

CONTRASTING THEORY WITH THE EMPIRICAL DATA OF SPECIES RECOGNITION

Terry J. Ord,^{1,2,3} Léandra King,³ and Adrian R. Young³

¹*Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington, New South Wales 2052, Australia*

²*E-mail: t.ord@unsw.edu.au*

³*Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138*

Received October 25, 2010

Accepted March 28, 2011

We tested hypotheses on how animals should respond to heterospecifics encountered in the environment. Hypotheses were formulated from models parameterized to emphasize four factors that are expected to influence species discrimination: mating and territorial interactions; sex differences in resource value; environments in which heterospecifics were common or rare; and the type of identity cues available for species recognition. We also considered the role of phylogeny on contemporary responses to heterospecifics. We tested the extent these factors explained variation among taxa in species discrimination using a meta-analysis of three decades of species recognition research. A surprising outcome was the absence of a general predictor of when species discrimination would most likely occur. Instead, species discrimination is dictated by the benefits and costs of responding to a conspecific or heterospecific that are governed by the specific circumstances of a given species. The phylogeny of species recognition provided another unexpected finding: the evolutionary relationships among species predicted whether courting males within species—but not females—would discriminate against heterospecifics. This implies that species recognition has evolved quite differently in the sexes. Finally, we identify common pitfalls in experimental design that seem to have affected some studies (e.g., poor statistical power) and provide recommendations for future research.

KEY WORDS: Animal communication, mate choice, meta-analysis, phylogeny, signal modality, territorial defense.

Whether an animal engages with another individual—during courtship, aggression, or some other social context—will often depend on an initial assessment of species identity. If animals are to produce viable offspring, mating should only occur between members of the same species, as hybrids will seldom have a selective advantage comparable to that of offspring from conspecific parents (see Pfennig 2007 for a rare exception). Animals that defend ecological resources are also less likely to interact with heterospecifics compared to conspecifics because the overlap in the type of resources exploited will generally be lower between individuals of different species than between members of the same species. Although intuition leads us to expect that animals in these situations will respond differently to conspecifics

and heterospecifics, recent evaluations of the empirical evidence of species discrimination have suggested this assumption is not supported as often as we might think. Two reviews have highlighted that animals frequently respond to heterospecific mating and territorial signals (Ord and Stamps 2009; Peiman and Robinson 2010). Both studies offer a number of biological explanations for why animals might respond to heterospecifics (especially for interspecific aggression; Peiman and Robinson 2010), but the generality of the phenomenon in so many species is surprising. Why animals should commonly respond to heterospecifics, especially in mating, still remains unclear. In this article, we used existing mathematical models to first demonstrate when animals might be more or less discriminatory against heterospecifics.

We then tested the predictions of these models in a formal meta-analysis of experimental studies of species discrimination published over the last three decades.

To understand why animals might “fail” to ignore heterospecifics during mating or territorial defense depends on identifying the benefits and costs associated with responding or not responding in different situations. As a basic example, if animals rarely encounter heterospecifics, it makes little sense to invest any time, energy, or even develop a recognition system for the assessment of species identity. In contrast, animals that constantly encounter heterospecifics should invest in species recognition to avoid the costs associated with superficial matings or aggressive disputes with heterospecifics that are not competing for the same ecological resources. Other variables that might influence the benefits or costs of animals responding to a heterospecific include the sex of the discriminating individual, whether discrimination is being made during courtship or territorial interactions, the type of cues available for discerning species identity, and whether animals have the required sensory or cognitive systems needed to detect and evaluate those identity cues.

We applied the optimization models developed by Reeve (1989) to visualize the effect of changes to the benefits and costs of responding to a given individual encountered in the environment. These models were originally formulated in the context of kin recognition, but are equally applicable to species recognition more broadly; in both situations there is a process of distinguishing “like” from “nonlike” individuals. As applied here, the recognition process is modeled on the assumption that a discriminating individual possesses a template of conspecific-identifying cues. This template is a set of conspecific cues that might have been learned or genetically inherited. This “conspecific” template is then matched against the set of cues exhibited by an encountered individual or “stranger.” The threshold at which an animal responds to the stranger is dependent on both the degree of overlap between the identity cues of the stranger and the conspecific template, and the situation in which assessment is being made. For example, if finding a conspecific mate is difficult either because they are scarce or because the cost of searching for a mate is high, an animal may choose to court any individual it encounters irrespective of the extent that conspecific-stranger cues might match (e.g., see Gowaty and Hubbell 2005).

For our purposes, Reeve’s models are especially useful because they can be used to evaluate changes in the optimal response threshold depending on the type of cues being used in species recognition. The accuracy of categorizing a cue or signal increases with the extent a signal resists degradation as it travels through the environment before reaching a receiver (Wiley 2006). The range over which signals degrade differs by modality (e.g., auditory, olfactory, or visual) and the level of redundancy in a signal (Dusenbery 1992; Bradbury and Vehrencamp 1998).

For example, acoustic signals generally have a greater potential to travel farther than volatile chemicals used as olfactory signals (Dusenbery 1992; Bradbury and Vehrencamp 1998). Signals constructed from multiple components encoding backup information on species identity in different modalities should also facilitate discrimination (Hebets and Papaj 2005). That is, errors in species recognition have the potential to vary between signal modalities, with multimodal signals presumably providing the best cues overall for accurate species discrimination. The properties of the conspecific template and distribution of heterospecific cues that might be encountered in an environment can be varied in Reeve’s models to represent contrasting levels of cue reliability: cues that are “noisy” or error prone (because of high environmental degradation) can be reflected by modeling broad cue distributions, whereas cues that are more reliable (because they are resistant to environmental degradation) are modeled with much narrower cue distributions (Fig. 1A).

We applied Reeve’s models to depict two classes of identity cues, those corresponding to cues of high reliability and those of low reliability, and considered their effect on changing response thresholds to potential mates and rivals. In the context of mating, we computed response thresholds for males and females separately because the costs of reproduction can differ between the sexes. The frequency of encounter rates with heterospecifics could also influence discrimination by determining the extent heterospecific phenotypes can be learnt or the likelihood that an appropriate recognition system can evolve (e.g., Svensson et al. 2010; Wellenreuther et al. 2010). We computed response probabilities under scenarios reflecting an encounter with a previously allopatric (unfamiliar) heterospecific and a current sympatric (familiar) heterospecific. Other details of Reeve’s models and descriptions of parameter values are presented in the methods section.

The predictions formulated from these models are summarized below and in Figure 1B, C:

- (1) *Social context*: Discrimination is generally similar during mating and territorial defense. Important exceptions occur when the cost of searching for a mate increases substantially and males begin to court any female encountered (always respond), and when the cost of defending a territory increases substantially and animals ignore rivals (never respond) and effectively stop being territorial.
- (2) *Sex differences*: In mating, females are generally more discriminating than males. However, males converge on similar discrimination thresholds as females when the cost of making a mistake in recognition is negligible (i.e., mistakenly rejecting a conspecific female), which allows males to become more selective in their responses to different females (conspecific vs. heterospecific).

