

Within-population variation in preference functions reveals substantial among-female disagreement in mate assessment

Kane Stratman and Gerlinde Höbel 

Behavioral and Molecular Ecology Group, Department of Biological Sciences, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, United States

Handling editor: Alejandro Gonzalez-Voyer, Associate editor: Trine Bilde

Corresponding author: Gerlinde Höbel, Behavioral and Molecular Ecology Group, Department of Biological Sciences, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, United States. Email: hoebel@uwm.edu

Abstract

The mate choice behaviours of females can greatly affect patterns of reproductive success in males and influence the evolution of sexually selected male traits. Population-level estimates of display preferences may provide an accurate estimate of the strength and direction of selection by female choice if all females in the population show homogeneous preferences. However, population-level estimates may yield misleading estimates if there is within-population variation in mate preferences. While it is increasingly clear that the latter situation is common in nature, empirical data on the magnitude of variation in female preferences are required to improve our current understanding of its potential evolutionary consequences. We explored variations in female preference functions for 3 male call properties in a treefrog. We document substantial within-population variation not only in peak preferences but also in preference function shape (open, closed, flat), with at best 62% of females sharing a preference function shape with the respective population curve. Our findings suggest that population curves may accurately capture the direction of sexual selection, but depending on the properties of the constituting individual functions they may over- or underestimate the strength of selection. Particularly population estimates suggesting weak selection may in fact hide the presence of individual females with strong but opposing preferences. Moreover, due to the high within-population variation in both peak preferences and preference function shapes, the population functions drastically underestimate the predicted variation in male mating success in the population.

Keywords: *Hyla versicolor*, gray treefrog, mate choice, sexual selection, call duration, call period, call frequency

Introduction

Mate choice is thought to play a pivotal role in the evolution of elaborate, conspicuous, and often costly displays (Andersson, 1994; West-Eberhard, 1983), and the degree of within-population variation in mate preferences should have major consequences for the evolution of sexually selected male traits. Sustained sexual selection arising from homogeneous female preferences is expected to reduce genetic variation in sexually selected male traits (Rowe & Houle, 1996), while extensive within-population variation in mate preferences is thought to maintain variation in male traits (Day, 2000; Jennions & Petrie, 1997; Widemo & Sæther, 1999). There is growing evidence that females frequently do not share mating preferences (e.g., Brooks & Endler, 2001; Feagles & Höbel, 2022a; Kelly, 2018; Neelon, et al., 2019; Qvarnström et al., 2000; Rodríguez et al., 2013), putting a premium on assessing the magnitude of variation in female preferences in order to improve our current understanding of its potential evolutionary consequences (Jennions & Petrie, 1997).

Preference functions, which illustrate how females rank display property values across a range of variations in display features are a useful approach for visualizing and quantifying female preferences (Andersson & Simmons, 2006; Ritchie, 1996; Shaw & Herlihy, 2000; Wagner, 1998). While preference curves can quickly reveal qualitative differences in *shape*

(open-ended curves that favour extreme trait values versus closed curves that favour intermediate ones), curve-fitting software (PFunc; Kilmer et al., 2017) can additionally provide quantitative metrics to explore preference functions in more detail (see Figure 1). Preference functions can be characterized by “*peak*” preference (the most preferred display trait value), “*tolerance*” (the range of relatively attractive display values), “*strength*” (the degree to which attractiveness decreases as display values change away from the peak), and “*responsiveness*” (average response across the entire function) (Kilmer et al., 2017).

Characterizations of female mate preference functions constitute hypotheses about the form of sexual selection on male displays. The peak of the preference function predicts the display value with the greatest reproductive success, while other aspects of the preference function predict how attractiveness varies as display values deviate from the peak. Relating the peak preference and overall shape of the function to the distribution of display values in the population predicts whether stabilizing or directional selection is expected (Gerhardt et al., 2000; Ritchie, 1996; Rodríguez et al., 2006; Shaw & Herlihy, 2000; Wagner, 1998).

While we can learn much about the strength and direction of sexual selection by female choice by establishing population preferences, these group-level functions work on the

Received May 26, 2023; revised July 2, 2024; accepted September 6, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the European Society of Evolutionary Biology. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

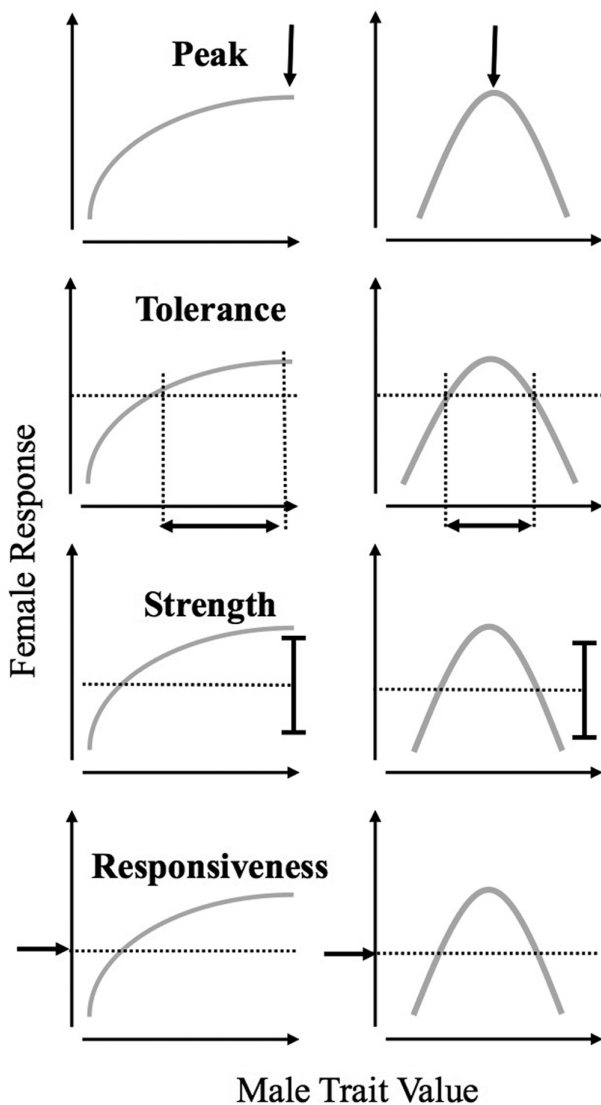


Figure 1. Preference functions illustrate variation in the attractiveness ranking of male traits. The shape of a preference function can be open (left), indicating that extreme values are preferred, or closed (right), indicating that intermediate values are most attractive. We extracted four preference function traits that describe the underlying curves: *peak* is the trait value eliciting the strongest response, *tolerance* describes the range of trait values within a 20% drop from peak attractiveness (i.e., relatively acceptable values), *strength* indicates the decline in attractiveness as values deviate from the peak, and *responsiveness* indicates the average response across all trait values.

implicit assumption that all females that contributed to the group function share the same function shape. Flattened population functions, which are interpreted as “weak” population preferences, may indeed be the result of all females having flat functions. But they may also be the outcomes of a mix of flat, closed, and open functions, or of a mix of strong open functions with opposing peaks that “cancel” one another out. Likewise, closed (or open) population functions might “hide” a certain number of others shapes provided that a sufficient number of strong closed (or open) functions are present. While within-population variation in peak preferences has received attention in the literature (Feagles & Höbel, 2022a; Kelly, 2018; Neelon, et al., 2019; Rodríguez et al., 2013), little consideration has been devoted to examining the presence or magnitude of within-population variation in function shapes.

