# Using song playback experiments to measure species recognition between geographically isolated populations: A comparison with acoustic trait analyses

Authors: Freeman, Benjamin G., and Montgomery, Graham A.

Source: The Auk, 134(4) : 857-870

Published By: American Ornithological Society

URL: https://doi.org/10.1642/AUK-17-63.1

The Auk
Ornithological Advances

AmericanOrnithology.org

Volume 134, 2017, pp. 857–870
DOI: 10.1642/AUK-17-63.1

*RESEARCH ARTICLE*

# Using song playback experiments to measure species recognition between geographically isolated populations: A comparison with acoustic trait analyses

**Benjamin G. Freeman[1,2]* and Graham A. Montgomery[3]**

[1] Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia, Canada
[2] Department of Zoology, University of British Columbia, Vancouver, Canada
[3] Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA
* Corresponding author: freeman@zoology.ubc.ca

Submitted April 7, 2017; Accepted June 20, 2017; Published September 13, 2017

## ABSTRACT

Geographically isolated populations of birds often differ in song. Because birds often choose mates on the basis of their song, song differentiation between isolated populations constitutes a behavioral barrier to reproduction. If this barrier is judged to be sufficiently strong, then isolated populations with divergent songs may merit classification as distinct species under the biological species concept. We used a dataset of 72 pairs of related but allopatric Neotropical passerines ("taxon pairs") to compare 2 methods for measuring song divergence between isolated populations: statistical analysis of 7 acoustic traits measured from spectrograms, and field playback experiments that "ask the birds themselves" if they perceive foreign song as conspecific or not. We report 4 main findings: (1) Behavioral discrimination (defined as failure to approach the speaker in response to allopatric song) is nonlinearly related to divergence in acoustic traits; discrimination is variable at low to moderate levels of acoustic divergence, but nearly uniformly high at high levels. (2) The same nonlinear relationship held for both song learners (oscines) and nonlearners (suboscines). (3) Song discrimination is not greater in taxon pairs ranked as species compared to taxon pairs ranked as subspecies. (4) Behavioral responses to allopatric song are symmetric within a taxon pair. We conclude (1) that playback experiments provide a stronger measure of species recognition relevant to premating reproductive isolation than do acoustic trait analyses, at least when divergence in acoustic traits is low to moderate; and (2) that playback experiments are useful for defining species limits and can help address the latitudinal gradient in taxonomy, which arises because species are defined more broadly in the tropics than in the temperate zone. To this end, we suggest that 21 Neotropical taxon pairs that are currently ranked as subspecies, but that show strong behavioral discrimination in response to allopatric song, merit classification as distinct biological species.

*Keywords:* allopatric speciation, biological species concept, premating reproductive isolation, signal evolution, speciation, species recognition, suboscines, tropical birds

**Usando experimentos de reproducción de cantos para medir el reconocimiento de especies entre poblaciones geográficamente aisladas: una comparación con análisis de rasgos acústicos**

## RESUMEN

Las poblaciones de aves geográficamente asiladas usualmente difieren en el canto. Debido a que las aves usualmente eligen sus parejas en base a sus cantos, la diferenciación de los cantos entre poblaciones aisladas constituye una barrera comportamental a la reproducción. Si se estima que esta barrera es suficientemente fuerte, entonces las poblaciones aisladas con cantos divergentes podrían merecer ser clasificadas como especies distintivas bajo el concepto biológico de especie. Usamos una base de datos de 72 parejas de paserinos neotropicales relacionados pero alopátricos ("pares de taxones") para comparar dos métodos para medir la divergencia del canto entre poblaciones aisladas—análisis estadístico de siete rasgos acústicos medidos a partir de espectrogramas, y experimentos de campo de reproducción de cantos que "le preguntan a las propias aves" si perciben los cantos foráneos como de la misma especie o no. Presentamos en este trabajo cuatro hallazgos principales: (1) La discriminación comportamental (definida como el fracaso de acercarse al parlante en respuesta a un canto alopátrico) no está relacionada linealmente a la divergencia en los rasgos acústicos; la discriminación es variable a niveles bajos a moderados de divergencia acústica, pero casi uniformemente alta a niveles altos; (2) Se mantiene la misma relación no lineal tanto para aprendices de canto (oscines) como para no aprendices (suboscines); (3) La discriminación del canto no es mayor en los pares de taxones clasificados como especies, comparada con los pares de taxones clasificados como subespecies, y 4) Las respuestas comportamentales al canto alopátrico son simétricas adentro del par de taxones. Concluimos que: (1) los experimentos de reproducción de cantos brindan una medida más fuerte de reconocimiento de especies, relevante

para el aislamiento reproductivo previo al apareamiento, que los análisis de rasgos acústicos, al menos cuando la divergencia en los rasgos acústicos es baja a moderada, y (2) los experimentos de reproducción de cantos son útiles para definir los límites de las especies y pueden ayudar en la consideración del gradiente latitudinal en la taxonomía, que surge debido a que las especies están definidas más ampliamente en los trópicos que en la zona templada. Con este fin, sugerimos que 21 pares de taxones neotropicales actualmente clasificados como subespecies, pero que muestran una fuerte discriminación comportamental al canto alopátrico, merecen ser clasificados como especies biológicamente distintas.