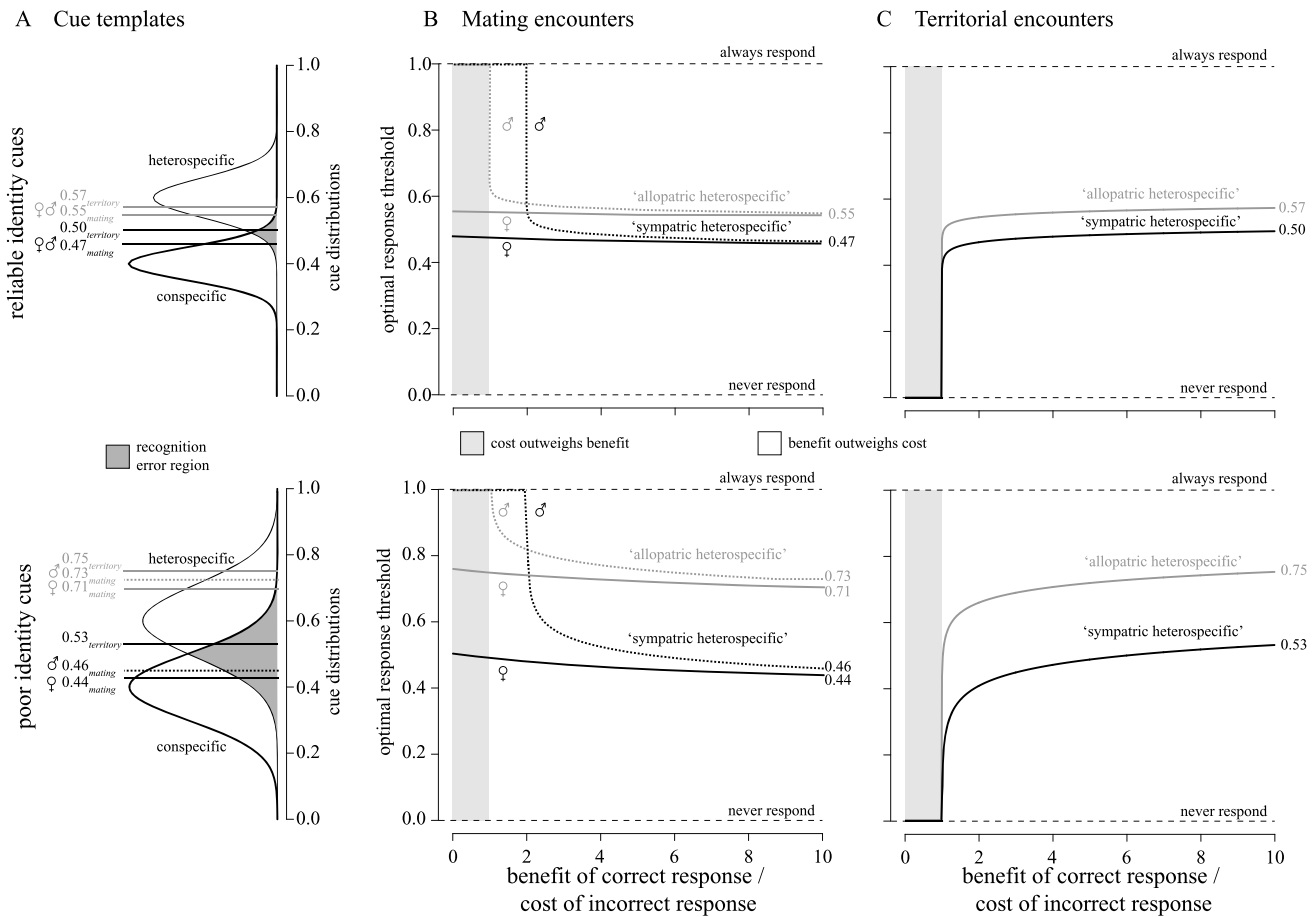


Figure 1. We parameterized Reeve’s (1989) models using a range of biologically intuitive values to visualize the circumstances in which an animal might or might not respond to a heterospecific encountered in the environment. The “decision” to respond—the response threshold—is dependent on the degree to which the conspecific template and the set of heterospecific cues differ (A). We modeled a situation in which an animal is presented with a reliable set of identity cues (upper row) by minimizing the degree of overlap between the conspecific template and distribution of heterospecific cues (i.e., the recognition error region was small). Conversely, to illustrate a situation in which only a poor set of identity cues was available for species recognition (bottom row), we increased the degree of overlap between templates (making the recognition error region large). The optimal response threshold is biologically defined as the level of dissimilarity that is “tolerated” between the set of cues of an encountered individual and the conspecific template that maximizes the benefit-cost ratio of responding or not responding given the particular circumstances in which discrimination is being made. Mathematically, the optimal response threshold is defined as the threshold value that maximizes the fitness function for a given set of parameter values. The model governing the shape of the optimal response threshold differs between mating (B) and territorial defense contexts (C; see eqs. 1 and 2 in the text, respectively). Models were parameterized to reflect environments where heterospecifics were abundant (sympatric; 50% of encountered individuals were heterospecifics) or rare (allopatric; 1% of encountered individuals were heterospecifics). In mating, the fitness cost of mistakenly responding to a heterospecific was assumed to be higher for females than males (i.e., the investment in reproduction differed between the sexes). An assumption was also made that finding a conspecific mate incurred a cost and that this cost was similar for both sexes (e.g., time away from other activities, exposure to predators, etc). In a territorial context, failing to respond to a conspecific rival was expected to incur a cost (e.g., stolen resources), as was mistakenly responding to a heterospecific (e.g., wasted energy from producing an unnecessary threat response). No sex-specific parameters were included in the application of the territorial model. In the framework of these costs for mating and territorial contexts, we varied the fitness benefits of correctly responding to a conspecific from no benefits to very high benefits (0–10). Parameter space in which the costs outweigh the benefits of responding is highlighted in gray.

(3) *Familiarity*: Animals generally discriminate more against heterospecifics that are sympatric (familiar) than those that are allopatric (unfamiliar) in both mating and territorial contexts. Again, as with (1), there are impor-

tant exceptions: when the cost of searching for mates increases substantially for males (and males only; leading to “always respond”), or when the cost of defending a territory becomes so high that animals (of

either sex) stop being territorial (resulting in “never respond”).

- (4) *Reliability of species cues*: As the reliability of identity cues increases, animals become more discriminatory in their responses, unless mate search for males or the defense of territories for either sex becomes very costly (see 1 and 3). We used the outcome of Reeve’s models in this context as a means of identifying in our meta-analysis whether identity cues in certain modalities (auditory, olfactory, or visual) provided more or less reliable cues for species recognition. For example, if acoustic signals were associated with higher levels of species discrimination than cues from olfactory signals, this would be consistent with differences in reliability between the two types of signals.

In addition to these model predictions, we might expect the cognitive and sensory systems used in species recognition to be shared between closely related species. The degree to which different species discriminate against heterospecifics in similar social and ecological settings might therefore depend on phylogeny. For example, the type of response exhibited by two bird species to heterospecifics might be similar if the properties of their conspecific-templates, recognition systems, or the type of cues assessed for species recognition are shared through common ancestry. Conversely, the type of response given by a bird and a frog to a heterospecific could be quite different because of phylogeny, despite both birds and frogs relying on acoustic communication for attracting mates. To explore the possibility that species recognition may have a phylogenetic component to it, we computed the phylogenetic signal of species recognition.

Finally, the experimental design and setting used by researchers will affect the precision of experiments testing discrimination responses of animals to different stimuli. A troubling trend appeared in an earlier meta-analysis (Ord and Stamps 2009) in which a study’s sample size, and whether animals were tested in captivity or the field, seemed to influence whether a study reported animals discriminating against heterospecifics. We revisited this issue in the current study and tested explicitly whether sample size and experimental setting had any impact on the nature of responses reported by a study.

Methods

THEORETICAL MODELS

Mating encounters

To model acceptance thresholds, t , during mating we used Reeve’s (1989) “search-and-settle” model (his eq. 20) described by the fitness function

$$W(t) = \frac{[FPG(t) + f(I - P)B(t) - C_s]}{[PG(t) + (I - P)B(t)]}. \quad (1)$$

F and f are the fitness consequences of responding to conspecifics and heterospecifics, respectively (positive values reflect a fitness benefit, whereas negative values reflect a fitness cost). We assumed that females would incur a fitness cost from interacting with heterospecifics and set f to -5 when modeling the female response threshold, while f was set to 0 for the male response threshold. We then explored the change in acceptance thresholds for both sexes as F ranged from 0 to 10 (i.e., from no fitness benefits ($F = 0$) to very high fitness benefits ($F = 10$) as a consequence of responding to a conspecific, relative to the sex-specific fitness cost of responding to a heterospecific (females, $f = -5$; males, $f = 0$). P is the proportion of conspecific individuals encountered. We simulated encounters with a sympatric heterospecific by setting P to 0.50 (i.e., 50% of individuals encountered were conspecifics and 50% were heterospecifics). Encounters with an allopatric heterospecific were modeled by setting P to 0.99 (i.e., 99% of all individuals encountered were conspecific, with very few, 1%, encounters with heterospecifics). C_s represents the cost of searching for a conspecific mate and was set to 1.

Finally, $G(t)$ and $B(t)$ are the probabilities of a response to a conspecific and heterospecific, respectively, given the set of cues available for evaluation by the discriminating individual and that individual’s discrimination threshold, t . These distributions were computed using a Gaussian cumulative distribution function bounded between 0 and 1. Conspecifics were centered on an arbitrary value of 0.4, with a standard deviation of 0.05 to simulate a set of identity cues of high reliability, or 0.10 to simulate a set of cues of low reliability (see Fig. 1A). Heterospecifics were centered on an arbitrary value of 0.60 and a standard deviation of 0.06 for a reliable set of identity cues, or 0.11 for a poor set of identity cues. Variance in heterospecific cues was assumed to be larger than those for conspecifics because members of heterospecific species encountered in the environment will most likely belong to more than one species.

Territorial encounters

To model acceptance thresholds (t) in a territorial context, we used Reeve’s (1989) “guard” model (his eq. 1)

$$W(t) = I(F + R)G(t) + i(f + r)B(t) - ir - IR. \quad (2)$$

I represents the average number of conspecific individuals encountered, whereas i is the average number of heterospecific individuals encountered. Encounters with sympatric “familiar” heterospecifics were simulated as $I = 1$ and $i = 1$ (equivalent to $P = 0.50$ for mating encounters modeled in the previous section), whereas encounters with an allopatric “unfamiliar” heterospecific were simulated as $I = 100$ and $i = 1$ (equivalent to $P = 0.99$). R and r represent the fitness costs associated with not responding to conspecifics and heterospecifics, respectively (r should not be

confused with the fitness cost incurred from “responding” to a heterospecific, which—by the way the model is structured—is represented by a negative value of f ; e.g., see next paragraph). We assumed that a territorial resident that does not respond to a conspecific intruder would incur a higher fitness cost than if that intruder were heterospecific and set R to -1 and r to 0 in all computations.

All other parameters and their values were identical as those described for mating encounters in the previous section, with the only exception being that we set values of f to -5 (i.e., we assumed that the fitness costs incurred from responding to heterospecifics would be identical for both sexes; more generally this model is only applicable if both sexes are territorial). We only modeled encounters in which heterospecifics were not in direct competition for resources with the focal species. There are, of course, important exceptions to this situation (reviewed by Grether et al. 2009 and Peiman and Robinson 2010). In situations where species are ecological competitors, animals should respond equally to both conspecific and heterospecific rivals. We discuss the consequences of ecological competitors in reference to specific studies that apparently lacked species discrimination in Table 4.

Our selection of parameter ranges was based on biological intuition (e.g., that females would incur a higher cost [$f = -5$] than males [$f = 0$] in responding incorrectly to a heterospecific during mating because of the energetic differences in gamete production between the sexes [egg vs. sperm production], or that searching for a mate is costly [$C_s = 1$] because of time away from other activities and the energy required for the search itself, and that this would likely be similar for both sexes). We experimented with a variety of realistic parameter settings for f , R , r , and C_s in the mating (eq. 1) and territorial models (eq. 2) to explore the sensitivity of Reeve’s models to changes in these parameters. Acceptance thresholds were qualitatively similar to those depicted in Figure 1B,C. It is important to note that the exact values of parameters are not especially important here, rather it is the ratio of parameter values that changes the shape and magnitude of acceptance thresholds.

EMPIRICAL DATA

Literature review

We searched the ISI Web of Science database using the search terms “species recognition” and “discriminat*” starting with articles appearing in 1980. All abstracts of primary research articles recovered in this search (i.e., not reviews or conference papers) were inspected (1157). Full-text articles that seemed to investigate species recognition in either mating or territorial (aggressive) contexts, and that were available electronically through the libraries of Harvard University, were downloaded for more thorough review (216 articles). For inclusion in the meta-analysis, studies had to test subject responses to both conspecific and nonconspecific

(heterospecific or novel) stimuli, as well as present information that could be used to calculate effect sizes and their confidence intervals. Specifically, we compiled information on the number of animals tested, and the means and standard deviations/errors of behaviors elicited by conspecific and nonconspecific stimuli (e.g., latency to respond or approach a stimulus; the number or duration of responses evoked, such as calls or displays; copulation or attempts to copulate), or the number of subjects that did or did not respond to a conspecific and nonconspecific stimulus. These data were gathered from the text, tables, and figures of a study, with data from figures measured digitally using Adobe Illustrator CS3, 13.0.0, Adobe Systems (Adobe Systems Incorporated, San Jose, CA).

Literature searches were completed on March 13 2009. We supplemented our survey with additional sources compiled as part of a smaller meta-analysis reported in Ord and Stamps (2009). The final dataset is presented in Table S4 and includes 92 studies published from 1980 to 2008, covering 111 species. This table also highlights 15 species for which effect size estimates were based on both pre-mating (courtship) and realized mating responses (copulation or attempts to copulate). Mating can sometimes occur as a result of harassment or forced copulation by heterospecific males, despite females exhibiting strong discriminatory responses against heterospecifics. An effect size estimate that includes realized matings might therefore underestimate species discrimination. We point out, however, that for no species was the assessment of heterospecific discrimination based exclusively on realized matings, and that the vast majority of species were assessed on behavioral responses exhibited during courtship (e.g., signal production, approach).