Simulations have explored the evolutionary impact of within-population variation in different aspects of female mate preference functions. Millan et al. (2020) used individual-based models to predict how variation in preference function shape, peak preference, and preference strength (conceptualized more like what we term *tolerance* in our analysis; see Kilmer et al., 2017), influences the opportunity for sexual selection and the evolution of sexually selected male traits. Their simulations suggest that predictions about the effect of within-population variation in female preference require knowledge about (i) the shape of the preference function, (ii) the magnitude of variation in peak preference, and (iii) the magnitude of difference between female peak preference variation and male trait variation. Among their main findings was that exaggeration of male traits is maximized when high within-population preference variation combines with an open preference function shape, while maintenance of male trait variation is maximized when high within-population variation in peak preference combines with a closed preference function shape. The lack of empirical data, especially whether open or closed functions are more frequent in nature, made it impossible for Millan et al. (2020) to gauge which of the simulated scenarios is the most common in natural populations. Also, while their models implemented variation in peak preference and preference strength, all function shapes were set to either open or closed. This again assumes that all females in a population share a given function shape, a conjecture that is yet untested.

Working with Eastern Gray Treefrogs (*Hyla versicolor*), — one of the case study examples of mate choice research—, we examined within-population preference variation for three properties of the species’ advertisement call: call duration, call period, and call dominant frequency. We paid particular attention to within-population variation in preference function shape. By generating preference functions for large samples of individual females, we aimed to assemble a high-resolution characterization of the within-population variation in the call preferences of a species with well-known population-based preferences. We use this data set to test the hypothesis that individual preference functions generally align with the population’s average preference. In accordance with the literature on eastern gray treefrogs (i.e., Gerhardt, 2005a, b; Gerhardt & Brooks, 2009; Gerhardt et al., 2000; Reichert & Höbel, 2015), this hypothesis predicts that individual call duration preference functions should be open-ended favouring longer duration calls, individual call period preference functions should be open-ended favouring shorter call periods, and individual call frequency preference functions should be closed favouring intermediate values. We also calculate the difference (“mismatch”) between both the population preference estimates and the individual female preference estimates with the male call trait value to compare whether either approach yields similar hypotheses about male mating success and the form of sexual selection on male traits. Finally, using data on within-population variation in function shape and the magnitude of variation in peak preference compared to the magnitude of variation in male call traits, we compare our empirical data with the simulated scenarios in Millan et al. (2020).

Methods

Focal species and collection site

Eastern gray treefrogs (*Hyla versicolor*) are a common North American hylid found across a broad range stretching from the Northeast and Upper Midwestern USA to Southeast Texas

(Elliot et al., 2009). Males produce pulsed advertisement calls comprising a series of short pulses with a duration of approximately 25 ms (at 20 °C) and that are repeated after a pause of 25 ms. At a given temperature, the duration of the call can be expressed in a number of pulses or in milliseconds. Average values (\pm SE) in our study population are call duration = 17 ± 0.5 pulses/call, call period = $5,936 \pm 3,027$ ms, first frequency peak = $1,071 \pm 99$ Hz, and second (dominant) frequency peak = $2,142 \pm 192$ Hz (data from $N = 54$ males recorded in 2011; Reichert & Höbel, 2015).

Mating decisions in this nocturnally chorusing animal are largely acoustically mediated (Augusto-Alves et al., 2024), with a relatively simple “best-of-n” sampling strategy (Stratman et al., 2021). Call duration has received the most experimental focus in these frogs, as it elicits the strongest binomial preference in two-choice assays (Gerhardt et al., 2000) and represents the strongest gradient of selection on male calls (Gerhardt & Brooks, 2009). Population-based preference function analysis also suggests that preferences are strongest for call duration, less strong for call period, and weakest for call frequency (Gerhardt, 2005b; Gerhardt et al., 2000; Reichert & Höbel, 2015).

Data collection spanned May and June of the 2019–2022 breeding seasons at two adjacent ponds at UWM’s Saukville field station in Saukville, WI. We collected females in amplexed pairs to ensure receptivity and transported them to a bioacoustics lab within 2 hr of capture. Pairs were kept in individual plastic containers (with 5 mm of water) in coolers on melting ice (4–6 °C) to prevent oviposition, after which females cease to respond to calls. Females were acclimatized to the testing temperature of 20 °C before phonotaxis trials began.

Playback setup and stimulus design

We performed phonotaxis experiments in a circular enclosure (2 m diameter) assembled from 45 cm high wire fencing and black fabric placed atop foam playmat floor pieces set up inside a $3 \times 3 \times 2.4$ m sound-attenuated playback chamber. A single speaker (JBL CONTROL 1Xtreme) was placed outside of the arena wall, 1 m from the centre. The amplitude of the speaker was adjusted to 80 dB SPL using a Lutron SL-4001 sound level meter (fast RMS, “C” weighting). Signals were broadcast from a laptop connected to a Behringer A500 Reference Amplifier. We released each female from the centre of the arena using a wire cage (8 cm in diameter and 4 cm deep) attached to a pulley that could be activated from outside the testing chamber. Using a stopwatch, we recorded the time from the moment the female was released until she entered a “choice zone”; a 20×9 cm rectangle directly in front of the speaker. A female was scored as “no choice” after 5 min of inactivity or taxis outside of the choice zone. After two “no-choice” trials, females were returned to their plastic containers to rest for at least 20 min. Testing on these individuals concluded after further inactivity/indecisiveness, and the females were removed from the data set ($N = 10$).

Testing preference functions

Our experimental design focused on characterizing how attractiveness varies across a range of advertisement call values for three distinct properties of an eastern gray treefrog call: its duration (number of pulses), its period (time elapsed from the start of a call to the start of the next call; inverse of calling rate), and its dominant frequency (in Hz). We assessed

preference functions at the individual level using a one-choice playback design. We used “latency to approach the speaker” as the measure of preference, assuming that a faster approach indicates that the call is deemed more attractive (Bush et al., 2002). For each call property, the range of tested values was set around the previously measured population mean, spanning the range of natural variation found in the population (± 2 SD; males sampled in 2011). Note that for clarity we use the term *trait* to refer to female preference function parameters (e.g., peak, tolerance, etc.), and the term *property* for the parameters of the male mating call (e.g., duration, period, frequency).

