Woe is the loner: Female treefrogs prefer clusters of displaying males over single "hotshot" males

Kane D. Stratman,[1,2] ID Emma A. Oldehoeft,[1] and Gerlinde Höbel[1] ID

[1]*Behavioral and Molecular Ecology Group, Department of Biological Sciences, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201*

[2]*E-mail: stratma8@uwm.edu*

Communal displays such as leks and choruses are puzzling phenomena, as it is not obvious why signalers or choosers should aggregate. It has been hypothesized that signalers in leks enjoy higher per-capita reproductive success because choosers prefer to sample among dense configurations ("clusters") that are easier to compare. Although female preferences as well as the signal features of attractive males are well characterized in many chorusing species, we know little about how mate sampling is influenced by the spatial dynamics within communal displays. Here, we ask how female Eastern Gray Treefrogs (*Hyla versicolor*) respond to isolated and clustered call stimuli in a simple one versus three playback design. We explored (i) whether females exhibit a general preference for call clusters, (ii) whether spatial preference is robust to call-feature preference, and (iii) how this affects the relative success of attractive and unattractive males in different spatial combinations. We found generalized spatial discrimination against lone callers but did observe fine-scale assessment of call features within clusters. The prominence of the spatial preference impacts the attractiveness of males, conferring particular advantage to attractive callers within clusters, while reducing attractiveness of isolated males regardless of their acoustic features. Our findings indicate that female frogs navigate complex choruses by initially orientating toward clusters of calling males, and then assess call features within them. This study provides novel insight into the mate choice heuristics involved in animal choruses.

KEY WORDS: Acoustic communication, female choice, mating systems, selection—sexual.

Mate-searching females of many species choose among males that aggregate in communal breeding displays such as leks or frog and insect choruses (Andersson 1994; Höglund and Alatalo 1995; Gerhardt and Huber 2002). Signaler success in these mating systems is often highly skewed, suggesting that sexual selection via female choice may be intense (Fincke 1982; Mackenzie et al. 1995; Widemo and Owens 1995). However, the spectacle created by the multitude of competing males generates numerous puzzles in sexual selection. How do choosers incorporate spatial location and display features of various signalers when navigating these breeding aggregations? Conversely, why should signalers form aggregations at all, and to what degree does display location within them matter?

A male's position within an aggregation can be a strong predictor of mating success (insects, Howard et al. 2011; birds, Hovi et al. 1994; frogs, Lea 2016). This suggests that spatial location (within the breeding aggregation as a whole or within smaller clusters of competitors) may be under sexual selection (Emlen and Oring 1977), and rather than simply describing male spatial distribution within leks it may be more informative to characterize aspects of female choice in the context of lek composition (Gibson et al. 1990). Because females set mate choice rules by which successful males are bound to play, parsing out which spatial cues females use may inform how these aggregations evolve in the first place. For example, a default "female attraction towards aggregations" may arise either to reduce predation risk

on the chooser, or to allow easier comparison of nearby options (Bradbury 1981; Murphy 2012). Such an attraction may yield the commonly observed higher per-capita success of males in larger leks (reviewed in Alexander 1975, Höglund and Alatalo 1995, Isvaran and Ponkshe 2013, and Alem et al. 2015, but see Kokko et al. 1998, Wong et al. 2018).

Females often show strong preferences for display features when selecting mates (Andersson 1994; Höglund and Alatalo 1995; Gerhardt and Huber 2002). However, a signaler's attractiveness is not a straightforward function of individual display features, but also of the phenotypes of nearby rivals and their spatial distribution (Bateson and Healy 2005). The interaction between these factors is of key interest in the study of sexual selection in communally displaying species. By their very nature, the aggregation of numerous displaying males of varying attractiveness generates socially complex environments where preferences for particular feature of a display alone may not predict a female's behavior.

Numerous "mate sampling rules" have been proposed to model how a female navigates the evaluation of potential mates, although they are unspecific in predicting how a signaler's display features and spatial location interact to determine her search path. Sequential mate sampling, a common strategy in lekking species, enables females to visit several options prior to mating, accepting or rejecting options based on an internal standard (Janetos 1980). In much of the lek literature, female visitation rates serve as a proxy for mating preference (Andersson 1989; Rintamäki et al. 1995; Isvaran and Ponkshe 2013). By contrast, in simultaneous sampling, a female assesses a pool of options and mates with the most attractive among them (Murphy 2012). The degree to which female visitations to males can be used as a proxy for mate preference is more limited in this model; these females are commonly observed mating with the first male that they approach (Arak 1988; Murphy and Gerhardt 2002).

Whether males gain higher reproductive success as a function of the combination of male location (typically "distance from lek center") and male advertisement (typically "display rate") has been studied in numerous classic lek taxa (Gibson and Bradbury 1985; Fiske et al. 1994; Jiguet and Bretagnolle 2006). We instead focus on how female choice is affected by signaler position and display features, by manipulating signaler cluster size and signaler composition. We do so in a species where choosers do not survey entire leks, but instead subsample from more immediate options (Murphy and Gerhardt 2002). Here, we are interested in parsing out how females assess male position and phenotype in a context where overall "lek centrality" is likely less important than more local spatial features.

First, we examined how small-scale spatial distribution of signalers determines mating decisions by providing females the choice between a cluster of identical signalers and an equidistant lone signaler. We manipulated the relative attractiveness of the cluster versus the lone signaler to test how females assess increasingly "lower quality" aggregates. Here, we developed several hypotheses. The *Local Density* hypothesis predicts that females first attend to differences in signaler number and reward clustered males irrespective of phenotypes. The *Nearest Hotshot* hypothesis predicts that females first attend to display features, preferring attractive males regardless of their relative location in the aggregation. Lastly, a *Clustered Hotshot* hypothesis involves females attending to both spatial and display factors when choosing a mate, with male success depending both on his location and relative attractiveness. This hypothesis makes the prediction that females should prefer clusters in general, but having reached the cluster they should select the most attractive male available within it. To test this prediction, we manipulated signal features both between and also *within* clusters. Finally, we devised a repeated-measure testing protocol (see *Individual consistency protocol* in methods) that allowed us to test for relevant among-female variation in spatial and/or signal feature preferences. Population-based single-measure choice designs may obscure relevant individual variation. For example, a mating preference held by a minority of females (about 30% in a binomial test) may go undetected if enough of the population chooses "randomly," or strong but conflicting preferences may cancel each other out. By repeatedly testing females in a small number of choice assays, and comparing individual preferences to a simulation, we are able to better detect meaningful patterns of preference that exist within populations.