Effect size calculations

For studies presenting means and standard deviations/errors, we calculated a standardized effect size as Cohen’s d using equations outlined in Ord and Stamps (2009). When studies collected binary data in the form of contingency tables (i.e., for χ^2 tests), we computed the log odds-ratio

$$\log OR = \ln \frac{N_{+c}N_{-h}}{N_{+h}N_{-c}},$$

and its variance

$$v_{OR} = \frac{1}{N_{+c}} + \frac{1}{N_{+h}} + \frac{1}{N_{-c}} + \frac{1}{N_{-h}}.$$

Here, N is the number of individuals that responded (+) or did not respond (–) to a conspecific (c) or heterospecific (h) stimulus. The log odds-ratio and variance was then converted into Cohen’s d (Borenstein et al. 2009),

$$d = \log OR \frac{\sqrt{3}}{\pi} \quad (3)$$

with

$$v_d = v_{OR} \frac{3}{\pi^2}. \quad (4)$$

There were a large number of mate choice tests using binomial proportions, p , in which researchers calculated the proportion of animals that responded to the conspecific stimulus, N_{+c} , out of the total number of animals tested (the number of animals that made no choice, N_{-c-h} , plus the number of animals that responded to either stimulus, leading to the following equation: $\frac{N_{+c}}{N_{+c}+N_{+h}+N_{-c-h}}$). The difficulty here was there were no previous published accounts of converting p into d . We contacted two meta-analysis experts and both recommended converting p first into a log odds-ratio, and then converting the logs odd-ratio into d . To do so, both experts independently formulated the following equations (L. V. Hedges, pers. comm.; D. B. Wilson, pers. comm.)

$$\log OR = \ln \frac{p}{(1-p)},$$

with the variance associated with this odds-ratio computed as

$$v_{OR} = \frac{1}{N_{total}p} + \frac{1}{N_{total}(1-p)}.$$

These equations effectively treat p as a special case of the odds-ratio statistic and assume the null response in a binomial proportion test is 0.5. That is, had a control been tested (and ignoring sampling error), a random distribution of responses would be reflected by an even split between the treatment and control stimuli. The $\log OR$ and v_{OR} were then transformed into Cohen's d using equations (3) and (4).

To provide a bounded effect size measure (-1 to 1) for presentation in figures and tables that would be familiar to most readers, Cohen's d and its values corresponding to upper and lower 95% confidence intervals (CI) were converted into r using (Borenstein et al. 2009)

$$r = \frac{d}{\sqrt{d^2 + 4}}.$$

Finally, to justify combining quantitative and qualitative data into a single metric (i.e., an overall effect size based on Cohen's d , odds ratio, and binomial proportions), we confirmed there was no statistical difference in the magnitude of effect sizes depending on the method authors used to measure subject responses (95% CIs, $r_{quantitative}$ 0.27–0.58, $r_{qualitative}$ 0.35–0.61). Many studies also reported several different types of behavioral responses to stimuli (e.g., time taken to approach a stimulus, time spent near a stimulus, number of calls/displays elicited by a stimulus) or recorded responses to several different heterospecifics (e.g., Table 4). We computed two effect sizes, the "overall" response of subjects based on all behaviors measured and all heterospecific comparisons made by authors, and a "maximum" response based on the largest effect size computed for a given species by

a study (see also Ord and Stamps 2009). In the end, conclusions drawn from analyses of both sets of effect sizes were identical. We report analyses and present figures based on overall responses in the main text, and provide the equivalent analyses based on maximum responses in the Supporting information.

Meta-analysis

Once effect sizes had been converted into a common statistic, a combined effect estimate could be calculated to test whether broad trends exist across studies in relation to a specified predictor variable. There are two statistical approaches that can be used to combine effect sizes (see Borenstein et al. 2009 for an excellent summary of these techniques). The fixed-effect model assumes that there is one true effect common to all studies and any differences between studies in the estimate of this true effect reflects sampling error. The random-effects model, on the other hand, allows the underlying effect size for each study to vary and computes a combined effect estimate assuming that each of these true effect sizes are distributed around a common mean. Deciding on the "correct" approach to use can be difficult. The decision is dependent on the question being addressed and the way in which the empirical data have been collected. In our study, we used both approaches at different stages in the analysis based on the following considerations.

It has been argued that for meta-analyses of ecology and evolution data, the assumption of the fixed-effect model (in which there is one true effect for all studies) is unrealistic (e.g., Gurevitch and Hedges 1993). However, there are some important statistical issues that need to be kept in mind when evaluating results from both a fixed-effect and random-effects model. In the fixed-effect model, the overall effect estimate is a weighted average, with weights specified by the variance associated with each effect size (which essentially reflects a study's sample size). Results from a fixed-effect model have a greater tendency to be skewed by outliers, that is, a study with very large sample sizes that reports an unusually large or small effect relative to other studies included in the analysis. A random-effects model also computes the overall effect estimate as a weighted average, but because the model assumes the collection of true effects for each study is normally distributed around a common mean, an additional parameter summarizing the variance of this distribution is included in calculations. Results from a random-effects model will tend to reduce the influence of an unusually large (or small) effect size from a study with a very large sample size. Although this limits the pull that an outlier study might have on the final overall effect estimate, reducing the influence of studies with larger sample sizes and more precise effect sizes might be philosophically unsatisfactory in some situations. In which case, preference for a fixed-effect model over a random-effects model has merits.

Our approach was to carefully evaluate the philosophy of combining effect sizes at three stages in the analysis—(i) within-studies; (ii) among studies of the same species; and (iii) across species in the same experimental condition—and apply the meta-analysis model that we considered to be more appropriate based on its underlying statistical assumptions. That is, we felt it reasonable to assume that there was one true effect size for all studies conducted on the same species (fixed-effect), but that the true effect size probably varied among species (random-effects). First, we used a fixed-effect model to combine effect sizes within-studies when more than one effect size was calculated for a given study and for a given species. For example, many studies recorded more than one response variable to quantify responses to stimuli, or in some cases replicated stimulus presentations several times. Next, this within-study effect measure was combined with those from other studies using a fixed-effect model to obtain an “among studies” effect measure whenever separate studies tested the same species in the same social context (mating or territorial) and under the same experimental conditions for the predictor variables being examined (sex of subject, familiarity of heterospecific, or modality of cue available for species identity assessment). Among studies effect measures for each species were used in calculations of phylogenetic signal, and for presentations in forest plots to allow readers to evaluate the distribution of effect sizes across different species and to identify species that might be of personal interest (see Supporting information). Finally, for hypothesis testing of model predictions we relied on a random-effects model to combine among study effect measures across species and experimental conditions.

Fixed-effect and random-effects models were applied using the “Meta-Analysis Package for R” version 0.5–4 (Viechtbauer 2009) implemented in R 2.8.1 (R Development Core Team). For random-effects models, a number of estimators are available for calculating the variance in the distribution of true effects around the common mean (the overall effect size being calculated; estimators evaluated were (Viechtbauer 2009): Hunter Schmidt, Hedges, DerSimonian Laird, Sidik Jonkman, maximum likelihood, restricted maximum likelihood, and empirical Bayes). In all cases, AIC values indicated a model using a maximum likelihood estimator fit the data best and we report results using this estimator for all random-effects models.

We relied on 95% confidence intervals to determine whether effect sizes differed significantly between predictor variables and from a value of zero (corresponding to subjects reacting equally to both conspecific and heterospecific cues).

Phylogeny

We used three approaches for estimating phylogenetic signal to evaluate the extent heterospecific discrimination was associated with phylogeny. First, the phylogenetic mixed model (Lynch

1991; Housworth et al. 2004) implemented in COMPARE 4.6b (Martins 2004) measures the relative contribution of phylogenetically heritable (h^2 ; those phenotypic changes inherited from evolutionary ancestors) and nonphylogenetically heritable factors ($1 - h^2$; those phenotypic changes that are not retained in descendants from evolutionary ancestors). Low phylogenetic signal is indicated by an h^2 value approaching 0 (i.e., phenotypic variance among present-day species is unrelated to phylogeny), whereas high phylogenetic signal is indicated by a value approaching 1 (i.e., closely related species are more similar in their responses to heterospecifics than distantly related species). Second, Hansen et al.’s (2008) method in the SLOUCH version 1.2 package for R relies on an Ornstein–Uhlenbeck model of “constrained” evolution that accounts for the extent that trait evolution has been free to vary adaptively ($t_{1/2}$; the phylogenetic half-life of a phenotypic characteristic) and the influence of stochastic factors (v_y) during evolutionary diversification. Low phylogenetic signal is reflected by $t_{1/2}$ values that approach 0 (i.e., phenotypic characteristics are not retained from evolutionary ancestors), whereas high phylogenetic signal corresponds to large values of $t_{1/2}$ ($t_{1/2}$ can range from 0 to ∞). Low values of v_y suggest that stochastic forces resulting in nonadaptive phenotypic variation have been weak (v_y has a range from 0 to ∞). We report the range of $t_{1/2}$ and v_y within two likelihood units that by convention are considered to be estimates that fit the data equally well. Third, the method implemented in the program BayesTraits relies on a Brownian motion model of evolution to compute the parameter λ (Pagel 1999), which can be used to interpret the extent present-day phenotypes reflect phylogenetic relationships between species. Low phylogenetic signal is indicated by a value of λ approaching 0, whereas values that approach 1 are indicative of high phylogenetic signal. To test the significance of the estimated λ , we reran the program with λ set to 0 and 1, and compared the likelihoods of these computations with the original estimate of λ . Significance was based on whether the difference in likelihood was greater than 2 units.

Each method offers a fundamentally different view of how evolution might have occurred and collectively provided a way of testing the phylogeny of species discrimination irrespective of the underlying evolutionary model used. That is, estimates of phylogenetic signal that are consistent across the three methods should reflect robust evolutionary trends in the data. We estimated phylogenetic signal separately for each sex and social context (mating or territorial) because species discrimination may differ enough in these situations to affect estimates of phylogenetic signal.

To control for the potential influence of phylogeny on species discrimination when testing for differences in effect size among species in the meta-analysis, we used Lajeunesse’s (2009) PhyloMeta, beta version 1.0. We report these estimates and associated AIC scores along side those from conventional random-effects models (i.e., those not incorporating phylogeny).