To assemble preference functions for *call duration*, we presented females with a randomized sequence of seven stimuli that ranged from 5 and 29 pulses per call (increasing in increments of four pulses; i.e., 5, 9, 13, 17, 21, 25, and 29 pulses). To assemble preference functions for *call period*, we presented females with a randomized sequence of seven stimuli that ranged from 1.36 to 13.73 s (i.e., one call every 1.36, 3.43, 5.48, 7.56, 9.51, 11.67, and 13.73 s). Finally, to assemble preference functions for *call frequency* we presented females with a randomized sequence of eight stimuli that ranged from 1,640 to 2,810 Hz (i.e., 1,640, 1,808, 1,974, 2,142, 2,308, 2,476, 2,642, and 2,810 Hz). The call of eastern gray treefrogs is comprised of a number of harmonics, with the second (higher) frequency peak being 10 dB louder, and hence the dominant frequency. Accordingly, all synthetic stimuli had two frequency components (i.e., $1,071 + 2,142$ Hz), with the higher one being 10 dB louder. Except for the acoustic property being manipulated (duration, period, or frequency, respectively), all other call parameters were set to values near the population averages; 17 pulses, a period of 6 s, pulse duration of 25 ms, rise and fall times of 12ms each and frequency composition of $1,071 + 2,142$ Hz. Synthetic call stimuli were generated using the R packages tuneR (Ligges et al., 2016) and Seewave (Sueur et al., 2008).

Mate choice decisions are the outcome of two components that vary independently: the preference function (the ranking of the attractiveness of prospective mates) and choosiness (the effort invested in obtaining the preferred mate type) (Feagles & Höbel, 2022a; Jennions & Petrie, 1997; Neelon et al., 2019). Choosiness is a repeatable trait in female eastern gray treefrogs (Feagles & Höbel, 2022b), but we currently do not know whether mate preferences are repeatable as well. However, we consider one-time estimates an appropriate measure for assessing individual preference functions, because most females in our population breed only once per year (within-season recapture < 10%) and they have a short lifespan (1–3 years) which makes it unlikely that they will breed in successive years (Höbel et al., 2021).

Statistical analyses

Analysis of within-population variation in preference functions

Preference functions represent responses to continually varying ornaments (Kilmer et al., 2017; Rodríguez et al., 2013; Stinchcombe & Kirkpatrick, 2012), and represent probability curves of a female’s likelihood of choosing a mate with a given trait. Preference functions were generated using PFunc (Kilmer et al., 2017), which fits non-parametric cubic spline regressions to the transformed latency responses of each female. In our assay, a stronger preference for a given call

property value is expressed by a faster approach (a shorter approach latency). However, because the interpretation of preference functions is more intuitive if a female's most preferred value is shown as the highest point in a curve, not the lowest, we converted raw approach latency (in seconds) to a response score = [200—approach latency] before generating preference functions. Note: Although we waited 300 s before terminating a trial for lack of response, the majority of trials (> 95%) ended in less than 200 s. Keeping 300 s as the maximum flattened all functions, and to maximize resolution, we opted to set all response times above 200 s to 200 before running PFunc.

The PFunc program makes no assumptions regarding function shape (quadratic, linear, etc.), only that the curve should have some degree of smoothness. Smoothing parameters were determined empirically by manually adjusting smoothing to maximize the fit of the curve with the underlying data points. This was particularly helpful in distinguishing open functions that peaked in a linear fashion at an extreme value from functions that plateaued at an earlier value. We also inspected data and function fit for all preference functions to ensure that function shape was not determined by potential “outlier” data points. PFunc provides standard error (*SE*) estimates with each preference function. We used those *SE* estimates as a safeguard against function shape being determined by spurious responses; we only included functions that had mean *SE* values of < 35. To verify that function fit in PFunc performed similarly well across preferences for different call properties (i.e., duration, period, frequency) as well as for the different function shapes (i.e., closed, open, flat), we computed the least squares mean model with mean *SE* as the test variable, and call property, function shape, and their interaction, as explanatory variables. Function fit was similar across the three tested call property preferences ($F_{2,307} = 2.76$, $p = 0.06$), but with an average mean *SE* of 15.9, closed functions were slightly “noisier” than flat and open functions (11.3 and 12.8, respectively; $F_{2,307} = 6.60$, $p = 0.002$; post-hoc Tukey HSD). The interaction term was not significant: $F_{4,307} = 0.54$, $p = 0.71$).

To compare within-population variation in preferences to population-based preference estimates, we averaged the approach latency values across all females that provided data for the individual preference functions. Then we used PFunc to obtain population curves as described above for the individual curves. For the population-based functions, we also computed 95% CI to inspect whether “noisier” group functions could suggest the presence and type of within-population variation in preferences.

Finally, we used PFunc to extract preference function traits (peak, tolerance, strength, and responsiveness; described in Figure 1) for each individual preference function as well as for the three population-based preference functions.

Before further statistical analysis, both authors visually inspected all functions and classified them as being “closed,” “open,” or “flat. Closed functions always featured an obvious peak at intermediate stimulus values. Open functions typically featured a peak at one extreme of the trait range, but we also applied this label for functions that peaked at an intermediate value and then plateaued towards one of the extremes (i.e., intermediate and extreme values are equally attractive). Flat functions showed no clear peak at extreme or intermediate values. Note that for a flat function, responsiveness is the only meaningful preference function trait. For flat shapes, the peak is arbitrary, as it could lay anywhere

along the curve, tolerance is essentially infinite, and strength is essentially zero. Consequently, for flat functions, we only included the preference function trait responsiveness for further analysis. Sample sizes of individual preference functions for each call trait were: duration: $N = 112$, period: $N = 104$, and frequency: $N = 101$. We used ANOVA and Tukey HSD to test whether these preference function traits differed by shape category (separately for each call property preference).

Magnitude of variation in female preferences and male calls

We computed the coefficient of variation (*CV*) to obtain a unitless measure that allowed us to compare the magnitude of variation of preference function traits between call property preferences. We also calculated the *CV*s of call duration, call period and call frequency produced by males in our study population. Male call data were obtained from 83 males recorded during the 2017 breeding season (Stratman & Höbel, 2019).

Preference-call relationship

To examine the strength and direction of selection that female preferences may exert on male calls, we calculated the preference-call mismatch, using the difference between the peak preference and the mean of the respective call property, divided by the mean call property [Mismatch = (Pref—Call)/Call]. Male call data were obtained from 83 males recorded during the 2017 breeding season (Stratman & Höbel, 2019). A mismatch value of 0 indicates a perfect match between what females most prefer and the most typical male call value. We obtained a mismatch estimate for the population average function as well as for each individual function for each of the three examined call properties (duration, period, frequency). Because the measure is unitless, it allowed for direct comparison between call properties. We used one-tailed independent *t*-tests to determine whether the observed mismatch values differed significantly from zero. All statistical analyses were performed using JMP software (SAS Institute Inc., Cary, NC).