Having explored how the interaction of chooser preferences for both spatial factors and signal features governs the mate sampling strategy of a female, we next tested how these "rules" may accrue benefits to males. We calculated males' "relative attractiveness" by comparing female choices for lone and clustered signals to the null assumption that the quantity of signalers alone predicts taxis (e.g., a given male among four total males is chosen 25% of the time). Resulting increases or decreases in the "attractiveness" of a given signal inform how the spatial and ornamental features of his display combine to affect his mating success.

To the extent that female choice is a source of selection on the formation of breeding aggregations, particularly *where* males settle within the aggregation, we expected that it would involve a mechanism that maximizes females' chance of localizing an attractive male, which stems from the "female attraction toward aggregations" hypothesis of lek evolution. This model should feature two key elements: (1) a female preference for a cluster of signalers with the capacity for fine call assessment within it, and consequently (2) an increase in the success of attractive, clustered males. Here, we use a classic communal mating aggregation, the frog chorus. These choruses are somewhat clumped aggregations with highly variable local densities (Stratman and Höbel 2019).

We know little about how males select positions within them, but their distribution within choruses is neither uniform nor random (Mitchell and Miller 1991; Berec 2017). How female frogs navigate such large, patchy choruses and discriminate among callers is equally mysterious. There is evidence that anurans use multiple cues in recognizing and localizing large conspecific choruses from a distance (Christie et al. 2010; Buxton et al. 2015; James et al. 2015), but we know much less about how choosers navigate them upon arrival (but see Murphy and Gerhardt 2002).

Methods

FOCAL SPECIES AND COLLECTION SITE
Eastern Gray Treefrogs (*Hyla versicolor*) are found in a geographic range extending from Southeast Texas to the Upper Midwest and Northeastern region of the United States and Canada (Elliott et al. 2009). Males have trilled mating calls consisting of a series of short pulses that, at our test temperature of 20°C, have a duration of approximately 25 ms and are repeated after a pause of 25 ms; thus, at a given temperature the duration of the call can be expressed in number of pulses or in ms. Average values ($\pm$ SE) in our study population are as follows: call duration $= 17 \pm 0.5$ pulses/call; call period $= 5936 \pm 3027$ ms; first frequency peak $= 1071 \pm 99$ Hz; and second (dominant) frequency peak $= 2142 \pm 192$ Hz (data from $N = 54$ males recorded in 2011). Females attend mostly to three call features: call duration, call rate, and pulse rate (Gerhardt et al. 2000; Schwartz et al. 2001; Reichert and Höbel 2015), and can discriminate differences in these features when the magnitude of difference exceeds 20% (Stratman and Höbel 2019). Chorus sizes vary widely across a season, and distances between nearest calling males in a chorus range from 0.25 to over 15 m (mean $= 3.7$ m; Stratman and Höbel 2019). Consequently, females encounter calling males in a multitude of configurations and proximities.

We collected frogs in the late springs of 2018–2020 at a private pond in Saukville, WI. We collected females in amplexed pairs to assure sexual receptivity and transported them to an acoustics lab. Pairs were kept in individual plastic containers in coolers on melting ice to prevent oviposition. All frogs were returned to the pond within 4 days of capture.

GENERAL METHODS
Stimulus generation
We used the R packages tuneR (Ligges et al. 2016) and seewave (Sueur et al. 2008) to generate three stimuli: a population mean stimulus (M), and attractive stimulus (A) and an unattractive stimulus (U). We used the digital audio editor Audacity (https://www.audacityteam.org) to combine call stimuli into the various combinations (Table 1) designed to test our hypotheses.
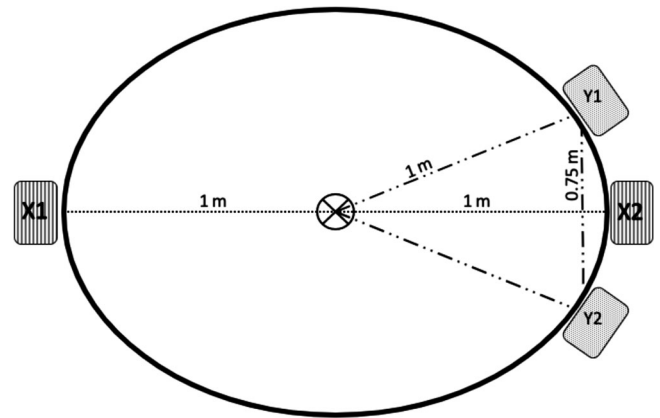

Figure 1. Diagram of the arena design. X1 represents a "lone" speaker and X2 represents the "cluster center" speaker. Y1 and Y2 indicate the position of the two "cluster flanking" speakers. Females were released at the location denoted with a crisscrossed circle.

To create the *population mean stimulus* (M), we generated a stimulus that consisted of a 17-pulse call, played once every 5.9 s, with a pulse rate of 20.5 pulses/s. To generate the *attractive stimulus (A)*, we increased duration, call rate, and pulse rate by 20% (21 pulses, played once every 4.5 s, with a pulse rate of 24.5 pulses/s; note that when varying pulse rate, we only varied the silent interval between the pulses, but kept the pulse duration constant). To generate the *unattractive stimulus (U)*, we decreased duration, call rate, and pulse rate by 20% (13 pulses, played every 7.3 s, with a pulse rate of 16.5 pulses/s). Note that although 20% adjustments for any single call parameter create a realistic but discriminable difference (Stratman and Höbel 2019), increasing (or decreasing) all three in tandem creates a composite call that is noticeably different in sound energy produced per unit time. Although calls similar to our *U* do exist in nature, the high-energy *A* combination was never observed in a survey of >80 males (Stratman and Höbel 2019). In all of our stimuli, the dominant and secondary frequency peaks were held at 2200 and 1100 Hz, respectively, and amplitude of the second peak (1100 Hz) was attenuated by 10 dB. Pulse duration was held at 25 ms, and rise and fall times were each 12 ms.

General playback procedure
We conducted female choice trials inside a semi-anechoic chamber. The testing arena was a near-circular enclosure measuring 2 m in diameter at its widest (Fig. 1). The floor surface was assembled from foam play mats, and the 45-cm-high arena walls were made of wire fencing covered in cloth. All speakers were placed outside the arena walls; speakers faced the center of the arena (release point of the frogs) and were all at a distance of 1 m from the center (see Fig. 1). We used two sets of speakers; two speakers (UBL CONTROL Xtreme, each driven by a Behringer

Table 1. Outline of the null expectations and arena designs in the various experiments, categorized by the corresponding hypotheses. Subscript C = center, F = flanks. Speaker stimuli are denoted by M (population mean call), A (attractive call), and U (unattractive call).

