



It's the effort that counts: female gray treefrogs ignore differences in call duration and rate when calling efforts are equal

Katie J. Krueger¹ · Alexander T. Baugh² · Mark A. Bee^{1,3}

Received: 10 October 2025 / Revised: 16 February 2026 / Accepted: 9 March 2026
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract

Females commonly exhibit directional preferences for males that produce more exaggerated, elaborate, or energetic signals. Key questions remain regarding individual variation in the expression of such preferences, particularly when performance related trade-offs in signaling occur in social environments (e.g., leks and choruses) where males dynamically compete for females. In this study of Cope's gray treefrog (*Hyla chrysoscelis*), we investigated directional female preferences for longer calls and faster call rates, two performance-related traits that negatively covary. Subjects repeatedly chose between two simulated males with different call durations and call rates, but with equal overall calling efforts (pulses per unit time). In half of tests, call duration and rate were dynamically swapped between the two simulated males prior to females reaching a speaker. Females chose randomly with respect to differences in call duration and call rate and ignored dynamic shifts in calling behavior. The repeatability of female choice was not different from zero, and females were highly committed to choosing the simulated male they initially approached, even when stimuli differing in call duration and rate were dynamically swapped between speakers as she approached. There was no evidence that subpopulations of females differentially weighted call duration or rate in making their choices. Individual differences in body size, condition, and corticosterone levels had no effects on female responses. Together, these results support an inter-signal interaction hypothesis in showing that females choose mates by attending to the emergent signal property of calling effort instead of its constituent components of call duration and rate.

Significance statement

Mate choice can impose selection on communication signals and signaling behaviors, but key questions remain regarding the expression of mate preferences, particularly when multiple signals or signal components related to a signaler's overall performance potentially interact to influence mating decisions. In this study of a treefrog, we used two-alternative choice tests to investigate female preferences for higher values of two performance-related traits in male advertisement signals (call duration and call rate) that negatively covary but together determine a male's overall signaling output (calling effort). We found that females chose randomly with respect to differences in both traits, suggesting they integrate temporal information across multiple signals to choose mates based on differences in their overall signaling output and not its constituent components. These results highlight the potential importance of inter-signal interactions in mating decisions.

Keywords Communication · Corticosterone · Female preferences · Mate choice · Repeatability · Sexual selection

Communicated by E. Ringler.

✉ Mark A. Bee
mbee@umn.edu

¹ Department of Ecology, Evolution, and Behavior, University of Minnesota Twin Cities, Saint Paul, MN 55108, USA

² Department of Biology, Swarthmore College, Swarthmore, PA 19081, USA

³ Graduate Program in Neuroscience, University of Minnesota Twin Cities, Saint Paul, MN 55108, USA

Introduction

Communication signals frequently form the basis for recognizing and discriminating among potential mates (Anderson 1994; Rosenthal 2017). In mating systems characterized by leks or choruses, males gather in a common area and dynamically compete for access to mates using repeated and metabolically demanding sexual advertisement signals (Prestwich 1994; Höglund and Alatalo 1995; Gerhardt and

Huber 2002; Clark 2012). Females assess signalers and tend to mate preferentially with males that engage in the costliest

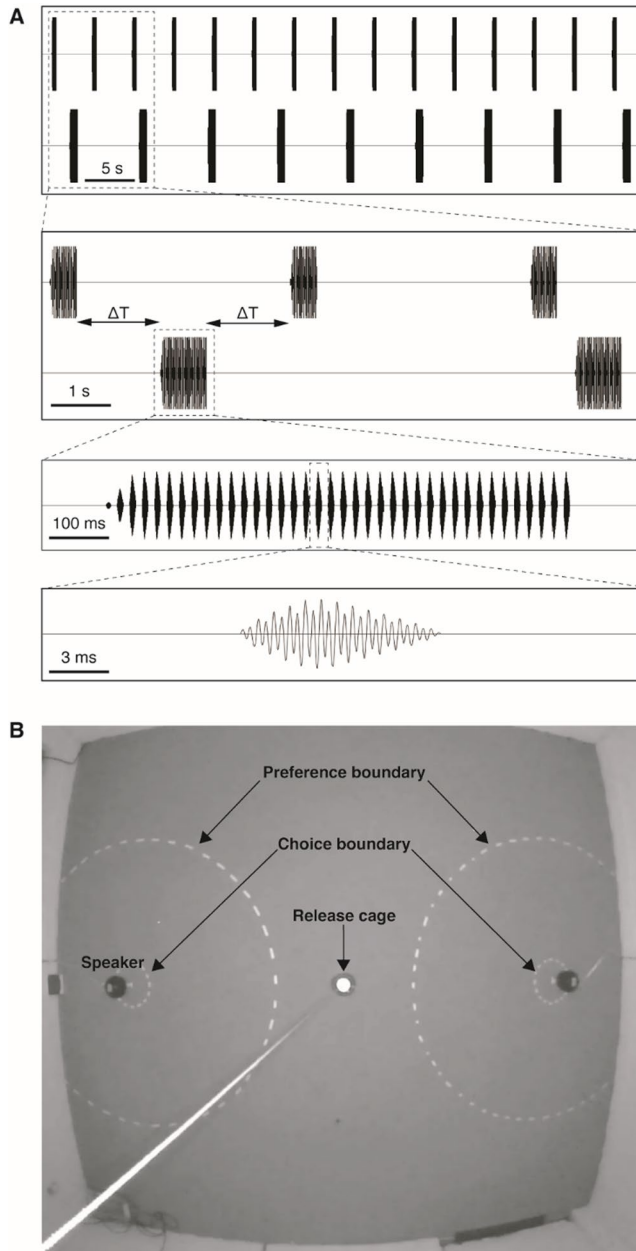


Fig. 1 Acoustic stimuli and test apparatus. **A** Example waveforms (top trace) showing the first 1 min of the stereo sound file used to present the ± 2 SD stimuli in some choice tests (22 pulses versus 38 pulses; see Table 1). The three insets depict (top to bottom) a 10-s segment showing three 22-pulse calls and two 38 pulse calls, a 1-s segment showing a single 38-pulse call, and a 30-ms segment showing one pulse from the middle of a call. In all choice tests, an equal period of silence (ΔT , top trace) separated the first three calls, but thereafter, calls were repeated according to their specified call rates. **B** A view of the sound chamber through the overhead camera showing the two speakers located 2 m and 180° apart from each other and 1 m from the central release cage, each with its associated preference boundary and choice boundary. In dynamic tests, stimuli were swapped between speakers when a female first crossed one of the two preference boundaries

or most exaggerated signaling behaviors (Ryan and Keddy-Hector 1992; Andersson 1994; Rosenthal 2017). Female mate preferences stem from sensory, perceptual, or cognitive traits that predispose individuals to choose mates differentially with respect to their signaling phenotypes (Ryan and Keddy-Hector 1992; Jennions and Petrie 1997; Ryan and Cummings 2013). Elucidating how female preferences give rise to mating decisions remains fundamental to understanding both the evolution of sexually selected traits and, more broadly, signaler-receiver co-evolution.

Acoustically signaling insects and anurans are among the best studied lek- and chorus-breeding species in terms of female preferences and mate choice (reviewed in Gerhardt and Huber 2002). In both taxa, signals often consist of temporally discrete pulses, and signals are repeated through time during bouts of signaling. Numerous playback studies of insects and frogs have revealed directional female preferences favoring longer and more intense acoustic signals repeated at faster rates, a pattern broadly consistent with those observed across other taxa and sensory modalities (Ryan and Keddy-Hector 1992; Gerhardt and Huber 2002; Rosenthal 2017). While directional biases for longer signals produced at faster rates are well documented, key questions remain regarding their phenotypic expression and evolution, as well as their underlying mechanisms. For example, how do directional preferences for signal duration and rate interact when choosing a mate, particularly if signalers face tradeoffs in performance-related signal attributes (Møller and Pomiankowski 1993; Candolin 2003; Hebets and Papaj 2005; Podos 2022)? How are preferences expressed in complex signaling environments where multiple signalers dynamically compete for mates (Bee and Micheyl 2008; Tanner and Hemingway 2025)? To what extent are receiver preferences expressed consistently both within and among individuals, populations, and related species (Bell et al. 2009; Rosenthal 2017)? And what mechanisms underlie preference expression (Adkins-Regan 1998; Leary and Baugh 2020)? We addressed these questions in the present study.

We investigated female preferences for differences in call duration and call rate in Cope's gray treefrog (*Hyla chrysoscelis*), the diploid member of a cryptic diploid-tetraploid species complex; the eastern gray treefrog (*Hyla versicolor*) is the morphologically indistinguishable tetraploid species (Holloway et al. 2006; Booker et al. 2022). Males of these two closely related species attract females in dense breeding choruses by repeatedly producing short (< 1 s) advertisement calls, each consisting of a series of nearly identical pulses (Fig. 1A; Gerhardt 2001). Each pulse within a call is produced separately by contraction of the laryngeal and trunk musculature (McLister et al. 1995; Girgenrath and Marsh 1997, 1999). The number of pulses per call is the

main determinant of call duration at a given temperature. Together, call duration (pulses/call) and call rate (calls/min) determine a male's overall "calling effort," defined here as the product of call duration and call rate (also termed "total call energy," Ryan and Keddy-Hector 1992; "call effort", Ward et al. 2013; Reichert et al. 2024; and "pulse effort," Schwartz et al. 2001, 2002). Calling is metabolically demanding, and call duration, call rate, and calling effort predict a male's metabolic investment in calling (Taigen and Wells 1985; Wells and Taigen 1986; Grafe 1997; McLister 2001). Males appear to be under performance limitations related to physiological constraints on their ability to produce relatively longer calls at relatively faster rates (Reichert and Gerhardt 2012; Welch et al. 2014), and calling effort reflects the overall performance of a male's cardiovascular and respiratory systems (Schwartz and Rahmeyer 2006). Both call duration and call rate are repeatable and exhibit moderate heritability, and there is negative phenotypic and genetic covariance between these two traits, such that males producing longer calls tend to do so at slower rates (Sullivan and Hinshaw 1992; Gerhardt et al. 1996; Schwartz et al. 2002; Welch et al. 2014; Reichert et al. 2024). In competitive signaling environments, males increase call duration (by adding pulses) and reduce in a compensatory way their rate of calling while maintaining relatively constant calling effort. The adaptive function or functions of these dynamic shifts in calling behavior are not yet well established (Schwartz et al. 2002, 2008, 2013; Love and Bee 2010).

Mate choice has been well studied in gray treefrogs at the population level (reviewed in Gerhardt 2001; Gerhardt and Huber 2002). It is well established for both species that females exhibit directional preferences for longer calls with more pulses when call rate is held constant, and for faster call rates when call duration is held constant (Klump and Gerhardt 1987; Gerhardt et al. 1996, 2000; Bush et al. 2002; Bee 2008; Ward et al. 2013; Tanner et al. 2017, 2025; Underhill and Höbel 2017; Feagles and Höbel 2022a, b; Li et al. 2022; Stratman and Höbel 2024). Interestingly, in the diploid (*H. chrysoscelis*), robust population-level preferences for longer calls and faster call rates were abolished or significantly attenuated when females chose between two alternatives differing in both properties simultaneously so long as the overall calling effort of the two alternatives was equal (Ward et al. 2013). This result suggested females might base their choice of mates on calling effort and not on call duration or call rate *per se*. Here, we tested two hypotheses that might explain this population-level result in the diploid.