RESULTS

Within-population variation in preference function shapes and preference function traits

Call duration

The population preference function shape for call duration was open-ended, favouring longer duration calls (Figure 2A); peak preference = 29 pulses, tolerance = 17.5, strength = 0.03, and responsiveness = 133.4. Of the three investigated preferences, this population preference function had the lowest tolerance and highest preference strength, indicative of a strong preference.

The most abundant individual function shape in the population was open (62%, all favouring longer calls), followed by closed (23%) and flat (15%) functions (Figure 2B). The magnitude of variation in peak preference was relatively small ($CV_{\text{peak}} = 14.9$); peak preferences ranged from 13.8 to 29 pulses per call (tested range was 5–29 pulses). Females with open-function shapes had significantly higher peak preference (favouring longer calls) than females with closed functions ($F_{1,93} = 174.9$, $p < 0.0001$), but tolerance ($F_{1,93} = 0.77$, $p = 0.38$) and preference strength ($F_{1,93} = 2.32$, $p = 0.13$) did not differ. Responsiveness, the only meaningful preference function trait provided by flat functions, differed among function shapes ($F_{2,109} = 4.27$, $p = 0.02$); females with flat

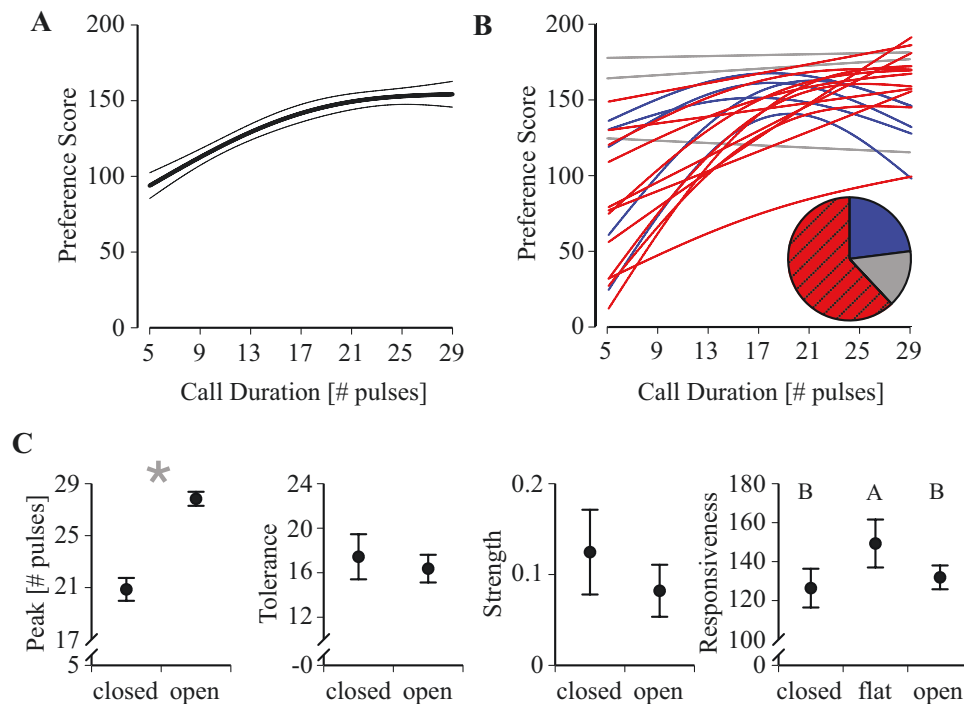


Figure 2. Preference functions for the property *call duration*. (A) Population mean function \pm 95% CI. (B) Individual female preference functions, selected to illustrate the range of shapes present in the population (proportionally here, $N = 20$ of the total 112 functions). Function shapes are color-coded, with red indicating open, blue indicating closed, and gray indicating flat. Inset pie chart shows the proportion of individual functions by shape assignment; hatching indicates direction of open functions. (C) Relationship between preference function shape and preference function traits. Shown are mean \pm 95% CI; gray asterisk/letters indicate statistically significant differences.

functions had significantly higher responsiveness than those with open or closed functions (Figure 2C).

Comparing data from population and individual functions shows that the population function captures tolerance (17.5_{pop} vs. $16.4_{\text{open}}/17.4_{\text{closed}}$) and responsiveness reasonably well (133.4_{pop} vs. $132_{\text{open}}/126_{\text{closed}}$), but overestimates peak preference (29_{pop} vs. $27.9_{\text{open}}/20.9_{\text{closed}}$) and underestimates preference strength (0.03_{pop} vs. $0.08_{\text{open}}/0.12_{\text{closed}}$).

Call period

The population preference function shape was open-ended towards shorter call periods (i.e., faster-repeated calls) (Figure 3A); peak preference = 3.1 s, tolerance = 12.4, strength = 0.007, and responsiveness = 127.7. Of the three investigated preferences, this population preference function had high tolerance and low preference strength, indicative of a weaker preference.

The most abundant individual function shape in the population was open (59%), followed by closed (36%) and flat (5%) functions (Figure 3B). The open-function group contained females that favoured either the low or the high extreme, but a significant majority favoured shorter call periods (55 of 62; binomial test $p < 0.0001$). The magnitude of variation in peak preference was large ($CV_{\text{peak}} = 82.0$); variation in peak preferences spanned the entire range of tested values (1.36–13.7 s). Females with open functions had on average significantly lower preference peaks (favouring faster calls) than females with closed functions ($F_{1,97} = 52.1$, $p < 0.001$). Open and closed functions did not differ in tolerance ($F_{1,97} = 0.003$, $p = 0.95$), but open-function females showed lower preference strength ($F_{1,97} = 6.8$, $p = 0.01$). Responsiveness was lowest in closed functions, intermediate in flat functions and highest in open functions ($F_{2,101} = 5.1$, $p = 0.008$; Figure 3C).

Comparing data from population and individual functions shows that the population function captures peak preference of open functions (3.1_{pop} vs. $2.1_{\text{open}}/6.3_{\text{closed}}$) as well as responsiveness (126.7_{pop} vs. $141_{\text{open}}/126_{\text{closed}}$) reasonably well, but overestimates tolerance (12.4_{pop} vs. $8.8_{\text{open}}/8.8_{\text{closed}}$), underestimates preference strength (0.007_{pop} vs. $0.05_{\text{open}}/0.12_{\text{closed}}$) and also underestimates peak preference of closed functions (3.1_{pop} vs. $2.1_{\text{open}}/6.3_{\text{closed}}$).

Dominant frequency

The population preference function for dominant frequency had a flat shape (Figure 4A); peak preference = 2,191 Hz, tolerance = 1,170 Hz, strength = 0.00, and responsiveness = 140.3. Of the three investigated preferences, this population preference function had the highest tolerance and lowest preference strength, indicative of a very weak preference.