Hypothesis	Null population preference	Hypothesis supported if…	Speaker array/ experimental design	Key	N
Local Density Hypothesis	50:50 (lone vs. cluster)	Females approach clustered calls regardless of features within it	M-MMM	lone mean vs. mean cluster	20
			A-MMM	lone attr. vs. mean cluster	21
			A-UUU	lone attr. vs. unattr. cluster	21
			Control trials:		
			$M_{3\times}$-MMM	lone mean (3× call rate) vs. mean cluster	20
			$M_{89\,dB}$-MMM	lone mean (89 dB) vs. mean cluster	20
Nearest Hotshot	50:50 (attr. lone vs.	Females approach	M-MMM	lone mean vs. mean cluster	20
			A-MMM	lone attr. vs. mean cluster	21
			A-UUU	lone attr. vs. unattr. cluster	21
			Trials establishing baseline attractiveness:		
			M-M	mean vs. mean	29
			A-M	attr. vs. mean	27
			A-U	attr. vs. unattr.	21
Clustered Hotshot Hypothesis	25:25:50 (lone vs. $cluster_C$ vs. $cluster_F$)	Females prefer clusters in general, but having reached the cluster they select the most attractive male within it	*Mixed call features within cluster:*		
			$M\text{-}M_F M_C$	lone mean vs. mean center, mean flanks	20
			$M\text{-}M_F A_C M_F$	lone mean vs. attr. center, mean flanks	20
			$M\text{-}U_F A_C U_F$	lone mean vs. attr. center, unattr. flanks	20

A500 Reference Amplifier) were placed directly opposite each other (at a 180-degree angle). These served as the speaker options during two-choice trials. Two additional speakers (OPTIMUS XTS 40, each driven by a PYLE PCA1 amplifier) were placed adjacent to one of the aforementioned speakers. This second pair was placed 0.75 m apart from each other and equidistant on either side of the central UBL speaker (Y1 and Y2, Fig. 1). This allowed us to create a four-speaker scenario in which a *lone stimulus* could be broadcast in opposition to a *cluster of three stimuli*. We broadcast stimuli through the UBL and OPTIMUS speaker pairs using two separate laptops. Unless otherwise specified, sound pressure levels of all four speakers were set to 85 dB SPL.

To test female preferences, we placed each female into a circular wire release cage in the center of the arena. After the stimuli had played four times each, we released the female using a pulley system that could be activated from outside the testing chamber. This allowed females to move freely around the arena. We scored a positive choice when a female entered a 20 × 9 cm rectangle "choice zone" directly in front of a speaker (and noted whether this was the lone speaker or part of a cluster). We recorded latency to choose using a stopwatch. Phono-

taxis was also recorded using a video camera (EQ150, EverFocus USA, Duarte, CA), mounted directly above the release cage. To analyze female approach paths, we employed Avidemux (http://fixounet.free.fr/avidemux/) to split videos into frame-by-frame JPGs. These photos were stacked into a composite image using the Extended Depth of Field plugin within ImageJ (Schneider et al. 2012). From these images, we could extract path length (cm).

Other considerations about the playback setup

In any experiment featuring identical stimuli (i.e., with equal call durations and calling rates; see Table 1), all stimuli were broadcast with no temporal overlap. In experiments featuring different stimuli (i.e., with differences in call duration and call period; see Table 1), stimuli cycled in and out of various degrees of temporal overlap. Males of *H. versicolor* do not avoid call overlap when displaying in close proximity (Schwartz et al. 2002; Reichert and Gerhardt 2013). Consequently, some of our trials had less, and some had more call overlap than mate-searching females would encounter in nature. We do not think that variation in call overlap biased females toward either the lone or clustered calls, however. Studies examining the call preferences of female *H. versicolor* show that (i) when overlap of the call as a whole is

considered, females tend to prefer nonoverlapped to overlapped alternatives. However, attractiveness of overlapped calls is only truly compromised under a limited set of relative timing relationships of the *pulses* that constitute the species' trilled call, namely, a 180-degree pulse phase shift that results in interdigitized pulses (which likely compromises a female's perception of the species-specific pulse pattern; Schwartz and Marshall 2006); (ii) when lead-lag relationships of the calls are considered, females prefer alternatives in which *individual pulses* are in leading position (Marshall and Gerhardt 2010), but they do not have a generalized leading call preferences. Hence, it is not call overlap that would bias female preferences, but the relative pulse timing relationships of those overlapped calls. None of the call overlap in our trials involved consistent pulse timing relationships of the type that affect female preferences.

To present the loner and cluster stimuli in the spatially restricted area of an indoor playback arena, we had to place cluster speakers at an angular separation of about 21 degrees from the female release point (Fig. 1). This is closer than males would typically position themselves in a natural chorus (pers. obs.). To our knowledge, it has not been empirically determined whether female treefrogs can differentiate between similarly closely spaced sound sources. However, experiments with females of two other treefrog species show that orientation errors during phonotactic approach markedly decrease when the sound separation angle increases from 15 to 30 degrees (Klump and Gerhardt 1989; Caldwell and Bee 2014). A 21-degree speaker separation should thus be sufficient for accurate sound source separation. This is further supported by our observation that females did preferentially approach an attractive alternative that was flanked by two closely bordering unattractive ones (see *Results*).

EXPERIMENTAL DESIGN
Do females prefer clusters?
To test the *Local Density* hypothesis (females prefer clustered calls regardless of the features of the presented calls), we tallied whether more females approached the cluster than the lone speaker in the uniform-cluster trials (M-MMM, A-MMM, A-UUU) (Table 1). Using the video recordings of phonotaxis experiments, we scored the following variables: (i) which stimulus was chosen (in case of a cluster choice, we only scored that a cluster speaker was selected, not which specific speaker within the cluster), (ii) the latency it took the females to reach the chosen speaker, and (iii) the distance the female covered to reach the chosen speaker.

Here, we conducted a population-based comparison for cluster preference, using the data from the first trial a female was tested in each of these trials. The null in these tests was 50/50 (lone vs. cluster). Because we wanted to explore potential among-female variation in cluster preference, we administered the three

cluster trials a total of five times to each female. To measure individual "consistency," we recorded how many females chose the same stimulus type (loner vs. cluster) in at least four out of the five trials ("a consistent" preference) and compared our results to the "two-option" simulation (see *Individual consistency protocol* below).

In addition, we conducted two control trials to differentiate whether an M-MMM cluster preference was due to an actual preference for more sound sources, or an artifact of the cluster being perceived as louder or containing a higher density of call stimuli. To test whether females prefer clusters because they are perceived as a sound source containing more calls, we maintained the call rate of the three cluster speakers (one call played every 5.9 s), but the lone speaker was tripled and now played a call every 2 s (equivalent to the three alternating calls opposite it). To test whether females prefer clusters because they are perceived as a louder sound source, we conducted a trial in which the amplitude of the three cluster speakers remained at 85 dB SPL, but the amplitude of the lone speaker was increased to 89 dB SPL (equivalent to adding three 85-dB-SPL sound sources).