First, the simultaneous assessment of multiple features or components of signals can interact in additive or multiplicative ways to influence a female's choice of a mate (Candolin 2003; Hebets and Papaj 2005). An *inter-signal interaction hypothesis* (Hebets and Papaj 2005) holds that

the preferences of individual females for longer calls and faster call rates – the two constituent components of calling effort – are epiphenomena stemming from a preference based on higher values of the emergent property of calling effort. Provided two (or more) males have sufficiently similar calling efforts, this hypothesis predicts individual females should choose a mate randomly with respect to differences in their call durations and rates. This hypothesis thus reduces to a null hypothesis of no preference based on call duration or call rate under conditions of equal calling effort. Furthermore, this hypothesis predicts individuals should exhibit little or no consistency in their choice of mates, and ignore real-time, dynamic variation in these two signal attributes, provided the calling efforts of available alternatives remain relatively constant. Under the scenario of equal calling efforts, the choices of individual females should mirror the lack of any preference observed at the population level (Ward et al. 2013).

Second, the preferences of individual females for different features of male signals and signaling behavior could vary reliably among individuals within populations (Jennions and Petrie 1997; Brooks and Endler 2001; Bell et al. 2009; Fowler-Finn and Rodriguez 2013). According to a *differential feature weighting hypothesis* (Jennions and Petrie 1997), individual phenotypes differ in how call duration versus call rate are weighted in choosing a mate. For *H. chrysoscelis*, this hypothesis predicts individual females should differ, on average, both in how consistently they choose males that produce *either* longer calls *or* calls at faster rates and in how willing they are to update their choice of a male in real time when call duration and call rate dynamically vary among alternatives, provided calling efforts remain similar. Moreover, because population-level preferences for both features largely disappeared when calling effort was held constant (Ward et al. 2013), this hypothesis would further predict that the proportions of individual females that differentially weight longer calls over faster rates, and vice versa, should be similar in the population.

Our primary goal was to test the inter-signal interaction and differential feature weighting hypotheses by examining the choices of individual females that were tested repeatedly. Using two-alternative choice tests and female phonotaxis behavior, we measured both the consistency of female choices of stimuli having different call durations and rates but equal calling efforts and female commitment to choosing an initially preferred signaler under conditions of dynamic changes in call duration and rate. Because differences among receivers in factors such as size, condition, or hormone levels can create individual variation in mate choice and preferences (Jennions and Petrie 1997; Hunt et al. 2005; Cotton et al. 2006; Rosenthal 2017; Leary and Baugh 2020), our secondary goal was to also investigate potential sources

of preference variation among female gray treefrogs. To this end we assessed the influences of body size, body condition, and circulating corticosterone (CORT), all of which can influence mating decisions in treefrogs (Ryan et al. 1992; Davis and Leary 2015; Kuczynski et al. 2017; Bastien et al. 2018; Baugh et al. 2021; Feagles and Höbel 2022a).

Materials and methods

Subjects

Our protocols for collecting and testing gravid female treefrogs followed those described in previous studies (Bee 2008; Ward et al. 2013; Tanner et al. 2017, 2025; LaBarbera et al. 2020). Briefly, we collected 115 gravid females of the Western clade of *H. chrysoscelis* (Holloway et al. 2006; Booker et al. 2022) as potential subjects for this study. Gravid female treefrogs collected in amplexus are as discriminating as females collected prior to making a mating decision in the field (Murphy and Gerhardt 1996). Subjects were collected in amplexus from wetlands near the Twin Cities campus of the University of Minnesota in the Tamarack Nature Center (Ramsey County, MN, USA), the Carver Park Reserve (Carver County, MN, USA), and the Hyland Lake Park Reserve (Hennepin County, MN, USA). Collections were made between 2100 h and 0100 h in May and June of 2022. Amplexed pairs were placed in separate small (700 mL) plastic containers and transported back to the University of Minnesota, where they were provided with clean, aged tap water and maintained at approximately 4 °C until testing. This protocol delays egg-laying; extends the period over which gravid females respond to acoustic stimuli; and does not impact endogenous (baseline) CORT levels (Gerhardt 1995; Gall et al. 2019). Females were tested within 24 h of collection and were housed with their chosen male at all times aside from behavioral testing, body size measurements, and blood sampling. Prior to testing, amplexed pairs were transferred to a darkened, temperature-controlled incubator, and their body temperatures were allowed to

reach approximately 20 °C during a 30 min acclimatization period. We released subjects at their site of capture within 48 h of the completion of testing.

Study design

Across a series of 10 choice tests, each female was allowed to choose between two simulated calling males having equal calling efforts and located 2 m apart. Only 104 (of 115) females that made a choice in all 10 tests were used as subjects in the study; data from the remaining 11 collected females was not used. Calling effort was always held constant for the two alternatives (i.e., speakers) across all 10 choice tests for both a control group and two experimental groups. In the control group (Table 1), call duration and call rate (and hence calling effort) were identical for both alternatives, and all three signal attributes were close to their population mean values. In the two experimental groups, the two alternatives had call durations that differed in standard deviation (SD) units by either ± 1 SD or ± 2 SD around the population mean (Table 1; Fig. 1A). Because call rates were adjusted to equalize calling effort across stimuli, longer stimulus calls repeated at slower rates, and shorter calls repeated at faster rates (Fig. 1A). For each female, five of her 10 choice tests were considered “static tests” because the differences between alternatives were maintained until she made a final choice. The other five of 10 choice tests were considered “dynamic tests.” In dynamic tests, females were allowed to move a portion of the way (Fig. 1B) toward one simulated calling male, at which point we dynamically swapped the signal attributes of the two simulated males. Females that stayed the course and chose the simulated male toward which they initially exhibited phonotaxis were considered committed to their initial choice of signaler (Baugh and Ryan 2009). Reversing course and choosing the initially preferred signal in the opposite location was interpreted as females temporally updating their mate choice decision based on a real-time assessment of signal attributes (Baugh and Ryan 2010). Static tests served as a negative control by allowing us to assess the number of times subjects reversed in the absence of swapping of stimuli.

Table 1 Experimental stimuli

Condition	Stimulus	Difference from population mean call duration (in SD units)	Call duration (pulses/call)	Call rate (calls/min)	Calling effort (pulses/min)
Control	Alternative 1	0	30	11.0	330
	Alternative 2	0	30	11.0	330
Experimental (± 1 SD)	Alternative 1	-1	26	12.7	330
	Alternative 2	+1	34	9.7	330
Experimental (± 2 SD)	Alternative 1	-2	22	15.0	330
	Alternative 2	+2	38	8.7	330

Values of call duration, call rate, and calling effort used in the two-alternative choice tests of the control and experimental treatments

In local populations, the mean \pm SD call duration, call rate, and calling effort are approximately 30 \pm 4 pulses, 11 \pm 3 calls/minute, and 334 \pm 65 pulses/min, respectively (see Supplementary Material in Ward et al. 2013). As illustrated in Table 1, these values were used to create acoustic stimuli that had equivalent calling efforts (330 pulses/min) but differed in call duration (either 0 SD, \pm 1 SD, or \pm 2 SD from the population mean) and call rate. The shortest call used in this study (22 pulses) exceeds the range of variation in the number of pulses required to elicit phonotaxis at 85 dB SPL (range: 4–16 pulses; median=7.5 pulses; Gupta et al. 2021). All values of call rate, which were determined based on call duration and calling effort, were within 1.4 SD of the population mean call rate. Differences in call duration and call rate on the order of 1 to 2 SD units around the mean are commonly present among males calling on the same night in the same chorus in gray treefrogs (Gerhardt et al. 1996). In the \pm 1 SD condition, the call durations of the two stimuli differed by 2 SD units, and in the \pm 2 SD condition, they differed by 4 SD units (Fig. 1A; Table 1). Females readily discriminate stimulus differences of this magnitude and smaller (e.g., Gerhardt et al. 2000; Stratman and Höbel 2019; LaBarbera et al. 2020).

Apparatus

Behavioral tests were conducted in a semi-anechoic sound chamber (IAC Acoustics, North Aurora, IL, USA) with a vibration isolating floor covered in low pile gray carpet. The inside walls and ceiling were lined with Sonex acoustic foam panels (Model VLW-60; Pinta Acoustic, Inc. Minneapolis, MN, USA), resulting in interior dimensions of 2.9 \times 2.7 \times 1.9 m (L \times W \times H). Stimuli were broadcast using Adobe Audition 3 (Adobe Systems, Inc. San Jose, CA, USA) running on a Dell 5520 computer (Dell Technologies Inc., Round Rock, TX, USA) equipped with a MOTU M4 sound card (MOTU, Inc., Cambridge, MA, USA). The sound card output was amplified using a Crown XLS 1000 amplifier (Crown International, Los Angeles, CA, USA) connected to two Orb Mod1 speakers (Orb Audio LLC, New York, NY, USA) located 2 m apart on the floor midway along opposite walls of the sound chamber (Fig. 1B). On the sound chamber floor, located 1 m from and centered between the two speakers, was a small (9-cm diameter) subject release cage that could be operated remotely from outside the sound chamber using a rope and pulley system (Fig. 1B). Using infrared-reflective tape, we marked two concentric circles around the face of each speaker having radii of 70 cm and 5 cm; these are denoted as the “preference boundary” and the “choice boundary,” respectively (Fig. 1B). The two stimuli used in each choice test (Table 1) were combined in a stereo sound file, and each stimulus was output through a

separate channel and delivered to a different speaker. Using the Swap Channels effect in Adobe Audition, each stimulus could be instantaneously reassigned to the opposite speaker during a choice test with a single mouse click.

Prior to testing each day, the sound pressure levels (SPL, LCF_{max} re 20 μ Pa) of stimuli were calibrated to 85 dB by placing the microphone of a Larson Davis 831 sound level meter (Larson Davis, Inc., Depew, NY, USA) at a distance of 1 m from the speaker at a height equivalent to that of a female frog's head in the release cage. At 1 m, 85 dB SPL approximates natural call amplitudes in *H. chrysoscelis* (Gerhardt 1975). All tests were performed under infrared light using an overhead mounted IR-sensitive camera (ELP, Shenzhen Ailipu Technology Co., Ltd, Shenzhen, Guangdong, China) and were viewed on a monitor located outside the chamber using OBS Studio (Open Broadcaster Software, version 27.2.4). Behavioral responses were video recorded and simultaneously scored in real time by two observers seated outside the sound chamber. Video recordings allowed us to subsequently verify all data using blinded methods.

