1.962	0.988	0.208	2.955
Premanipulation−blue	5.334	0.665	14.766	2.859	0.281	11.208

Differences in colour space between each manipulated colour and the nearest premanipulated colour, given by Euclidean distance and expressed in JND units (just noticeable differences), with values below 1 JND thought not to be discriminable (Kelber et al., 2003).

Table A3
Female and male matrix correlations of PCA xyz chromatic coordinates of the cheek before and after the manipulation for experiment 2

		All		Female		Male	
		Correlation (r)	P	Correlation (r)	P	Correlation (r)	P
Back	Cheek before	0.27	0.03	0.69	<0.001	0.14	0.49
	Cheek after	0.61	<0.001	0.75	<0.001	0.05	0.80
Head	Cheek before	0.56	<0.001	0.711	<0.001	0.21	0.28
	Cheek after	0.65	<0.001	0.77	<0.001	0.17	0.39
Shoulder	Cheek before	−0.29	0.02	0.51	0.001	0.02	0.92
	Cheek after	0.04	0.7	0.58	<0.001	0.02	0.92
Tail	Cheek before	0.39	0.001	0.64	<0.001	0.09	0.65
	Cheek after	0.28	0.02	0.72	<0.001	0.02	0.92
Body size	Cheek before	0.53	<0.001	0.16	0.34	0.18	0.36
	Cheek after	0.26	0.03	0.18	0.30	0.19	0.35

Female N = 35, male N = 29.

Table A4
Loadings for principal component analysis (PC1$_{chroma}$) of xyz colour for experiments 1 and 2 and both experiments combined, with sexes together (used in the main analysis)

	Variables	Experiment 1 PC1$_{chroma1}$	Experiment 2 PC1$_{chroma2}$	Experiment 1+2 PC1$_{chroma1+2}$
Cheek	x	0.112	0.063	0.073
	y	−0.708	−0.684	−0.711
	z	−0.697	−0.727	−0.699
	Variance (%)	83	83	85
Back	x	0.099		−0.067
	y	−0.771		−0.093
	z	−0.629		−0.993
	Variance (%)	92		92
Head	x	0.0524		−0.0519
	y	−0.605		−0.173
	z	−0.795		−0.983
	Variance (%)	90		92
Tail	x	−0.234		−0.047
	Y	−0.032		−0.018
	z	−0.972		−0.999
	Variance (%)	92		99
Shoulder	x	0.607		0.441
	y	−0.603		−0.462
	z	−0.518		−0.769
	Variance (%)	92		58

Table A4 (*continued*)

	Variables	Experiment 1 PC1$_{chroma1}$	Experiment 2 PC1$_{chroma2}$	Experiment 1+2 PC1$_{chroma1+2}$
Cheek postmanipulation	x		−0.046	
	y		−0.313	
	z		−0.949	
	Variance (%)		89	
Cheek difference	x		−0.065	
	y		−0.232	
	z		−0.971	
	Variance (%)		95	

Cheek difference: difference between before and after treatment. Variable x represents the relative stimulation of the S cone in relation to the VS cone; y represents relative stimulation of the M cone in relation to the VS and S cones; and z axis represents the relative stimulation of the L cone in relation to the VS, S and M cones. For all blue colours (cheek, back, head, tail), negative values explained variation in y and z; spectra with high PC1$_{chroma}$ values provide low stimulation of L cone relative to VS + S + M cones, and M relative to S + VS, i.e. high PC1 values had colours richer in shorter wavelengths (UV/blue) and poorer in longer wavelengths (red). For red/brown colours (shoulder), positive values explain variation in x, and negative values explain variation in y. Thus, low PC1 values represent high stimulation of M relative to S + VS.

Table A5
Correlation between colour patches (PCA xyz chromatic coordinates) in females and males

		Female		Male	
		Correlation (r)	P	Correlation (r)	P
Cheek	Back	0.61	<0.0001	0.31	0.03
	Head	0.62	<0.0001	0.37	0.01
	Tail	0.53	<0.0001	0.34	0.01
	Shoulder	0.46	<0.001	0.11	0.46
Back	Head	0.96	<0.0001	0.52	<0.0001
	Tail	0.91	<0.0001	0.48	<0.0001
	Shoulder	0.74	<0.0001	0.03	0.81
Head	Tail	0.88	<0.0001	0.89	<0.0001
	Shoulder	0.69	<0.0001	0.09	0.55
Tail	Shoulder	0.49	<0.0001	0.22	0.14

Female $N = 58$, male $N = 48$.