Do females prefer more attractive calls?
To test the *Nearest Hotshot* hypothesis (females prefer calls with attractive features regardless of spatial location), we conducted a population-based comparison and tallied whether more females approached an attractive lone speaker over a less attractive cluster (Table 1). Here, the results from the trials that presented call alternatives in lone spatial configuration (M-M, A-M, A-U) describe call attractiveness in the absence of spatial/cluster information, whereas the three uniform-cluster trials (M-MMM, A-MMM, A-UUU) test whether this call preference is correspondingly robust to spatial information.

Do females differentiate call quality within a cluster?
To test the *Clustered Hotshot* hypothesis that predicts females attend to both spatial (preference for cluster) and display (attractiveness *within* a cluster) factors when choosing a mate, we compared a lone signal with mean features against various combinations of clustered signals (M-MMM, M-MAM, M-UAU) (Table 1). Here, we scored the exact stimulus chosen by the female (lone, cluster-center, cluster-flanking; note that we did not differentiate between choices for the "left" or "right" flanking speaker, because they broadcast identical stimuli). We again explored how consistent females are in these choices and repeated each of these trials five times per female. The "four out of five" criterion of consistency could be met with regard to any of the *three* stimulus types (lone, cluster-center, cluster-flanking) presented, and observed choices were compared to the "three-option" simulation.

What does female preference mean for per-capita male attractiveness?

For group formation to be an advantageous behavior for males, groups must not only attract more females than lone callers; the per-capita attractiveness in the group has to be larger than for single males. In a four-speaker scenario featuring a three versus one setup, the null assumption is that each male should expect a per-capita success rate of 0.25, with approaches to the cluster as a whole being 0.75 and approaches to the lone speaker being 0.25. To study how the interaction of clustering and variation in call features affects male success, we explored whether different configurations of attractive and unattractive alternatives improved or reduced per-capita male success of lone and clustered callers. For each four-speaker experiment (see Table 1), we first calculated the *per-capita attractiveness* of the clustered calls by dividing the number of females that selected a cluster speaker by three (the *n* of the group). Next, we subtracted the *equal chance attractiveness* (sample size divided by four) from each corresponding per-capita attractiveness value. Positive values would therefore indicate a boost in attractiveness, and negative values would signify a reduction.

STATISTICAL ANALYSIS

To test whether females differentiate between two lone call stimuli, or between clustered over lone spatial presentation, we computed binomial tests. To compare whether female phonotaxis behavior differed during the solo/cluster experiments, we computed mixed models in JMP Pro version 13.1.0 (SAS Institute Inc. 2015). We entered approach latency and path length during the phonotactic approach as test variables, and Experiment and Stimulus nested in Experiment as predictor variables. We also entered female ID as a random term, to account for some females having been tested in several experiments.

"Individual consistency" protocol

The preferences underlying female choices may well vary within populations, but such preferences could be masked in a single-measure assay of *N* females. We quantified patterns of female preference by first repeating choice assays five times for each female. We then established "four out of five" consistent choices as a criterion for actual preference, allowing for individuals to demonstrate a repeated preference for any call type. This particular criterion enables us to detect significant patterns of preference with modest sample sizes while allowing each individual to perform multiple five-repetition experiments. To test in detail whether females approached certain stimuli more consistently than expected by chance, we ran two simulations to distinguish meaningful patterns of behavior from statistical noise (R software, version 3.50). The first replicated a "two-option" (e.g., lone call or three-call cluster, choice probabilities of 0.25 and 0.75,

respectively). The second included three-options (lone call, two-flank call, and a center call, with choice probabilities of 0.25, 0.5, and 0.25, respectively). Simulation codes are available upon request. These scripts repeatedly applied a basic random sample function to one of these sets of options five times. A customizable number of truly random "females" cycle through the test and are labelled with a preference type when a "four out of five" criterion is met (else, "inconsistent"). The outcomes for each individual are stored, and the test repeats until the sample size is reached, generating a table that summarizes the quantity of females meeting the consistency criterion for all possible choice outcomes. It loops each simulation 10,000 times and compiles all summary tables into a single data frame. This allows direct comparisons for any given performance in our actual dataset (e.g., consistent preference for *Attractive center*). To determine whether an observed pattern of consistency is meaningful, we simply calculate the proportion of 10,000 iterations exhibiting a "passing rate" pattern that is at least as extreme as the observed. We can also generate a summary table of the simulation data frame that provides the average number of "females" that achieve all possible preference criteria by chance.

Results

FEMALE *H. versicolor* STRONGLY PREFER CLUSTERS

With all call features held equal (M-MMM), females preferentially chose the cluster in the first trial measure (Fig. 2A). Females did not differ in how they approach a clustered speaker versus a lone speaker, taking similar time to travel similar distances (latency: $F_{2,38.4} = 0.5$, $P = 0.61$; path length: $F_{2,35.9} = 0.06$, $P = 0.94$). The control trials showed that this is a true preference for a higher number of callers, not an artifact of a perceived increase in calling rate (Fig. 3, middle) or higher amplitude (Fig. 3, right). Latency to choose a stimulus also did not differ in these control trials ($F_{3,50.8} = 0.88$, $P = 0.46$), further indicating that an increase in amplitude or call rate alone does not explain the cluster preference. Later repeated experiments (*consistency protocol*) confirmed that there were no biases for call location within this homogenous cluster (Fig. 5C).

Cluster preferences were mostly robust to differences in call feature, depending upon the magnitude of difference in attractiveness. When a cluster of average calls was tested against a lone attractive call, preference for the clustered calls remained, although they were relatively less attractive (A-MMM, Fig. 2B). However, significant preference was not observed at the population level when this cluster was composed of unattractive calls (A-UUU, Fig. 2C). Note that in trials involving only two (lone) speakers, females choose the attractive alternative as expected in experiments A-M and A-U (Fig. 2B, C). This partly supports the *Clustered Hotshot* hypothesis, and indicates that a general
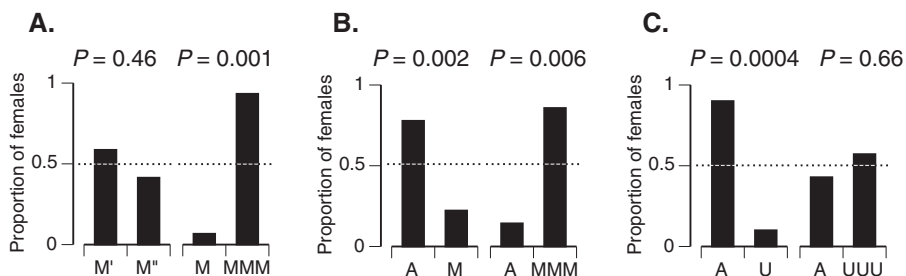
Figure 2. Stimulus choices made in corresponding two and four speaker experiments (first trial only). Bars illustrate the proportion choosing either of two options per experiment; the dotted line indicates the null expectation of nonpreference. Significance values are from binomial tests. M = mean, A = attractive, U = unattractive.
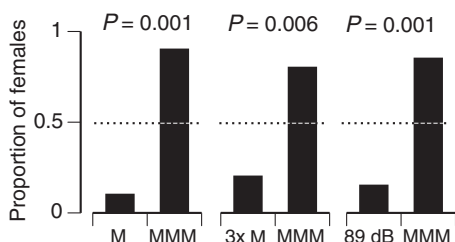

