

A Cognitive Framework for Mate Choice and Species Recognition

Steven M. Phelps,^{1,2,*} A. Stanley Rand,² and Michael J. Ryan^{2,3}

1. Department of Zoology, University of Florida, Gainesville, Florida 32605;
2. Smithsonian Tropical Research Institute, Smithsonian Institution, Gamboa, Panama;
3. Section for Integrative Biology, University of Texas, Austin, Texas 78712

*Submitted December 15, 2004; Accepted July 26, 2005;
Electronically published November 7, 2005*

Online enhancement: zip archive of Matlab files.

Keywords: mate choice, sexual selection, animal communication, animal cognition, decision theory, perceptual error.

Few decisions are as important to an individual as choosing when and with whom to mate. Accordingly, the study of sexual selection, concerned with the evolutionary consequences of mating decisions, occupies a prominent place in the evolutionary analysis of behavior. Over the past three decades, a tremendous diversity and depth of work has explored how the fitness consequences of mate choice shape animal decisions (Andersson 1994) and how these decisions in turn influence the pace of evolutionary diversification. Much of this work suggests that sexual selection does not simply produce elaborate courtship displays: it shapes the emergence of reproductive isolation and, perhaps, speciation itself (Lande 1981; Higbie et al. 2000; Kirkpatrick and Ravigne 2002).

In light of these findings, it is not surprising that researchers are particularly concerned with the potentially conflicting demands of mate choice and species recognition (Gerhardt 1982; Ryan and Rand 1993b; Boake et al. 1997; Pfennig 1998; Hankison and Morris 2003). It is surprising, however, that relatively little theoretical and empirical work addresses the general perceptual processes that underlie mate selection (Endler 1992; Johnstone 1994; Wiley 1994; Bradbury and Vehrencamp 1998; Phelps and Ryan 1998). Neglect of the topic limits our ability to formulate empirically grounded models of sexual selection, to understand how mating decisions are made in different contexts, and to relate mate choice to the evolution of natural diversity.

While mating decisions are of clear evolutionary significance, they are also interesting examples of natural classification. How do animals process such profoundly important information? Researchers sometimes find that different aspects of courtship displays predict decisions made in the contexts of species recognition and mate choice (Doherty 1985; Schul et al. 1998; Ryan and Rand 2001). Perhaps the decisions rely on dissociable perceptual mechanisms. In one such scenario, females might perform a nested classification, first assessing species status and only

ABSTRACT: Mating decisions contribute to both the fitness of individuals and the emergence of evolutionary diversity, yet little is known about their cognitive architecture. We propose a simple model that describes how preferences are translated into decisions and how seemingly disparate patterns of preference can emerge from a single perceptual process. The model proposes that females use error-prone estimates of attractiveness to select mates based on a simple decision rule: choose the most attractive available male that exceeds some minimal criterion. We test the model in the túngara frog, a well-characterized species with an apparent dissociation between mechanisms of mate choice and species recognition. As suggested by our model results, we find that a mate attraction feature alters assessments of species status. Next, we compare female preferences in one-choice and two-choice tests, contexts thought to emphasize species recognition and mate choice, respectively. To do so, we use the model to generate maximum-likelihood estimators of preference strengths from empirical data. We find that a single representation of preferences is sufficient to explain response probabilities in both contexts across a wide range of stimuli. In this species, mate choice and species recognition are accurately and simply summarized by our model. While the findings resolve long-standing anomalies, they also illustrate how models of choice can bridge theoretical and empirical treatments of animal decisions. The data demonstrate a remarkable congruity of perceptual processes across contexts, tasks, and taxa.

* Corresponding author. Address for correspondence: Department of Zoology, B30 Bartram Hall, University of Florida, Gainesville, Florida 32605; e-mail: phelps@zoo.ufl.edu.

then assessing attractiveness. In another, species recognition might occur when an animal evaluates a single stimulus, and mate choice when it compares multiple stimuli. Although both are plausible, the two scenarios are neither mutually exclusive nor exhaustive. Without a model that describes how to compare preference strengths in the varied contexts, such interpretations are difficult to assess.

We present a framework that enables us to evaluate the relationship between mate choice, species recognition, and the experimental paradigms proposed to measure these decision processes. The resulting model provides a general approach to animal perception, of which the classification of courtship displays is a particularly interesting example. We present results from three related studies. In the first, we develop a formal model of mating decisions and ask whether it can resolve anomalous findings in the literature. The resulting analysis leads to a novel explanation for the apparent dissociation of species recognition and mate choice, calling into question the hypothesis that such features are assessed hierarchically. In the second study, we use phonotaxis experiments to test this hypothesis directly. In the third and final study, we ask whether the model can successfully relate response probabilities in one-choice and two-choice tasks when stimuli are allowed to vary in features related to mate attraction, species status, and signal fidelity. We find that our model neatly synthesizes seemingly disparate data, withstands strong empirical tests, and contributes to the logical integration of proximate and ultimate approaches to mating decisions.

Study 1: Defining the Basic Model

We begin with the assumption that an animal considering a prospective mate translates the diverse attributes of courtship into a single estimate of attractiveness. We next propose that stochasticity inherent in the nervous system causes attractiveness to vary, even when signal form and receiver motivation are constant. The animal uses this imperfect estimate of attraction to choose a mate. Mate selection follows a single, simple rule: choose the most attractive mate available, provided that this mate meets some minimal criterion. This model has the advantage of being plausible, simple, and general. It must contend, however, with significant anomalies in the literature.

The model must explain why different attributes seem to govern attraction in different contexts (Gerhardt 1982; Doherty 1985; Schul et al. 1998; Wagner 1998; Ryan and Rand 2001; Bush et al. 2002). If there is a single decision rule and a single representation of attraction, why do many animals appear to emphasize some traits in species recognition and others in mate choice? How do such decisions relate to evaluating a single prospective mate versus choosing among several? Either there are multiple decision rules

or we simply have not understood the underlying architecture of choice. To address these concerns we must define our model mathematically.

The Formal Model

We can describe attractiveness as the “internal response” ($R_i[a]$) of individual i to stimulus a ; this response is the sum of deterministic and stochastic components:

$$R_i(a) = \psi_i(a) + \varepsilon_i(a). \quad (1)$$

The term $\psi_i(a)$ is the deterministic component representing the mean attractiveness of a signal, a value we call the “preference strength.” For a group of stimuli varying in some continuous parameter, $\psi_i(\cdot)$ is an “internal preference function”; when we want to consider a pattern of preference strengths without reference to any particular parameter, we call the set of preferences $\{\psi_i(a_1), \psi_i(a_2), \dots, \psi_i(a_j)\}$ a “preference scale.” The second term in the equation, $\varepsilon_i(a)$, is the stochastic component of perception. It has a probability density function $f_i(\cdot)$ with an expectation 0; the internal response has a probability density function $r_i(\cdot)$ with expectation $\psi_i(a)$. We allow R and ψ to range over positive and negative values.

When an internal response exceeds some criterion, the subject classifies the prospective mate as acceptable. The stimulus that evokes a response exceeding this threshold is said to be “recognized.” For convenience, we define R so that the threshold lies at 0. As depicted in figure 1a, the probability that a stimulus a will be recognized is

$$\begin{aligned} \Pr(i \text{ chooses } a) &= \Pr[R_i(a) > 0] \\ &= \int_0^{\infty} r_i(x) dx \\ &= \int_{-\psi_i(a)}^{\infty} f_i(x) dx. \end{aligned} \quad (2)$$

Finally, if an individual is tested with two stimuli (e.g., whine vs. whine-chuck), the subject will choose stimulus a over stimulus b if two conditions are met: the internal response evoked by a must exceed the threshold, and it must exceed the response evoked by stimulus b :

$$\begin{aligned}
 & \Pr(i \text{ chooses } a \text{ when } b \text{ is present}) \\
 &= \Pr[R_i(a) > 0 \cap R_i(a) > R_i(b)] \\
 &= \int_0^{\infty} \int_{-\infty}^x r_i(x) r_i(y) dy dx \\
 &= \int_{-\psi_i(a)}^{\infty} \int_{-\infty}^{x+\psi_i(a)-\psi_i(b)} f_i(x) f_i(y) dy dx.
 \end{aligned} \tag{3}$$

We refer to this unforced, two-stimulus decision as “discrimination” (fig. 1b; Littlejohn and Michaud 1959). Note that when $\Pr[R_i(a) > R_i(b)] = 1$, discrimination is equivalent to recognition; for similar reasons, both recognition and discrimination can be considered special cases of a multinomial decision process. Unforced choice is not only a necessary attribute of decision making in a natural context, it allows one to posit a single decision rule for both recognition and discrimination.