Stimuli

We used SynSing (v1.0; Tanner et al. 2020) to synthesize 6-min long stimuli that simulated calling males (44 kHz, 16-bit). Each stimulus consisted of a series of identical repeated calls having a call duration and call rate specified by the given control or experimental conditions (Table 1). Each call consisted of a series of identical pulses modeled after the average pulse produced by males in local populations (Ward et al. 2013). Call rise time was shaped by linearly increasing the amplitude of the first three pulses from baseline to full amplitude. Each pulse within a call was constructed from two phase-locked and harmonically related sinusoids with frequencies [and relative amplitudes] of 1250 Hz [-11 dB] and 2500 Hz [0 dB]. While temperature has a strong influence on pulse rate within a call ($r^2=0.67$), the effects of temperature on call duration ($r^2<0.01$), call rate ($r^2=0.06$), and calling effort ($r^2=0.06$) are much less pronounced (see Table S1 in Ward et al. 2013). Because temperature inside the sound chamber varied with ambient room temperature, we generated three replicate sets of stimuli having pulse rates (range: 50 to 58.8 pulses/s) specified for three different temperature ranges: 19.0 to 20.9 °C, 21.0 to 22.9 °C, and 23.0 to 24.9 °C. These values are based on an analysis of 1000 calls (50 calls from each of 20 males) recorded in local populations across a range of temperatures (Ward et al. 2013), and they comport with the ranges of temperature-dependent pulse rates observed in other populations (McLister et al. 1995; Reichert et al. 2024). The stimuli constructed for the 20.0 to 21.9 °C temperature range consisted of 10-ms pulses separated by 10-ms interpulse

intervals (20-ms pulse period, 50% pulse duty cycle). Pulse rise time was 3.4 ms and pulse fall time was 5.8 ms. Faster pulse rates for higher temperatures were created by proportionally decreasing pulse duration, interpulse interval, and pulse rise and fall times while maintaining a constant pulse duty cycle of 50%. We selected the stimuli presented to each subject based on the temperature inside the sound chamber at the time testing of that subject commenced; these temperatures varied between 21.2 °C and 24.2 °C (\bar{x} = 23.0 °C) across subjects.

Protocol

Subjects were randomly assigned to the control condition (0 SD; $n=33$) or one of the two experimental conditions (± 1 SD: $n=37$; ± 2 SD: $n=34$). Because some subjects were collected from a population that is syntopic with a population of the morphologically indistinguishable tetraploid species, *H. versicolor*, the first phonotaxis test for all subjects was a test of species identification (Fig. S1). In this initial test, we gave females a choice between alternating synthetic *H. chrysoscelis* calls and *H. versicolor* calls broadcast from the two speakers (after Gupta et al. 2021; Kalra et al. 2025; Bee and Kalra 2025; Kalra and Bee 2025). Only females that were motivated to respond by choosing the *H. chrysoscelis* stimulus in this test were used as subjects in subsequent behavioral tests.

After completing the initial test of species identification, subjects in the control and experimental conditions were tested individually in the series of 10 choice tests (five static, five dynamic) determined according to their assigned treatment. For each subject, we randomly determined whether the first test was a static or dynamic test, and test order alternated between the two types of tests thereafter (Fig. S1). The purpose of alternating test types was to sequentially block one static test and one dynamic test into each of five consecutive test pairs to examine and control for any changes that might have occurred separately within each test type (static or dynamic) over the course of repeated testing. For the control condition, we randomized which speaker played the first of the two identical calls. When identical calls were played through each speaker in this condition, subjects exhibited no bias toward either speaker (Fig. S2). In the experimental conditions, we randomized which speaker initially played the stimulus having the longer call duration. In dynamic tests, we used the Swap Channel effect in Adobe Audition to instantaneously change the speaker from which each stimulus played as soon as possible after the subject first crossed a preference boundary (Fig. 1B). Upon making a channel swap, the stimulus that had played from the left speaker started playing from the right speaker, and vice versa. For static tests, we performed a “sham” channel

swap using the same protocol, but stimuli were reassigned to their original channels and, therefore, were not redirected to opposite speakers. Note that in the control condition, in which the two calls were identical, static and dynamic tests were acoustically and functionally equivalent. Channel swaps were always made during the silent intervals between calls to avoid interrupting a stimulus call.

To begin both static and dynamic tests, subjects were placed in the release cage at the center of the sound chamber. Subjects were allowed to freely move inside the release cage. Each subject experienced a 1-min acclimation period in the release cage after which we commenced stimulus broadcasts. The two stimuli were broadcast for an additional 1-min listening period while the female remained in the release cage. The computed call efforts of the stimuli were approximately equal and near 330 pulses/min by the end of the 1-min listening period prior to subject release (Fig. S3). Following the listening period, the release cage was opened, and the subject was able to move freely in the sound chamber. Subjects were given up to 5 min to respond by crossing the choice boundary in front of one of the speakers. We recorded each subject's initial preference as the stimulus towards which it was moving when it first crossed one of the preference boundaries (see Supplementary Material), and we noted the elapsed time required to do so. Each subject's choice and latency to choose (hereafter “choice latency”) were recorded, respectively, as the stimulus playing from the speaker associated with the choice boundary it first crossed and the elapsed time between being released and first crossing a choice boundary. Subjects were tested one at a time, and after each test, the subject was immediately placed back into the release cage, and the next test began with a 1-min acclimation period. There is little evidence that repeat phonotaxis testing in female frogs has carryover effects that would influence female preferences and choices between tests separated by short time intervals (Gerhardt 1981; Gerhardt et al. 2000; Baugh and Ryan 2009). Subjects typically finished all tests within 1 h.

Hormone collection and morphological measurements

Immediately after a subject completed its last behavioral test, approximately 50 μ L of whole blood was collected via cardiac puncture for later analysis of CORT concentrations using a heparin rinsed, 30-gauge insulin syringe (BD Micro-fine U-100, 0.3 mL). At the time of blood collection, we also measured each subject's tibia-fibula length (TFL) to the nearest 0.1 mm using dial calipers (SPI Polymid Plastic Dial Caliper 150 m, MSC Industrial, Melville, NY, USA). Subjects were then reunited with their chosen mates in their original container, given fresh aged tap water, and placed in

a dark incubator at a temperature of 20 °C. Usually within 6–8 h (and no longer than 10 h) females oviposited eggs within the container (Feagles and Höbel 2022a). Following oviposition, we measured each subject's body mass to the nearest 0.1 g using a digital balance (Smart Weigh SWS100; Smart Weigh Instrument Co., LTD, Changzhou, China). All blood draws and measures of body size and mass were taken by the same person to ensure uniformity. We extracted and measured CORT following methods previously validated in gray treefrogs (Baugh et al. 2019, 2021; Gall et al. 2019; Baugh 2024; Rodriguez-Santiago et al. 2024), and these methods are described in detail in the Supplemental Material. We used TFL and post-oviposition mass to compute the scaled mass index (SMI) as a measure of body condition (Peig and Green 2009). Complete data for CORT, TFL, and SMI were available for 90 of the 104 subjects, and descriptive statistics for these measures are included in the Supplemental Material.

Statistical analyses

All statistical analyses were performed using R version 4.4.1 (R Core Team 2024) and RStudio v2023.6.2.561 (Posit team 2023). Unless indicated otherwise, we used a significance criterion of $\alpha=0.05$.

Female choice at the population-level

We assessed female choice at the population level using binomial tests, as is customary in statistical analyses of two-alternative choice tests of phonotaxis behavior (e.g., Gerhardt 1995). For each of 10 choice tests, we separately tested the null expectation that a proportion equal to 0.50 would choose each alternative. For these binomial analyses (and for these analyses only), we assumed repeated tests of a female were independent. Given the large number of repeated tests per individual (10 tests pooled across static and dynamic tests in each condition), we separately

corrected for multiple comparisons (10 total) within each experimental condition (± 1 SD, ± 2 SD) using the Holm-Šidák method.

We further analyzed female choices in the experimental conditions by fitting a set of generalized linear mixed-effects models (GLMMs) (*lme4* package, Bates et al. 2015). The binary response variable was whether the female chose the longer call produced at a slower rate (yes = 1, no = 0). We compared candidate models (Table 2) that included an intercept only model and models incorporating various experimental design factors and the individual subject-related variables of TFL, SMI, and $\log_{10}(\text{CORT})$. CORT concentrations were \log_{10} -transformed to reduce right skew and better meet assumptions of linearity and homogeneity. Subject ID was included as a random effect in all models. Two primary factors of interest were the magnitude difference in call duration (and hence call rate) between the two alternatives in a choice test ("Stimulus"; ± 1 SD versus ± 2 SD difference in call duration) and whether the test was a static or dynamic test ("Test Type"). We also included a factor ("Test Pair") to account for sequential test order by blocking tests into five sequential pairs, each consisting of a static and dynamic test. To improve model stability, we rescaled subject-related variables as z scores ($\bar{x}=0$, $\text{SD}=1$). Model selection was based on the corrected Akaike Information Criterion (AIC_c), which balances model fit and complexity, and all models with a ΔAIC_c value less than or equal to 2.0 were retained (Burnham and Anderson 2002). Additionally, if retained models were nested (i.e., differed by a single predictor term), we assessed whether additional predictors statistically improved model fits using likelihood ratio tests (LRTs). Finally, model fits were evaluated using conditional and marginal coefficients of determination (R^2) (Nakagawa and Schielzeth 2013; Nakagawa et al. 2017).

We analyzed choice latency using a similar GLMM approach and the same candidate models (Table 2) with the following differences. Because choice latencies were continuous, positive, and right skewed, models were fit using

Table 2 Candidate GLMM models

Model Fixed Effects

Intercept only
Stimulus
Test Type
Stimulus+Test Type
Stimulus+Test Type+Stimulus:Test Type
Stimulus+Test Type+Stimulus:Test Type+Test Pair
Stimulus+Test Type+Stimulus:Test Type+Test Pair+Stimulus:Test Pair
Stimulus+Test Type+Stimulus:Test Type+TFL+SMI+CORT
Stimulus+Test Type+Stimulus:Test Type+SMI+CORT+SMI:CORT

Factors used in the GLMM analyses of female choice, choice latencies, and reversals. The reference conditions for each predictor were as follows: Stimulus (± 1 SD for choices; 0 SD for choice latencies and reversals), Test Type (static tests), Test Pair (pair 1). All models included an intercept and a random effect of subject ID

a gamma distribution with a log link function (*glmmTMB* package, Brooks et al. 2017). The predictor of Stimulus included all three levels of difference in call duration, which allowed us to compare latencies in the choice tests involving differences in call duration and rate (± 1 SD and ± 2 SD) to those in the control conditions, in which these two properties did not differ (0 SD).

Individual consistency of choice

We evaluated individual consistency of female choices in two ways. First, we determined Repeatability (R) from the best GLMM models (*rptR* package, Stoffel et al. 2017). We report adjusted repeatability, which takes into account the influences of both random and fixed effects in the GLMM model, as well as unadjusted repeatability, which reflects solely the variance explained by the random effect of subject ID (Nakagawa and Schielzeth 2010; Dingemanse and Dochtermann 2013). Repeatability values for female choices and other response variables are reported on the latent scale and have confidence intervals [2.5%, 97.5%] derived from 1000 bootstrap iterations and 1000 permutations. The statistical significance of repeatability was assessed using both LRTs and permutation tests.