Table A6
Comparison of female and male plumage homologous colour patches

	Pillai	F	df	P
Cheek	0.85	196.22	3,10	<0.0001
Back	0.82	160.12	3,10	<0.0001
Head	0.88	239.56	3,10	<0.0001
Tail	0.14	5.3375	3,10	<0.001
Shoulder	0.86	205.66	3,10	<0.0001

The table shows the results of a multivariate analysis of chromatic coordinates (xyz) for each colour patch.

Table A7
Predictors of aggressive behaviour of females and males in response to the mirror image stimulation test

Model order	Predictors of aggression	Intercept	Loglik	AIC$_c$	ΔAICc	$\dot{\omega}_i$
1	Cheek colour+Sex	1.23	−63.06	137.78	0	0.38
2	Sex	0.36	−65.48	140.04	2.26	0.12
3	Body size+Cheek colour+Sex	0.87	−60.57	141.51	3.73	0.06
4	Density+Cheek colour+Sex	−0.17	−62.11	141.51	3.73	0.06
5	Breeding stage+Cheek colour+Sex	1.21	−63.89	142.18	4.39	0.04

Summary of model selection showing the main predictors of aggression score, using the Akaike information criterion corrected for small sample size (AICc) in package 'MuMIn' (Barton, 2017). Loglik: restricted log-likelihood; ΔAICc: difference between the best model and the given model; $\dot{\omega}_i$: Akaike weight of a given model. Group identity was also included as a random factor. The models with the lowest AICc and ΔAICc <2 were retained and considered to be the best supported (Burnham & Anderson, 2002).

Table A8
Loading for principal component analysis (PC1$_{chroma}$) of xyz colour for experiments 1 and 2 and both experiments for each sex

	Variables	Experiment 1 PC1$_{chroma1}$		Experiment 2 PC1$_{chroma2}$		Experiment 1+2 PC1$_{chroma1+2}$	
		F	M	F	M	F	M
Cheek	X	−0.005	0.109	−0.038	0.048	−0.038	0.058
	y	−0.485	−0.534	−0.288	−0.562	−0.305	0.529
	z	−0.874	−0.838	−0.956	−0.826	−0.951	0.847
	Variance (%)	87	88	95	78	92	81
Back	x	−0.109	0.063			−0.063	−0.022
	Y	−0.263	−0.583			−0.102	0.479
	z	−0.959	−0.810			−0.993	0.877
	Variance (%)	90	91			99	89

(*continued on next page*)

Table A8 (*continued*)

| | Variables | Experiment 1 PC1$_{chroma1}$ | | Experiment 2 PC1$_{chroma2}$ | | Experiment 1+2 PC1$_{chroma1+2}$ | |
		F	M	F	M	F	M
Head	x	−0.108	−0.031			−0.120	−0.026
	y	−0.263	−0.353			−0.072	−0.114
	z	−0.959	−0.935			−0.990	−0.993
	Variance (%)	93	91			98	99
Tail	x	−0.209	−0.266			−0.098	−0.045
	y	−0.039	−0.012			−0.043	−0.016
	z	−0.977	−0.964			−0.994	−0.998
	Variance (%)	96	87			99	99
Shoulder	x	0.674	0.752			−0.310	0.397
	y	−0.649	−0.565			0.034	0.284
	z	−0.352	−0.338			0.949	−0.872
	Variance (%)	63	99			96	54
Cheek postmanipulation	x			−0.052	−0.073		
	y			−0.226	−0.317		
	z			−0.973	−0.946		
	Variance (%)			98	92		
Cheek difference	x			−0.061	−0.085		
	Y			−0.197	−0.360		
	z			−0.978	−0.928		
	Variance (%)			97	93		

F: female; M: male. Cheek difference: difference between before and after treatment. Variable x represents the relative stimulation of the S cone in relation to the VS cone; y represents relative stimulation of the M cone in relation to the VS and S cones; and z represents the relative stimulation of the L cone in relation to the VS, S and M cones. For all blue colours (cheek, back, head, tail), high PC1 values had colours richer in shorter wavelengths (UV/blue) and poorer in longer wavelengths (red). For red/brown colours (shoulder), low PC1 values represent high stimulation of M relative to S + VS.