The most abundant individual function shape in the population was flat (43%), followed by closed (31%) and open (26%) functions (Figure 4B). Unlike with open-function shapes observed for call duration (all open in the same direction, favouring longer duration calls) and period (significant majority in one direction, favouring faster-repeated calls), for call frequency we observed a statistically even split in the orientation of the curves (15 preferred low-frequency calls, 12 preferred high-frequency calls; binomial test $ps = 0.70$). The magnitude of variation in peak preference was small ($CV_{\text{peak}} = 17.4$); variation in peak preferences spanned the entire range of tested values (1,640–2,810 Hz). None of the preference function traits (peak, tolerance, strength, or responsiveness) differed among function shapes (Figure 4C).

Comparing data from population and individual functions shows that the population function captures peak preference ($2,191_{\text{pop}}$ vs. $2,230_{\text{open}}/2,184_{\text{closed}}$) and responsiveness (140.3_{pop}

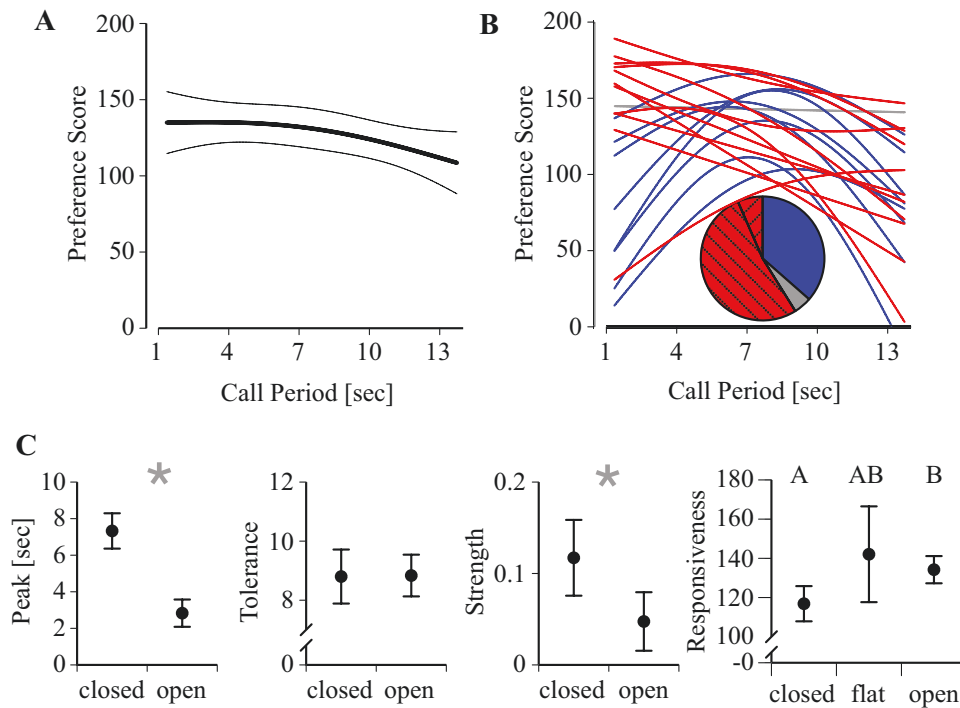


Figure 3. Preference functions for the property *call period*. (A) Population mean function \pm 95% CI. (B) Individual female preference functions, selected to illustrate the range of shapes present in the population (proportionally here, $N = 20$ of the total 104 functions). Function shapes are color-coded, with red indicating open, blue indicating closed, and gray indicating flat. Inset pie chart shows the proportion of individual functions by shape assignment; hatching indicates direction of open functions. (C) Relationship between preference function shape and preference function traits. Shown are mean \pm 95% CI; gray asterisk/letters indicate statistically significant differences.

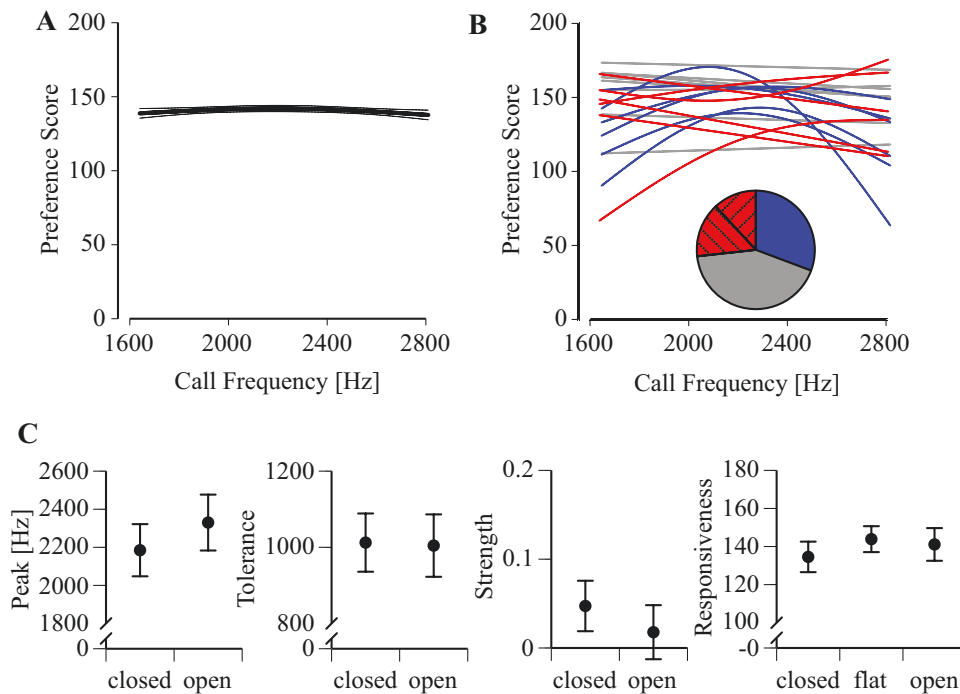


Figure 4. Preference functions for the property *call dominant frequency*. (A) Population mean function \pm 95% CI. (B) Individual female preference functions, selected to illustrate the range of shapes present in the population (proportionally here, $N = 20$ of the total 101 functions). Function shapes are color-coded, with red indicating open, blue indicating closed, and gray indicating flat. Inset pie chart shows the proportion of individual functions by shape assignment; hatching indicates direction of open functions. (C) Relationship between preference function shape and preference function traits. Shown are mean \pm 95% CI.

vs. $141_{\text{open}}/135_{\text{closed}}$) reasonably well, but overestimates tolerance ($1,170_{\text{pop}}$ vs. $1,005_{\text{open}}/1,012_{\text{closed}}$) and underestimates preference strength (0.00_{pop} vs. $0.02_{\text{open}}/0.05_{\text{closed}}$).