Second, we compared observed distributions of the number of times real subjects chose the same stimulus across sequential tests in the experimental conditions to expected distributions that modeled different levels of individual consistency in simulated subjects. In static and dynamic tests, both real and simulated subjects ($n=71$) could choose the longer call between zero times (i.e., *never* chose the longer call) and 5 times (i.e., *always* chose the longer call). We manipulated the degree of individual consistency in simulated subjects by incorporating a memory effect (hereafter, \widehat{M}), implemented as $0.0 \leq \widehat{M} \leq 1.0$ in 0.2 steps. The outcome of a simulated subject's first choice test (i.e., chose the longer call or the shorter call) was always determined as a random "coin flip" (i.e., with a 50/50 chance of each possible outcome). On all subsequent choice tests, the outcome was either the same as the immediately preceding test (with probability \widehat{M}) or again determined as another random coin flip (with probability $1-\widehat{M}$). Hence, when $\widehat{M} = 0.0$, each test outcome for each simulated subject was randomly determined and independent of all other tests; when $\widehat{M} = 1.0$, all test outcomes for a simulated subject were the same after being randomly determined in its first test. We compared observed distributions to expected distributions for $0.0 \leq \widehat{M} \leq 0.80$ using Chi-square goodness-of-fit tests with a Monte Carlo approximation (10,000 simulations) to estimate p -values. Statistical comparisons for $\widehat{M} = 1.0$ were not possible because all simulated subjects, by design, chose one or the other stimulus 100% of the time resulting in bins with no simulated subjects, which violates the assumption that the

sampling distribution of the test statistic follows the Chi-square distribution. We explored streamlining the number of comparisons by using Chi-square tests of homogeneity to determine whether the observed distributions could be aggregated across the two levels of difference in stimulus call duration (± 1 SD and ± 2 SD). We aggregated observed distributions that were not significantly different.

Commitment to initial preferences

We investigated an individual's commitment to the simulated signaler (i.e., speaker) associated with the first preference boundary crossed by scoring whether it reversed course after crossing the preference boundary. A reversal in both static and dynamic tests indicated a subject ultimately chose the *speaker* on the opposite side of the arena from the speaker it initially approached. Recall that in dynamic tests, but not in static tests, the two alternative stimuli were swapped between speakers when the subject first crossed a preference boundary. Hence, a reversal in a dynamic test corresponds to a subject that ultimately chose the *stimulus* it was initially approaching prior to the stimulus swap. In a static test, in which stimuli were not swapped, a female that reversed chose the alternative to the stimulus it initially approached. A non-reversal corresponded to a subject that did not reverse course and, thus, ultimately chose the *speaker* (not necessarily the *stimulus*) associated with the first preference boundary it crossed. A non-reversal also corresponds to a choice of the stimulus with the higher relative SPL at the position of a frog that had just crossed the preference boundary (and was thus closer to one speaker than the other). To assess commitment, we used the GLMM models and model selection approach described above for female choices (see Table 2) with the binary response variable of reversal during the trial (reversal=1; non-reversal=0). All candidate models used a binomial distribution with a logit link function and included subject ID as a random effect. In contrast to our previous analyses, the factor of Stimulus in models of reversals also included the control condition (0 SD), such that all three levels of difference in call duration (0 SD, ± 1 SD, and ± 2 SD) were tested. By including the 0 SD condition, we were able to assess whether reversals were more common when stimuli differed in call duration and rate compared to a control in which these call properties did not change.

Results

Female choice at the population-level

At the population level, we found little evidence to suggest that females' initial preferences (Table S1, S2; Fig.

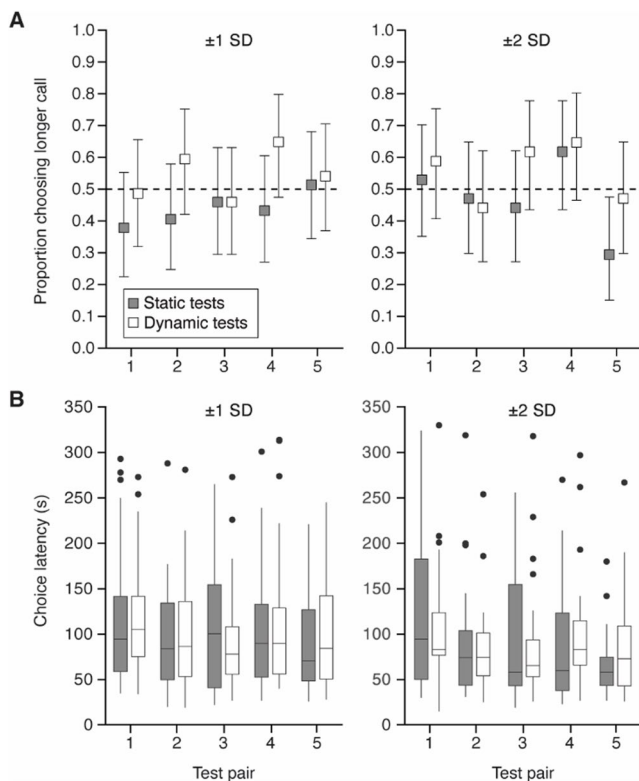


Fig. 2 Female choices and choice latencies. **A** Proportions ($\pm 95\%$ exact binomial confidence intervals) of subjects that chose the call with the longer duration (produced at a slower rate), shown separately for static tests (gray symbols) and dynamic tests (white symbols), when the difference in call duration was ± 1 SD ($n=34$) and ± 2 SD ($n=37$) around the population mean call duration. **B** Boxplots show choice latencies in static tests (gray boxes) and dynamic tests (white boxes) when the difference in call duration was ± 1 SD ($n=34$) and ± 2 SD ($n=37$). Plots depict group latency medians (line), IQRs (box), and $1.5 \times$ IQR (whiskers). Latencies greater than $1.5 \times$ IQR are depicted as individual points

S4) or their final choices were strongly influenced by differences in call duration or call rate when calling effort was held constant. Across the five test pairs, the proportions of subjects choosing the longer call at a slower rate ranged from 0.38 (14 of 37 subjects) to 0.65 (24 of 37

subjects) in the ± 1 SD condition and from 0.29 (10 of 34 subjects) to 0.65 (22 of 34 subjects) in the ± 2 SD condition (Fig. 2A). With only one exception (static test of ± 2 SD in test pair 5; Fig. 2A), the 95% exact binomial confidence intervals for all tests included 0.50. The proportion of subjects choosing the longer call was not significantly different from 0.50 in any binomial test after corrections for multiple comparisons.

Model selection based on AIC_c scores identified two competitive GLMM models, neither of which explained much of the variation in the choices females made (conditional and marginal $R^2 \leq 0.03$; Table 3). The best GLMM model included a significantly negative intercept and a significantly positive effect of Test Type as the only predictor (Tables 3 and S3). These results suggest, overall, that the proportions of subjects choosing the longer call were below 0.50 and higher in dynamic tests compared to static tests after controlling for other effects in the model (Fig. 2A). The other model ($\Delta AIC_c \leq 2$) included the effects of Stimulus and Test Type, but only the intercept and the effect of Test Type were significant (Tables 3 and S3). Adding Stimulus as a predictor did not significantly improve model fit compared to the best model with Test Type as the only predictor ($\chi^2_{(1)} = 0.028, p = 0.867$).

Median choice latencies (Fig. 2B) ranged between 58 s and 106 s across the 20 factorial combinations of Stimulus (± 1 SD, ± 2 SD), Test Type (static, dynamic), and Test Pair (1–5). The best GLMM for choice latency included Stimulus, Test Type, their interaction, and Test Pair as predictors. There were no other competitive models (Tables 4 and S4). After controlling for other effects in the best model, choice latencies were significantly longer when the difference in call duration was ± 1 SD ($p = 0.017$) and significantly shorter in all test pairs after the first test pair ($p \leq 0.014$) (Table 4). Overall, the best model explained about 23% of the variation in choice latencies (conditional $R^2 = 0.23$); however, little variance in latency was explained by the predictors in the model (marginal $R^2 = 0.05$). Thus, even the best model was a poor predictor of choice latencies after accounting for individual differences.

Table 3 GLMM results for female choice

Model	Term	Estimate	SE	Wald z	p	Cond. R^2	Marg. R^2
Test Type	Intercept	-0.323	0.122	-2.638	0.008	0.03	0.02
	Test Type (dynamic)	0.519	0.169	3.074	0.002		
Stimulus + Test Type	Intercept	-0.336	0.146	-2.303	0.021	0.03	0.02
	Stimulus (± 2 SD)	0.030	0.177	0.167	0.867		
	Test Type (dynamic)	0.519	0.169	3.074	0.002		

The two top GLMM models ($AIC_c \leq 2.0$) for female choice based on differences in call duration and call rate. The reference for Test Type was static tests, and the reference for Stimulus was ± 1 SD. The estimate, standard error (SE), statistic (Wald z) and p -value are shown for each term in each model, and the conditional R^2 and marginal R^2 are shown for each model. Results from model selection are included in the Supplementary Material

Table 4 GLMM results for choice latency

Model	Term	Estimate	SE	Wald z	p	Cond. R ²	Marg. R ²
Stimulus + Test Type + Stimulus: Test Type + Test Pair	Intercept	4.508	0.070	64.387	<0.001	0.23	0.05
	Stimulus (± 1 SD)	0.202	0.085	2.379	0.017		
	Stimulus (± 2 SD)	0.099	0.090	1.108	0.268		
	Test Type (dynamic)	0.033	0.060	0.544	0.587		
	Stimulus: Test Type (±1 SD: dynamic)	0.042	0.085	0.492	0.623		
	Stimulus: Test Type (±2 SD: dynamic)	0.018	0.090	0.205	0.838		
	Test Pair (pair 2)	-0.191	0.057	-3.386	0.001		
	Test Pair (pair 3)	-0.164	0.057	-2.897	0.004		
	Test Pair (pair 4)	-0.139	0.056	-2.459	0.014		
	Test Pair (pair 5)	-0.225	0.057	-3.974	<0.001		

The best GLMM model (AIC_c ≤ 2.0) for choice latency based on differences in call duration and call rate. The reference conditions for the predictors of Stimulus, Test Type, and Test Pair were 0 SD, static tests, and pair 1, respectively. The estimate, standard error (SE), statistic (Wald z) and p-value are shown for each term in each model, and the conditional R² and marginal R² are shown for each model. Results from model selection are included in the Supplementary Material

Individual consistency of choice

Females were largely inconsistent in the choices they made, but there was some consistency in how long it took them to make a choice. For female choice, repeatability was low (0.008 ≤ R ≤ 0.010) and not significantly different from zero (Table 5). In contrast, repeatability estimates for choice latency (0.142 ≤ R ≤ 0.148) were more than an order of magnitude higher and were significantly nonzero (Table 5).