In conventional two-choice tests, sexual selection researchers may disregard data in which a subject does not choose. We refer to this special case as the “traditional

discrimination” test. The probability that a will be chosen over b in this test is simply

$$\Pr(i \text{ chooses } a \text{ over } b \text{ given one is chosen}) = \frac{\Pr[i \text{ chooses } a \text{ over } b]}{\Pr([i \text{ chooses } a \text{ over } b] + \Pr[i \text{ chooses } b \text{ over } a])}. \tag{4}$$

This framework shares important features with decision theory developed in other disciplines (Luce 1959; Thurstone 1959; Green and Swets 1966; Manski 1977). In a signal-detection model from psychophysics, for example, the preference strength ψ is analogous to the average magnitude of a sensory event—the perceived brightness of a light or the volume of a tone (Green and Swets 1966). In a random-utility model from economics, ψ would correspond to the expected utility of a product (Manski 1977). We build on conventional models in these disciplines and related models of mate choice (Reeve 1989; Johnstone and Grafen 1992; Johnstone 1994, 1998; Wiley 1994; Bradbury and Vehrencamp 1998, 2000) first by incorporating unforced choice, a modification that enables us to describe a more natural and diverse set of choices by a single decision rule. We then use this extension to develop novel

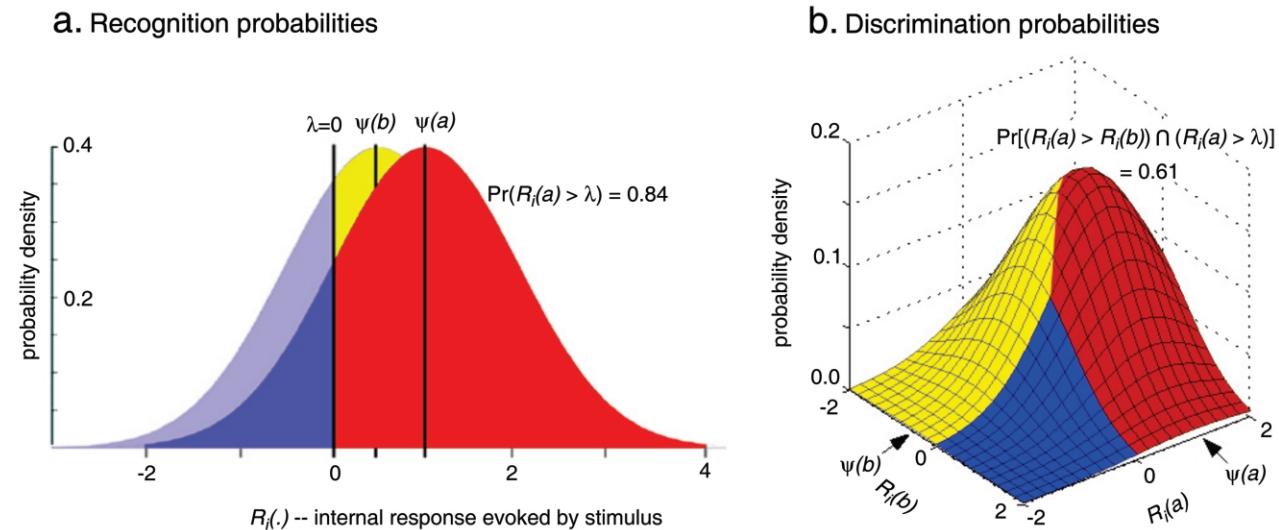


Figure 1: Calculating response probabilities in recognition and discrimination tests when preference strengths are known. *a*, Recognition. Stimulus a elicits a variable internal response with a mean of $\psi_i(a)$, defined as the preference strength. The recognition probability for stimulus a is the probability that an internal response $R_i(a)$ drawn from the distribution $r_i(\cdot)$ exceeds the threshold ($\lambda = 0$). Here, we assume that $r_i(\cdot)$ is a normal distribution with a standard deviation of unity. If $\psi_i(a) = 1.0$, this probability is 0.84 (area in red). The area of the yellow curve corresponds to the probability of choosing b in a recognition task given $\psi_i(b) = 0.5$. The dark blue and light blue areas correspond to the probability of not choosing a or b , respectively. The difference in the preference strengths $\psi_i(a)$ and $\psi_i(b)$ is simply $\Delta\psi_{a,b} = 1.0 - 0.5 = 0.5$. *b*, Discrimination. The probability that a will be chosen in a discrimination paradigm is the probability that $R_i(a)$ will be both greater than threshold ($\lambda = 0$) and greater than $R_i(b)$. This is the volume of the region in red. The volume of the region in yellow is the probability that a female will choose b . The volume of the region in blue is the probability that both a and b will elicit responses below threshold and the subject will choose neither.

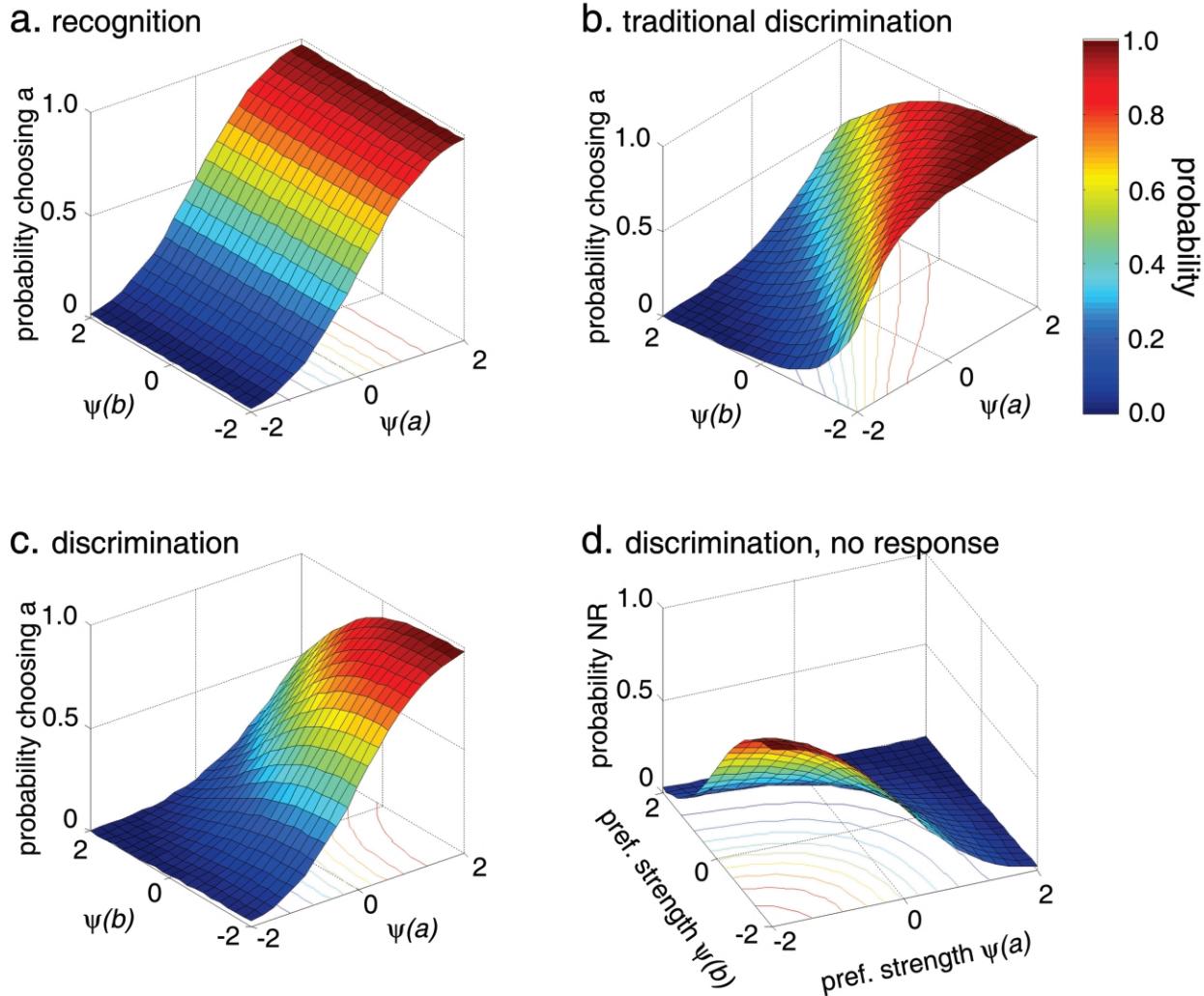


Figure 2: Response probabilities as a function of preference strength. *a*, Given the model described in figure 1, we calculate the probability of responding to stimulus a in the recognition paradigm as a function of preference strengths $\psi_i(a)$ and $\psi_i(b)$. *b*, Probability of choosing a in a traditional discrimination paradigm, in which no-responses are disregarded, as a function of $\psi_i(a)$ and $\psi_i(b)$. *c*, Probability of choosing a in a discrimination paradigm. In this version of the discrimination task, females have the option of withholding responses. Note that as $\psi_i(b)$ gets weak, the response to stimulus a approaches that in a recognition task. *d*, Probability of choosing neither a nor b in a discrimination paradigm. As the preference strengths of a and b increase, the probability of not responding drops precipitously. The colored scale bar in the upper right corresponds to the probability of responding indicated in each panel and to the probability isolines projected onto the floor of each plot. The color bar is provided to help visualize these three-dimensional plots.

theoretical and empirical tools that address paradoxical findings in the mate-choice literature.