Figure 3. General preference against the lone speaker remains when amplitude (89 dB) or number of calls (3× rate) is held constant. Choice proportions are shown alongside M-MMM for comparison. The dotted line indicates the null expectation of nonpreference.
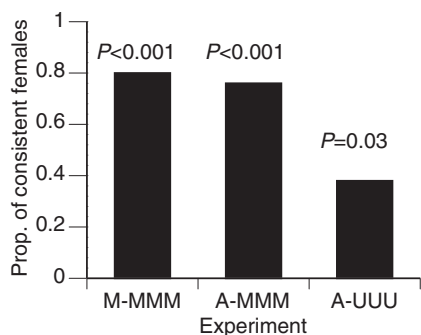

Figure 4. Individual consistency (four out of five trials) shows that cluster preference exists among females when call differences are not too large (black bar). Significance values indicate the proportion of 10,000 simulations producing consistency as frequently as observed.

preference for clustered calls can depend upon the degree to which a lone alternative is more attractive (although preference is never fully reversed in favor of the lone attractive caller).

Choice consistency varied between experiments and was affected by the combination of attractive and unattractive alternatives presented to the females. In the M-MMM and A-MMM trials, the majority of females consistently chose the cluster (Fig. 4, left and center bars). None of the 10,000 iterations of our simu-lation returned consistencies as extreme as the consistent cluster preferences observed in M-MMM and A-MMM. By contrast, in the A-UUU trials only 38% of females chose the cluster consistently. Note that in the population-level analysis there appeared to be no preference for the unattractive cluster over the lone attractive option (Fig. 2C), whereas this individual-based analysis revealed that a sizable minority of females do indeed exhibit such a preference (Fig. 4; $P = 0.03$). Two females (9.5%) did consistently choose the lone but attractive alternative over the unattractive cluster; however, this pattern was nonsignificant (almost always produced by chance simulation; $P = 0.93$).

FEMALES CAN CONSISTENTLY DIFFERENTIATE CALL FEATURES WITHIN A CLUSTER

Where call quality within the cluster differed (M-MAM, M-UAU), females maintained the cluster preference, but generally selected the more attractive stimulus within it, again providing strong support for the *Clustered Hotshot* hypothesis. In the M-MAM trials, consistent females chose the attractive (center) speaker more often than expected in a random simulation (six out of 20 females, $P < 0.001$; Fig. 5A); but we did not find repeated preference for average (flanking) stimuli at a rate higher than seen in the simulation (five out of 20 females, $P = 0.32$). In the M-UAU trials, consistent females picked the attractive (center) speaker much more frequently than expected by chance (indeed, nine out of 20 females is more extreme than all 10,000 simulations), whereas the choice of unattractive (flanking) speakers did not deviate from random (one out of 20 females, $P = 0.99$; Fig. 5B).

Where call quality was held constant (M-MMM), females did not approach center or flanking speakers with a consistency that surpassed the expected (one out of 20 for center, $P = 0.27$; three out of 20 for flanks, $P = 0.75$; Fig. 5C). Note that on average a simulated sample size of ~20 will typically produce a few individuals with consistent "flanking" preference. Chance preference for a single-option stimulus (lone or center) is much more rarely observed (Fig. 5, gray bars). In none of the "mixed call
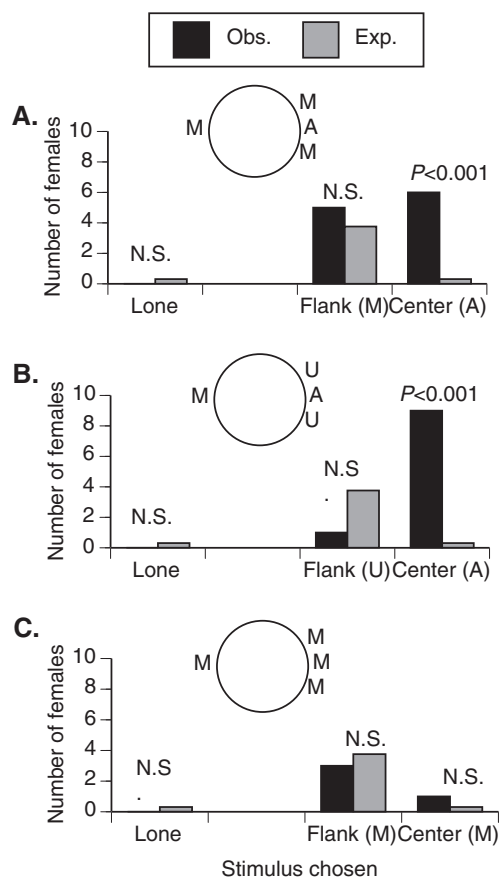
Figure 5. Comparison of observed performance of female preference consistency to chance performance predicted by random-choice simulations. Bars show the number of females passing the "four out of five" criterion for consistency in the observed data (dark) and simulated (light, averaged across 10,000 iterations). Significance values indicated the proportion of iterations with results as extreme as those observed. Chosen stimuli (*x*-axis) correspond to the arena diagrams in M-MAM (A), M-UAU (B), and M-MMM (C) experiments.

Figure 6. The difference between observed and expected per-capita attractiveness of clustered (black) and lone (gray) calls across various configurations. Positive values indicated a higher than expected approach, and negative numbers indicated lower.

feature" experiments did a female consistently prefer the lone speaker, as predicted by the simulation (a sample of 20 will produce a consistent lone preference just 27% of the time).