The observed distributions of the number of times subjects consistently chose the longer call did not differ between the ±1 SD and ±2 SD distributions for either static tests (Fig. S5A, n=71; χ²₍₅₎ = 5.35, p=0.375) or dynamic tests (Fig. S5B, n=71; χ²₍₅₎ = 1.11, p=0.953). Consequently, we separately pooled observed data for static and dynamic tests across the two differences in call duration before comparing with simulated data. The observed distributions for these pooled data were centered around two or three (out of five) choices of the longer call in static and dynamic tests, respectively (static tests: mode=2, median=2, Fig. 3A; dynamic tests mode=3, median=3, Fig. 3B). Comparisons of observed data from real subjects (Figs. 3A, B) to the expected distributions from simulated subjects across varying degrees of individual consistency indicate females exhibited little or no consistency in making the same choice across multiple tests. Expected distributions (Fig. 3C) shifted between a Gaussian distribution and a completely bimodal distribution as the individual consistency of simulated subjects increased from $\widehat{M} = 0.0$ (completely random) to $\widehat{M} = 1.0$ (initially random but completely consistent thereafter). For static tests, the observed and expected distributions were not different for \widehat{M} equal to 0.0, 0.2, and 0.4 (Fig. 3C), but differed significantly (p < 0.05) for values of \widehat{M} equal to 0.6 and 0.8. In dynamic tests, the observed and expected distributions were not different for \widehat{M} equal to 0.0 and 0.2 (Fig. 3C) but differed significantly (p < 0.05) for values of \widehat{M} equal to 0.4, 0.6, and 0.8.

Commitment to initial preferences

Most subjects were committed to choosing their initial preference; that is, they ultimately chose the speaker associated with the first preference boundary they crossed (Fig. 4A). Moreover, they did so regardless of any differences in call duration and rate between the two alternatives and whether stimuli were swapped between speakers. Across all factorial combinations of Stimulus (0 SD, ±1 SD, and ±2 SD) and Test Type (static and dynamic), the proportions of tests in which subjects reversed course (Fig. 4B) after crossing the initial preference boundary were uniformly low, varying between 0.07 and 0.18 (Fig. 4C). Model selection yielded

Table 5 Repeatability

Response variable	Type	R [95% CI]	<i>p</i> (LRT)	<i>p</i> (permutation test)
Choice	Adjusted	0.010 [0, 0.053]	0.314	0.254
	Unadjusted	0.008 [0, 0.047]	0.347	0.272
Choice latency	Adjusted	0.142 [0.087, 0.205]	<0.001	0.001
	Unadjusted	0.148 [0.090, 0.207]	<0.001	0.001
Reversals	Adjusted	0.066 [0.013, 0.113]	<0.001	0.001
	Unadjusted	0.081 [0.028, 0.135]	<0.001	0.001

Estimates of adjusted and unadjusted repeatability (R), 95% confidence intervals, and results from likelihood-ratio tests (LRT) and permutation tests derived from the best GLMM models for the response variables of choice, choice latency, and reversals

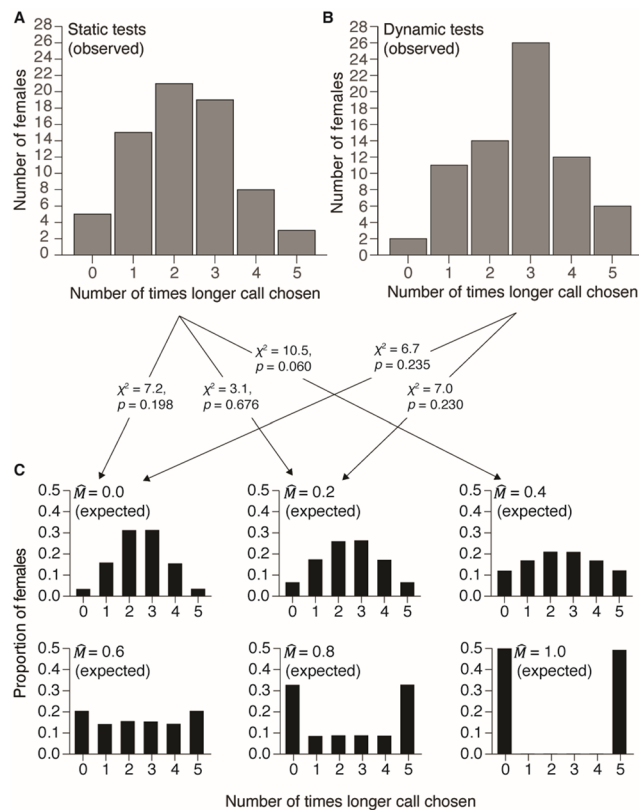


Fig. 3 Observed and expected distributions of individual consistency. Observed distributions of the number of times real subjects consistently chose the longer call in **A** static tests and **B** dynamic tests. **C** Expected distributions of the number of times simulated subjects consistently chose the longer call as a function of the degree of consistency (\hat{M}). Arrows (with associated χ^2 and *p* values) connect observed and expected distributions that did not differ significantly

two top GLMM models (Table S5), but the marginal coefficients of determination for these models indicate the fixed effects included were poor predictors of reversals (marginal $R^2=0.04$; Table 6). Stimulus was the only predictor included in the best GLMM model for reversals (Tables 6 and S5). Compared with the 0 SD difference in call duration, reversals were significantly more likely when the difference in call duration was ± 1 SD ($p=0.003$) but not ± 2 SD ($p=0.061$), though the latter was marginally significant

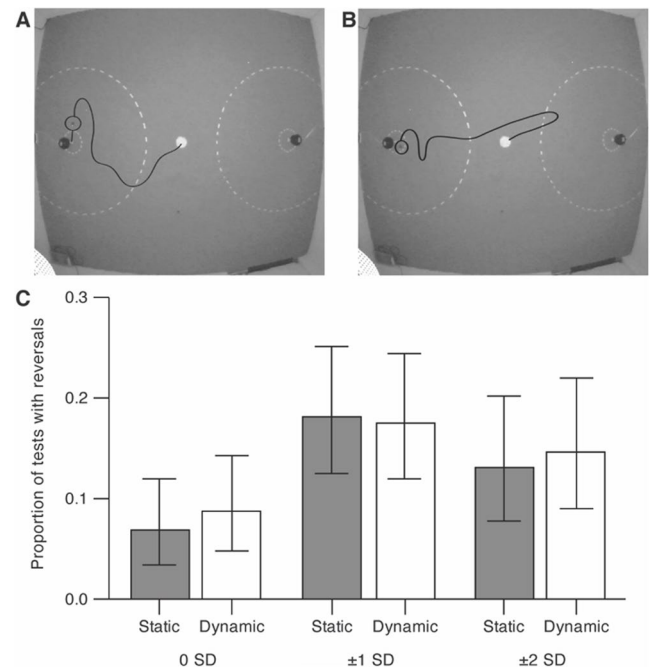


Fig. 4 Reversals in choice tests. **A** and **B** show two examples of traced phonotaxis paths (with circled frog) as monitored through the overhead camera. **A** Most females chose the speaker associated with the first preference boundary they crossed. **B** A minority of females reversed course and ultimately chose the speaker on the opposite side of the arena from the preference boundary they first crossed. **C** The proportions ($\pm 95\%$ exact binomial confidence intervals) of tests in which subjects reversed in the 0 SD, ± 1 SD and ± 2 SD conditions, shown separately for static and dynamic tests

(Table 6; Fig. 4C). The only other competitive model included Stimulus and Test Type as predictors (Table S5). The effect of Test Type in this second-best model was not significant ($p=0.680$).

There was variation among individuals in reversal rates. The difference between marginal and conditional coefficients of determination from the best GLMM models indicated the random effect of subject ID was relatively important compared to fixed effects (marginal: $R^2=0.04$; conditional: $0.19 \leq R^2 \leq 0.20$; Table 6). The repeatability of reversals was significantly nonzero and fell between estimates of repeatability for choice and choice latency (Table 5).

Table 6 GLMM results for reversals

Model	Term	Estimate	SE	Wald z	p	Cond. R^2	Marg. R^2
Stimulus	Intercept	-2.712	0.275	-9.857	<0.001	0.19	0.04
	Stimulus (± 1 SD)	0.985	0.336	2.934	0.003		
	Stimulus (± 2 SD)	0.673	0.359	1.877	0.061		
Stimulus+ Test Type	Intercept	-2.756	0.296	-9.325	<0.001	0.20	0.04
	Stimulus (± 1 SD)	0.985	0.336	2.934	0.003		
	Stimulus (± 2 SD)	0.673	0.359	1.877	0.061		
	Test Type (dynamic)	0.085	0.207	0.413	0.680		

The two top GLMM models ($AIC_c \leq 2.0$) for reversals based on differences in call duration and call rate. The reference conditions for the predictors of Stimulus and Test Type were 0 SD and static tests, respectively. The estimate, standard error (SE), statistic (Wald z) and p -value are shown for each term in each model, and the conditional R^2 and marginal R^2 are shown for each model. Results from model selection are included in the Supplementary Material

Discussion

Our results support an inter-signal interaction hypothesis and refute the differential feature weighting hypothesis. At the population level, there was no evidence females chose between simulated males based on differences in call duration and call rate when calling effort was held constant. Instead, choices were random with respect to differences in call duration and call rate. Estimates of repeatability of choice were low and non-significant, and the outcomes of repeated tests were similar to expectations based on repeatedly flipping a coin. Hence, the choices of individual females mirrored the lack of female choice observed at the population level. Finally, females were highly committed to the simulated male they initially approached, even when calls differing in duration and rate were dynamically swapped between speakers. Based on these results, we conclude there is a multiplicative interaction between call duration and call rate, such that individual females choose mates by attending to the emergent property of calling effort instead of its constituent components. This conclusion rests on two caveats.

First, as noted earlier, an inter-signal interaction hypothesis equates to a null hypothesis of “no difference” given our chosen study design. We held calling effort constant because there is already robust evidence female gray treefrogs prefer longer calls and faster call rates when calling effort is allowed to covary. But could our results represent a failure to reject a false null hypothesis? Such an interpretation rests on the unlikely outcome of obtaining the same null result across ten sequential tests. Moreover, our results replicate earlier work by Ward et al. (2013) showing that females largely ignore differences in call duration and rate when calling effort is held constant. Beyond providing greater confidence in negative results, replicating behavioral studies – particularly those using wild or field-collected animals – is fundamental to overcoming bias associated with the selective reporting of positive results or results consistent with theory (Palmer 2000; Kelly 2006, 2019; Ryan 2011; Fraser et al. 2020).