Within-population variation in peak preferences and male call properties

For call duration, the within-population variation for peak preference of females ($CV_{\text{peak}} = 14.9$) was smaller than the within-population variation in male calls ($CV_{\text{call}} = 18.0$). For the call period, the variation in female peak preferences ($CV_{\text{peak}} = 82.0$) was much larger than the variation in male calls ($CV_{\text{call}} = 33.0$). For call frequency, the variation in female peak preferences ($CV_{\text{peak}} = 17.4$) was larger than the variation in male calls ($CV_{\text{call}} = 6.6$).

Estimating selection from population-based vs. individual preference functions

The mismatch between the peak preference and the average male call property value can provide an estimate of the strength of selection by female choice. Figure 5 illustrates how the strength of selection would be estimated when based on the average population function compared to individual functions. Patterns are quite different across call property preferences.

Call duration shows the largest mismatch between female preferences and male call properties, both based on population as well as individual functions (Figure 5, left gray square). The population estimate is aligned with many females with open preference functions, but not with those with closed preferences. The mismatch values of both open and closed functions are significantly different from the null of zero (open: $t = 40.3$, $df = 68$, $p < 0.0001$; estimate \pm SD: 0.55 ± 0.11 ; closed: $t = 5.04$, $df = 25$; $p < 0.0001$; estimate: 0.16 ± 0.16).

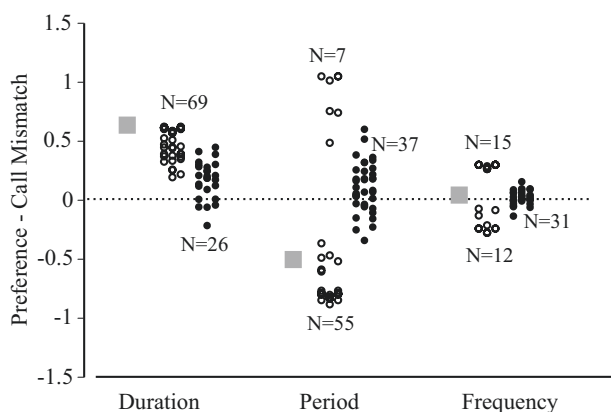


Figure 5. Mismatch between peak preference and mean male call property values (duration, period, and frequency). Gray squares represent population-based peak preference estimates, derived from a function that was fitted to averaged response scores across all females. Filled circle represent peak preference estimates from individual-based closed preference functions and open circle represent peak preference estimates from individual-based open preference functions; note that open functions can favour either extreme. The dotted line represents a perfect match between the female peak preference and the corresponding male call property (i.e., mismatch of zero). For context, a ratio of ± 0.5 is equivalent to a difference of 9 pulses, 3.4 s, and 1,084 Hz, respectively. Individual flat functions were omitted from this figure, since they do not provide a peak preference estimate; however, individual flat functions did contribute to the population average function estimate.

For call period, the population estimate suggests a relatively large mismatch between female preference and the male call property. The population-derived estimate is in agreement with the open-function females preferring faster calls but is completely misaligned with the not insignificant number of females with closed functions or those with open functions that prefer slower calls. The mismatch values of both open and closed functions were significantly different from the null of zero (open_{faster}: $t = 11.1$, $df = 6$, $p < 0.0001$; estimate: 0.89 ± 0.21 , open_{slower}: $t = -58.5$, $df = 54$, $p < 0.0001$; estimate: -0.77 ± 0.10 ; closed: $t = 2.67$, $df = 36$, $p = 0.01$; estimate: 0.09 ± 0.21).

Call frequency shows the smallest mismatch between female preferences and the male call property, both based on population and individual functions. The population estimate is aligned with the closed-function females, but misaligned with the open-function females (both those that prefer lower or higher extremes). The mismatch values of closed functions did not differ from the null of zero (closed: $t = 0.78$, $df = 30$, $p = 0.44$; estimate: 0.01 ± 0.06), but the mismatch values of open functions did (open_{higher}: $t = 390.9$, $df = 14$, $p < 0.0001$; estimate: 0.30 ± 0.003 , open_{lower}: $t = -10.6$, $df = 11$, $p < 0.0001$; estimate: -0.20 ± 0.07).

Discussion

A major finding of our survey of within-population variation of mate preferences is the diversity of function shapes that can lie buried within the population preference. The prediction that individual females would share the population preference function shape was not supported in any of the three preferences we examined—at best 62% of females shared the group shape. We also observed that a mix of different function shapes (flat, open, and closed) was present for all three call preferences (albeit with different proportions), and for two call preferences (period, frequency) we also observed open functions with opposing peaks. This high degree of within-population variation in female mating preferences suggests that there is less directional or stabilizing sexual selection on particular male call parameters, even for those where population patterns would strongly predict it (i.e., call period and particularly call duration).

Predicting sexual selection from population-based functions vs. within-population variation in preference functions

General hypotheses about the direction of sexual selection derived from population-based functions largely align with hypotheses that take into account within-population variation, but estimates about the strength of selection may differ (see below). Moreover, due to the high within-population variation in peak preferences and preference function shapes, the population functions drastically underestimate the predicted variation in male mating success in the population. Higher variation in preference function shape of contributing females can result in “noisier” population functions (i.e., wider 95% CI; Figure 3A and B) where the mean shape is less predictive of variation in male mating success, but even population functions with low variation around the mean can hide substantial within-population variation (i.e., Figure 4A and B).

For call frequency, the rather flat population curve and the lack of mismatch between the population peak preference

and the average male call suggest little selection on male call frequency. The population curve did overestimate tolerance and underestimate preference strength of the females in the population, likely due to the high proportion of flat functions (43%) that contributed to it. Yet, the remaining 57% of females showed a diversity of preferences, from open functions favouring low frequencies, closed functions favouring intermediate frequencies, and open functions favouring high frequencies. Overall, variation in male mating success is expected to be low; not only from “random” mating due to lack of preference but also because the diversity of female preferences means that a male that is highly attractive to few females in the population may have an equal or even higher mating success than a male that is moderately preferred to all females (Jennions & Petrie, 1997). Consequently, the selection on male call frequency is expected to be similarly low as suggested by the population function, but not only because of a lack of preference (many flat functions) but also because of the diversity of open and closed individual functions.

For call duration, the open shape and comparatively low tolerance and high preference strength of the population curve, as well as the large mismatch between the population peak preference and the average male call, suggests that call duration is under strong directional selection.

The population curve did overestimate the peak preference and underestimate preference strength of the females in the population, particularly for females with closed functions. About a third of the open functions plateaued at intermediate values (i.e., these females deem intermediate call durations as attractive as extreme ones), and there are closed-function females who prefer intermediate values and reject very long calls. This pattern of within-population variation in preferences shows that males with very long calls are not as universally preferred as the population curve would suggest, and this should increase variation in male mating success. Consequently, directional selection and male trait exaggeration are expected to be weaker than suggested by the population function.