Numerical Exploration of the Model

We now illustrate our basic model with some numerical examples. For simplicity, we assume that the stochastic component $\varepsilon_i(a)$ is drawn from a normal distribution with a variance of unity. We calculate response probabilities in three contexts: the recognition test (one-choice; eq. [2],

fig. 1*a*), the discrimination test (two-choice; eq. [3], fig. 1*b*), and the traditional discrimination test (eq. [4]). These calculations and those that follow were performed using routines written in Matlab. Code is available as a zip archive in the online edition of the *American Naturalist* or by request from the corresponding author.

We begin by plotting response probabilities as functions of stimulus preference strengths $\psi_i(a)$ and $\psi_i(b)$ in figure 2. We consider these response probabilities in three behavioral contexts: the recognition paradigm (one-choice;

fig. 2a), the discrimination paradigm (with no-response data included; fig. 2b), and the traditional discrimination paradigm (no-response data excluded; fig. 2c). Simply comparing how response probabilities change as a function of preference strength and behavioral paradigm provides insights into patterns of data in the sexual selection literature. To begin with, the model demonstrates how a single preference scale can be transformed into a wide range of response probabilities. Although each transformation preserves the rank order of preferences, no transformation is linear. This has two important consequences. First, response patterns that rank the same stimuli differently would refute the hypothesis that the responses result from the same preference scale. (We return to this point in the third study.) Second, it demonstrates a limitation of methods that assume linear relations between stimulus attributes and response probabilities. Finding that different stimulus attributes are correlated with response probabilities in different paradigms, for example, does not indicate that different scales underlie the decisions.

More specifically, we can see that the response probability in a discrimination test (fig. 2c) approaches that in a recognition test as the comparison stimulus becomes very weakly preferred. This reflects the fundamental unity of decision processes formalized by the model. The response probability in a traditional discrimination task, however, does not reduce to that of recognition because the response measures have been distorted by the exclusion of no-response data. This is reflected in the curvature of the response isolines for stimuli near or below the threshold (fig. 2b). In a true forced-choice test, animals must choose between two alternatives (2AFC), and it is not possible to withhold a response. The probability of responding in a 2AFC task is well defined (e.g., Green and Swets 1966), and a graph of response isolines yields a set of parallel lines through this space (data not shown). In other words, the probability of responding to a stimulus in 2AFC depends only on the difference in preference strength between the two stimuli and not on the absolute preference for either stimulus. This demonstrates that traditional discrimination is not a proper analog of the 2AFC test. Nevertheless, traditional discrimination does remain a valid predictor of the direction of preference.

The Power of Choice Tests

Because each combination of preference strengths specifies response probabilities in both recognition and discrimination tests, we may also calculate the statistical power of tests as a function of preference. We define power as the probability of correctly rejecting the null hypothesis that two preference strengths are equal. Researchers in sexual selection frequently investigate preferences for a series of

stimuli using either recognition or discrimination tasks. Because it is laborious to present all possible pairs of stimuli in a discrimination task, researchers often choose a single referent stimulus; all other stimuli are presented paired with this referent in a traditional discrimination paradigm (Gerhardt 1982; Doherty 1985; Ritchie 1996; Wagner 1998; Ryan and Rand 2001; Bush et al. 2002). We calculate the power of the recognition and referent-discrimination paradigms using Fisher's exact test, one-tailed P values, a criterion for statistical rejection of $\alpha = 0.05$, and a sample size of 20. We determined the rejection region (the possible experimental outcomes that would lead to a correct rejection of the null hypothesis) for these parameters and summed the probabilities of these outcomes. This sum is the power of the test for any given pair of response probabilities. In the discrimination test, we assume the referent stimulus has a 0.90 probability of eliciting a recognition response ($\psi[a] = 1.24$). The exact values of our power calculations were influenced by these parameters, but the relationship between power, testing paradigm, and preference strength was consistent.

Plots of the power of recognition and referent discrimination (fig. 3a, 3b) reveal that both are very good at detecting differences when one stimulus elicits a response well above threshold and the other does not. The recognition paradigm, however, is unable to detect subtle differences when both stimuli are well above threshold, but one evokes a stronger preference than the other. We refer to this as the "zone of effective stimuli" and suggest that studies of intraspecific assessment favor stimuli in this region.

Similarly, the recognition test is profoundly more powerful near threshold, a region we refer to as the "zone of marginal stimuli" and anticipate will include displays on the outer boundaries of intraspecific variation. Thus, differences in power alone (fig. 3c) may explain the association of choice tasks with species recognition and mate choice.

Model Summary

The model we have outlined suggests that mate choice and species recognition are not fundamentally distinct processes. Instead, preferences can be regarded as varying along a single continuous scale. These preferences are non-linearly transformed into responses in various contexts. Furthermore, the extent to which species recognition and mate choice seem to map onto recognition and discrimination does not indicate dissociable decision mechanisms. It can be more parsimoniously interpreted as a difference in the statistical power of the two paradigms across preference space. This suggests an underlying uniformity in

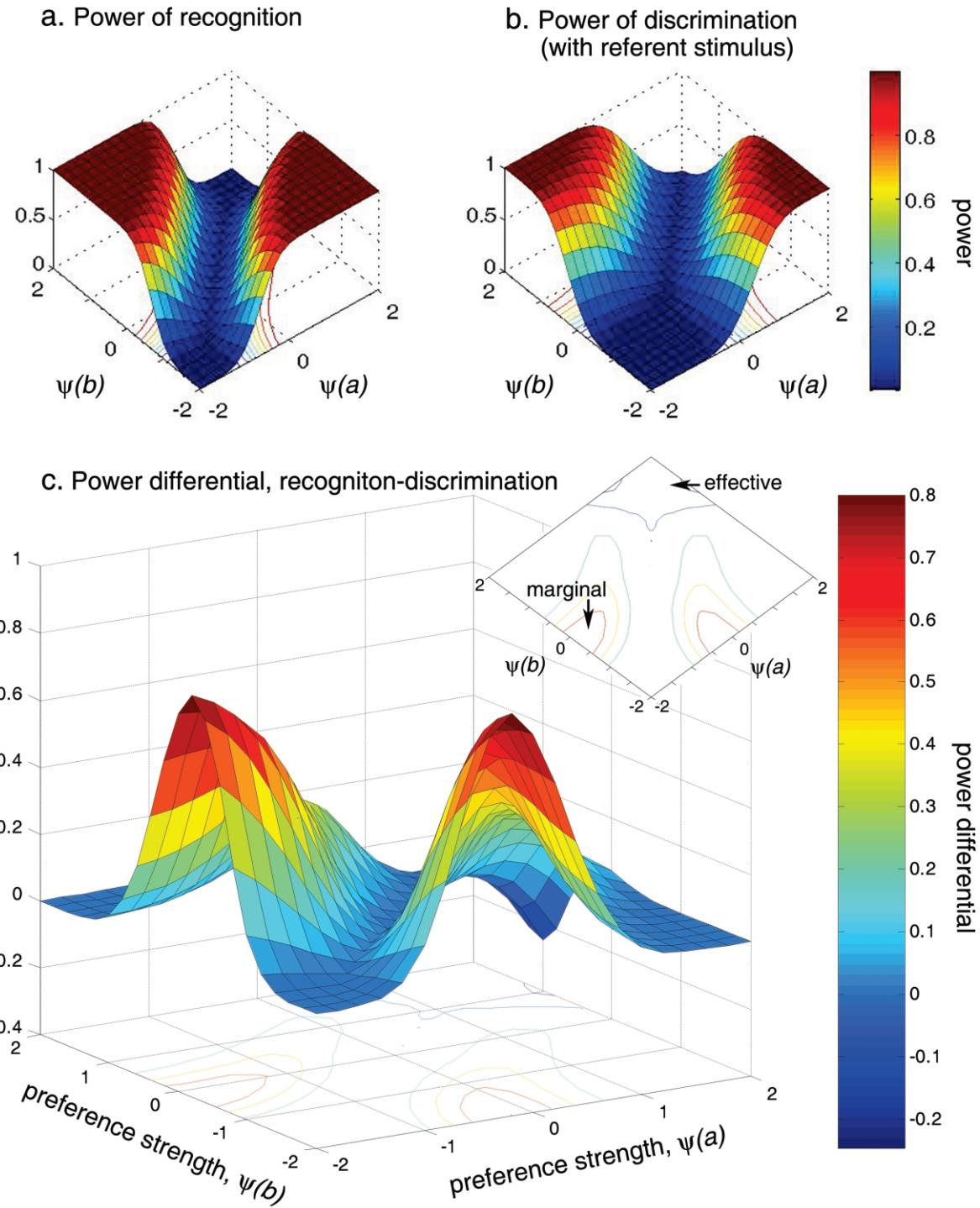


Figure 3: Power of recognition and traditional discrimination tasks. *a*, Recognition power, the probability of correctly rejecting the null hypothesis that the preference strength $\psi(a) = \psi(b)$. *b*, Discrimination power, the probability of correctly rejecting the null hypothesis ($\text{preference strength } \psi_a[a] = \psi_b[b]$) using traditional discrimination tasks and a common referent stimulus. *c*, The difference between the power calculated for recognition and discrimination tasks. Recognition tests are better able to detect differences in preference strength in the regions near threshold ($\lambda = 0$). Where both $\psi_a(a)$ and $\psi_b(b)$ are at least 1 SD above threshold, discrimination tasks are better at resolving differences in preference strengths. Lines represent isolines of power (panels *a*, *b*) or the power differential (panel *c*), with values corresponding to the color given in the scale bar on the right. The upper right corner of panel *c* includes a top view of the contour map for the power differential. The zones of marginal and effective stimuli are indicated with arrows.

seemingly disparate data and makes strong, falsifiable predictions.