Table A9
Results of principal component analysis for body size (aggression, time spent close to the mirror, time spent far from the mirror), for each sex

	Female	Male
Tarsus length	0.40	0.52
Wing length	0.65	0.62
Head-bill length	0.64	0.59
Eigenvalues (% of variance)	1.16 (45)	1.21 (49)

Table A10
Results of principal component analysis (PCA) of activity variables (aggression, time spent close to the mirror, time spent far from the mirror) in the mirror image stimulation test, after uncovering the mirror for experiments 1 and 2, and both experiments combined, for each sex

| | Experiment 1 PC1$_{aggression1}$ | | Experiment 2 PC1$_{aggression2}$ | | Experiment 1 and 2 PC1$_{aggression1+2}$ | |
	Female	Male	Female	Male	Female	Male
Aggression (pecks+swoops+displays)	0.56	0.49	0.60	0.52	0.58	0.48
Time spent close to the mirror	0.58	0.78	0.62	0.63	0.61	0.67
Time spent far from the mirror	−0.60	−0.39	−0.50	− 0.59	−0.54	−0.56
Eigenvalues (% of variance)	62	41	70	65	65	55

Table A11
Experiment 1: summary of linear model showing all parameters of aggression score (PC1$_{aggression1}$), for each sex

| Parameter | Female | | | | Male | | | |
	β	SE	t	P	B	SE	t	P
Fixed effects								
Intercept	−155	158	−0.98	0.35	7.91	162.24	0.05	0.96
Body size	−1.97e-02	0.33	−0.06	0.95	−0.14	0.42	−0.33	0.75
Cheek colour	−0.59	0.40	−1.47	0.17	−0.35	0.46	−0.76	0.47
Head colour	0.10	0.24	0.42	0.68	0.04	0.29	0.15	0.89
Shoulder colour	0.44	0.46	0.95	0.36	0.011	0.06	0.18	0.86
Tail colour	−2.11e-02	0.14	−0.15	0.88	4.28e-03	0.12	0.04	0.97
Breeding stage[a]								
Nest	−0.65	1.93	−0.33	0.74	0.18	1.66	0.11	0.92
Fledglings	−2.32	2.20	−1.06	0.31	−0.39	2.27	−0.17	0.87
Unknown	−0.29	1.31	−0.22	0.83	0.53	1.89	0.28	0.79
No. of neighbours	−5.81e-02	0.28	−0.21	0.84	0.25	0.43	0.58	0.58
Julian day	1.01e-02	1.04e-02	0.98	0.35	−6.73e-04	0.01	−0.06	0.95
Time of day	4.16e-04	1.49e-03	0.28	0.78	1.68e-03	2.18e-03	0.77	0.47

Female $N = 22$, male $N = 19$.
Global model: Aggression score ~ Cheek colour + Head colour + Shoulder colour + Tail colour + Body size + Breeding stage + Nr. Neighbours + Julian day + Time of day.
[a] Breeding stage is a categorical term with four levels: no nest, unknown, nest and fledglings; the reference is no nest.

Table A12
Experiment 2: effect of fixed predictors on aggression score (PC1$_{aggression2}$), models built for each sex

Parameter	Female					Male				
	β	SE	df	t	P	β	SE	df	t	P
Natural colour										
Intercept	−0.89	0.25	1,33	0	1.0	7.18e-11	0.26	1,27	0	1.0
Cheek colour	0.14	0.10	1,33	1.39	0.18	0.04	0.21	1,27	0.19	0.84
Treatment										
Intercept	−0.82	0.39	2,32	−2.05	0.04	−0.44	0.45	2,26	−0.98	0.34
Treatment[a]										
Blue	1.29	0.59	2,32	2.19	0.03	0.78	0.68	2,26	1.15	0.26
Grey	1.21	0.55	2,32	2.18	0.03	0.60	0.63	2,26	0.96	0.34
Colour after treatment										
Intercept	−8.32e-11	0.25	1,33	0	1.0	5.18e-11	0.24	1,27	0	1.0
Cheek colour	0.04	0.03	1,33	1.16	0.25	0.16	0.07	1,27	2.15	0.04
Colour difference										
Intercept	−8.70e-11	0.24	1,33	0	1.0	5.55e-11	0.24	1,27	0	1.0
Cheek colour	0.04	0.05	1,33	0.90	0.37	0.19	0.08	1.27	2.30	0.02

Female $N = 35$, male $N = 29$.