For the call period, the population function indicates an intermediate to large mismatch with the mean male value; however, the broad tolerance and low strength of this open-function shape suggests only a weak preference for faster calls.

The population curve did overestimate tolerance and underestimate preference strength of the females in the population, and obscures the striking among-female variation in call period peak preferences that span the entire range of presented stimuli. The population curve was open but rather flat towards faster calls, likely because some females strongly favoured such values while others disfavoured them entirely (i.e., strong “open” preference for a 1.36 s call period versus the strong “closed” preference against it, Figures 3B and 5). Moreover, flat individual functions were exceedingly rare (5%), suggesting that females generally care a lot about this call property, but disagree about what they prefer. Because different females find certain males either highly attractive or highly unattractive, variation in male mating success is expected to be lower than suggested by the open population curve favouring faster calls. Weak directional selection towards faster calls thus arises not because females have weak preferences. Rather, there is a relative abundance (53%) of females that strongly prefer fast calls (generating the direction of selection) but a solid minority (42%) that strongly

prefer intermediate or slow calls (generating the strength of selection).

With respect to the two temporal traits investigated here, call duration is generally considered the primary acoustic property mediating mate choice decisions in the eastern gray treefrog literature (i.e., Gerhardt & Brooks, 2009; Gerhardt et al., 2000; Reichert & Höbel, 2015). Yet, more females attend to variation in call period than call duration, as suggested by the relative proportion of flat functions (5% for call period, 15% for call duration). Moreover, realized mate choice decisions in nature are not only determined by female preferences but also by the availability of potential mates and the ability of females to discriminate between males with similar displays. Call period is by far the most variable temporal property between calls of neighbouring males, i.e., the property showing sufficiently large inter-male differences to be discriminated by mate-searching females. For this reason, Stratman and Höbel (2019) proposed call period to be more likely to mediate preference-based mate choice in eastern gray treefrogs under natural chorus conditions. The present study further underscores the likely dominant role of this preference for determining mate choice decisions, especially in highly competitive chorus conditions where short call periods (fast calling rates) may strongly attract some females (those with open preferences for fast calls) but strongly deter others (those with open preferences for slow calls and with closed preferences)

Simulations and empirical data

Millan et al. (2020) used individual-based models to predict how within-population variation in female preferences influences the opportunity for sexual selection and the evolution of sexually selected male traits. Their simulations suggest that the exaggeration of the sexually selected trait or its maintenance of variation depends on both the shape of the preference function and the magnitude of variation in female peak preferences (but not preference strength). When females were modelled to choose their mates according to an open-ended function and there was within-population preference variation, the mean male trait became more exaggerated and its variation decreased. By contrast, when females were modelled to choose their mates according to a closed function, mean male trait values did not change, irrespective of the magnitude of within-population preference variation, and male trait variation decreased more in the simulations with low within-population preference variation compared to simulations with high within-population preference variation.

The high within-population diversity of preference function shapes observed for all investigated call property preferences complicates the comparison of our empirical data with these simulations. Both call duration and call period had a slight majority of open preference functions (62% and 59%, respectively) combined with within-population variation in peak preference, the conditions suggested by Millan et al. (2020) to be favourable for the exaggeration of male traits. Preference variation was substantially larger for the call period ($CV_{\text{peak}} = 82.0$) than duration ($CV_{\text{peak}} = 14.9$), suggesting that directional sexual selection should be stronger for a period than for duration. Then again, while the open functions for call duration uniformly preferred longer calls, the open functions for call period were split between a majority preferring faster calls (55 of 62) but also a minority preferring slower calls (7

of 62), lowering the estimate of the strength of selection on call period.

The simulation by Millan et al. (2020) also suggested that maintenance of male trait variation is maximized when high within-population variation in peak preference combines with closed preference function shapes. None of the three preferences investigated here had predominantly closed-function shapes (range: 23–36%). Moreover, because closed functions in our study were not spread uniformly across the full range of tested trait values but centred around intermediate values, the resulting within-population variation in peak preference of those closed functions was relatively low (irrespective of the call property in question). Rather, in our data set it was the variation in function shapes that generated within-population variation in peak preference.

The high within-population variation in preference function shape uncovered in our study highlights a previously underappreciated aspect of within-population preference variation: to more accurately reflect nature and to be able to predict the effect of preference variation on sexual selection, future simulations should model the relative proportion of function shapes that are required to affect the opportunity for sexual selection and the evolution of sexually selected male traits.

Data availability

The data that support the findings of this study are openly available in Hoebel, Gerlinde; Stratman, Kane (Forthcoming 2024). Within-population variation in treefrog preference functions [Dataset]. Dryad. <https://doi.org/10.5061/dryad.z612jm6m9>

Author contributions

Kane Stratman (Conceptualization [equal], Data curation [supporting], Formal analysis [equal], Investigation, Writing—original draft [lead], Writing—review & editing [equal]) and Gerlinde Höbel (Conceptualization [equal], Data curation [lead], Formal analysis [equal], Project administration, Supervision [lead], Writing—original draft [supporting], Writing—review & editing [equal])

Acknowledgments

We thank O. Feagles and E. Oldehoeft for their assistance with data collection. We also thank N. and M. Byers for granting access to their property, and the staff at the University of Wisconsin–Milwaukee Saukville Field Station for logistical support.

Funding

Funding for K.S. was generously provided by the James and Dorathea Levenson Ecology and Field Biology Fellowship.

Conflicts of interest

None declared.

Ethical approval

All applicable international, national, and/or institutional guidelines for the use of animals were followed. The

Institutional Animal Care and Use Committee of UWM approved the use of these treefrogs (protocol # 19-20#27).