In addition to power analyses, the mapping of preference strengths onto response probabilities permits maximum-likelihood estimates of $\psi(.)$ from empirical data. We can use these estimates to investigate the biological significance of stimuli (sensu Nelson and Marler 1990) and to probe how representations of preference have been shaped by evolutionary forces. The model not only explains patterns in existing data, it generates a statistical framework for estimating and comparing preference strengths measured in different contexts.

We next determine whether the model can resolve a well-documented discrepancy between species recognition and mate choice in a focal species, the túngara frog (*Physalaemus pustulosus*). Lastly, we use maximum-likelihood estimates of preference to test whether distinct perceptual processes influence attraction in recognition and discrimination tasks. The resulting data address both the external validity of the model and the logical unity of mating preferences.

Study 2: Species Recognition, Mate Choice, and Hierarchical Assessment

The túngara frog is a small Neotropical frog that has been the subject of more than 25 years of sexual selection studies (Rand and Ryan 1981; Ryan 1985). Females visit male choruses and select a mate by approaching within 10 cm of a calling male (phonotaxis). Receptive females are identified by finding pairs in which a female has allowed a male to mount and clasp her. The pair can be intercepted before mating occurs and the females isolated for testing in phonotaxis experiments (Rand and Ryan 1981; Ryan 1985).

The túngara frog call has a descending frequency sweep, called the whine, followed by 0–6 repetitions of a broadband, amplitude-modulated sound called a chuck (fig. 4a). The acoustic energy in the two sounds falls predominantly on different auditory organs (Ryan et al. 1990), and they seem to convey different information. When the whine alone is broadcast to females, it predictably elicits female phonotaxis. The chuck alone does not. If one adds a chuck to a whine, however, the compound stimulus is strongly preferred to the whine alone. The whine, it seems, is both necessary and sufficient for species recognition. The chuck is neither necessary nor sufficient but does make the call considerably more attractive (Rand and Ryan 1981; Ryan 1985; A. S. Rand and M. J. Ryan, unpublished data: 359 whine vs. 2,265 whine chuck, $P < .000000001$). The same seems to be true of other species in the genus. Even species that lack chucks altogether prefer conspecific whines with túngara chucks appended. (Ryan and Rand 1993a). There

is perhaps no clearer distinction between stimulus attributes important to species recognition and mate choice, or between the demands of one-choice and two-choice tasks. It appears as though females are performing a nested classification. Females seem to judge species status first, using the whine, and attractiveness second, using the chuck.

The hierarchical assessment of features would imply a complex context specificity that could be achieved in more than one way. In the most intuitive scenario, females faced with a single option would assess features related to species recognition, and higher levels of assessment would never be called on; females faced with multiple options would assess the species of each male and then evaluate additional features to determine attraction. This scenario implies that species-recognition features will be relevant to both one-choice and two-choice tasks, but mate-attraction features will be relevant only when there is more than one conspecific (outlined in Wilczynski et al. 1995). In this case, the preference strength associated with a complex stimulus like the whine chuck will be different in one-choice and two-choice tasks. This fundamentally contradicts our model. In a second scenario, the components of a display undergo a complex, hierarchical assessment that yields an estimate of preference. This preference, however, can then be transformed into one-choice and two-choice responses in a manner consistent with our model. Both hierarchical scenarios seem consistent with published data, and both predict that mate-attraction features will be ineffective if paired with stimuli that are unable to signal conspecific status.

Our power analysis suggests a more parsimonious explanation for the data. Perhaps the whine and whine chuck lie in the region of preference space in which recognition tests poorly predict preference strength—the zone of effective stimuli. The chuck may add preference in a recognition task, but its contribution has not been detected because of power constraints. This makes a novel and unambiguous prediction. If the chuck is added to a stimulus whose preference strength is just below threshold, a whine within the zone of marginal stimuli, the combination should elicit recognition, though neither stimulus evokes reliable responses alone. This hypothesis and the hierarchical alternatives are mutually exclusive.

We tested female phonotactic responses to stimuli made progressively less like the conspecific call and more like the call of a congener, *Physalaemus enesefae* (fig. 4b), with and without the chuck. If species status is assessed first and mate attractiveness second, then a stimulus that does not signal species status should never elicit responses. If, however, the chuck contributes to attraction in a more general way, it may enable the recognition of marginal stimuli that are otherwise unable to convey species status.

This latter outcome would be consistent with our interpretation of the power analysis but would be inconsistent with a hierarchical analysis of the whine and chuck.

Methods

Stimulus Synthesis. We synthesized a set of whines that were intermediate between the call of the túngara frog and that of an unrecognized congener, *Physalaemus enesefae*, using a previously described procedure (data and methods in Ryan et al. 2003). Briefly, a set of seven call variables is sufficient to describe the frequency and amplitude modulation of the túngara whine as well as that of *P. enesefae* (nsf). By going 20% or 40% of the acoustic distance from the túngara call to the nsf call on each of the seven variables, we could synthesize two intermediate stimuli—the 20nsf and 40nsf calls—which vary in their similarity to the conspecific (fig. 4b). By appending a chuck to these calls, we could put traits associated with interspecific and intraspecific assessment into direct conflict. Both the 20nsf and 40nsf stimuli were outside the range of normal variation in the túngara whine (Ryan et al. 2003). All whines were normalized to have the same peak amplitude. The chuck was a synthetic stimulus constructed to match the population mean (fig. 4a).

Behavioral Testing. Tests were performed in a dark, sound-resistant chamber. We scored the behavioral tests by means of an infrared camera connected to a monitor we viewed from outside the chamber. We broadcast each of the stimuli in a recognition paradigm in which the test stimulus is broadcast from one speaker and a null stimulus (white noise in a matching amplitude envelope) from a second. This null stimulus allows one to assess whether phonotactic responses are indeed specific to the stimulus under study. In discrimination tests, two calls were played antiphonally from opposite speakers. Tests were preceded and followed by a positive control test that consisted of a choice between the whine and a whine plus chuck. Data from females who responded to both control tests were considered informative for the intervening tests, even if they did not respond in an intervening test. We recorded the number of females that approached the speaker within 10 cm during 10 min of stimulus playback. In total, we report >500 informative choice tests gathered over 2 years of work. Additional details regarding testing procedures have been previously published (Rand and Ryan 1981; Phelps et al. 2001; Ryan and Rand 2001).

Results and Discussion

We found that a stimulus 20% of the acoustic distance to the *P. enesefae* call (20nsf) elicited recognition when played