[a] Treatment is a categorical term with three levels: 'blender', 'grey' and 'blue'; the reference is 'blender'.

Table A13
Effect of fixed predictors on aggression score (PC1$_{aggression2}$) in experiment 2 for treatment and comparison between treatments, excluding individuals that were outside the natural range and had detectable differences in colour (see Tables A1 and A2)

| Parameter	β	SE	df	t	P
Intercept	−1.10	0.39	48.86	−2.76	**0.007**
Treatment[a]					
Blue	1.30	0.73	48.98	1.77	0.08*
Grey	1.39	0.54	45.48	2.55	**0.01**
Sex[b]	1.08	0.55	30.17	1.96	0.059*
Treatment[a]∗Sex[b]					
Blue∗Sex	−1.23	1.00	36.13	−1.22	0.22
Grey∗Sex	−1.13	0.79	32.84	−1.43	0.16

Significant predictors are in bold; an asterisk indicates marginal significance. Female $N = 30$, male $N = 25$. Models were performed using packages lme4 (Bates & Maechler, 2010) and the significance of factors and degrees of freedom were assessed using the 'lmerTest' package (Kuznetsova et al., 2017).

[a] Treatment is a categorical term with three levels: 'blender', 'grey' and 'blue'; the reference is 'blender'.

[b] Sex is a categorical term; the reference is female.

Figure A1. Male and female lovely fairy-wrens are dichromatic. Blue colours are a result of coherent light scattering of feather barbs (structural colours; Prum, 2006), and the red and blacks are a result of melanin pigment deposition, identified based on the shape of reflectance spectra described in Delhey (2015). Photo credits Ana V. Leitão.

Table A14
Effect of fixed predictors (colour postmanipulation and the difference between natural colour and resultant colour) on aggression score (PC1$_{aggression2}$) in experiment 2 for both sexes

| Parameter	Colour after treatment					Colour difference				
	β	SE	df	t	P	β	SE	df	t	P
Intercept	0.01	0.27	59.66	0.04	0.97	0.34	0.25	59.93	1.34	0.19
Cheek colour	0.19	0.08	54.96	2.37	**0.02**	0.23	0.09	52.29	2.63	**0.01**
Sex[a]	−0.18	0.34	37.78	−0.52	0.61	−0.56	0.31	34.89	−1.78	0.08*
Cheek colour∗Sex[a]	−1.15	0.09	56.26	−1.71	0.09*	−0.19	0.09	52.81	−1.89	0.06*

Significant predictors are in bold; an asterisk indicate marginal significance. Female $N = 35$, male $N = 29$. Models were performed using packages lme4 (Bates & Maechler, 2010) and the significance of factors and degrees of freedom (Satterthwaite's method for approximating degrees of freedom) were assessed using the 'lmerTest' package (Kuznetsova et al., 2017).