CLUSTERING GENERALLY BOOSTS PER-CAPITA MALE ATTRACTIVENESS

In the majority of experiments, the change in per-capita attractiveness was positive for signals clustered in a group (Fig. 6, black bars), and negative for lone signals opposite a given group (Fig. 6, gray bars). The exception was the A-UUU experiment, where clustering did not improve the per-capita attractiveness of unattractive signals (Fig. 6, far right black bar). Note also that in the four experiments where clusters attracted more females, the three to one imbalance results in a greater decrease in per-capita attractiveness of the lone signals compared to the in-
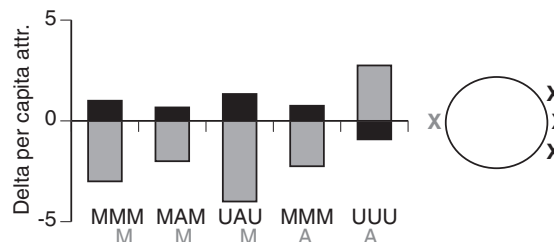
crease in the per-capita attractiveness of clustered signals. When clustered call features were mixed, the boost in attractiveness primarily fell to the attractive stimulus. In experiment M-MAM, the attractive alternative in the cluster was approached by 40% of females, whereas only 22.5% approached each flanking mean alternatives; the remaining 15% of females chose the lone M. In experiment M-UAU, 65% of females chose the attractive stimulus in the cluster, 15% approached each flanking unattractive one, and only 5% of the females selected the lone M.

Discussion

We show that female Eastern Gray Treefrogs have a strong preference toward aggregates of callers, and that this arises from the attractiveness of a higher number of sound sources rather than from differences in amplitude or calling rates. This spatial preference does not, however, negate the call-feature preferences. Having approached the cluster, females do secondarily enact repeatable call-feature discrimination in ways that significantly advantage attractive callers (our *Clustered Hotshot* hypothesis). Consequently, per-capita attractiveness of clustered males appears to broadly exceed the success predicted by the number of callers (three out of four speakers), with a stark reduction in the attractiveness of any caller positioned in isolation. Taken together, these results provide new insights into the mate sampling strategies employed by females of lekking species and offer twofold support for the "female attraction" hypothesis of lek evolution.

The proximate mechanism resulting in the preference for clustered callers remains unclear, because our control trials eliminated the most obvious potential causes (i.e., composite amplitude or calling rates). But the cluster preference observed in initial phonotaxis can indeed culminate in choices for less attractive males because this preference is maintained almost independent of call quality; although the A-UUU scenario does not yield a cluster bias at the population level (Fig. 2C), tests for within-female consistency indicate that such a preference does

exist in almost 40% of females (Fig. 4). This is especially unexpected given that our "attractive stimulus" was 20% more energetic than an average call for multiple key parameters, making it an unusually attractive as well as costly display (Reichert and Gerhardt 2012). In a sample of 84 males (Stratman and Höbel 2019), pulse numbers or call rates as attractive as in our stimulus were observed in 14% and 17% of call recordings, respectively, but never simultaneously in one caller (these traits are negatively correlated; Klump and Gerhardt 1987). Although longer or faster calls exist in nature, the combination creates a supernormal stimulus, yet females do not favor such extreme call attractiveness in the presence of an equidistant cluster. The A-M and A-U tests confirm that our attractive stimulus does overwhelmingly attract females, all spatial factors held equal.

Despite the apparently maladaptive consequence of a cluster preference in some scenarios (i.e., choice of mean or unattractive over an available attractive call; Fig. 2), it is possible that general cluster preferences could be more widely advantageous. Initiating phonotaxis toward clusters regardless of composition may increase a female's likelihood of encountering an attractive mate, provided that she can distinguish him upon arrival to the cluster. Male social behavior could provide such a source of selection on phonotaxis; temporal components of male advertisements are socially plastic, and the features of a male's acoustic signal are affected by both the features of neighboring calls and the number of males in an aggregation (Wells and Taigen 1986; Schwartz et al. 2002). Additionally, the prolonged nature of anuran mating behavior (amplexus and oviposition may take hours to many days; Wells 1977) itself may select for females to use decisive and time-efficient strategies: should another female reach the preferred male first and enter in amplexus with him, he will not be available for another mating for the rest of the night. Approaching a cluster assures that there will at least be another male close-by, regardless of the behavior of other females.

At least in cases of dense clustering with a spatial caller imbalance of three versus one, our study shows that males in groups enjoy higher per-capita attractiveness (Fig. 6). Female bias for clusters is somewhat generalized across lekking species, and per-capita male success tends to correlate with lek size (Isvaran and Ponkshe 2013; Alem et al. 2015). This is in line with the "female attraction" hypothesis of lekking; males cluster because they accrue higher mating success rates in larger aggregations (Lank and Smith 1992; Höglund and Alatalo 1995). Further in line with this hypothesis is the fact that the success of males in a cluster will likely depend on their neighbors. In situations where clusters show high call feature variation (e.g., MAM, UAU), many females consistently chose the most attractive call within the cluster. Although per-capita benefit accrues mostly to more attractive callers, a caller of average attractiveness increases his chance of being approached 1.5× by joining a cluster. The predicted success of an attractive male in a cluster is particularly striking when contrasted with the broad discrimination against equally attractive lone callers (Fig. 2). Although central signalers are often most successful in avian leks (Wiley 1973; Bradbury et al. 1985), note that we do not consider the preference for this caller to be an artifact of "center preference" (sensu Fiske et al. 1998, Howard et al. 2011); females did not disproportionately approach the central speaker in the M-MMM experiment (Fig. 5C). Moreover, female approaches to a central attractive signal are a function of its relative attractiveness within the cluster, not just its location: more females went consistently to the attractive (center) signal when it was flanked by unattractive calls than when it was flanked by merely average ones (Figs. 5A vs. 5B).

Chorusing insects and frogs are among the best-known case studies of female choice and sexual selection, but much of what we know comes from experiments using one-choice and two-choice designs (Ryan 2001; Greenfield 2002; Gerhardt and Huber 2002). Although informative, such tests do not take into account the spatially complex nature of communal displays, or the mate sampling rules different taxa may employ. Lehmann (2007) showed that male density influences female choices in bush crickets that employ sequential mate sampling, a strategy observed in many classic lekking species (Gibson and Langen 1996). It appears that relative signaler density also plays a pivotal role in the behavior of choosers that use "best of *n*" sampling (simultaneously assessing a multitude of proximate options; Janetos 1980). Our study suggests that, having reached a chorus with varying signaler densities, female anurans employ a two-step mate sampling strategy. In the first step, they orient toward a nearby cluster of sound sources without spending much time discriminating call features of isolated options. In the second step, they employ a simultaneous comparison to localize the most attractive signal within the cluster (Murphy 2012). We confirm that having reached a cluster, many females are fully capable of selecting the most attractive option. This two-step tactic has been proposed for another treefrog species (*Hyla gratiosa*; Murphy and Gerhardt 2002), and here we find evidence that females are not only predisposed to seek subgroups that facilitate easy call assessment, but they will do so almost regardless of the quality of unclustered signalers. Provided that choosers can discriminate display differences in dense clusters, such a tactic would save energy travelling among options, remove the need to "store" information about previously attended signalers, and reduce predation risk.