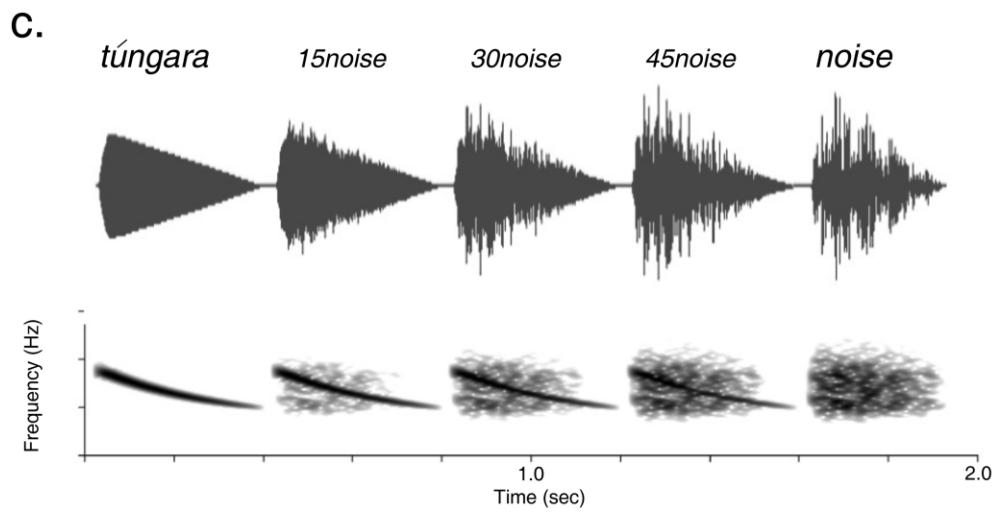
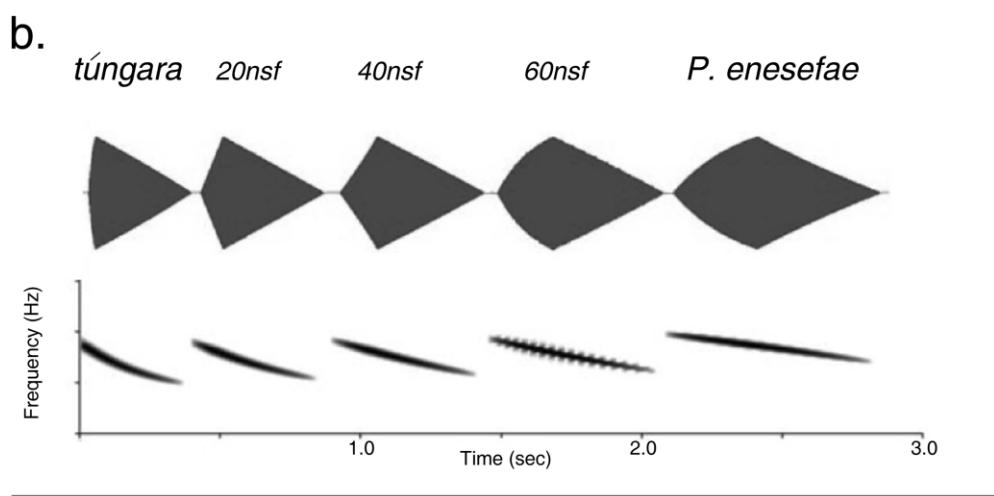
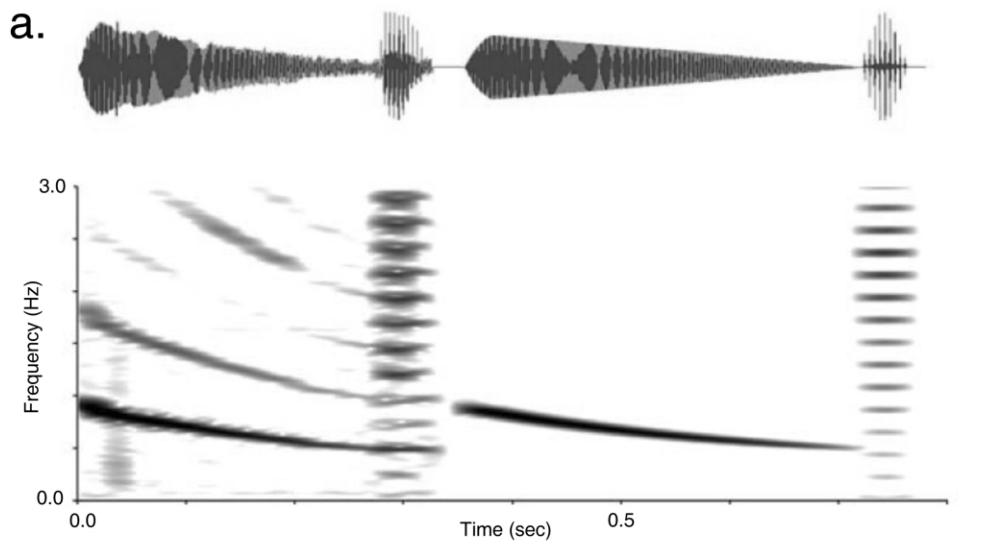
alone ($P = .032$; fig. 5a). Adding a chuck (20+ch), however, greatly increased the probability of recognition (17 : 3 vs. 8 : 12, $P = .004$). Similarly, females showed a strong preference for 20+ch over 20nsf alone in a discrimination paradigm ($P < .001$; fig. 5b). More compellingly, although the females preferred the species-specific whine to 20nsf, the preference was reversed by adding a chuck to 20nsf ($P = .024$; fig. 5c). The chuck is able to increase female responses in both recognition and discrimination tasks, even when paired with a stimulus that is only marginally able to convey species status.

A stimulus 40% of the acoustic distance to the *P. enesefae* call (40nsf) is clearly insufficient to elicit species recognition: no more females approach it than a silent speaker ($P = .50$; fig. 5d). Nevertheless, it was highly effective at eliciting recognition if paired with a chuck ($P < .001$). Females asked to choose between 40nsf and 40+ch in a discrimination paradigm showed strong preferences for 40+ch ($P < .001$; fig. 5e). When 40nsf and 40+ch were each tested against the whine in a discrimination paradigm, the chuck was able to increase the probability that a female would choose the 40nsf stimulus over the species-specific whine ($P = .021$; fig. 5f).

The data demonstrate that female túngara frogs are not assessing whines and chucks in hierarchically distinct stages. The females instead behave as though both stimuli are contributing to a single assessment of attractiveness, as though judgments of species status and mate attraction are derived from a common estimate of preference. The model thus passes a strong test of its external validity. If the model is to be more than heuristic, however, it should make quantitative predictions as well. Next, we test a more subtle distinction between species recognition and mate choice—that they are separable perceptual processes weighted differently in recognition and discrimination experiments.

Study 3: Species Recognition, Mate Choice, and the Number of Available Options

We have reported, as have others (Doherty 1985; Ryan and Rand 1995, 2001; Schul et al. 1998; Wagner 1998; Bush et al. 2002), that different stimulus attributes are associated with response probabilities in recognition and discrimination tasks, results suggesting the two paradigms emphasize distinct demands of species recognition and mate choice. We refer to this more subtle dissociation of species recognition and mate choice as task-specific “feature weighting.” Feature weighting can be defined by a stimulus-specific change in preferences estimated in the two behavioral tasks. If discrimination places a higher premium on mate-assessment cues, for example, stimuli including a chuck should exhibit larger preferences in dis-



crimination tasks than would be predicted based on recognition performance.

In contrast to the feature-weighting interpretation of task differences, the cognitive architecture we propose attributes apparent differences to the nonlinear transformation of preference strengths into response probabilities. If we estimate preferences in a common currency—the preference strength, ψ —preference scales derived from recognition tasks should be strongly and linearly related to those from discrimination tasks. Equally importantly, the resulting relationship should be consistent across stimulus classes.

We can test these predictions by systematically varying preference strength in response to multiple stimulus dimensions, testing the stimuli in one-choice and two-choice tasks, and observing whether there is a single, linear relationship between estimates of differences in preference strength ($\psi_{a,b}$). Our null hypothesis is that the model does not effectively describe the relationship between recognition and discrimination, and so preferences estimated using the model should be poorly related across tasks. Our model predicts a strong and linear relationship between tasks once preference strengths have been appropriately estimated. Finally, the feature-weighting hypothesis predicts that stimuli relevant for mate attraction will exhibit a selective increase in preference strength estimated in discrimination tests.

Methods

Stimulus Synthesis. We constructed two sets of auditory stimuli that resemble the whine of the túngara frog to varying degrees. In the nsf set, as in the previous experiment, we synthesized whines that are progressively more like the call of *P. enesefae* (fig. 4b), including 20nsf, 40nsf, and 60nsf. In the “noise” set, we overlay a whine with noise in a matching amplitude and spectral envelope at various signal-to-noise ratios (fig. 4c). A standard white noise stimulus was shaped to the amplitude envelope of the túngara whine, then this stimulus was bandpass filtered (500–800 Hz) to ensure a spectral envelope that approximated that of the whine. The túngara whine and the filtered noise were then normalized to have the same root mean square (RMS) call energy. To construct a noise/tún-

gara stimulus, the noise was multiplied by a constant (e.g., 0.15) and the túngara whine by 1 less the same constant (e.g., 0.85). The two waveforms were added, then renormalized to have the same call energy as the original túngara call. This was done to generate stimuli that had 15%, 30%, and 45% of the energy attributable to noise (fig. 4c). All stimuli from these sets were tested in recognition tasks, and all possible pairwise combinations were tested using discrimination tasks. A third stimulus set, nsf+ch, included those eight tests described in the preceding experiments as well as 20nsf versus 40nsf and 20+ch versus 40+ch. Four tests in this set directly compared whines with and without chucks (20nsf vs. 20+ch, 20+ch vs. túngara, 40nsf vs. 40+ch, 40+ch vs. túngara). The tests were conducted as described in the preceding experiment, and our analysis includes the data we report above.