The phylogenetic hypothesis we used was created by positioning the major taxonomic groups—insects, crustaceans, fish, amphibians, mammals, birds, turtles and other reptiles—using the classification scheme of the Encyclopedia of Life (eol.org). Genus-level relationships within these major groups were then positioned using the most comprehensive phylogenies available in the PhyLoTA database (phyloTA.net), with finer resolution of some relationships based on primary sources (Schluter 1989; Coddington 1991; Cannatella et al. 1998; Marcus and McCune 1999; Ranwez et al. 2007; Zhu et al. 2007). Species were left as polytomies in the few cases where phylogenetic relationships could not be fully resolved. Although we were not able to include information on branch lengths when creating our phylogeny, we explored alternative modes of evolution by manipulating branch lengths artificially to create two phylogenies: a “speciational” phylogeny in which all branch lengths were set to 1 to simulate a scenario that assumed evolutionary divergence in species discrimination was concentrated during speciation events (at phylogenetic nodes); and a “gradual” phylogeny in which branch lengths were scaled ultrametrically using the program Mesquite version 2.72 (Maddison and Maddison 2009) to simulate a scenario that assumes evolutionary divergence in species discrimination occurs gradually over evolutionary time. Figure S1 illustrates the ultrametric phylogeny used.

Results

BIOLOGICAL PREDICTORS OF SPECIES DISCRIMINATION

Meta-analyses of species effect sizes that did not incorporate phylogeny were of significantly better fit, although the interpretation of results would be similar regardless of whether models did or did not assume an underlying phylogenetic structure to the data (Table 1). This should not be considered a reflection of the phylogenetic signal in the data (see next section), rather AIC assesses the fit of incorporating phylogeny in computations of the combined effect size as a function of the predictor variable being evaluated.

There were no significant differences in effect size as a function of the social context of discrimination (mating vs. territorial), the sex of the discriminating individual, familiarity with the heterospecific (sympatric vs. allopatric), or modality of species cues used for assessment (acoustic, visual, olfactory or multimodal; Table 1; Figs. S2 and S3). The same was true for analyses based only on the maximum effect size computed for a study (Table S1). We also assessed the extent focal species were phylogenetically related to the heterospecific influenced species discrimination and again found no difference in our results (e.g., Table S2; phylogenetic relatedness was determined as whether the heterospecific cues tested belonged to species within the same or different genus to the focal species).

Although Reeve’s models did predict cases where similarities in species discrimination thresholds would be expected under some circumstances (e.g., similar levels of discrimination in mating and territorial contexts), the fact that no variable accounted for any variance in discrimination among species is still unexpected. Based on these results alone, there appears to be no universal predictor of whether a species will or will not discriminate against heterospecifics. It suggests that the species we surveyed lie collectively along the full breadth of parameter space illustrated in Figure 1. Species discrimination is therefore dictated by the benefits and costs of responding to a conspecific or heterospecific that are specific to the circumstances of a given species being studied.

In several cases the same species had been tested under different conditions or with several types of stimuli (either by the same or separate studies). These within species “paired” comparisons offered a potentially more powerful means of testing predictions from Reeve’s models as the subset of species included in the analysis contributed an effect measure for all the conditions of a given predictor (e.g., an effect size for both “mating” and “territorial defense” in an analysis of social context). This should minimize variance due to differences among species in responses that are otherwise unrelated to the predictor variable being tested. Yet focusing only on the subset of species for which all conditions of a predictor variable had been tested still revealed no prominent trends associated with any predictor variable (Fig. 2):

- (i). *Social context*: The consistency of responses between mating and territorial contexts continued to follow predictions from Reeve’s models (Fig. 1B), but most of the species included in this analysis were of the same genus (Darwin’s finches; *Geospiza*), making broader inferences beyond this group difficult (Fig. 2A).
- (ii). *Sex differences*: Reeve’s models also predicted that females would either be the more discriminating sex during mating, or that males and females would tend to be equally discriminating in some mating scenarios (Fig. 1B). Inspection of paired estimates for single species in Figure 2B shows several species in which males were generally indiscriminate of species identity during mating compared to highly discriminating females (Birds: *Uraeginthus cyanocephalus* and *U. gongalus*; Fish: *Poecilia latipinna* and *Xiphophorus birchmanni*; Insects: *Leptidea reali* and *L. sinapis*), and a handful of species in which the responses of the sexes were very similar (Tortoise: *Testudo hermanni*; Fish: *Pseudotropheus zebra*; Lizard: *Eumeces laticeps*; Insect: *Gryllus rubens*). Nevertheless, there were also several species in which males were the more discriminating sex, whereas females were equally likely to respond to either a conspecific or heterospecific

Table 1. Predictors of species recognition.

Predictors	N_{species}	No phylogeny		“Speciational” phylogeny		“Gradual” phylogeny	
		r (lower CI, upper CI)	AIC	r (lower CI, upper CI)	AIC	r (lower CI, upper CI)	AIC
A. Social context							
Mating	90	0.36 (0.27, 0.43)	351.1	0.14 (0.08, 0.18)	449.1	0.19 (0.12, 0.24)	455.6
Territorial	28	0.43 (0.25, 0.56)	94.8	0.41 (0.24, 0.54)	136.6	0.40 (0.24, 0.53)	138.4
B. Sex differences							
<i>Mating</i>							
Males	57	0.40 (0.24, 0.52)	231.8	0.09 (0.02, 0.17)	265.0	0.11 (0.04, 0.18)	275.4
Females	55	0.41 (0.30, 0.50)	204.3	0.21 (0.09, 0.32)	261.1	0.25 (0.14, 0.35)	263.7
<i>Territorial</i>							
Males	22	0.44 (−0.50, 0.84)	77.6	0.44 (0.27, 0.58)	106.9	0.44 (0.27, 0.57)	109.0
Females	3	0.87 (0.62, 0.94)	–	–	–	–	–
C. Familiarity							
<i>Mating</i>							
Sympatric	62	0.31 (0.23, 0.38)	235.6	0.20 (0.13, 0.27)	295.3	0.22 (0.16, 0.28)	297.4
Allopatric	27	0.24 (0.09, 0.37)	89.1	0.22 (0.11, 0.33)	120.4	0.22 (0.11, 0.32)	125.8
<i>Territorial</i>							
Sympatric	24	0.47 (0.31, 0.59)	84.1	0.47 (0.30, 0.60)	117.0	0.46 (0.31, 0.58)	122.0
Allopatric	3	0.36 (−0.21, 0.70)	–	–	–	–	–
D. Modality							
<i>Mating</i>							
Acoustic	22	0.38 (0.09, 0.58)	89.4	0.15 (−0.03, 0.33)	92.3	0.13 (−0.05, 0.30)	93.1
Visual	17	0.38 (0.07, 0.59)	51.2	0.44 (0.25, 0.58)	66.8	0.44 (0.26, 0.57)	68.1
Olfactory	29	0.36 (0.09, 0.56)	106.8	0.40 (0.19, 0.56)	130.1	0.35 (0.15, 0.51)	137.2
Multimodal	35	0.49 (0.34, 0.60)	150.7	0.28 (0.09, 0.44)	173.1	0.29 (0.11, 0.44)	178.0
<i>Territorial</i>							
Acoustic	15	0.48 (0.14, 0.69)	64.7	0.36 (0.12, 0.54)	67.3	0.37 (0.17, 0.53)	68.6
Visual	8	0.32 (−0.12, 0.63)	23.3	0.36 (0.03, 0.59)	31.3	0.36 (0.04, 0.58)	31.6
Olfactory	3	0.07 (−0.49, 0.58)	9.8	0.18 (−0.75, 0.83)	12.6	0.18 (−0.75, 0.83)	12.6
Multimodal	8	0.56 (0.32, 0.71)	26.1	0.84 (0.67, 0.91)	33.3	0.84 (0.68, 0.91)	33.5

Combined effect sizes and 95% confidence intervals (CI) of species responses to conspecifics relative to nonconspecifics. Positive r -values correspond to greater levels of response to a conspecific stimulus, whereas negative values reflect greater levels of response to a nonconspecific stimulus.

AIC, Akaike information criterion.

male (Lizards: *Podarcis hispanica* and *P. bocagei*; Bird: *Uraeginthus angolensis*; Amphibians: *Plethodon jordani* and *P. teyahalee*).

(iii). *Familiarity*: There was some support for the prediction (Fig. 1B,C) that animals would be more discriminating of sympatric heterospecifics compared to allopatric heterospecifics, but when effect sizes were combined, estimates for the two categories were virtually identical (Fig. 2C).

(iv). *Reliability of species cues*: Assuming that multimodal signals provide multiple backup cues of species identity and are therefore the most reliable signals for discriminating species (e.g., Fig. 1A), there was little evidence supporting an association between the assessment of multimodal

cues and strong discrimination (Fig. 2D). More generally, no modality stood out as being more or less likely to be associated with high levels of species discrimination.

PHYLOGENETIC SIGNAL OF SPECIES DISCRIMINATION

Estimates of phylogenetic signal were generally consistent across the three methods (Phylogenetic Mixed Model, BayesTraits and SLOUCH) and phylogenies used (“speciational” vs. “gradual”; Table 2).

There was virtually no phylogenetic signal in the species discrimination of females during mating, whereas the level of species discrimination exhibited by courting males showed high phylogenetic signal (this was consistent across methods and