*Palabras clave:* aislamiento reproductivo previo al apareamiento, aves tropicales, concepto biológico de especie, especiación, especiación alopátrica, evolución de la señal, reconocimiento de especies, suboscines

## INTRODUCTION

Percy Bysshe Shelley wrote in 1821 that the nightingale "sits in darkness and sings to cheer its own solitude with sweet sounds." However, unlike English romantic poets, ornithologists know that nightingales sing not to stave off despair, but for sex—a primary reason birds sing is to attract and maintain mates (Edwards et al. 2005, Price 2008). As a consequence, song is often an important barrier to reproduction. Populations that diverge in song may fail to recognize each other as conspecific, which leads to assortative mating and, potentially, speciation (Irwin and Price 1999, Slabbekoorn and Smith 2002, Uy et al. 2009, Tobias et al. 2010a, McEntee et al. 2016).

Most avian taxonomists follow the biological species concept (BSC), which defines species as populations that are reproductively isolated from one another (Price 2008). In this context, "reproductive isolation" is defined as assortative mating (i.e. nonrandom mate selection); thus, biological species may occasionally hybridize with each other. Because differences in song can constitute a premating barrier to reproduction, taxonomists often measure vocal divergence when deciding whether related but geographically isolated (allopatric) populations should be classified as distinct biological species (Payne 1986, Isler et al. 1998, Alström and Ranfft 2003, Remsen 2005, Rheindt et al. 2008, Tobias et al. 2010b). In general, species rank is supported when allopatric populations differ in song to a similar or greater extent than do co-occurring (sympatric) populations within the same lineage (Isler et al. 1998). Differences in vocalizations are thought to be particularly taxonomically informative in lineages in which song development is innate (e.g., the suboscine passerines), because vocal divergence reflects genetic divergence in these groups (Touchton et al. 2014). For example, studies of vocal differences among tapaculos in the genus *Scytalopus* (suboscines in the family Rhinocryptidae) have more than tripled the number of recognized species in this genus in the past quarter-century, from 13 species in 1990 (Sibley and Monroe) to 42 species in 2017 (Remsen et al. 2017). By contrast, cultural evolution alone may drive song divergence in song-learning clades, such as the oscine passerines. Nevertheless, because song learners have genetic predispositions to learn species-specific song and often discriminate behaviorally against foreign song (Peters et al. 1980, Baker and Baker 1990, Soha and Marler 2000), vocalizations are also used to determine species limits among allopatric populations of oscine passerines (e.g., Cadena and Cuervo 2010).

There are two principal methods to quantify whether song differences between allopatric populations are sufficient that song would likely be a barrier to reproduction if populations were to come into contact. In the first method, researchers measure a suite of acoustic traits of songs (e.g., mean frequency, number of notes) from spectrograms of audio recordings, then quantitatively compare acoustic traits between populations (e.g., Isler et al. 1998, Chaves et al. 2010). This approach is increasingly popular, facilitated by software that permits detailed acoustic analyses (Sueur et al. 2008, Bioacoustics Research Program 2014, Araya-Salas and Smith-Vidaurre 2017) and by the remarkable growth of audio recordings of birdsong archived in publicly available collections (e.g., Macaulay Library at the Cornell Lab of Ornithology and xeno-canto.org). However, a potential drawback of acoustic trait analyses is that the vocal cues birds use when making mating decisions may not be the same characteristics that researchers measure from audio recordings (i.e. statistical acoustic differences between populations may not be biologically relevant; Nelson 1998, Soha et al. 2016).

In the second method, researchers use field playback experiments to measure how individuals respond behaviorally to playback of song from an allopatric population (Lanyon 1978). Playback experiments "ask the birds themselves" if they perceive the song of a related, allopatric population as conspecific or not, and so they may be a better proxy for reproductive isolation based on song. However, playback experiments require fieldwork that can be logistically difficult, and interpreting behavioral responses to playback requires its own set of assumptions (see below). It is thus desirable to know how the conclusions of statistical acoustic trait analyses compare to the conclusions of field playback experiments. For example, if the results of acoustic trait analyses and playback experiments are tightly correlated, then the 2 methods provide interchangeable inference, and the effort expended to undertake field playback experiments would

be unnecessary. It is currently unknown how the results of acoustic trait analyses and playback experiments compare.