References

- Andersson, M. (1994). *Sexual selection*. Princeton University Press.
- Andersson, M., & Simmons, L. S. (2006). Sexual selection and mate choice. *Trends in Ecology and Evolution*, 21, 296–302.
- Augusto-Alves, G., Feagles, O. S., Toledo, L. F., & Höbel, G. (2024). Visual cues do not function in a multimodal signaling context for mate attraction in eastern gray treefrogs. *Biological Journal of the Linnean Society*. In press.
- Brooks, R., & Endler, J. A. (2001). Female guppies agree to differ: phenotypic and genetic variation in mate-choice behavior and the consequences for sexual selection. *Evolution*, 55(8), 1644–1655. <https://doi.org/10.1111/j.0014-3820.2001.tb00684.x>
- Bush, S. L., Gerhardt, H. C., & Schul, J. (2002). Pattern recognition and call preferences in treefrogs (Anura: Hylidae): a quantitative analysis using a no-choice paradigm. *Animal Behaviour*, 63(1), 7–14.
- Day, T. (2000). Sexual selection and the evolution of costly female preferences: Spatial effects. *Evolution*, 54(3), 715–730. <https://doi.org/10.1111/j.0014-3820.2000.tb00074.x>
- Elliott, L., Gerhardt, H. C., & Davidson, C. (2009). *The frogs and toads of North America: A comprehensive guide to their identification, behavior, and calls*. Houghton Mifflin Harcourt.
- Feagles, O., & Höbel, G. (2022a). Mate preferences and choosiness are distinct components of mate choice in eastern gray treefrogs (*Hyla versicolor*). *The American Naturalist*, 200(4), 506–517. <https://doi.org/10.1086/720730>
- Feagles, O. S., & Höbel, G. (2022b). Female gray treefrogs maintain mate choice decisions under predation threat, but adjust movements to reduce conspicuousness during mate approach. *Behavioral Ecology and Sociobiology*, 76(1), 17.
- Gerhardt, H. C. (2005b). Acoustic spectral preferences in two cryptic species of grey treefrogs: Implications for mate choice and sensory mechanisms. *Animal Behaviour*, 70(1), 39–48. <https://doi.org/10.1016/j.anbehav.2004.09.021>
- Gerhardt, H. C. (2005a). Advertisement-call preferences in diploid-tetraploid treefrogs (*Hyla chrysoscelis* and *Hyla versicolor*): Implications for mate choice and the evolution of communication systems. *Evolution*, 59(2), 395–408.
- Gerhardt, H. C., & Brooks, R. (2009). Experimental evidence of multivariate female choice in gray treefrogs (*Hyla versicolor*): Evidence for directional and stabilizing selection. *Evolution*, 63(10), 2504–2512. <https://doi.org/10.1111/j.1558-5646.2009.00746.x>
- Gerhardt, H. C., Tanner, S. D., Corrigan, C. M., & Walton, H. C. (2000). Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). *Behavioral Ecology*, 11(6), 663–669. <https://doi.org/10.1093/beheco/11.6.663>
- Höbel, G., Kolodziej, R., Nelson, D., & White, C. (2021). Effect of body size, age and timing of breeding on clutch and egg size of female eastern gray treefrogs, *Hyla versicolor*. *Amphibia-Reptilia*, 43(1), 25–35. <https://doi.org/10.1163/15685381-bja10077>
- Jennions, M. D., & Petrie, M. (1997). Variation in mate choice and mating preferences: A review of causes and consequences. *Biological Reviews of the Cambridge Philosophical Society*, 72(2), 283–327. <https://doi.org/10.1017/s0006323196005014>
- Kelly, C. D. (2018). The causes and evolutionary consequences of variation in female mate choice in insects: The effects of individual state, genotypes and environments. *Current Opinion in Insect Science*, 27, 1–8. <https://doi.org/10.1016/j.cois.2018.01.010>
- Kilmer, J. T., Fowler-Finn, K. D., Gray, D. A., ... Rodríguez, R. L. (2017). Describing mate preference functions and other function-valued traits. *Journal of Evolutionary Biology*, 9, 1658–1673.
- Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2016). tuneR: Analysis of music. <http://r-forge.r-project.org/projects/tuner/>
- Millan, C. H., Machado, G., & Muniz, D. G. (2020). Within-population variation in female mating preference affects the opportunity for sexual selection and the evolution of male traits, but things are

- not as simple as expected. *Journal of Evolutionary Biology*, 33(11), 1579–1592. <https://doi.org/10.1111/jeb.13694>
- Neelon, D. P., Rodríguez, R. L., & Höbel, G. (2019). On the architecture of mate choice decisions: Preference functions and choosiness are distinct traits. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 286(1897), 20182830. <https://doi.org/10.1098/rspb.2018.2830>
- Qvarnström, A., Pärt, T., & Sheldon, B. C. (2000). Adaptive plasticity in mate preference linked to differences in reproductive effort. *Nature*, 405(6784), 344–347. <https://doi.org/10.1038/35012605>
- Reichert, M. S., & Höbel, G. (2015). Modality interactions alter the shape of acoustic mate preference functions in gray treefrogs. *Evolution*, 69(9), 2384–2398. <https://doi.org/10.1111/evo.12750>
- Ritchie, M. G. (1996). The shape of female mating preferences. *Proceedings of the National Academy of Sciences of the United States of America*, 93(25), 14628–14631. <https://doi.org/10.1073/pnas.93.25.14628>
- Rodríguez, R. L., Hallett, A. C., Kilmer, J. T., & Fowler-Finn, K. D. (2013). Curves as traits: Genetic and environmental variation in mate preference functions. *Journal of Evolutionary Biology*, 26(2), 434–442. <https://doi.org/10.1111/jeb.12061>
- Rodríguez, R. L., Ramaswamy, K., & Cocroft, R. B. (2006). Evidence that female preferences have shaped male signal evolution in a clade of specialized plant-feeding insects. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 273(1601), 2585–2593. <https://doi.org/10.1098/rspb.2006.3635>
- Rowe, L., & Houle, D. (1996). The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 263(1375), 1415–1421.
- Shaw, K. L., & Herlihy, D. P. (2000). Acoustic preference functions and song variability in the Hawaiian cricket *Laupala cerasina*. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 267(1443), 577–584. <https://doi.org/10.1098/rspb.2000.1040>
- Stinchcombe, J. R., & Kirkpatrick, M.; Function-valued Traits Working Group. (2012). Genetics and evolution of function-valued traits: Understanding environmentally responsive phenotypes. *Trends in Ecology and Evolution*, 27(11), 637–647. <https://doi.org/10.1016/j.tree.2012.07.002>
- Stratman, K., & Höbel, G. (2019). Opportunity for female choice in the wild is frequently curtailed by low male trait variation in *Hyla versicolor*. *Behavioral Ecology and Sociobiology*, 73(5), 59.
- Stratman, K. D., Oldehoeft, E. A., & Höbel, G. (2021). Woe is the loner: Female treefrogs prefer clusters of displaying males over single “hotshot” males. *Evolution*, 75(12), 3026–3036. <https://doi.org/10.1111/evo.14376>
- Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave: A free modular tool for sound analysis and synthesis. *Bioacoustics*, 18(2), 213–226. <https://doi.org/10.1080/09524622.2008.9753600>
- Wagner, W. E. (1998). Measuring female mating preferences. *Animal Behaviour*, 55(4), 1029–1042. <https://doi.org/10.1006/anbe.1997.0635>
- West-Eberhard, M. J. (1983). Sexual selection, social competition, and speciation. *Quarterly Review of Biology*, 58(2), 155–183. <https://doi.org/10.1086/413215>
- Widemo, F., & Sæther, S. A. (1999). Beauty is in the eye of the beholder: Causes and consequences of variation in mating preferences. *Trends in Ecology & Evolution*, 14(1), 26–31. [https://doi.org/10.1016/S0169-5347\(98\)01531-6](https://doi.org/10.1016/S0169-5347(98)01531-6)