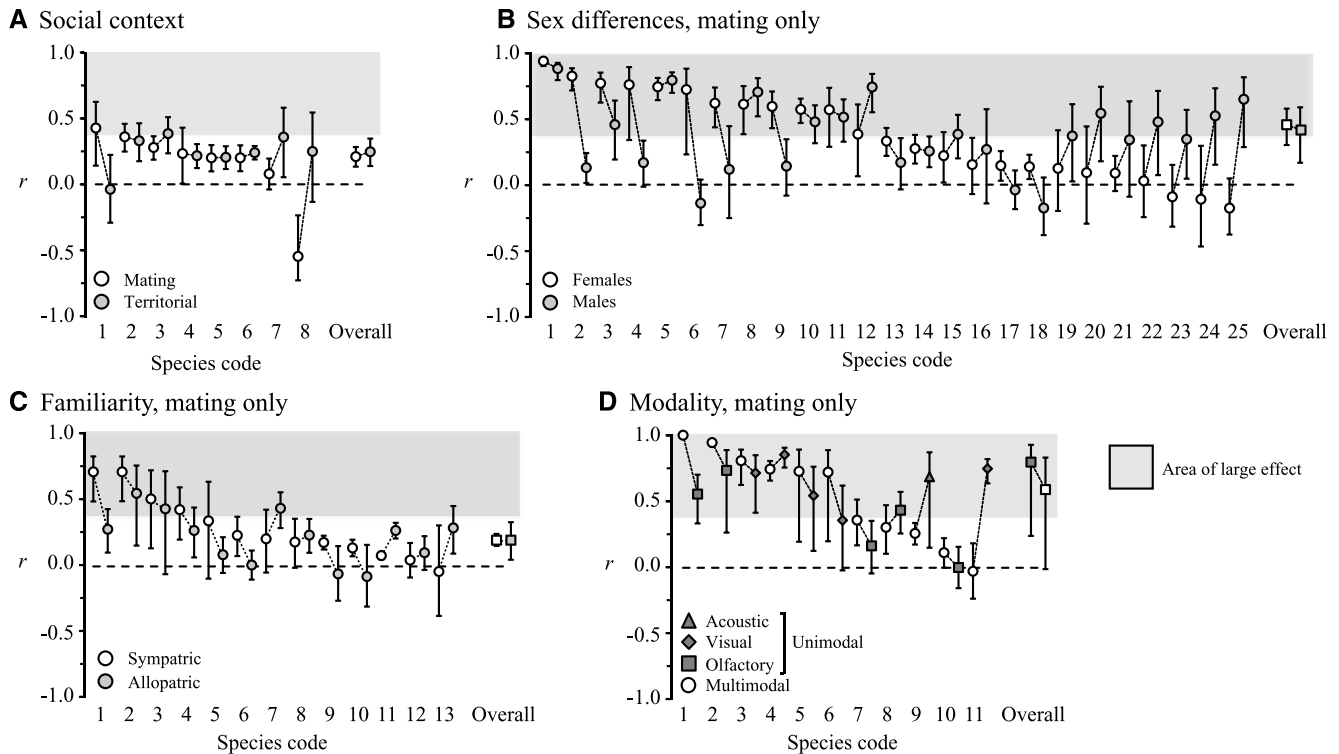


Figure 2. The magnitude of species discrimination for those species in which responses had been tested for both: mating and territorial contexts (A); males and females (B); sympatric and allopatric heterospecifics (C); and unimodal and multimodal cues (D). Multimodal cues were a set of cues from two or more modalities. Effect sizes are presented as r -values with 95% confidence intervals. All experiments testing a given species were combined to produce a single effect size measure for each species. These estimates were in turn combined into an “overall” effect size measure across all species tested for a given predictor variable. Positive values of r indicate responses were greater to the conspecific stimulus than the nonconspecific stimulus, whereas negative values indicate responses were greater to the nonconspecific stimulus. A value of zero (dotted line) indicates the level of response was the same to conspecific and nonconspecific stimuli. The shaded region corresponds to the conventional interpretation of a large effect ($r \geq 0.37$). For example, an r -value of 0.44 indicates that the magnitude of response evoked by the conspecific stimulus was one standard deviation greater than the level of response evoked by the nonconspecific stimulus. Species codes are: (A) 1 = *Melospiza georgiana* (bird); 2 = *Geospiza difficilis* (bird); 3 = *Geospiza fuliginosa* (bird); 4 = *Cyprinodon variegatus* (fish); 5 = *Geospiza scandens* (bird); 6 = *Geospiza fortis* (bird); 7 = *Serinus serinus* (bird); and 8 = *Hemidactylus frenatus* (lizard); (B) 1 = *Calopteryx splendens* (insect); 2 = *Poecilia latipinna* (fish); 3 = *Uraeginthus cyanocephalus* (bird); 4 = *Leptidea reali* (insect); 5 = *Pseudotropheus callainos* (fish); 6 = *Leptidea sinapis* (insect); 7 = *Uraeginthus bengalus* (bird); 8 = *Testudo hermanni* (turtle); 9 = *Xiphophorus birchmanni* (fish); 10 = *Pseudotropheus zebra* (fish); 11 = *Eumeces laticeps* (lizard); 12 = *Prokelisia dolus* (insect); 13 = *Gryllus texensis* (insect); 14 = *Gryllus rubens* (insect); 15 = *Dipsosaurus dorsalis* (lizard); 16 = *Zaprionus sepsoides* (insect); 17 = *Poecilia mexicana* (fish); 18 = *Serinus canaria* (bird); 19 = *Prokelisia marginata* (insect); 20 = *Podarcis hispanica* (lizard); 21 = *Photinus greeni* (insect); 22 = *Uraeginthus angolensis* (bird); 23 = *Plethodon jordani* (amphibian); 24 = *Podarcis bocagei* (lizard); and 25 = *Plethodon teyahalee* (amphibian); (C) 1 = *X. birchmanni* (fish); 2 = *Uraeginthus cyanocephalus* (bird); 3 = *Uraeginthus bengalus* (bird); 4 = *Teleogryllus taiwanemma* (insect); 5 = *Photinus greeni* (insect); 6 = *Spea multiplicata* (amphibian); 7 = *Alectoris rufa* (bird); 8 = *P. latipinna* (fish); 9 = *Lemmus sibiricus* (mammal); 10 = *Dicrostonyx groenlandicus* (mammal); 11 = *Schizocosa ocreata* (insect); 12 = *Poecilia mexicana* (fish); and 13 = *Alectoris graeca* (bird); (D) 1 = *Calopteryx virgo* (insect); 2 = *Calopteryx splendens* (insect); 3 = *Cochliomyia macellaria* (insect); 4 = *Cochliomyia hominivorax* (insect); 5 = *Laticauda frontalis* (snake); 6 = *Laticauda colubrina* (snake); 7 = *Poecilia petenensis* (fish); 8 = *Poecilia velifera* (fish); 9 = *Gryllus rubens* (insect); 10 = *Poecilia mexicana* (fish); and 11 = *Xiphophorus pygmaeus* (fish).

phylogeny used; Table 2). This indicates that effect sizes in the degree to which females did or did not discriminate against heterospecifics during mating differed among species independently of phylogeny. In contrast, males of closely related species were more likely to exhibit similar levels of species discrimination during mating than males of more distantly related species.

There were too few species in which females were tested in a territorial context to warrant phylogenetic analysis (only three species; see Table 1). However, species discrimination by males in a territorial context showed virtually no phylogenetic signal (regardless of method or phylogeny used; Table 2).

Table 2. The phylogeny of species recognition.

Sex, context	N_{species}	Phylogenetic mixed model			BayesTraits		SLOUCH		
		Speciational h^2 (likelihood)	Gradual h^2 (likelihood)	Speciational λ (likelihood)	Gradual λ (likelihood)	Gradual only supportregion	$t_{1/2}$, supportregion	v_y , supportregion (likelihood)	
Females, mating	55	Estimated	0.00 (-37.9)	0.14 (-37.8)	0.00 (-99.0)	0.00 (-93.1)	0.13, 0.01-0.57	0.97, 0.5-2.0	(-85.7)
		Set to 0	-	-	0.00 (-99.0)	0.00 (-93.1)	-	-	-
		Set to 1	-	-	1.00 (-102.2)	1.00 (-99.3)	-	-	-
Males, mating	57	Estimated	0.87 (-60.0)	0.69 (-55.8)	1.00 (-120.4)	0.67 (-114.4)	10, 10-∞	10, 0-∞	(-84.1)
		Set to 0	-	-	0.00 (-132.4)	0.00 (-118.8)	-	-	-
		Set to 1	-	-	1.00 (-120.4)	1.00 (-123.1)	-	-	-
Males, territorial	22	Estimated	0.28 (-10.2)	0.00 (-10.3)	0.00 (-29.0)	0.00 (-31.8)	0.05, 0.00-0.30	1.0, 0.5-1.8	(-31.7)
		Set to 0	-	-	0.00 (-29.0)	0.00 (-31.8)	-	-	-
		Set to 1	-	-	1.00 (-35.1)	1.00 (-40.22)	-	-	-

SLOUCH can only be applied using an ultrametric phylogeny.

A phylogeny with ultrametric branch lengths simulating a “gradual” mode of evolution significantly fit male mating responses better than a “speciational” phylogeny in which branch lengths were set to 1 (Table 2). There was very little difference in fit between phylogenies for female mating responses and male territorial responses.

EXPERIMENTAL PRECISION OF SPECIES DISCRIMINATION STUDIES

Despite an overall trend showing that most species reacted more strongly to conspecific than heterospecific stimuli (Table 1; Figs. S2 and S3), there were still a surprising number of cases in which estimates could not be considered significantly different from zero, especially during mating (29 of 90 species; Fig. S2A). That is, many species were reported to respond equally to both conspecifics and heterospecifics in mating. There were even species that tended to react more strongly to heterospecifics than conspecifics, but these cases were rare (e.g., a salamander: *Hemidactylus frenatus*; a bird: *Sylvia atricapilla*; and a fish: *Xiphophorus continens*).

Aggressive responses among ecologically competing heterospecifics might be adaptive, but there is little fitness benefit to responding to a heterospecific during mating for either sex and, as such, positive responses to courting heterospecifics should be rare (e.g., Fig. 1B). The large number of cases in which species lacked heterospecific discrimination mirrors findings from an earlier meta-analysis, where it was suggested that many studies suffered from low statistical power or complications associated with experimental setting (laboratory vs. field; Ord and Stamps 2009). To examine this issue, we tested the influence of sample size and experimental setting on effect size calculations, and then conducted a detailed review of the experiments conducted on those species apparently lacking species discrimination (or were found to react more strongly to heterospecifics).

The precision (variance) of effect sizes was heavily dependent on both the experimental setting and the number of subjects tested (Table 3; Fig. 3): studies testing fewer animals, and animals held in captivity, had significantly lower precision (higher variance) in effect size calculations. Many of these studies were also those computed to have effect sizes that could not be considered statistically different from zero.

Generally, however, our detailed review of all the studies found to have nonsignificant effects (or effects skewed toward stronger responses to heterospecifics) showed that many experiments (30 of 55 experiments; Table 4) were well designed, had high statistical power, and authors provided clear biological interpretations of their results that were consistent with predictions from Reeve’s models. In most cases, lack of species discrimination was attributed to similarity between conspecific and heterospecific signals leading to misidentification (corresponding to

Table 3. The precision of species recognition studies.

Overall model: $F_{3,117} = 6.87, P < 0.001, R^2 = 0.15$			
Variable	β	T	P -value
Intercept	-0.50	-2.23	0.028
Sample size	-0.68	-3.89	<0.001
Social context	0.11	0.76	0.451
Experimental setting	-0.28	-2.02	0.046

Model applied: $\log(\text{effect variance}) = \log(\text{sample size}) + \text{social context} + \text{experimental setting}$. Sample size reflects the average number of subjects tested for a given species; social context was scored as 0 for mating and 1 for territorial defense; experimental setting was scored as 0 for laboratory/captive based studies and 1 for field experiments. Two-way interaction terms between sample size and context or setting were not statistically significant and were excluded from the model.

a scenario in which the distribution of conspecific and heterospecific cues have a large degree of overlap; e.g., Fig. 1A). Yet, for every robust study we examined, we found an almost equal number of experiments (25 of 55; Table 4) that were poorly designed (10 of 55), lacked statistical power (13 of 55 experiments tested ≤ 10 subjects) or authors concluded species discrimination based on a subset of significant tests out of a larger set of nonsignificant results (2 of 55; Table 4). To assess the impact of these studies on our hypothesis tests, we excluded them from a second series of meta-analyses and obtained virtually identical results to those reported in Table 1 (see Table S3).