Estimating Preference Strength. Because the model provides an explicit description of how preference strengths are translated into response probabilities (eqq. [2], [3]), we can use maximum likelihood methods to estimate population preferences from empirical data. In recognition tests, we interpret the number of females (k) approaching the speaker as the number of “successful” outcomes in a binomial experiment of sample size N . Using the numerical estimates given in figure 2, we calculate the likelihood of a population-level preference strength as the probability of observing the results in an experiment given the response probability (p) associated with that preference strength:

$$L[\text{preference strength } \psi(a)] = \left(\frac{N!}{k! \times (N-k)!} \right) \times p^k \times (1-p)^{N-k}. \quad (4)$$

Next, we presented pairs of stimuli in a discrimination paradigm in which both speakers broadcast a test stimulus. For each discrimination experiment, we recorded the number of females approaching each stimulus and the number of females not responding (k_a , k_b , and k_{NR}). (As described in study 2, positive control tests were used to ensure female receptivity.) We used the trinomial distri-

Figure 4: Natural and synthetic calls. *a*, The upper left panel depicts a sonogram and waveform of a natural túngara frog call, including the frequency-modulated whine and broadband chuck. The upper right panel depicts a synthetic whine and chuck of the sorts used in our experiments. Females do not show a preference for the natural call. *b*, Waveforms and sonograms of calls made progressively more like the call of *Physalaemus enesefae* (nsf set). The call on the left is the túngara whine, the call on the far right *P. enesefae*. The intermediate calls are 20%, 40%, and 60% of the acoustic distance between *Physalaemus pustulosus* and *P. enesefae*. In some experiments, the 20% and 40% calls were presented with and without synthetic chucks (nsf+ch set). *c*, Waveforms and sonograms of calls made by adding noise to the túngara whine (noise set). The first call is the túngara whine, the last the filtered noise. Intermediate calls contain 15%, 30%, and 45% of total call energy from noise.

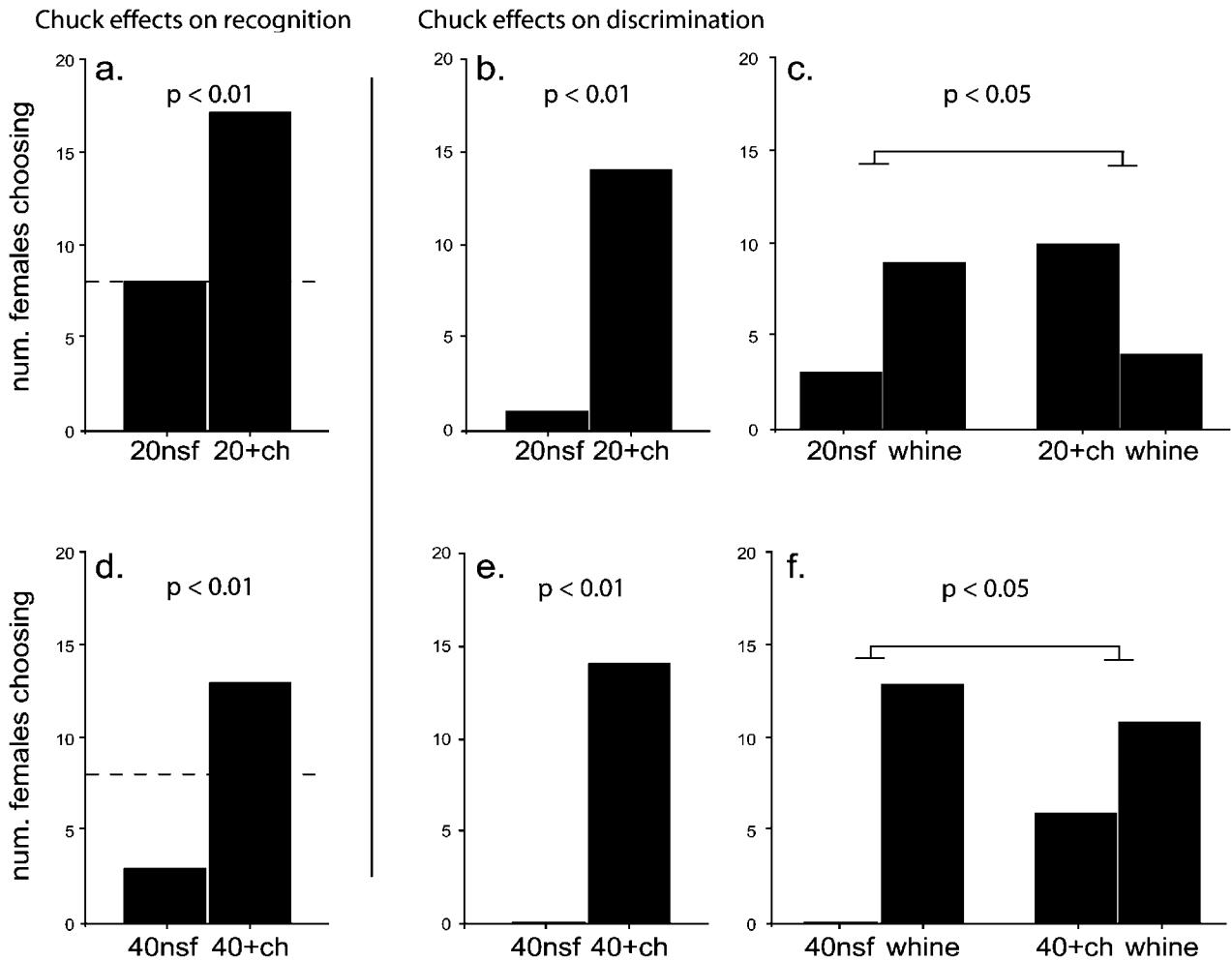


Figure 5: Contribution of a putative mate attraction signal, the chuck, to recognition and discrimination. *a*, Number of females out of 20 approaching a call 20% of the acoustic distance between the conspecific whine and the call of another species, the congener *Physalaemus enesefae*, with and without a chuck (20nsf, 20+ch). The dashed line represents the criterion for significant recognition compared with the number of females approaching a silent speaker (2 of 20; Fisher's exact one-tailed, $P < .05$). Both stimuli are recognized, but the 20+ch elicits significantly more frequent recognition responses (20nsf, 8 : 12; 20+ch, 17 : 3; Fisher's exact, $P = .004$). *b*, 20nsf versus 20+ch in a discrimination paradigm ($n = 20$). Females approach the 20+ch significantly more often than 20nsf (14 : 1; sign test, $P < .001$). *c*, Comparisons of the 20nsf stimuli to the species-specific whine in a discrimination task ($n = 20$). Females prefer the whine to 20nsf (20nsf : whine, 3 : 9) but prefer 20+ch to the whine (20+ch : whine, 10 : 4; Fisher's exact, $P = .024$). *d*, Number of females of 20 approaching a call 40% of the acoustic distance to the call of *P. enesefae* (40nsf), with and without a chuck. The 40nsf stimulus alone is not recognized (3 : 17 vs. 2 : 18; $P = .50$), while the 40+ch is strongly recognized (13 : 7 vs. 2 : 18; $P < .001$). These differences are significant (13 : 7 vs. 3 : 17; $P = .002$). *e*, 40nsf versus 40+ch in a discrimination paradigm ($n = 20$). Females exhibit strong preferences for the call with the chuck (14 : 0; sign test, $P < .001$). *f*, Comparisons of 40nsf and 40+ch to the species-specific whine in a discrimination paradigm ($n = 20$). On average, females prefer the species-specific whine to both stimuli, but 40+ch is significantly more likely to be chosen over the whine than is 40nsf (0 : 13 vs. 6 : 11; $P = .021$).

bution to calculate the likelihood of observing the data given each possible pair of preference strengths:

$$\begin{aligned} L[\text{preference strengths } \psi(a), \psi(b)] = \\ \left(\frac{N!}{k_a! \times k_b! \times k_{NR}!} \right) \times p_a^{k_a} \times p_b^{k_b} \times p_{NR}^{k_{NR}}. \end{aligned} \quad (5)$$

In this expression, p_a , p_b and p_{NR} are the probabilities of choosing *a* or *b*, or of not responding, for any pair of population preference strengths $\psi(a)$ and $\psi(b)$. For a pair of stimuli, we can compare the difference in preference strengths estimated in two recognition experiments with the difference estimated in a single discrimination experiment. (By comparing the differences between preference

strengths, we avoid treating pairs of discrimination preferences as though they were estimated independently.)

Results and Discussion. As predicted, the experiments reveal a robust relation between $\Delta\psi_{a,b}$ estimates from recognition and discrimination experiments (fig. 6; $R^2 = 0.88$, $P < .001$). Moreover, it does not matter whether we compare noise stimuli with other noise or with nsf stimuli. Even preference-strength estimates from stimuli with and without chucks show a common relationship between recognition and discrimination tests. To formally test for feature weighting, we first used the 95% confidence intervals (CIs) defined by regression of recognition and discrimination $\Delta\psi_{a,b}$ estimates. We asked whether the four discrimination tests that directly compared whines with and without chucks produced $\Delta\psi_{a,b}$ values greater than those predicted by the entire data set. No single stimulus pair lay outside these CIs. Using a permutation test, we took the mean of these four residuals and asked whether it was significantly greater than expected of four samples drawn without replacement from the remaining data: the $\Delta\psi_{a,b}$ values for pairs of stimuli with and without chucks were not significantly greater than predicted (one-tailed, $P = .249$).

Despite substantial stimulus heterogeneity—including variation in signal fidelity, species status, and mate attraction—we find no evidence for multiple preference scales. Responses in recognition tests explain nearly 90% of the variation in discrimination tests. Since neither stimulus type nor stimulus number changes the pattern of female preferences, there is no need to postulate multiple modules for mate choice and species recognition. We find no evidence for feature weighting.