[a] Negative coefficients indicate that females are less aggressive.
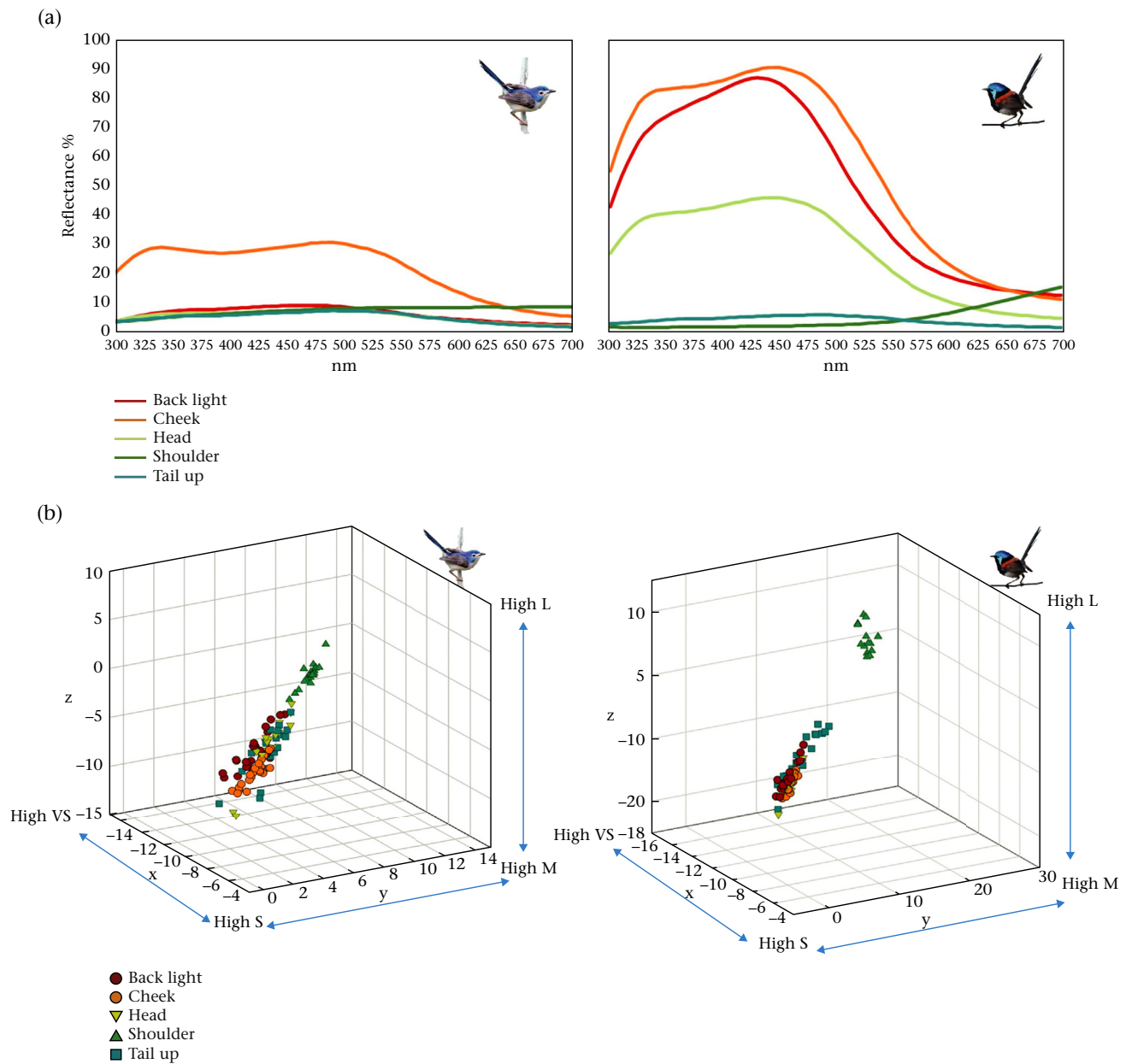
Figure A2. Natural variability of each plumage patch in female and male lovely fairy-wrens ($N = 23$ females and 19 males). (a) The average reflectance spectra for each colour patch and for each sex. (b) Graphical representation of chromatic coordinates (xyz) that define their position in the avian visual space. Here, x represents the relative stimulation of the S cone in relation to the VS cone, y represents relative stimulation of the M cone in relation to the VS and S cones and z represents the relative stimulation of the L cone in relation to the VS, S and M cones. Sample from experiment 1.
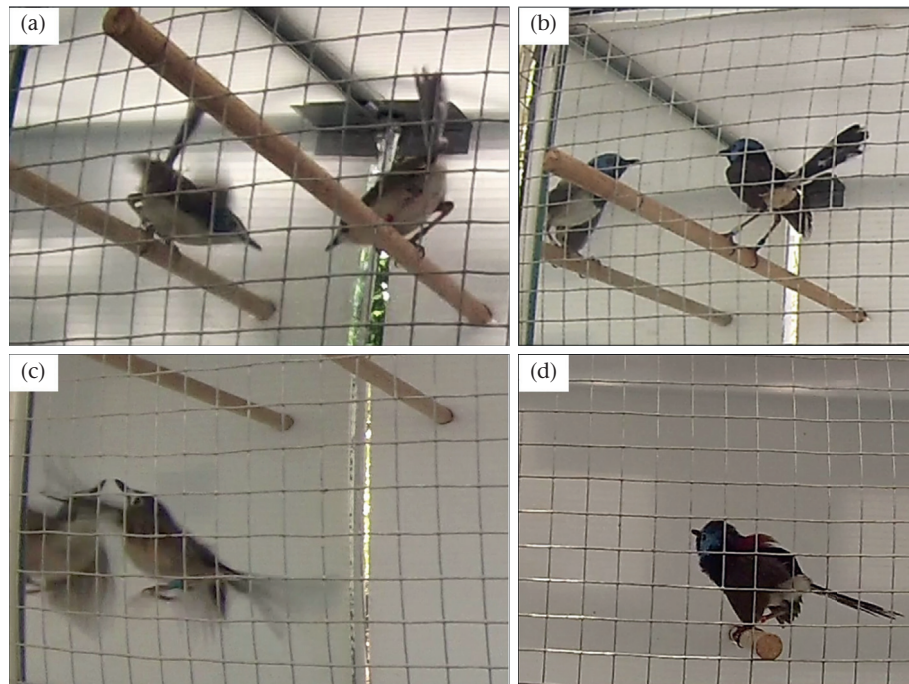