Second, an inter-signal interaction hypothesis, as it pertains to calling effort, makes the same predictions as the influential sensory bias hypothesis (sensu Ryan and Keddy-Hector 1992), which holds that females prefer signals with more total acoustic energy because such signals elicit greater sensory stimulation. While “energy” and “sensory stimulation” are certainly important considerations in studies of signal design, these terms possess a certain vagueness that does not always connect in clear ways to underlying neuro-sensory mechanisms. Both insects and frogs, for example, possess well-described neural mechanisms for processing their temporally structured signals (Gerhardt and Huber 2002; Hartbauer and Römer 2014; Rose 2018). In frogs, including Cope’s gray treefrog (Edwards et al. 2002, 2007; Rose et al. 2015; Kamath et al. 2026), electrophysiological studies have ruled out temporal summation of acoustic energy as a mechanism for how pulsatile calls are processed by the central nervous system. Specifically, pulse rate is not processed by summing acoustic energy, but rather by neural computations that integrate precisely-timed excitation and inhibition to “count” discrete interpulse intervals (reviewed in Rose 2018). Carried out by so-called “interval-counting neurons” in the midbrain, these computations mediate species recognition as well as discrimination among different conspecific call types (e.g., advertisement versus aggressive calls). Such findings raise the intriguing possibility that receiver preferences based on calling effort also engage counting mechanisms that are more complex than summation of acoustic energy. While our results cannot rule out a sensory bias explanation, we prefer a hypothetical framework that makes fewer (implicit) assumptions about sensory mechanisms.

Different classes of inter-signal interaction hypotheses have been proposed to explain aspects of communication from the separate perspectives of signalers and receivers (Candolin 2003; Hebets and Papaj 2005). From the signaler’s perspective, the negative phenotypic and genetic covariance between call duration and call rate are consistent with a *multitasking hypothesis* (Hebets and Papaj 2005; Ward et

al. 2013). According to this hypothesis, a signaler's ability to produce one signal or signal component is constrained by, and thus negatively correlated with, their ability to produce another. In gray treefrogs, males are constrained, almost certainly by performance costs, in terms of how fast they can repeat long calls (Taigen and Wells 1985; Wells and Taigen 1986; Grafe 1997; McLister 2001; Reichert and Gerhardt 2012). But what about the perspective of female receivers? Our data are broadly in line with the *emergence hypothesis* (Hebets and Papaj 2005), which holds that receivers are better able to discriminate among signalers based on properties that emerge from combining multiple signals or signal components in such a way that a new, unique signal is formed. Within the experimental parameters of our study, we demonstrated that females chose simulated mates randomly and ignored differences in call duration and call rate provided calling effort was equal. This result suggests females integrate information about the number of pulses males produce over multiple calls before making their final choice of a mate. But over what time window? Our best estimate comes from a study of the tetraploid gray treefrog by Schwartz et al. (2004). That study combined field and laboratory tests of female choice to estimate the most likely assessment time window as being on the order of 58 s to 161 s. While females may listen longer to calling males in a chorus, their final choice of a mate appears influenced by differences in calling effort that occur on the order of tens to hundreds of seconds just prior to choosing. At present, the mechanistic basis of how females assess calling effort is unknown. An intriguing direction for future studies of the anuran auditory system would be to compare the neural basis of temporal information processing over the different times scales of mate choice. For example, how does recognizing a conspecific by counting pulses within calls over the course of tens to hundreds of milliseconds compare to discriminating among conspecifics based on differences in their calling effort over the course of tens to hundreds of seconds (e.g., Hartbauer and Römer 2014)? As discussed next, such studies should be conducted in a carefully considered comparative framework.

The present study highlights the potential for evolutionary lability in female preferences among closely related lineages (Ritchie 1996; Boughman et al. 2005; Rodriguez et al. 2006). Extant populations of the diploid Cope's gray treefrog (*H. chrysoscelis*) can be assigned to two genetically distinct lineages, a Western clade and an Eastern+Central clade, with a monophyletic Eastern clade nested within the Central lineage (Holloway et al. 2006; Booker et al. 2022). Subjects in the present study were from the Western clade. The tetraploid eastern gray treefrog (*H. versicolor*) arose through a single autopolyploid genome duplication event within a now-extinct clade of the diploid. Previous

studies have identified differences between the two species, or between lineages within one species, in female preferences for pulse shape (Gerhardt and Doherty 1988; Gerhardt 2005), pulse rate (Gerhardt 1994, 2008; Schul and Bush 2002), spectral content (Gerhardt et al. 2007; Gupta and Bee 2023), and spectral cues related to body size (Morris and Yoon 1989; Schrode et al. 2012). In contrast to the present study of the Western Clade of the diploid in Minnesota, females of both the tetraploid (Klump and Gerhardt 1987; Gerhardt et al. 1996; Schwartz et al. 2001) and the Central lineage of the diploid in Missouri (Gerhardt et al. 1996) prefer longer calls at slower rates even when calling effort remains constant. The apparent lineage difference within the diploid deserves additional study, but also additional comment here. Recall that our stimuli differed symmetrically in call duration around the population mean (± 1 SD, ± 2 SD). In their tests of the diploid, Gerhardt et al. (1996) used longer and shorter call alternatives that were -1 SD and -3 SD, respectively, below the mean call duration in local Missouri populations. This is relevant, because both diploid and tetraploid females exhibit nonlinear preferences that strongly favor longer over shorter calls when both are of lower-than-average duration (Gerhardt et al. 2000; Bee 2008; LaBarbera et al. 2020). Additional work is needed to determine whether the lineage difference described here is real or reflects heretofore unidentified nonlinearities in preferences for greater calling effort.

The study of individual variation in mate choice gained attention following the influential review by Jennions and Petrie (1997). The repeatability of female choice based on differences in call duration and call rate in the present study was not significantly different from zero. This result underscores the role of calling effort as an emergent property on which females based their choices, which were random with respect to call duration and call rate. Repeatability varies substantially across different behaviors, with mate choice being among the most studied and often least repeatable, and the low repeatability for female choice reported here is consistent with values reported in other vertebrates (Bell et al. 2009). A similar lack of repeatability in frog mate choice was observed in experiments with túngara frogs (Kime et al. 1998). Notably, our study and that by Kime et al. (1998) demonstrated low repeatability despite what are ideal conditions for detecting it: a controlled laboratory assay using call alternatives of similar attractiveness at the population level and back-to-back repeat testing of individual females. When mate choice alternatives vary dramatically in attractiveness such that one is strongly preferred by females, there can be limited variation to parse (and hence low repeatability); shorter time intervals between repeat tests are also associated with higher repeatability (Bell et al. 2009). Hence, studies to date suggest population-level preferences in female frogs may

largely reflect within-individual variation; that is, evidence is weak for strong polymorphisms in female preferences for male call characteristics. In a recent study of the tetraploid gray treefrog, Stratman and Höbel (2024) documented considerable within-population variation in female preferences for call duration, call rate, and dominant frequency. However, females were not tested repeatedly with the same stimuli, so the extent to which the observed within-population variation reflects within-individual variation remains unclear. Within-individual variation in frog mate choice might be due to a variety of factors, including rapidly shifting female condition, and therefore behavior, during reproductive readiness (Baugh and Ryan 2009), which in turn could be driven by positive feedback activity in gonadal hormones like estrogens (Chakraborty and Burmeister 2009). Such rapidly fluctuating reproductive states could obscure among-individual differences in mate choice behavior.

Other aspects of mate choice behavior in gray treefrogs may, in fact, be highly repeatable. Evidence from the tetraploid gray treefrog suggests a female's mate preferences are distinct from the *effort* she is willing to invest to acquire a preferred mate (Feagles and Höbel 2022a). This investment of effort is taken to be an indicator of a female's "choosiness" (Jennions and Petrie 1997; Kuczynski et al. 2017; Feagles and Höbel 2022a, b; also referred to as "preference strength," Gerhardt et al. 2000). In gray treefrogs, choosiness for longer calls has been measured as the extent to which the SPL of a preferred call can be attenuated – thereby simulating a preferred male calling from a greater distance – and still elicit a preference. Measured in this manner, there is often substantial repeatability in female choosiness for longer calls (Gerhardt et al. 2000; Feagles and Höbel 2022a, b). Importantly, these previous metrics of choosiness are functionally analogous to reversal behavior in the present study: both quantify the degree of persistence or effort a female invests when pursuing the preferred alternative under conditions that impose real or perceived locomotor costs. By holding calling effort constant, our design removed strong attractiveness differences, thereby limiting the opportunity for females to express stable differences in choice. Yet, like Baugh (2009) and consistent with previous findings with the tetraploid gray treefrog (Bastien et al. 2018), we found significant repeatability in how females executed their decisions—specifically, in choice latency and reversal behavior. Hence, consistent among-individual differences in mate choice may be more likely to appear in components of the decision process itself, especially under conditions that expose tradeoffs and allow preferences and choosiness to be expressed quantitatively.

None of the individual subject variables we measured (body size [TFL], condition [SMI], and endogenous baseline glucocorticoid levels [CORT]; Table S6) predicted choice

latency or reversals, nor did any of these variables predict the choices females ultimately made. This outcome provides a mixed level of support and corroboration for findings in earlier studies. In the tetraploid gray treefrog, Kuczynski et al. (2017) found a positive relationship between female body size and the likelihood of choosing a more attractive (and more distant) male having a higher calling effort associated with producing longer calls. In túngara frogs, individual differences in reversal rates were associated with body condition (based on pre-oviposition mass); higher condition females reversed less often (Baugh and Ryan 2009). This result led to the hypothesis that greater oviposition urgency (more hydrated egg masses leading to higher SMI) is linked to reduced female reversals. In contrast to túngara frogs, which were collected and tested before they had commenced construction of the foam nest that coincides with oviposition, we had no knowledge about how many, if any, of each subject's eggs had been oviposited prior to collection. We found no relationship between SMI (based on post-oviposition mass) and reversals, which is consistent with a similar finding in the tetraploid gray treefrog (Kuczynski et al. 2017). Lastly, in a related study of the tetraploid, there was no relationship between oviposition timing and reversals (Bastien et al. 2018). Together, these studies suggest body size and condition may have limited explanatory power as determinants of features of mate choice in some frogs, such as latency to choose and commitment.

The influence of glucocorticoids on female choice has received increased attention in recent years (Leary and Baugh 2020). In a previous study by Baugh et al. (2021) that experimentally manipulated CORT levels in the diploid gray treefrog (*H. chrysoscelis*), moderate exogenous increases in circulating CORT increased female reversal rates in dynamic two-choice tests in which alternatives differed simultaneously in call duration and calling effort; low and high CORT elevations resulted in nominal declines in reversal rate. This inverted-U hormone-behavior relationship was further explained by individual variation in the expression of the glucocorticoid receptor (GR) in the forebrain (Rodríguez-Santiago et al. 2024). Given the nature of the dynamic mate choice assay used in these studies (which has a locomotor cost associated with choosy behavior), it is not surprising that metabolic hormones such as glucocorticoids are involved in modulating costly decision making. However, these results contrast with those of the present study, in which calling effort was held constant and endogenous levels of plasma CORT failed to predict female choice, choice latency, consistency of choice, or commitment to initial preferences, despite a large range of endogenous CORT concentrations across individuals (Table S6). In addition to the difference in whether calling effort was held constant, these contrasting effects of CORT on commitment might

also reflect fundamental differences of action between rapidly elevated ('stress-induced') CORT levels, as measured by Baugh et al. (2021), and baseline endogenous levels, as measured in this study.