A good theoretical model summarizes data and suggests novel experiments. It also makes assumptions that provide null expectations for empirical data, helping identify interesting phenomena that remain to be explained. Our data suggest the first limitation of the basic model. Although recognition responses and discrimination responses are clearly related, the slope of this relationship is substantially greater than 1.0 (1.58, 95% CI = 1.35, 1.80). Across all stimulus classes, females are making much more subtle distinctions in discrimination tests than anticipated. Furthermore, females are more likely to withhold responses in discrimination tasks than one would predict from recognition data (24 of 29 discriminations, $P < .001$; fig. 7a). This contradicts our model predictions. Because the difference is not specific to mate-attraction stimuli, it also contradicts the feature-weighting hypothesis. Females seem to contend with uncertainty by raising their thresholds and postponing choice. Indeed, we find that the more difficult the discrimination, the longer the latency to choose ($R^2 = 0.34$, $P < .001$; fig. 7b; also see Bosch et al.

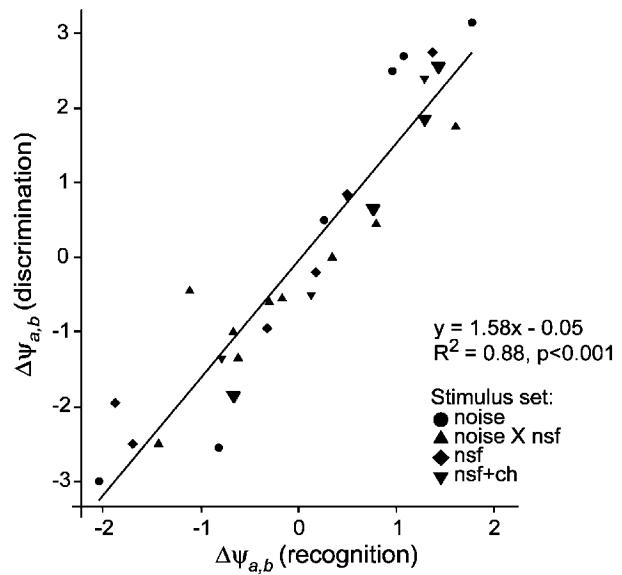


Figure 6: Comparing preference scales measured in recognition and discrimination paradigms. Responses of female túngara frogs were tested in recognition and discrimination paradigms for several series of stimuli. Estimates of $\Delta\psi_{a,b}(\psi_i[a] - \psi_i[b])$, where stimuli were assigned status of *a* or *b* at random) between members of the noise set (circles), the nsf set (diamonds), between noise and nsf sets, noise \times nsf (triangles), and between members of the nsf+ch set (inverted triangles). For the nsf and noise stimulus sets, all possible pairwise combinations of stimuli were tested in discrimination tasks. In the third data set, nsf+ch, the 20nsf and 40nsf stimuli were presented with and without a chuck in a total of eight discrimination tasks (six are given in fig. 5; the remaining are 20nsf vs. 40nsf [9 : 1] and 20+ch vs. 40+ch [14 : 1]). Those four stimulus pairs in which calls with chucks were compared with those without are marked by an enlarged nsf+ch symbol (inverted triangle). The regression relating recognition and discrimination $\Delta\psi_{a,b}$ estimates across all stimulus types is highly significant ($y = 1.58x - 0.05$, $R^2 = 0.88$, $P < .0001$).

2000). A relationship between decision thresholds, task difficulty, and response latency has commonly been found in the psychological literature (Birdsall and Roberts 1965; Green and Swets 1966; Pike and Ryder 1973; Luce 1986; Dooling and Hulse 1989). Perhaps in mate selection, as in psychological experiments (Birdsall and Roberts 1965), this could be interpreted as a Bayesian decision strategy (also see Luttbeg 1996). This finding, like our others, reinforces the extent to which models of human judgments inform inquiry into domains as distant as anuran mate choice.

Conclusion

As the interaction between sexual selection and reproductive isolation continues to gain prominence, so does the need to explain discrepancies in the empirical literature. We demonstrate that a simple choice model resolves

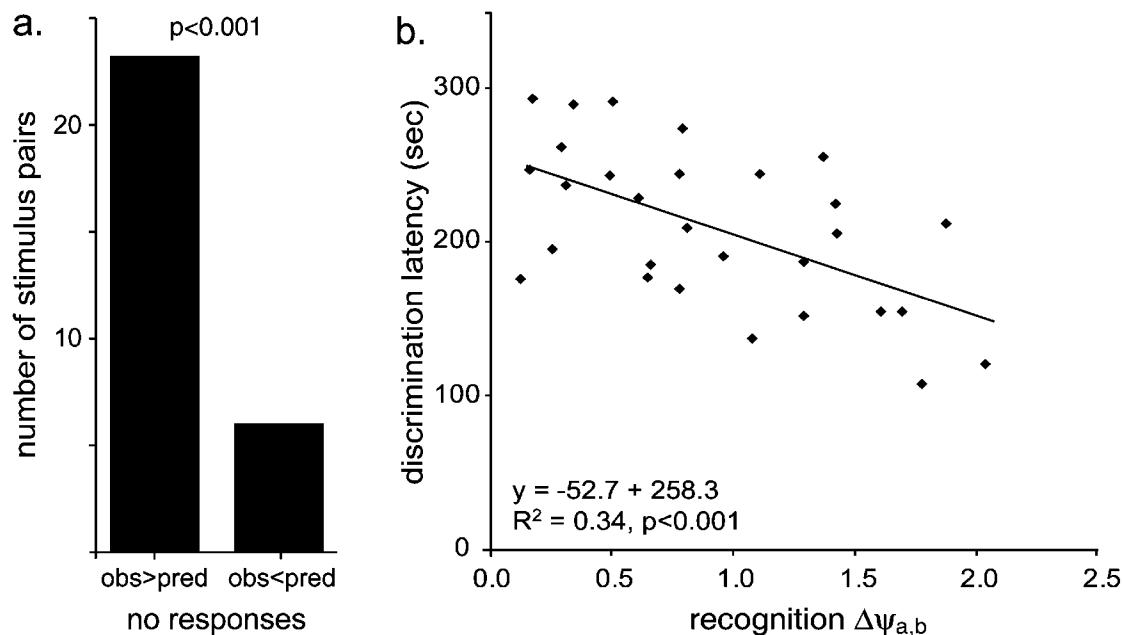


Figure 7: Evidence for extended sampling in the discrimination paradigm. *a*, The number of discrimination experiments in which the observed frequency of no response (NR) was greater than that predicted ($P < .001$), suggesting a raised threshold to respond. The predicted frequency of NR was simply the probability of no response estimated from recognition tests times the sample size of the discrimination tests. *b*, The average latency to choose in a discrimination task increases as the difference in preference strength ($\Delta\psi_{a,b}$) decreases. The value of $\Delta\psi_{a,b}$ was estimated in independent recognition tasks.

common disparities between species recognition and mate choice. It does so by clarifying the relationship between these assessments and the recognition and discrimination behavioral paradigms. Intuition suggests that the full expression of mating preferences requires the assessment of multiple prospective mates; in the absence of such options, criteria used in species recognition ought to predominate. Somewhat surprisingly, the model suggests that one-choice and two-choice responses do reflect species recognition and mate choice, respectively, but not for the reasons supplied by conventional wisdom. Recognition is a better predictor of interspecific assessment, and discrimination of intraspecific assessment, not because they tap distinct perceptual processes or solve unique ecological problems but because statistical power varies with both choice task and preference strength.

We tested the model using the túngara frog, a species in which apparent dissociations between species recognition and mate choice are particularly well documented (Rand and Ryan 1981; Ryan 1985; Ryan and Rand 1995, 2001; Wilczynski et al. 1995). Prior data from túngara frogs suggested a distinction between species recognition and mate choice based on call attributes and experimental paradigms. Our power analysis, however, provided a more parsimonious interpretation of the data. The model led us

to question whether females evaluated stimuli hierarchically based on the components of calls, or distinctly based on the number of available options. In study 2, we found that calls that fail to convey species status (e.g., 40nsf) can become potent stimuli if paired with a mate-attraction feature (e.g., 40+ch) in both behavioral paradigms. Clearly, female túngara frogs do not separate mate assessment into hierarchical stages of species recognition and mate choice. In study 3, we compared preference strengths estimated in recognition and discrimination tasks to assess whether the two paradigms placed different weights on call characters relevant to species recognition and mate choice. We found that the preferences estimated in the two tasks were strongly and linearly related, indicating that recognition responses were very good predictors of discrimination responses across a broad range of stimuli. Although females were consistently more discriminating when given two choices than when given one, this increase corresponded to changes in threshold and sampling common to all stimulus types. We could detect no differences in how these paradigms led females to assess features related to species status, signal fidelity, or mate attraction. In túngara frogs, species recognition and mate choice seem to emerge from a common perceptual process. The model thus provides a cogent and empirically sound explanation

for the relationship between mating decisions, stimulus attributes and choice paradigms.