Exploring patterns of mate choice at the individual level is inherently relevant, because fitness in the context of mating success is a not a feature of populations. We suggest that many single-measure designs report no general preference when in fact biologically relevant preferences exist within a population. A repeatable preference in a minority of females could be easily obscured in a binomial test. For example, the population analysis

suggested that there was no preference for the UUU cluster over the A loner, yet individual repeatability analysis revealed that 38% of the females consistently went to the cluster (well above the null expectation). Likewise, different sets of females, each with strong but conflicting preferences, could cancel each other out. Although it can be difficult to detect among-individual variation with simultaneous stimulus presentation (Wagner 1998), our consistency protocol proves to be robust to subpopulation patterns of dichotomous (or especially trichotomous) choice. That "call-feature preference" is expressed secondarily to "spatial preference" is an insight that would not have been revealed through population-level analysis alone. The simulations of two-choice and three-choice outcomes allow for detailed calculation of expected "performance," even at modest sample sizes. The primary limitation of such repetition is inefficiency; although females of high-yield species like gray treefrogs can give >15 responses each, the number of experiments in which a given female can be used is quite small with this protocol. Nonetheless, we recommend using a criterion of "four out of five," as it is stringent enough to minimize statistical noise and allow us to distinguish preference from chance.

Conclusion

Here, we provide novel evidence that mate sampling in chorusing frogs involves (1) preferentially localizing clusters within a chorus and then (2) employing individual call discrimination within them. We find indications that the per-capita success of clustered males is increased (particularly for attractive callers), and that the success of lone males of nearly any quality is greatly reduced. This provides consistent support for the "female attraction" hypothesis of lek evolution; males broadly benefit from aggregating because it best facilitates call-feature assessment by females. Our findings elucidate the tactics with which females navigate a large chorus to enact a finer-scale mating decision, a critical component of sexual selection in chorusing animals.

AUTHOR CONTRIBUTIONS

KS designed and performed the experiments and wrote the manuscript. EO assisted in running the experiments. GH assisted in designing the experiments and co-wrote the manuscript.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

LITERATURE CITED

Alem, S., C. Clanet, V. Party, A. Dixsaut, and M. D. Greenfield. 2015. What determines lek size? Cognitive constraints and per capita attraction of females limit male aggregation in an acoustic moth. Anim. Behav. 100:106–115.

Alexander, R. D. 1975. Natural selection and specialized chorusing behavior in acoustical insects. Pp. 35–77 *in* D. Pimentel, ed. *Insects*, science and society. Academic Press, New York.

Andersson, S. 1989. Sexual selection and cues for female choice in leks of Jackson's widowbird *Euplectes jacksoni*. Behav. Ecol. Sociobiol. 25:403–410.

Andersson, M. 1994. Sexual selection. Princeton Univ. Press, Princeton, NJ.

Arak, A. 1988. Female mate selection in the natterjack toad: active choice or passive attraction? Behav. Ecol. Sociobiol. 22:317–327.

Bateson, M., and S. D. Healy. 2005. Comparative evaluation and its implications for mate choice. Trends Ecol. Evol. 20:659–664.

Berec, M. 2017. Where is my place? Quick chorus structure assembly in the European tree frog. Acta Herpetol. 12:109–112.

Bradbury, J. W. 1981. The evolution of leks. Pp. 138–169 *in* R. D. Alexander and D. W. Tinkle, eds. Natural selection and social behavior. Chiron Press, New York.

Bradbury, J. W., S. L. Verhencamp, and R. M. Gibson. 1985. Leks and the unanimity of female choice. Pp. 301–314 *in* P. J. Greenwood, P. H. Harvey, and M. Slatkin, eds. Evolution: essays in honour of John Maynard Smith. Cambridge Univ. Press, Cambridge, U.K.

Buxton, V. L., M. P. Ward, and J. H. Sperry. 2015. Use of chorus sounds for location of breeding habitat in 2 species of anuran amphibians. Behav. Ecol. 26:1111–1118.

Caldwell, M. S., and M. A. Bee. 2014. Spatial hearing in Cope's gray treefrog: I. Open and closed loop experiments on sound localization in the presence and absence of noise. J. Comp. Phys. A 200:265–284.

Christie, K., J. Schul, and A. S. Feng. 2010. Phonotaxis to male's calls embedded within a chorus by female gray treefrogs, Hyla versicolor. J. Comp. Phys. A 196:569–579.

Elliott, L., H. C. Gerhardt, and C. Davidson. 2009. The frogs and toads of North America: A comprehensive guide to their identification, behavior, and calls. Houghton Mifflin Harcourt.

Emlen, S. T., and L. W. Oring. 1977. Ecology, sexual selection, and the evolution of mating systems. Science 197:215–223.

Fincke, O. M. 1982. Lifetime mating success in a natural population of the damselfly, *Enallagma hageni* (Walsh) (Odonata: Coenagrionidae). Behav. Ecol. Sociobiol. 10:293–302.

Fiske, P., J. A. Kalas, and S. A. Saether. 1994. Correlates of male mating success in the lekking great snipe (*Gallinago media*): results from a four year study. Behav. Ecol. 5:210–218.

Fiske, P., P. T. Rintamaeki, and E. Karvonen. 1998. Mating success in lekking males: a meta-analysis. Behav. Ecol. 9:328–338.

Gerhardt, H. C., and F. Huber. 2002. Acoustic communication in insects and anurans: common problems and diverse solutions. The Univ. of Chicago Press, Chicago.

Gerhardt, H. C., S. D. Tanner, C. M. Corrigan, and H. C. Walton. 2000. Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). Behav. Ecol. 11:663–669.

Gibson, R. M., and J. W. Bradbury. 1985. Sexual selection in lekking sage grouse: phenotypic correlates of male mating success. Behav. Ecol. Sociobiol. 18:117–123.

Gibson, R. M., and T. A. Langen. 1996. How do animals choose their mates? Trends Ecol. Evol. 11:468–470.

Gibson, R. M., C. E. Taylor, and D. R. Jefferson. 1990. Lek formation by female choice: a simulation study. Behav. Ecol. 1:36–42.

Godwin, G. J., and S. Roble. 1983. Mating success in male treefrogs, *Hyla chrysoscelis* (Anura: Hylidae). Herpetologica 39:141–146.

Greenfield, M. D. 2002. Signalers and receivers: mechanisms and evolution of arthropod communication. Oxford Univ. Press, Oxford, U.K.