Figure A3. Female and male aggressive displays towards the mirror 'rival': (a) female pecking the mirror, (b) male displaying (wings extended) and pecking, (c) female swooping at the mirror and (d) male displaying towards the mirror (feathers erected, wings extended). Photo credits: Ana V. Leitão.
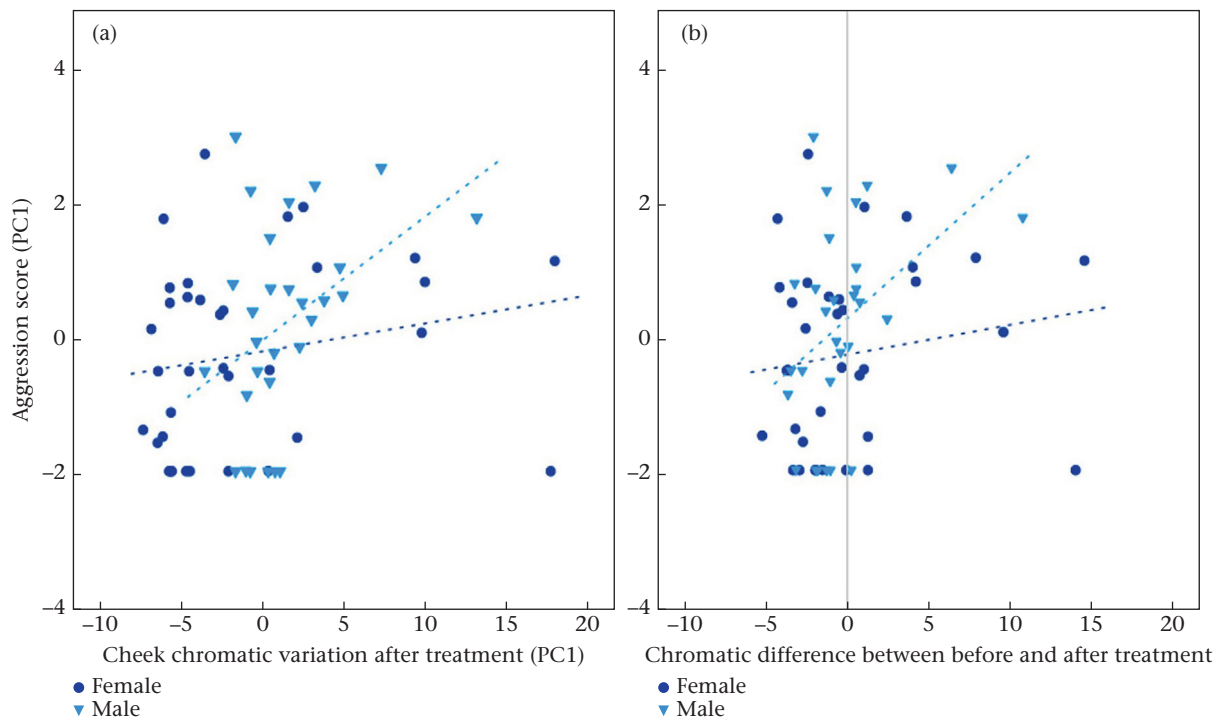
Figure A4. Relation between aggression score and chromatic variation in experiment 2: (a) chromatic variation in cheek colour (PCAxyz) post-treatment; (b) within-individual colour changes between natural colour and treatment. Grey dashed vertical line at 0 indicates no difference in colour (between natural and treatment), negative values indicate decreased colour and positive values indicate increased colour. In both plots, the x axis scale is represented in JNDs (just noticeable differences). Dashed lines depict the linear regression lines.