We address this data gap by quantifying the relationship between the results of acoustic trait analyses and of playback experiments for 72 taxon pairs of allopatric Neotropical passerines. These taxon pairs were chosen because they are closely related (often sister taxa) but geographically isolated populations; in some cases, whether the taxa should be classified as separate species or subspecies of the same species is controversial. This dataset includes both song learners (oscines; $n = 43$) and species with innate song (suboscines; $n = 29$) and contains populations currently classified as distinct species ($n = 27$; 16 oscines and 11 suboscines) or as subspecies ($n = 45$; 27 oscines and 18 suboscines; taxonomy follows Chesser et al. 2017, Remsen et al. 2017). For each taxon pair, we (1) calculated acoustic divergence by measuring a set of acoustic traits commonly used in comparative analyses of birdsong (e.g., Mason et al. 2017) and (2) conducted playback experiments to measure behavioral song discrimination toward allopatric song. We then examined the correlation between these 2 metrics to test the correspondence between results of acoustic trait analyses and results of playback experiments. We use our results to argue that playback experiments are a useful method for estimating premating reproductive isolation, and we suggest that taxon pairs that are currently ranked as subspecies but that largely ignore song from allopatric populations merit recognition as distinct biological species. Although it is widely recognized that temperate zone species are more narrowly defined than tropical species (the latitudinal gradient in taxonomy), efforts to address this bias remain controversial (Gill 2014, Remsen 2015, Toews 2015, Collar et al. 2016). We suggest that using playback experiments to measure species recognition—an approach that has been termed "behavioral systematics" (Pegan et al. 2015)—holds promise as a broadly applicable method to flatten the latitudinal gradient in taxonomy.

## METHODS

We have conducted playback experiments on >150 geographically isolated taxon pairs of Neotropical passerines as part of a larger project studying the evolutionary tempo of song evolution in geographic isolation in Neotropical birds. Playback experiments took place in Costa Rica, Panama, and Ecuador. We have measured acoustic traits for 72 of these taxon pairs (see below). Here, we (1) analyze the 72 taxon pairs for which we have matched playback experiments and acoustic trait analyses and (2) consider the taxonomic implications of our results for our broader dataset of taxon pairs (including those for which we have conducted playback experiments only). We focus on allopatric populations because speciation in birds is initiated in geographic isolation (Barraclough and Vogler 2000, Coyne and Price 2000), and several million years typically elapse before range expansions bring related populations into secondary contact (Weir and Price 2011). We note that some taxon pairs in our dataset are largely allopatric but do have contact zones; in all cases, playback experiments occurred far from contact zones. Only data for taxon pairs that do not have contact zones are relevant to species limits.

### Playback Experiments

Our playback experiments simulated secondary contact between geographically isolated populations and thus provide insight into the degree to which signal divergence in allopatry affects species recognition. For the 72 taxon pairs that are the main focus of the present study, we conducted a mean (± SD) of 12.1 ± 2.8 playback experiments population$^{-1}$ (range: 5–21; see Supplemental Material Table S1). Sample sizes of the territories tested are similar for taxon pairs for which we have conducted playback experiments but for which we have not measured acoustic traits (Tables 1–3). Playback experiments followed a standard methodology (McGregor et al. 1992; see also Pegan et al. 2015). Briefly, each experiment measured the behavioral response of a territorial bird to 2 treatments: (1) song from the local population (sympatric treatment) and (2) song from the allopatric population (allopatric treatment). We alternated treatment order between territories; a previous analysis showed that treatment sequence does not influence behavioral response to allopatric song in this dataset (B. G. Freeman et al. personal observation). We used large banks of natural vocalizations in experiments, archived at xeno-canto.org and the Macaulay Library (Cornell Lab of Ornithology), to maximize independence of replicates ($n = 6.6 \pm 1.7$ recordings treatment$^{-1}$), and we used a single recording in each treatment. We used recordings made within the geographic distribution of the population in question; sympatric recordings came from the vicinity of the region where we conducted playback experiments, and allopatric recordings came from the most geographically proximate portion of the allopatic populations' distribution. For example, for comparisons of Andean taxa found on either side of the Marañon Gap, we used recordings made near this biogeographic barrier (i.e. in southeast Ecuador or northern Peru) and did not use recordings made far from the barrier (even if they are classified as the same subspecies). We used good-quality recordings in playback experiments (most recordings were graded "A" on xeno-canto, or rated ≥3 stars on Macaulay Library) and did not normalize amplitudes.