Discussion

An important step in the speciation process is believed to be the formation of behavioral mechanisms that establish or reinforce reproductive isolation between populations (Streelman and Danley 2003; Ritchie 2007; Sobel et al. 2010). The most obvious barrier to reproduction that will limit gene flow between populations is one associated with assortative mating. Here, preferences for certain cues in mates can lead to discrimination against members from foreign populations. Although the role of territorial defense in promoting reproductive isolation is less intuitive, it can also enforce segregation between populations by excluding foreign individuals from establishing residence in certain areas (Grether et al. 2009; Peiman and Robinson 2010). How species discriminate between conspecifics and heterospecifics can provide valuable clues on the nature of these behavioral isolating mechanisms, and this is often the motivation underlying many of the studies included in our meta-analysis.

With this broader context of speciation in mind, our analysis suggests species discrimination is equally likely during mating and territorial contexts, is not dependent on previous familiarity

with the heterospecific cue or the modality of cues used for assessment, or the sex of the individual assessing identity cues (e.g., species recognition is not primarily driven by female mate choice decisions). By extrapolation, this would imply the formation of reproductive isolation between populations is not constrained by the type of social system, encounter rate, or the modality of social communication used by animals. Even so, there was considerable variation among species in the magnitude of discrimination reported by studies. Whether animals do or do not respond to heterospecifics is more likely to depend on the specific circumstances and natural history of the species in question (e.g., the spacing patterns of conspecifics, the intensity of sexual selection, predation pressure, and a host of other factors affecting the benefits/costs of responding to an individual in a particular environment). Furthermore, we found some evidence that phylogeny may play a role (Table 2). Although inferring evolutionary process from estimates of phylogenetic signal is difficult (Revell et al. 2008), the sex difference in the phylogenetic signal of species discrimination warrants further investigation in future comparative and genetic studies. It implies that, across a large and diverse group of species, the genetic correlation between the sexes in discrimination phenotypes (e.g., the conspecific-template) may be low, and that some aspects of the evolution of species recognition may have subsequently differed between the sexes. For example, it is possible that males rely on a small number of very similar, evolutionary conserved cues to assess species identity, whereas females tend to assess a variety of “back up” cues that vary widely among species. Alternatively (or in addition), the evolutionary lability in female responses to heterospecifics might reflect a learning component to species recognition in which ecological or social differences between the sexes has led to species variation in learning identity cues by sex. Sex differences in phylogenetic signal might consequently reflect a genetic component to species recognition in males, and a plastic component to species recognition in females (e.g., Svensson et al. 2010).

It needs to be made clear that, as with any meta-analysis (or any qualitative review for that matter), broad generalizations will be complicated by differences among studies in objectives and methodology. Not all the studies included in our meta-analysis were necessarily concerned with testing the species recognition abilities of their subjects. Some studies included a heterospecific stimulus as a control, but were mainly interested in determining whether test animals could distinguish between conspecific individuals (e.g., Fornasieri and Roeder 1992). The extent this might influence our results is unclear. Of potentially bigger consequence were the frequent differences among studies in experimental design (Table 4). The effect sizes of many studies were small ($r < 0.24$; Figs. S2 and S3) and sometimes of such poor precision that effects were not statistically different from zero (see Table 4). There are certain circumstances in which the cost of failing to

Table 4. Species that lacked heterospecific discrimination or showed discrimination in favor of heterospecifics.

Context, taxon	Species	Study	Sex	Stimulus type	Stimulus modality	Stimulus	Non-conspecific stimulus	Design	Setting	r (lower CI, upper CI)	N _{subjects}	Study conclusion	Comments*	
Amphibian	<i>Bufo gargarizans</i>	1	M	Live	Multi		<i>Rana catesbeiana</i> (S)	Two choice	Laboratory	-0.17 (-0.39, 0.08)	20	Misidentification	Scramble competition possible; models predict males to be indiscriminate.	
		2	F	Playback	Acoustic		<i>Hyla chrysoxcelis</i> (S)	Two choice	Laboratory	-0.09 (-0.49, 0.34)	11	Misidentification (cue similarity)		
	<i>Physalaemus</i> <i>enesefae</i>	3	F	Playback	Acoustic		Call manipulation of <i>Physalaemus pustulosus</i> (S), <i>P. freibergi</i> (A), <i>P. pustulatus</i> (A)	Two choice	Laboratory	0.08 (-0.03, 0.19)	20-27	Misidentification (cue similarity or lack of cues)		
		4	F	Scented airflow	Olfactory		<i>Plethodon taylorae</i> (S)	Two choice	Laboratory	-0.09 (-0.32, 0.15)	21-24	Discrimination (some populations)	Potential problem with experimental design highlighted by authors that may account for low effect sizes.	
	<i>Plethodon taylorae</i>	4	M	Scented airflow	Olfactory		<i>Plethodon taylorae</i> (S)	Two choice	Laboratory	0.34 (0.04, 0.57)	20-37			
		4	F	Scented airflow	Olfactory		<i>Plethodon jordani</i> (S)	Two choice	Laboratory	-0.18 (-0.38, 0.05)	16-27			
	<i>Plethodon shermani</i>	5	M	Scented airflow	Olfactory		<i>Plethodon jordani</i> (S)	Two choice	Laboratory	0.65 (0.28, 0.81)	12-19			
		5	M	Live / Scented	Olfactory		<i>Plethodon yonahlossee</i> (A), <i>P. montanus</i> (A)	No choice	Laboratory	0.03 (-0.19, 0.25)	12-18	Misidentification (cue similarity)	All presentations used a conspecific female, which may have resulted in conflicting cues irrespective of olfactory signals tested.	
	Bird	<i>Spea multiplicata</i>	6	F	Playback	Acoustic		<i>Spea bombifrons</i> (S), <i>Scaphiopus couchii</i> (S)	Two choice	Laboratory	0.12 (0.03, 0.21)	8-54	Discrimination	Heightened responses to heterospecifics probably reflect a preference in allopatric females for exaggerated call characteristics.
			6	F	Playback	Acoustic		<i>Spea bombifrons</i> (A)	Two choice	Laboratory	-0.33 (-0.55, -0.04)	18	Misidentification	
<i>Serinus canaria</i> (domestic)		7	F	Playback	Acoustic		<i>Carduelis spinus</i> (A), <i>Serinus canaria</i> (wild-type)	Sequential choice	Laboratory	0.23 (0.08, 0.37)	9	Discrimination		
	8	M	Playback	Acoustic		<i>Carduelis spinus</i> (A)	Sequential choice	Laboratory	-0.18 (-0.38, 0.05)	18	Discrimination	Lower responses were elicited by a conspecific stimulus than a heterospecific stimulus.		
Bird	<i>Serinus serinus</i>	9	F	Playback with dummy	Acoustic		<i>Serinus</i> spp (19 congeneric species; all A)	Two choice	Laboratory	-0.05 (-0.17, 0.08)	24-27	Discrimination	Conclusion is based on one of five comparisons that showed greater responses to conspecific calls. Multiple species were pooled for the heterospecific stimulus and may have reduced treatment effects. All presentations were conducted with a taxidermic mount of a conspecific, which may have also resulted in conflicting identity cues.	
		9	F	Playback	Acoustic		<i>Serinus</i> spp (19 congeneric species; all A)	Sequential choice	Laboratory	0.13 (-0.18, 0.40)	15			

(Continued)

Table 4. (Continued)

Context, taxon	Species	Study	Sex	Stimulus type	Stimulus modality	Non-conspecific stimulus	Design	Setting	<i>r</i> (lower CI, upper CI)	<i>N</i> _{subjects}	Study conclusion	Comments*
	<i>Uraeginthus angolensis</i>	10	F	Live	Multi	<i>Uraeginthus bengalis</i> (?), <i>U. cyanocephalus</i> (?)	Two choice	Laboratory	0.01 (-0.34, 0.36)	6-7	No discrimination (lack of recognition system)	Poor statistical power; biological interpretation of nonsignificant results questionable.
		11	F	Live	Multi	<i>Uraeginthus bengalis</i> (A), <i>U. cyanocephalus</i> (A)	Two choice	Laboratory	0.06 (-0.36, 0.44)	6-7	No discrimination	Poor statistical power; biological interpretation of nonsignificant results questionable.
			M	Live	Multi	<i>Uraeginthus bengalis</i> (A), <i>U. cyanocephalus</i> (A)	Two choice	Laboratory	0.48 (0.07, 0.71)	4	Discrimination	Authors conclude that discrimination is a byproduct of male preferences for larger females. However, sample size is very low making any interpretation of results problematic.
Fish												
	<i>Cyprinodon pecosensis</i>	12	F	Live	Visual	<i>Cyprinodon variegatus</i> (SP)	Two choice	Laboratory	-0.29 (-0.54, 0.03)	18-20	Misidentification	Preference for an ancestral phenotype that resembles current heterospecifics (specific cue not identified).
			F	Live	Visual	<i>Cyprinodon</i> hybrid	Two choice	Laboratory	-0.13 (-0.42, 0.19)	18-20		
	<i>Gasterosteus aculeatus</i> (benthic)	13	F	Scented water	Olfactory	<i>Gasterosteus aculeatus</i> (limnetic; S)	Two choice	Laboratory	0.58 (-0.09, 0.83)	6	Discrimination	A significant conspecific preference is reported in females using a Wilcoxon sign-rank test. However, study lacks power; biological interpretation of nonsignificant result questionable.
			F	Scented water	Olfactory	<i>Gasterosteus aculeatus</i> (benthic; S)	Two choice	Laboratory	0.00 (-0.40, 0.40)	6	No discrimination	
	<i>Poecilia orri</i>	14	F	Live	Multi	<i>Poecilia mexicana</i> (A), <i>Poecilia orri</i> (A)	Two choice	Laboratory	0.19 (-0.12, 0.45)	12	Misidentification (cue similarity)	
	<i>Poecilia mexicana</i>	14	F	Live	Multi	<i>Poecilia latipinna</i> (A), <i>Poecilia orri</i> (A)	Two choice	Laboratory	0.32 (0.05, 0.54)	12	Misidentification (cue similarity)	
			M	Live	Multi	<i>Poecilia latipinna</i> (A), <i>Poecilia orri</i> (A)	Two choice	Laboratory	0.09 (-0.11, 0.28)	24		
	<i>Poecilia mexicana</i>	15	F	Live	Multi	<i>Xiphophorus hellerii</i> (S)	Two choice	Laboratory	0.11 (-0.09, 0.29)	13	Discrimination	
			F	Live	Visual	<i>Xiphophorus hellerii</i> (S)	Two choice	Laboratory	0.31 (0.04, 0.52)	13		
	<i>Poecilia mexicana</i> (cave)	15	F	Live	Multi	<i>Xiphophorus hellerii</i> (A)	Two choice	Laboratory	-0.01 (-0.26, 0.23)	15	No discrimination	Models predict that females will generally have a higher error rate leading to positive responses to unfamiliar heterospecifics
			F	Live	Visual	<i>Xiphophorus hellerii</i> (A)	Two choice	Laboratory	0.01 (-0.33, 0.34)	15		
	<i>Poecilia mexicana</i>	16	M	Live	Visual	<i>Poecilia formosa</i> (S)	Two choice	Laboratory	-0.20 (-0.39, 0.02)	40	No discrimination	Models predict indiscriminate mating by males whenever the benefit of responding to a conspecific is double that of the cost of wrongly responding to a heterospecific.