The mechanisms of mate selection seem to us inherently interesting phenomena. Precisely how such decisions are implemented, however, can also have profound evolutionary consequences. Treatments of animal communication, for example, suggest that perceptual error may permit deception (Wiley 1994; but see Johnstone and Grafen 1992), favor the emergence of stereotyped displays, or alter the number of evolutionarily stable strategies available (Johnstone 1994). We provide a formulation of perceptual error, decision thresholds, and mating preferences that allows these parameters to be estimated from behavioral data. Such an approach can inform studies of sexual selection and nascent speciation, but it also raises novel questions in animal cognition. Differences in "choosiness," for example, influence many types of decisions—selecting a meal, a mate, or a nesting site, to name a few. Are evolutionary changes in selectivity due to alterations in decision threshold, preference resolution, or both? Similarly, we can ask whether shifts in preference acuity or threshold reflect changes in the number or nature of neurons that process sensory information or assign it affective value. Such questions transcend the domain of sexual selection and challenge us to consider animal decisions in a broad context.

In animal behavior, as in all of contemporary biology, the integration of proximate and ultimate explanations forms the foundation of an exciting and ongoing synthesis. The model we have outlined, with its well-explored precedents in psychology and economics (Luce 1959; Thurstone 1959; Green and Swets 1966; Manski 1977), its clear ties to evolutionary models of choice (Reeve 1989; Johnstone and Grafen 1992; Johnstone 1994, 1998; Luttbeg 1996; Bradbury and Vehrencamp 2000), and its compelling empirical utility, seems particularly well suited to this endeavor.

Acknowledgments

We thank M. Kirkpatrick, who suggested much of the notation in our model as well as substantial revisions in its presentation. M. Hauser, R. Haven Wiley, L. Higgins, and K. Hoke provided valuable criticism on early drafts of the manuscript. Finally, we thank B. Bolker, E. Braun, C. Osenberg, and the faculty and students in the University of Florida's animal behavior group in the Zoology Department.

Literature Cited

- Andersson, M. 1994. Sexual selection. Princeton University Press, Princeton, NJ.
- Birdsall, T. G., and R. A. Roberts. 1965. Theory of signal detectability-deferred-decision theory. *Journal of the Acoustical Society of America* 37:1067–1074.
- Boake, C. R. B., M. P. DeAngelis, and D. K. Andreadis. 1997. Is sexual selection and species recognition a continuum? mating behavior of the stalk-eyed fly *Drosophila heteroneura*. *Proceedings of the National Academy of Sciences of the USA* 94:12442–12445.
- Bosch, J., A. S. Rand, and M. J. Ryan. 2000. Signal variation and call preferences for whine frequency in the túngara frog, *Physalaemus pustulosus*. *Behavioral Ecology and Sociobiology* 49:62–66.
- Bradbury, J. W., and S. L. Vehrencamp. 1998. Principles of animal communication. Sinauer, Sunderland, MA.
- . 2000. Economic models of animal communication. *Animal Behaviour* 59:259–268.
- Bush, S. L., H. C. Gerhardt, and J. Schul. 2002. Pattern recognition and call preferences in treefrogs (anura: Hylidae): a quantitative analysis using a no-choice paradigm. *Animal Behaviour* 63:7–14.
- Doherty, J. A. 1985. Phonotaxis in the cricket, *Gryllus-bimaculatus degeeri*: comparisons of choice and no-choice paradigms. *Journal of Comparative Physiology A* 157:279–289.
- Dooling, R. J., and S. H. Hulse. 1989. The comparative psychology of audition: perceiving complex sounds. Erlbaum, Hillsdale, NJ.
- Endler, J. A. 1992. Signals, signal conditions, and the direction of evolution. *American Naturalist* 139(suppl.):S125–S153.
- Gerhardt, H. C. 1982. Sound pattern-recognition in some North American treefrogs (Anura, Hylidae): implications for mate choice. *American Zoologist* 22:581–595.
- Green, D. M., and J. A. Swets. 1966. Signal detection theory and psychophysics. Wiley, New York.
- Hankison, S. J., and M. R. Morris. 2003. Avoiding a compromise between sexual selection and species recognition: female swordtail fish assess multiple species-specific cues. *Behavioral Ecology* 14: 282–287.
- Higbie, M., S. Chenoweth, and M. W. Blows. 2000. Natural selection and the reinforcement of mate recognition. *Science* 290:519–521.
- Johnstone, R. A. 1994. Honest signaling, perceptual error and the evolution of all-or-nothing displays. *Proceedings of the Royal Society of London B* 256:169–175.
- . 1998. Conspiratorial whispers and conspicuous displays: games of signal detection. *Evolution* 52:1554–1563.
- Johnstone, R. A., and A. Grafen. 1992. Error-prone signaling. *Proceedings of the Royal Society of London B* 248:229–233.
- Kirkpatrick, M., and V. Ravigne. 2002. Speciation by natural and sexual selection: models and experiments. *American Naturalist* 159(suppl.):S22–S35.
- Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences of the USA* 78:3721–3725.
- Littlejohn, M. J., and T. C. Michaud. 1959. Mating call discrimination by females of Strecker's chorus frog (*Pseudacris streckeri*). *Texas Journal of Science* 11:87–89.
- Luce, R. D. 1959. Individual choice behavior: a theoretical analysis. Wiley, New York.
- . 1986. Oxford Psychology Series. Vol. 8. Response times: their role in inferring elementary mental organization. Oxford University Press, New York.
- Luttbeg, B. 1996. A comparative Bayes tactic for mate assessment and choice. *Behavioral Ecology* 7:451–460.
- Manski, C. 1977. The structure of random utility models. *Theory and Decision* 8:229–254.
- Nelson, D. A., and P. Marler. 1990. The perception of birdsong and

- an ecological concept of signal space. Pages 443–478 in W. C. Stebbins and M. A. Berkley, eds. Comparative perception. Vol. 2. Complex signals. Wiley, New York.
- Pfennig, K. S. 1998. The evolution of mate choice and the potential for conflict between species and mate-quality recognition. *Proceedings of the Royal Society of London B* 265:1743–1748.
- Phelps, S. M., and M. J. Ryan. 1998. Neural networks predict response biases of female túngara frogs. *Proceedings of the Royal Society of London B* 265:279–285.
- Phelps, S. M., M. J. Ryan, and A. S. Rand. 2001. Vestigial preference functions in neural networks and túngara frogs. *Proceedings of the National Academy of Sciences of the USA* 98:13161–13166.
- Pike, R., and P. Ryder. 1973. Response latencies in yes/no detection task: an assessment of two basic models. *Perception and Psychophysics* 13:224–232.
- Rand, A. S., and M. J. Ryan. 1981. The adaptive significance of a complex vocal repertoire in a Neotropical frog. *Zeitschrift für Tierpsychologie* 57:209–214.
- Reeve, H. K. 1989. The evolution of conspecific acceptance thresholds. *American Naturalist* 133:407–435.
- Ritchie, M. G. 1996. The shape of female mating preferences. *Proceedings of the National Academy of Sciences of the USA* 93: 14628–14631.
- Ryan, M. J. 1985. The túngara frog: a study in sexual selection and animal communication. University of Chicago Press, Chicago.
- Ryan, M. J., and A. S. Rand. 1993a. Sexual selection and signal evolution: the ghosts of biases past. *Philosophical Transactions of the Royal Society of London B* 340:187–195.
- . 1993b. Species recognition and sexual selection as a unitary problem in animal communication. *Evolution* 47:647–657.
- . 1995. Female responses to ancestral advertisement calls in túngara frogs. *Science* 269:390–392.
- . 2001. Feature-weighting in signal recognition and discrimination by túngara frogs. Pages 86–101 in M. J. Ryan, ed. Anuran communication. Smithsonian Institution, Washington, DC.
- Ryan, M. J., J. H. Fox, W. Wilczynski, and A. S. Rand. 1990. Sexual selection for sensory exploitation in the frog *Physalaemus pustulosus*. *Nature* 343:66–67.
- Ryan, M. J., W. Rand, P. L. Hurd, S. M. Phelps, and A. S. Rand. 2003. Generalization in response to mate recognition signals. *American Naturalist* 161:380–394.
- Schul, J., D. von Helversen, and T. Weber. 1998. Selective phonotaxis in *Tettigonia cantans* and *T. viridissima* in song recognition and discrimination. *Journal of Comparative Physiology A* 182:687–694.
- Thurstone, L. L. 1959. The measurement of values. University of Chicago Press, Chicago.
- Wagner, W. E. 1998. Measuring female mating preferences. *Animal Behaviour* 55:1029–1042.
- Wilczynski, W., A. S. Rand, and M. J. Ryan. 1995. The processing of spectral cues by the call analysis system of the tungara frog, *Physalaemus pustulosus*. *Animal Behavior* 49:911–929.
- Wiley, R. H. 1994. Errors, exaggeration, and deception in animal communication. Pages 157–189 in L. A. Real, ed. Behavioral mechanisms in evolutionary ecology. University of Chicago Press, Chicago.

Associate Editor: Eldridge S. Adams
Editor: Jonathan B. Losos