In conclusion, we suggest multitasking and emergence are two complementary inter-signal interaction hypotheses that explain the directional preferences of female *H. chrysoscelis* for two attributes of signaling males – call duration and call rate – that negatively covary due to performance-related constraints. Future neuroethological studies are needed to determine how the emergent property of calling effort is coded in receivers' nervous systems. We speculate that the observed absence of individual variation in female choice, but the presence of individual variation in choice latency and commitment, may reflect differences in underlying modulatory mechanisms involving the forebrain. Dorsal thalamic lesions that target forebrain processing do not impact responsiveness and preferences during mate choice in gray treefrogs (Endepols et al. 2003). But real-time plasticity in preference expression may involve forebrain processing. Therefore, we might expect individual differences in how choices are made to result from differences among females in physiological factors such as hormone levels (e.g. elevated CORT; Baugh et al. 2021) and neural receptor expression in the forebrain (e.g., GR; Rodriguez-Santiago et al. 2024). In other words, basic preferences and acoustically guided locomotion might be coded in more 'hard-wired' midbrain sensory and sensorimotor circuits, which might then be modulated by more plastic (hormone-dependent) forebrain circuits (reviewed in Wilczynski and Lynch 2011).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00265-026-03717-5>.

Acknowledgements We thank the Minnesota Department of Natural Resources, Ramsey County Parks and Recreation, and the Three Rivers Park District for access to wetlands and animals; Shoshana Altman, Kaitlyn Bonnema, Sarah Bonnema, Ava Brandewie, Jadon Chiu, Callie Cho, Madison Deile, Luke Drummer, Alyssa Grutt, Rishi Gulati, Adam Hartman, Isabelle Hoversten, James Landsman, Daniel Oakes, Alice Onyango-Opiyo, Adam Rohlader, and Annika Ruppert for help collecting and testing frogs; Lata Kalra, Lea Pollack, Jon Richardson, and Chris Wojan for numerous helpful discussions about this work; and two anonymous reviewers for helpful feedback.

Author contributions KJK: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft. ATB: Conceptualization, Methodology, Resources, Formal analysis, Investigation, Writing - original draft, Funding acquisition. MAB: Conceptualization, Methodology, Resources, Formal analysis, Investigation, Writing - original draft, Funding acquisition, Supervision, Project administration.

Funding This work was supported by the National Science Foundation (IOS-2154204); by the University of Minnesota Graduate Program in Ecology, Evolution, and Behavior; and by Swarthmore College.

Data availability Data is provided within the manuscript or supplementary information files.

Declarations

Ethical approval This work was approved by the University of Minnesota Institutional Animal Care and Use Committee (#2301–40692A). All applicable international, national, and/or institutional guidelines for the use of animals were followed.

Competing interests The authors declare no competing interests.

References

- Adkins-Regan E (1998) Hormonal mechanisms of mate choice. *Am Zool* 38:166–178
- Andersson M (1994) Sexual selection. Princeton University Press, Princeton, NJ
- Bastien B, Farley G, Ge F, Malin JS, Simon-Plumb CL, Pulley DM, Yang C, Baugh AT (2018) The waiting and mating game: condition dependent mate sampling in female gray treefrogs (*Hyla versicolor*). *Front Ecol Evol* 6:140
- Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48
- Baugh AT (2009) The development and execution of mate choice in túngara frogs. PhD Dissertation, The University of Texas at Austin
- Baugh AT (2024) Male Cope's gray treefrogs (*Hyla chrysoscelis*) in amplexus have elevated and correlated steroid hormones compared to solitary males. *Gen Comp Endocrinol* 345:114391
- Baugh AT, Ryan MJ (2009) Female túngara frogs vary in commitment to mate choice. *Behav Ecol* 20:1153–1159
- Baugh AT, Ryan MJ (2010) Mate choice in response to dynamic presentation of male advertisement signals in túngara frogs. *Anim Behav* 79:145–152
- Baugh AT, Bee MA, Gall MD (2019) The paradox of hearing at the lek: auditory sensitivity increases after breeding in female gray treefrogs (*Hyla chrysoscelis*). *J Comp Physiol A* 205:629–639
- Baugh AT, Gall MD, Silver SC, Bee MA (2021) Moderately elevated glucocorticoids increase mate choosiness but do not affect sexual proceptivity or preferences in female gray treefrogs. *Horm Behav* 130:104950
- Bee MA (2008) Parallel female preferences for call duration in a diploid ancestor of an allotetraploid treefrog. *Anim Behav* 76:845–853
- Bee MA, Kalra L (2025) Stereotyped but irrelevant: frequency modulation does not promote sound pattern recognition or auditory grouping in grey treefrogs (*Hyla versicolor*). *Bioacoustics* 34:30–49
- Bee MA, Micheyl C (2008) The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *J Comp Psychol* 122:235–251
- Bell AM, Hankison SJ, Laskowski KL (2009) The repeatability of behaviour: a meta-analysis. *Anim Behav* 77:771–783
- Booker WW, Gerhardt HC, Lemmon AR, Ptacek MB, Hassinger ATB, Schul J, Lemmon EM (2022) The complex history of genome duplication and hybridization in North American gray treefrogs. *Mol Biol Evol* 39:msab316
- Boughman JW, Rundle HD, Schluter D (2005) Parallel evolution of sexual isolation in sticklebacks. *Evolution* 59:361–373
- Brooks R, Endler JA (2001) Female guppies agree to differ: phenotypic and genetic variation in mate-choice behavior and the consequences for sexual selection. *Evolution* 55:1644–1655
- Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, Nielsen A, Skaug HJ, Mächler M, Bolker B (2017)

- glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R J* 9:378–400
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York
- Bush SL, Gerhardt HC, Schul J (2002) Pattern recognition and call preferences in treefrogs (Anura: Hylidae): a quantitative analysis using a no-choice paradigm. *Anim Behav* 63:7–14
- Candolin U (2003) The use of multiple cues in mate choice. *Biol Rev* 78:575–595
- Chakraborty M, Burmeister SS (2009) Estradiol induces sexual behavior in female túngara frogs. *Horm Behav* 55:106–112
- Clark CJ (2012) The role of power versus energy in courtship: what is the ‘energetic cost’ of a courtship display? *Anim Behav* 84:269–277
- R Core Team (2024) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>
- Cotton S, Small J, Pomiankowski A (2006) Sexual selection and condition-dependent mate preferences. *Curr Biol* 16:R755–R765
- Davis AG, Leary CJ (2015) Elevated stress hormone diminishes the strength of female preferences for acoustic signals in the green treefrog. *Horm Behav* 69:119–122
- Dingemanse NJ, Dochtermann NA (2013) Quantifying individual variation in behaviour: mixed-effect modelling approaches. *J Anim Ecol* 82:39–54
- Edwards CJ, Alder TB, Rose GJ (2002) Auditory midbrain neurons that count. *Nat Neurosci* 5:934–936
- Edwards CJ, Leary CJ, Rose GJ (2007) Counting on inhibition and rate-dependent excitation in the auditory system. *J Neurosci* 27:13384–13392
- Endepols H, Feng AS, Gerhardt HC, Schul J, Walkowiak W (2003) Roles of the auditory midbrain and thalamus in selective phonotaxis in female gray treefrogs (*Hyla versicolor*). *Behav Brain Res* 145:63–77
- Feagles O, Höbel G (2022a) Mate preferences and choosiness are distinct components of mate choice in eastern gray treefrogs (*Hyla versicolor*). *Am Nat* 200:506–517
- Feagles OS, Höbel G (2022b) Female gray treefrogs maintain mate choice decisions under predation threat, but adjust movements to reduce conspicuousness during mate approach. *Behav Ecol Sociobiol* 76:17
- Fowler-Finn KD, Rodriguez RL (2013) Repeatability of mate preference functions in *Enchenopa* treehoppers (Hemiptera: Membracidae). *Anim Behav* 85:493–499
- Fraser H, Barnett A, Parker TH, Fidler F (2020) The role of replication studies in ecology. *Ecol Evol* 10:5197–5207
- Gall MD, Bee MA, Baugh AT (2019) The difference a day makes: breeding remodels hearing, hormones and behavior in female Cope’s gray treefrogs (*Hyla chrysoscelis*). *Horm Behav* 108:62–72
- Gerhardt HC (1975) Sound pressure levels and radiation patterns of vocalizations of some North American frogs and toads. *J Comp Physiol* 102:1–12
- Gerhardt HC (1981) Mating call recognition in the green treefrog (*Hyla cinerea*): importance of two frequency bands as a function of sound pressure level. *J Comp Physiol* 144:9–16
- Gerhardt HC (1994) Reproductive character displacement of female mate choice in the grey treefrog *Hyla chrysoscelis*. *Anim Behav* 47:959–969
- Gerhardt HC (1995) Phonotaxis in female frogs and toads: execution and design of experiments. In: Klump GM, Dooling RJ, Fay RR, Stebbins WC (eds) *Methods in comparative psychoacoustics*. Birkhäuser, Basel, pp 209–220
- Gerhardt HC (2001) Acoustic communication in two groups of closely related treefrogs. *Adv Stud Behav* 30:99–167
- Gerhardt HC (2005) Advertisement-call preferences in diploid-tetraploid treefrogs (*Hyla chrysoscelis* and *Hyla versicolor*): implications for mate choice and the evolution of communication systems. *Evolution* 59:395–408
- Gerhardt HC (2008) Phonotactic selectivity in two cryptic species of gray treefrogs: effects of differences in pulse rate, carrier frequency and playback level. *J Exp Biol* 211:2609–2616
- Gerhardt HC, Doherty JA (1988) Acoustic communication in the gray treefrog, *Hyla versicolor*: evolutionary and neurobiological implications. *J Comp Physiol A* 162:261–278
- Gerhardt HC, Huber F (2002) Acoustic communication in insects and anurans: common problems and diverse solutions. Chicago University Press, Chicago
- Gerhardt HC, Dyson ML, Tanner SD (1996) Dynamic properties of the advertisement calls of gray tree frogs: patterns of variability and female choice. *Behav Ecol* 7:7–18
- Gerhardt HC, Tanner SD, Corrigan CM, Walton HC (2000) Female preference functions based on call duration in the gray tree frog (*Hyla versicolor*). *Behav Ecol* 11:663–669
- Gerhardt HC, Martinez-Rivera CC, Schwartz JJ, Marshall VT, Murphy CG (2007) Preferences based on spectral differences in acoustic signals in four species of treefrogs (Anura: Hylidae). *J Exp Biol* 210:2990–2998
- Girgenrath M, Marsh RL (1997) In vivo performance of trunk muscles in tree frogs during calling. *J Exp Biol* 200:3101–3108
- Girgenrath M, Marsh RL (1999) Power output of sound-producing muscles in the tree frogs *Hyla versicolor* and *Hyla chrysoscelis*. *J Exp Biol* 202:3225–3237
- Grafe U (1997) Use of metabolic substrates in the gray treefrog *Hyla versicolor*: implications for calling behavior. *Copeia* 1997:356–362
- Gupta S, Bee MA (2023) Female preferences for the spectral content of advertisement calls in Cope’s gray treefrog (*Hyla chrysoscelis*). *J Comp Physiol A* 209:31–45
- Gupta S, Alluri RK, Rose GJ, Bee MA (2021) Neural basis of acoustic species recognition in a cryptic species complex. *J Exp Biol* 224:jeb243405
- Hartbauer M, Römer H (2014) From microseconds to seconds and minutes-time computation in insect hearing. *Front Physiol* 5:1–11
- Hebets EA, Papaj DR (2005) Complex signal function: developing a framework of testable hypotheses. *Behav Ecol Sociobiol* 57:197–214
- Höglund J, Alatalo RV (1995) *Leks*. Princeton University Press, Princeton, NJ
- Holloway AK, Cannatella DC, Gerhardt HC, Hillis DM (2006) Polyploids with different origins and ancestors form a single sexual polyploid species. *Am Nat* 167:E88–E101
- Hunt J, Brooks R, Jennions Michael D (2005) Female mate choice as a condition-dependent life-history trait. *Am Nat* 166:79–92
- Jennions MD, Petrie M (1997) Variation in mate choice and mating preferences: a review of causes and consequences. *Biol Rev* 72:283–327
- Kalra L, Bee M (2025) Auditory streaming and rhythmic masking release in Cope’s gray treefrog. *J Acoust Soc Am* 157:2319–2329
- Kamath VG, Mukhopadhyay A, Alluri RK, Acord AS, Rose GJ, Bee MA (2026) Evidence that interval-counting neurons play a critical role in call recognition by Cope’s gray treefrogs. *J Comp Physiol A* 212(1):87–101. <https://doi.org/10.1007/s00359-025-01783-0>
- Kelly CD (2006) Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. *Q Rev Biol* 81:221–236
- Kelly CD (2019) Rate and success of study replication in ecology and evolution. *PeerJ* 7:e7654
- Kime NM, Rand AS, Kapfer M, Ryan MJ (1998) Consistency of female choice in the túngara frog: a permissive preference for complex characters. *Anim Behav* 55:641–649