(Continued)

Table 4. (Continued)

Context, taxon	Species	Study	Sex	Stimulus type	Stimulus modality	Non-conspecific stimulus	Design	Setting	<i>r</i> (lower CI, upper CI)	<i>N</i> _{subjects}	Study conclusion	Comments*	
Insect	<i>Xiphophorus contians</i>	17	F	Scented water	Olfactory	<i>Xiphophorus montezumae</i> (A)	Two choice	Laboratory	-0.55 (-0.70, -0.52)	24	Discrimination	Authors attribute heightened responses to heterospecifics as a preference for a novel odor cue.	
	<i>Bullacris membracioides</i>	18	F	Playback	Acoustic	<i>Bullacris intermedia</i> (S/P)	No choice test	Laboratory	-0.21 (-0.38, -0.03)	18	Misidentification (cue similarity)		
	<i>Drosophila persimilis</i>		19	F	Playback	Acoustic	<i>Bullacris serrata</i> (S/P)	No choice test	Laboratory	0.40 (0.23, 0.53)	18		
			19	M	Live	Multi	<i>Drosophila pseudoobscura</i> (S)	No choice, two choice	Laboratory	0.11 (-0.03, 0.24)	25	No discrimination	Models predict indiscriminate mating by males, whenever the benefit of responding to a conspecific is double that of the cost of wrongly responding to a heterospecific.
	<i>Drosophila pseudoobscura</i>	19	M	Live	Multi	<i>Drosophila persimilis</i> (S)	No choice, two choice	Laboratory	-0.10 (-0.23, 0.04)	25	No discrimination	As for <i>D. persimilis</i>	
	<i>Photinus greeni</i>	20	M	Scented swabs	Olfactory	<i>Photinus obscurus/illus</i> (S)	Two choice	Laboratory	0.34 (-0.09, 0.63)	10	No discrimination (lack of cues)	Poor statistical power; biological interpretation of nonsignificant results questionable.	
	<i>Photinus obscurus/illus</i>		21	F	Playback	Visual	Manipulated flash interval and duration	No choice test	Laboratory	0.09 (-0.05, 0.22)	53	Misidentification (cue similarity)	
			20	M	Scented swabs	Olfactory	<i>Photinus greeni</i> (S)	Two choice	Laboratory	0.06 (-0.40, 0.48)	8	No discrimination (lack of cues)	As for male <i>P. greeni</i>
	<i>Elychnia connexa</i>	20	M	Scented swabs	Olfactory	<i>Lucidota atra</i> (?)	Two choice	Laboratory	0.32 (-0.14, 0.63)	9	Discrimination	A significant conspecific preference is reported using a Wilcoxon sign-rank test. However, study lacks power; biological interpretation of results questionable.	
	<i>Lucidota atra</i>	20	M	Scented swabs	Olfactory	<i>Elychnia connexa</i> (?)	Two choice	Laboratory	-0.02 (-0.46, 0.42)	6	No discrimination	Poor statistical power; biological interpretation of nonsignificant results questionable.	
	<i>Leptidea sinapis</i>		22	F	Live	Multi	<i>Leptidea reali</i> (S)	No choice test	Laboratory	0.72 (0.23, 0.88)	55	Discrimination	
				M	Live	Multi	<i>Leptidea reali</i> (S)	No choice test	Laboratory	-0.13 (-0.30, 0.06)	15-21	Misidentification (cue similarity)	Poor statistical power for field component of study.
	<i>Pieris rapae crucivora</i>	23	M	Live, with surgery	Multi	<i>Pieris melete</i> (S)	Observation	No choice test	Field	-0.27 (-0.68, 0.35)	7	Misidentification	Poor statistical power; biological interpretation of nonsignificant results questionable.
	<i>Prokelisia marginata</i>		24	F	Playback	Acoustic	<i>Prokelisia dolus</i> (S)	Sequential choice	Laboratory	0.12 (-0.20, 0.41)	18	Discrimination	Authors conclude female discrimination based on a behavioral response not quantified.
			25	F	Dead	Olfactory	<i>Zaprius tuberculatus</i> (S)	No choice test	Laboratory	0.18 (-0.02, 0.36)	8-13	Discrimination	Discrimination is reported under specific circumstances. Authors highlight a problem with the quality of odor cues used as stimuli that may have reduced treatment effects.

(Continued)

Table 4. (Continued)

Context, taxon	Species	Study	Sex	Stimulus type	Stimulus modality	Non-conspecific stimulus	Design	Setting	r (lower CI, upper CI)	N_{subjects}	Study conclusion	Comments*
Lizard	<i>Zapriionus tuberculatus</i>	M	Dead	Olfactory	<i>Zapriionus tuberculatus</i> (S)	No choice test	Laboratory	0.27 (-0.14, 0.57)	10–13			
		F	Dead	Olfactory	<i>Zapriionus sepsoides</i> (S)	No choice test	Laboratory	-0.08 (-0.39, 0.26)	8	Discrimination	As for <i>Z. sepsoides</i> .	
		M	Dead	Olfactory	<i>Zapriionus sepsoides</i> (S)	No choice test	Laboratory	0.13 (-0.11, 0.34)	10–14			
Lizard	<i>Hemidactylus frenatus</i>	M	Live	Multi	<i>Hemidactylus garnotii</i> (S)	No choice, two choice	Laboratory	-0.55 (-0.73, -0.24)	14	Misidentification (lack of cues or recognition system)	Species have only recently come into contact and identity cues or recognition systems may have yet to evolve. Alternatively, models predict indiscriminate mating by males whenever the benefit of responding to a conspecific is double that of the cost of wrongly responding to a heterospecific.	
		F	Scented substrate	Olfactory	<i>Podarcis hispanica</i> (S/P)	No choice test	Laboratory	-0.11 (-0.47, 0.29)	11	In conclusive	Inappropriate behavioral assay and low sample size.	
Territorial Bird	<i>Circus pygargus</i>	M	Dummy	Visual	<i>Circus cyaneus</i> (S)	No 'choice'	Field	0.23 (-0.32, 0.63)	6	Discrimination	A significantly higher attack rate to conspecifics was reported based on a Wilcoxon sign-rank test. However, study lacks power; biological interpretation of results questionable. Species are inferred not to be in competition for ecological resources.	
		M	Playback	Acoustic	<i>Melospiza melodia</i> (S)	Sequential 'choice'	Field	0.32 (-0.15, 0.64)	9	Discrimination	Significantly closer approaches to conspecific song were reported based on a Wilcoxon sign-rank test. However, study lacks power; biological interpretation of results questionable. Species do not seem to compete ecologically.	
		M	Playback with dummy	Multi	<i>Melospiza melodia</i> (S) <i>Sylvia borin</i> (S)	No 'choice' Two 'choice'	Field	0.73 (0.49, 0.84) -0.30 (-0.50, -0.04)	12 31–39	Discrimination Discrimination	Authors were able to demonstrate discrimination occurred, and that subjects were aggressive to heterospecifics. Species were assumed to compete for ecological resources.	

(Continued)

Table 4. (Continued)

Context, taxon	Species	Study	Sex	Stimulus type	Stimulus modality	Stimulus	Non-conspecific stimulus	Design	Setting	r (lower CI, upper CI)	N _{subjects}	Study conclusion	Comments*
Lizard	<i>Hemidactylus frenatus</i>	26	F	Live	Multi	<i>Hemidactylus garnotii</i> (S)		No 'choice'	Laboratory	0.25 (-0.13, 0.54)	12-14	No discrimination	Species have only recently come into contact and identity cues or recognition systems may have yet to evolve. Alternatively, models predict males to be less discriminating in aggressive responses to allopatric (unfamiliar) heterospecifics. This is especially true when identity cues overlap or are error prone, and the benefits of correctly responding to an intruder greatly outweigh the cost of failing to respond to an intruder. Species may compete for ecological resources, although the authors do not state this explicitly.
Mammal	<i>Lemur fulvus</i>	32	F/M	Scent marks	Olfactory	<i>Lemur macaco</i> (?)		No 'choice'	Captive	-0.04 (-0.33, 0.26)	6	Discrimination probable	Conclusion is based on observations that subjects responded more to unfamiliar conspecifics than familiar conspecifics or heterospecifics. However, study lacks power; biological interpretation of results questionable.
	<i>Lemur macaco</i>	32	F/M	Scent marks	Olfactory	<i>Lemur fulvus</i> (?)		No 'choice'	Captive	0.03 (-0.33, 0.38)	4	Discrimination probable	As for <i>L. fulvus</i>

Study: 1. Cheong et al. (2008); 2. Marshall et al. (2006); 3. Tarano and Ryan (2002); 4. Dawley (1987); 5. Rollmann et al. (2003); 6. Pfennig (2000); 7. Nagle et al. (2002); 8. Parisot et al. (2002); 9. Cardoso et al. (2007); 10. Collins and Luddem (2002); 11. Luddem et al. (2004); 12. Rosenfield and Kodric-Brown (2003); 13. Rafferty and Boughman (2006); 14. Ptacek (1998); 15. Riesch et al. (2006); 16. Plath et al. (2008); 17. McLennan and Ryan (2008); 18. Coullidge and van Staaden (2006); 19. Noor (1996); 20. South et al. (2008); 21. Michaelidis et al. (2006); 22. Friberg et al. (2008); 23. Ohguchi and Hidaka (1988); 24. Heady and Denno (1991); 25. Lee (1983); 26. Dame and Petren (2006); 27. Barbosa et al. (2006); 28. Garcia (2003); 29. Peters et al. (1980); 30. Nowicki et al. (2001); 31. Matyjasiak (2005); 32. Fornasieri and Roeder (1992). Sex: F = females; M = males. Stimulus modality: Multi = multi-modal. Nonconspecific stimulus: A = allopatric; S = sympatric.