Höglund, J., and R. V. Alatalo. 1995. Leks. Princeton Univ. Press, Princeton, NJ.

Höglund, J., T. Johansson, and C. Pelabon. 1997. Behaviourally mediated sexual selection: characteristics of successful male black grouse. Anim. Behav. 54:255–264.

Hovi, M., R. V. Alatalo, J. Höglund, A. Lundberg, and P. T. Rintamäki. 1994. Lek centre attracts black grouse females. Proc. R. Soc. Lond. B Biol. Sci. 258:303–305.

Howard, D. R., N. Lee, C. L. Hall, and A. C. Mason. 2011. Are centrally displaying males always the centre of female attention? Acoustic display position and female choice in a lek mating subterranean insect. Ethology 117:199–207.

Isvaran, K., and A. Ponkshe. 2013. How general is a female mating preference for clustered males in lekking species? A meta-analysis. Anim. Behav. 86:417–425.

James, M. S., M. P. Stockwell, J. Clulow, S. Clulow, and M. J. Mahony. 2015. Investigating behaviour for conservation goals: conspecific call playback can be used to alter amphibian distributions within ponds. Biol. Conserv. 192:287–293.

JMP®, Version 10.1.0. SAS Institute Inc. 2015. Cary, NC, 1989-2019.

Janetos, A. C. 1980. Strategies of female mate choice: a theoretical analysis. Behav. Ecol. Sociobiol. 7:107–112.

Jiguet, F., and V. Bretagnolle. 2006. Manipulating lek size and composition using decoys: an experimental investigation of lek evolution models. Am. Nat. 168:758–768.

Klump, G. M., and H. C. Gerhardt. 1987. Use of non-arbitrary acoustic criteria in mate choice by female gray tree frogs. Nature 326:286–288.

———. 1989. Sound localization in the barking treefrog. Naturwissenschaften 76:35–37.

Kokko, H., W. J. Sutherland, J. Lindstro¨m, J. D. Reynolds, and A. MacKenzie. 1998. Individual mating success, lek stability, and the neglected limitations of statistical power. Anim. Behav. 56:755–762.

Lank, D. B., and C. M. Smith. 1992. Females prefer larger leks: filed experiments with ruffs (*Philomachus pugnax*). Behav. Ecol. Sociobiol. 20:323–329.

Lea, A. M. 2016. Sexual selection in complex choruses: the interplay of male signal variation, social structure, and female mate choice. Ph.D. diss., University of Texas at Austin, Austin, TX.

Lehmann, G. U. C. 2007. Density-dependent plasticity of sequential mate choice in a bushcricket (Orthoptera: Tettigoniidae). Aust. J. Zool. 55:123–130.

Ligges, U., S. Krey, O. Mersmann, and S. Schnackenberg. 2016. tuner: analysis of music. Available via http://r-forge.r-project.org/projects/tuner/.

Mackenzie, A., J. D. Reynolds, V. J. Brown, and W. J. Sutherland. 1995. Variation in male mating success on leks. Am. Nat. 145:633–652.

Marshall, V. T., and H. C. Gerhardt. 2010. A precedence effect underlies preferences for calls with leading pulses in the grey treefrog, *Hyla versicolor*. Anim. Behav. 80:139–145.

Mitchell, S. L., and G. L. Miller. 1991. Intermale spacing and calling site characteristics in a Southern Mississippi chorus of *Hyla cinereal*. Copeia 2:521–524.

Murphy, C. G. 2012. Simultaneous mate-sampling by female barking treefrogs (*Hyla gratiosa*). Behav. Ecol. 6:1162–1169.

Murphy, C. G., and H. C. Gerhardt. 2002. Mate sampling by female barking treefrogs (*Hyla gratiosa*). Behav. Ecol. 13:472–480.

Reichert, M. S., and G. Höbel. 2015. Modality interactions alter the shape of acoustic mate preference functions in gray treefrogs. Evolution 69:2384–2398.

Reichert, M. S., and H. C. Gerhardt. 2012. Trade-offs and upper limits to signal performance during close-range vocal competition in gray tree frogs *Hyla versicolor*. Am. Nat. 180:425–437.

———. 2013. Socially mediated plasticity in call timing in the gray tree frog, *Hyla versicolor*. Behav. Ecol. 24:393–401.

Rintamäki, P. T., R. V. Alatalo, J. Höglund, and A. Lundberg. 1995. Mate sampling behaviour of black grouse females (*Tetrao tetrix*). Behav. Ecol. Sociobiol. 37:209–215.

Ryan, M. J. 2001. Anuran communication. Smithsonian Institution Press, Washington, D.C.

Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH image to ImageJ: 25 years of image analysis. Nat. Methods 9:671–675.

Schwartz, J. J., and V. T. Marshall. 2006. Forms of call overlap and their impact on advertisement call attractiveness to females of the gray treefrog, *Hyla versicolor*. Bioacoustics 16:39–56.

Schwartz, J. J., B. W. Buchanan, and H. C. Gerhardt. 2001. Female mate choice in the gray treefrog (*Hyla versicolor*) in three experimental environments. Behav. Ecol. Sociobiol. 49:443–455.

Schwartz, J. J., B. Buchanan, and H. C. Gerhardt. 2002. Acoustic interactions among male gray treefrogs (*Hyla versicolor*) in a chorus setting. Behav. Ecol. Sociobiol. 53:9–19.

Stratman, K. D., and G. Höbel. 2019. Opportunity for female choice in the wild is frequently curtailed by low male trait variation in *Hyla versicolor*. Behav. Ecol. Sociobiol. 73:59.

Sueur, J., T. Aubin, and C. Simonis. 2008. Seewave: a free modular tool for sound analysis and synthesis. Bioacoustics 18:213–226.

Wagner, W. E. 1998. Measuring female mating preferences. Anim. Behav. 55:1029–1042.

Wells, K. D. 1977. The social behavior of anuran amphibians. Anim. Behav. 25:666–693.

Wells, K. D., and T. L. Taigen. 1986. The effect of social interactions on calling energetics in the gray treefrog (*Hyla versicolor*). Behav. Ecol. Sociobiol. 19:9–18.

Widemo, F., and I. P. F. Owens. 1995. Lek size, male mating skew and the evolution of lekking. Nature 373:148–151.

Wiley, R. H. 1973. Territoriality and nonrandom mating in sage grouse *Centrocercus urophasianus*. Anim. Behav. Monogr. 6:87–169.

Wong, B. B. M., T. K. Lehtonen, and K. Lindström. 2018. Spatial and temporal patterns of nest distribution influence sexual selection in a marine fish. Oikos 127:1104–1112.

Associate Editor: R. M. Tinghitella
Handling Editor: T. Chapman