- Klump GM, Gerhardt HC (1987) Use of non-arbitrary acoustic criteria in mate choice by female gray tree frogs. *Nature* 326:286–288
- Kuczynski MC, Getty T, Gering E (2017) Larger females are choosier in the gray treefrog (*Hyla versicolor*). *Behav Process* 135:29–35
- LaBarbera K, Nelson PB, Bee MA (2020) Mate choice and the ‘opposite miss’ to Weber’s law: proportional processing governs signal preferences in a treefrog. *Anim Behav* 168:199–209
- Leary CJ, Baugh AT (2020) Glucocorticoids, male sexual signals, and mate choice by females: implications for sexual selection. *Gen Comp Endocrinol* 288:113354
- Li H, Schrode KM, Bee MA (2022) Vocal sacs do not function in multimodal mate attraction under nocturnal illumination in Cope’s grey treefrog. *Anim Behav* 189:127–146
- Love EK, Bee MA (2010) An experimental test of noise-dependent voice amplitude regulation in Cope’s grey treefrog, *Hyla chrysoscelis*. *Anim Behav* 80:509–515
- McLister JD (2001) Physical factors affecting the cost and efficiency of sound production in the treefrog *Hyla versicolor*. *J Exp Biol* 204:69–80
- McLister JD, Stevens ED, Bogart JP (1995) Comparative contractile dynamics of calling and locomotor muscles in three hylid frogs. *J Exp Biol* 198:1527–1538
- Møller AP, Pomiankowski A (1993) Why have birds got multipel sexual ornaments? *Behav Ecol Sociobiol* 32:167–176
- Morris MR, Yoon SL (1989) A mechanism for female choice of large males in the treefrog *Hyla chrysoscelis*. *Behav Ecol Sociobiol* 25:65–71
- Murphy CG, Gerhardt HC (1996) Evaluating the design of mate-choice experiments: the effect of amplexus on mate choice by female barking treefrogs, *Hyla gratiosa*. *Anim Behav* 51:881–890
- Nakagawa S, Schielzeth H (2010) Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol Rev* 85:935–956
- Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* 4:133–142
- Nakagawa S, Johnson PCD, Schielzeth H (2017) The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J R Soc Interface* 14:20170213
- Palmer AR (2000) Quasireplication and the contract of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annu Rev Ecol Syst* 31:441–480
- Peig J, Green AJ (2009) New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. *Oikos* 118:1883–1891
- Podos J (2022) Costs, constraints and sexual trait elaboration. *Anim Behav* 184:209–214
- Posit team (2023) RStudio: integrated development environment for R (version 2023.6.2.561). Posit Software. PBC, Boston, MA. <https://www.posit.co/>
- Prestwich KN (1994) The energetics of acoustic signaling in anurans and insects. *Am Zool* 34:625–643
- Reichert MS, Gerhardt HC (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
- Reichert MS, de la Hera I, Moiron M (2024) Partitioning variance in a signaling trade-off under sexual selection reveals among-individual covariance in trait allocation. *Evolution* 78:1150–1160
- Ritchie MG (1996) The shape of female mating preferences. *P Natl Acad Sci USA* 93:14628–14631
- Rodriguez RL, Ramaswamy K, Crocroft RB (2006) Evidence that female preferences have shaped male signal evolution in a clade of specialized plant-feeding insects. *Proc R Soc Lond B* 273:2585–2593
- Rodriguez-Santiago M, Ruppert A, Gall MD, Hoke K, Bee MA, Baugh AT (2024) In your COURT: corticosterone and its receptors in the brain underlie mate choosiness in female Cope’s gray treefrogs (*Hyla chrysoscelis*). *Horm Behav* 159:105477
- Rose GJ (2018) The numerical abilities of anurans and their neural correlates: insights from neuroethological studies of acoustic communication. *Phil Trans R Soc B* 373:20160512
- Rose GJ, Hanson JL, Leary CJ, Graham JA, Alluri RK, Vasquez-Opazo GA (2015) Species-specificity of temporal processing in the auditory midbrain of gray treefrogs: interval-counting neurons. *J Comp Physiol A* 201:485–503
- Rosenthal GG (2017) Mate choice: the evolution of sexual decision making from microbes to humans. Princeton University Press, Princeton, NJ
- Ryan MJ (2011) Replication in field biology: the case of the frog-eating bat. *Science* 334:1229–1230
- Ryan MJ, Cummings ME (2013) Perceptual biases and mate choice. *Annu Rev Ecol Evol S* 44:437–459
- Ryan MJ, Keddy-Hector A (1992) Directional patterns of female mate choice and the role of sensory biases. *Am Nat* 139:S4–S35
- Ryan MJ, Perrill SA, Wilczynski W (1992) Auditory tuning and call frequency predict population-based mating preferences in the cricket frog, *Acris crepitans*. *Am Nat* 139:1370–1383
- Schrode K, Ward JL, Vélez A, Bee MA (2012) Female preferences for spectral call properties in the western genetic lineage of Cope’s gray treefrog (*Hyla chrysoscelis*). *Behav Ecol Sociobiol* 66:1595–1606
- Schul J, Bush SL (2002) Non-parallel coevolution of sender and receiver in the acoustic communication system of treefrogs. *Proc R Soc Lond B* 269:1847–1852
- Schwartz JJ, Rahmeyer KM (2006) Calling behavior and the capacity for sustained locomotory exercise in the Gray Treefrog (*Hyla versicolor*). *J Herpetol* 40:164–171
- Schwartz JJ, Buchanan BW, Gerhardt HC (2001) Female mate choice in the gray treefrog (*Hyla versicolor*) in three experimental environments. *Behav Ecol Sociobiol* 49:443–455
- Schwartz JJ, Buchanan BW, Gerhardt HC (2002) Acoustic interactions among male gray treefrogs, *Hyla versicolor*, in a chorus setting. *Behav Ecol Sociobiol* 53:9–19
- Schwartz JJ, Huth K, Hutchin T (2004) How long do females really listen? Assessment time for female mate choice in the grey treefrog, *Hyla versicolor*. *Anim Behav* 68:533–540
- Schwartz JJ, Brown R, Turner S, Dushaj K, Castano M (2008) Interference risk and the function of dynamic shifts in calling in the gray treefrog (*Hyla versicolor*). *J Comp Psychol* 122:283–288
- Schwartz JJ, Crimarco NC, Bregman Y, Umeoji KR (2013) An investigation of the functional significance of responses of the gray treefrog (*Hyla versicolor*) to chorus noise. *J Herpetol* 47:354–360
- Stoffel MA, Nakagawa S, Schielzeth H (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol Evol* 8:1639–1644
- Stratman KD, Höbel G (2019) Opportunity for female choice in the wild is frequently curtailed by low male trait variation in *Hyla versicolor*. *Behav Ecol Sociobiol* 73:59
- Stratman K, Höbel G (2024) Within-population variation in preference functions reveals substantial among-female disagreement in mate assessment. *J Evol Biol* 37:1278–1287
- Sullivan BK, Hinshaw SH (1992) Female choice and selection on male calling behavior in the gray treefrog *Hyla versicolor*. *Anim Behav* 44:733–744
- Taigen TL, Wells KD (1985) Energetics of vocalization by an anuran amphibian (*Hyla versicolor*). *J Comp Physiol B* 155:163–170
- Tanner JC, Hemingway CT (2025) Choice overload and its consequences for animal decision-making. *Trends Cogn Sci* 29:403–406
- Tanner JC, Ward JL, Shaw RG, Bee MA (2017) Multivariate phenotypic selection on a complex sexual signal. *Evolution* 71:1742–1754

- Tanner JC, Justison J, Bee MA (2020) SynSing: open-source MATLAB code for generating synthetic signals in studies of animal acoustic communication. *Bioacoustics* 29:731–752
- Tanner JC, Shein-Idelson M, Hoke KL, Bee MA (2025) Complex choice environments shelter unattractive signallers from sexual selection. *Proc R Soc B* 292:20250585
- Underhill VA, Höbel G (2017) Variation in nocturnal light levels does not alter mate choice behavior in female eastern gray treefrogs (*Hyla versicolor*). *Behav Ecol Sociobiol* 71:151
- Ward JL, Love EK, Vélez A, Buerkle NP, O'Bryan LR, Bee MA (2013) Multitasking males and multiplicative females: dynamic signalling and receiver preferences in Cope's grey treefrog. *Anim Behav* 86:231–243
- Welch AM, Smith MJ, Gerhardt HC (2014) A multivariate analysis of genetic variation in the advertisement call of the gray treefrog, *Hyla versicolor*. *Evolution* 68:1629–1639
- Wells KD, Taigen TL (1986) The effect of social interactions on calling energetics in the gray treefrog (*Hyla versicolor*). *Behav Ecol Sociobiol* 19:9–18
- Wilczynski W, Lynch KS (2011) Female sexual arousal in amphibians. *Horm Behav* 59:630–636

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.