*Model outcomes are shown in Figure 1.

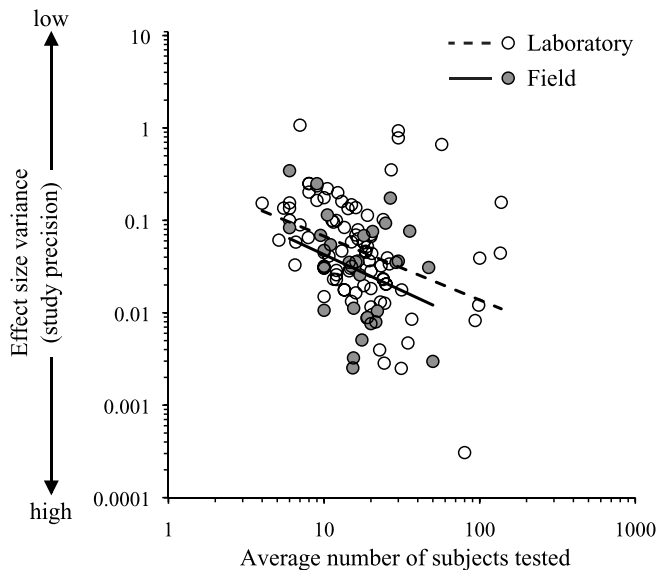


Figure 3. The statistical power, or precision, of species recognition studies as a function of experimental setting (laboratory vs. field studies) and the number of animals tested. Each point represents the variance associated with the combined effect size and average sample size of all experiments conducted by a single study.

respond appropriately to a conspecific might lead to a situation where the optimal strategy to adopt is one of responding to any animal encountered in the environment (Fig. 1; see also introduction). There were also a number of reports where identity cues assessed during mating overlapped enough between species that misidentification occurred (these species listed in Table 4 will likely be of special interest to speciation biologists). Yet the number of studies that seemed to suffer methodological and statistical problems is cause for concern (Table 4).

It might be argued that the heterogeneity in the data and the extent it reflects true biological variation among species makes it difficult to identify common factors that govern when animals will or will not discriminate against heterospecifics. We note, however, that the majority of the studies we surveyed were well designed and had good statistical power. Our results were also consistent when studies suffering methodological problems were excluded from analyses (Table S3). Furthermore, the empirical data were broadly consistent with the initial predictions that we formulated using Reeve's models, once the specific circumstances of discrimination for a given species were identified (Table 4). For example, the magnitude of species discrimination was consistent between mating and territorial contexts. Although females did not seem to be the primary sex discriminating against nonconspecifics, as was generally implicated by our parameterization of Reeve's models, poor discrimination was expected when identity cues overlapped extensively among species (Fig. 1). Misidentification resulting from similarities (overlap) between mating signals did explain

why females in several species failed to discriminate against heterospecifics (Table 4).

Perhaps it is not too surprising that the primary finding of our meta-analysis is that species discrimination is likely to be highly context specific and the product of a complex interaction of a range of competing factors. Unfortunately, few species have been tested for all the predictor variables of interest and our analyses were subsequently restricted to largely univariate approaches. The application of multivariate models that include interaction terms will become possible as more data become available (e.g., to examine whether the sexes differ in response to heterospecifics depending on the modality of cues being assessed, the phylogenetic relatedness of the heterospecific, and frequency of interaction with the heterospecific). A clearer picture might then emerge on the precise combination of factors that underlies the extensive variation among species in species discrimination that we document. We recommend that future studies include tests of free-ranging animals in natural settings as these experiments generally obtained results of higher precision. Captive experiments offer a degree of control not possible in experiments conducted in the field, but our results imply this control comes at a cost to the biological relevance of stimulus presentations compared to similar experiments done in natural settings (Table 3; Fig. 3), presumably because conspecific and heterospecific cues could be presented to animals in a more relevant biological context than experiments conducted in captivity. The link between sample size and statistical power will be familiar to readers, as will the practical constraints of obtaining large sample sizes for some taxa relative to others. We hope that the data compiled in the supporting information will assist future researchers in conducting a priori power analyses to determine the most appropriate sample size for a given taxa, in a given context.

ACKNOWLEDGMENTS

We thank M. Borenstein, L. Hedges, W. Shadish, and D. Wilson for advice on meta-analytic techniques, H. B. Shaffer for clarification on the phylogenetic positioning of turtles and H. Muller for help resolving relationships among the *Plethodon* salamanders, and J. Losos, K. Peiman, J. Stamps, E. Svensson, and three anonymous reviewers for comments on a previous version of the manuscript. This work was supported by a National Science Foundation grant (IOB-0517041/0516998) to TJO, Judy A Stamps and Jonathan B Losos.

LITERATURE CITED

- Borenstein, M., L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein. 2009. Introduction to meta-analysis. Wiley, Chichester, UK.
- Bradbury, J. W., and S. L. Vehrencamp. 1998. Principles of animal communication. Sinauer, Sunderland.
- Cannatella, D. C., D. M. Hillis, P. T. Chippindale, L. Weight, A. S. Rand, and M. J. Ryan. 1998. Phylogeny of frogs of the *Physalaemus pustulosus* species group with an examination of data incongruence. *Syst. Biol.* 47:311–335.

- Coddington, J. A. 1991. Cladistics and spider classification: araneomorph phylogeny and the monophyly of orb weavers (Araneae: Araneomorphae; Orbiculariae). *Acta Zool. Fennica* 190:75–87.
- Dusenbery, D. B. 1992. *Sensory ecology: how organisms acquire and respond to information*. Freeman, New York.
- Fornasieri, I., and J.-J. Roeder. 1992. Behavioral responses to own and other species' scent marks in *Lemur fulvus* and *Lemur macaco*. *J. Chem. Ecol.* 18:2069–2082.
- Gowaty, P. A., and S. P. Hubbell. 2005. Chance, time allocation, and the evolution of adaptively flexible sex role behavior. *Integr. Comp. Biol.* 944:931–944.
- Grether, G. F., N. Losin, C. N. Anderson, and K. Okamoto. 2009. The role of interspecific inference competition in character displacement and the evolution of competitor recognition. *Biol. Rev.* 84:617–635.
- Gurevitch, J., and L. V. Hedges. 1993. Meta-analysis: combining results of independent experiments. Pp. 378–398 in S. M. Scheiner and J. Gurevitch, eds. *Design and analysis of ecological experiments*. Chapman & Hall, New York.
- Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* 62:1965–1977.
- Hebets, E. A., and D. R. Papaj. 2005. Complex signal function: developing a framework of testable hypotheses. *Behav. Ecol. Sociobiol.* 57:197–214.
- Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *Am. Nat.* 163:84–96.
- Lajeunesse, M. J. 2009. Meta-analysis and the comparative phylogenetic method. *Am. Nat.* 174:369–381.
- Lynch, M. 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45:1065–1080.
- Maddison, W. P., and D. R. Maddison. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.72. Available free from <http://mesquiteproject.org> (accessed January 3, 2009).
- Marcus, J. M., and A. R. McCune. 1999. Ontogeny and phylogeny in the Northern swordtail clade of *Xiphophorus*. *Syst. Biol.* 48:491–522.
- Martins, E. P. 2004. COMPARE 4.6b: statistical analysis of comparative data. Available free from <http://compare.bio.indiana.edu/> (accessed January 6, 2006).
- Ord, T. J., and J. A. Stamps. 2009. Species identity cues in animal communication. *Am. Nat.* 174:585–593.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.
- Peiman, K. S., and B. W. Robinson. 2010. Ecology and evolution of resource-related heterospecific aggression. *Q. Rev. Biol.* 85:133–158.
- Pfennig, K. S. 2007. Facultative mate choice drives adaptive hybridization. *Science* 318:965–967.
- Ranwez, V., V. Berry, A. Criscuolo, P.-H. Fabre, S. Guillemot, C. Scornavacca, and E. J. P. Douzery. 2007. PhySIC: a veto supertree method with desirable properties. *Syst. Biol.* 56:798–817.
- Reeve, H. K. 1989. The evolution of conspecific acceptance thresholds. *Am. Nat.* 133:407–435.
- Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* 57:591–601.
- Ritchie, M. G. 2007. Sexual selection and speciation. *Ann. Rev. Ecol. Evol. Syst.* 38:79–102.
- Schluter, D. 1989. Bridging population and phylogenetic approaches to the evolution of complex traits. Pp. 79–95 in D. B. Wake and G. Roth, eds. *Complex organismal functions: integration and evolution in vertebrates*. Wiley, New York.
- Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. *Evolution* 64:295–315.
- Streelman, J. T., and P. D. Danley. 2003. The stages of vertebrate evolutionary radiation. *Trends Ecol. Evol.* 18:126–131.
- Svensson, E. I., F. Eroukhmanoff, K. Karlsson, A. Runemark, and A. Brodin. 2010. A role for learning in population divergence of mate preferences. *Evolution* 64:3101–3113.
- Viechtbauer, W. 2009. Meta-analysis package for R. Version 0.5–4. Available free from <http://www.wvbauer.com/> (accessed September 15, 2009).
- Wellenreuther, M., K. Tynknyen, and E. I. Svensson. 2010. Simulating range expansion: male species recognition and loss of premating isolation in damselflies. *Evolution* 64:242–252.
- Wiley, R. H. 2006. Signal detection and animal communication. *Adv. Stud. Behav.* 36:217–247.
- Zhu, X. Y., M. W. Chase, Y. L. Qiu, H. Z. Kong, D. L. Dilcher, J. H. Li, and Z. D. Chen. 2007. Mitochondrial matR Sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evol. Biol.* 7: 217.

Associate Editor: E. Svensson

Supporting Information

The following supporting information is available for this article:

Figure S1. The phylogeny of species recognition.

Figure S2. Species discrimination as a function of social context (mating vs. territorial interactions; A) and the sex of the discriminating animal (B).

Figure S3. Species discrimination as a function of the familiarity with the nonconspecific cue(s) (familiar, “sympatric” heterospecific versus unfamiliar, “allopatric” heterospecific; A), and the type of cues that provide potential information on species identity (B).

Table S1. Predictors of species recognition based on the maximum response reported by a study for a given species.

Table S2. Species recognition as a function of the phylogenetic relatedness between the focal species and heterospecific.

Table S3. Predictors of species recognition based on analyses in which data were excluded from those studies that suffered methodological problems (Table 4).

Table S4. Source data.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.