Each treatment consisted of placing a wireless speaker (UE Roll or JBL Charge 2+) within a territory, broadcasting song at natural amplitudes (~80 dB at 1 m from the

**TABLE 1.** Behavioral song discrimination in 9 allopatric taxon pairs that have recently been split, in part on the basis of divergent vocalizations. Song discrimination is the proportion of territories of the first-listed species that ignored song from the second species (sample size of territories tested is given in parentheses). These cases provide a yardstick for how much song discrimination is "enough" to merit classifying allopatric populations as distinct biological species when using our methodology. We include in this list 2 taxon pairs (*Zimmerius* and *Cyanocompsa*) for which proposals to define populations as distinct biological species are currently under consideration by the South American Classification Committee of the American Ornithological Society.

| Species pair | | | |
|---|---|---|---|
| Common names | Scientific names | Clade | Song discrimination |
| Dull-mantled Antbird–Magdalena Antbird | *Myrmeciza laemosticta–M. palliata* | Suboscine | 0.71 (14) |
| Zeledon's Antbird–Blue-lored Antbird | *Hafferia zeledoni–H. immaculata* | Suboscine | 0.91 (11) |
| Streak-headed Antbird–East Andean Antbird | *Drymophila striaticeps–D. caudata* | Suboscine | 0.3 (10) |
| Blackish Antbird–Riparian Antbird | *Cercamacroides nigrescens–C. fuscicauda* | Suboscine | 0.3 (10) |
| Long-tailed Tapaculo–Rufous-vented Tapaculo | *Scytalopus micropterus–S. femoralis* | Suboscine | 0.2 (10) |
| Paltry Tyrannulet (Central America)–Paltry Tyrannulet (Venezuelan Andes) | *Zimmerius vilissimus–Z. improbus* | Suboscine | 0.73 (15) |
| Northern Schiffornis–Russet-winged Schiffornis | *Schiffornis veraepacis–S. stenorhyncha* | Suboscine | 1 (5) |
| Costa Rican Warbler–Three-striped Warbler | *Basileuterus melanotis–B. tristriatus* | Oscine | 0.56 (9) |
| Blue-black Grosbeak (west of the Andes)–Blue-black Grosbeak (east of the Andes) | *Cyanocompsa cyanoides–C. rothschildii* | Oscine | 0.71 (14) |

speaker) for 2 min, and observing behavioral responses during both the 2 min of playback and a subsequent 5 min observation period. We recorded multiple behavioral responses to playback for each treatment. Here, we focus only on closest approach to the speaker (in meters), a reliable indicator of behavioral response to playback (Martin and Martin 2001, Jankowski et al. 2010, Freeman and Montgomery 2016, Freeman et al. 2016). At the beginning of each treatment, territory owners were within hearing distance and out of sight (i.e. >15 m distant from the speaker) or, more uncommonly, visible >15 m distant from the speaker. If a bird was still responding to playback at the conclusion of the first treatment—for example, if the territorial bird(s) remained within 15 m of the speaker or continued to vocalize at an elevated rate—then we waited until 2 min after it stopped responding (i.e. moved >15 m away but stayed within hearing distance, or ceased vocalizing at an elevated rate) before initiating the second treatment.

In our experimental design, the sympatric treatment served as a positive control. That is, we expected territorial birds to respond aggressively to sympatric song playback and included only experiments in which this was indeed the case (i.e. birds approached to within 15 m of the speaker—typically to within 5 m—in response to sympatric playback). Our aim was to use behavioral response to playback experiments as a proxy for premating reproductive isolation based on song. Thus, we define "song discrimination" here as instances in which the territory owner(s) ignored allopatric song, defined as a failure to approach within 15 m of the speaker in response to the allopatric treatment (i.e. we distinguished between "response" and "failure to respond," but not between "weak" and "strong" responses). We calculated song discrimination for each taxon pair as the percentage of territories that failed to approach the speaker in response to allopatric song. For example, a song discrimination score of 0.8 indicates that 80% of territorial birds (e.g., 8 of 10) ignored allopatric song while simultaneously actively defending a territory (as described above, all territories responded to sympatric song by approaching the speaker).

In our experiments, we played songs of populations A and B (where populations A and B comprise a taxon pair) to territorial birds of population A. Most taxon pairs were tested in only one direction. That is, in the majority of cases we asked whether population A discriminated against song from population B but not the reverse. To date, we have conducted reciprocal playback experiments for 23 taxon pairs (13 oscines and 10 suboscines; see Supplemental Material Table S2) in which we measured both discrimination of population A to song from population B and also discrimination of population B to song from population A, and in which we conducted at least 4 experiments on each population ($n = 11.5 \pm 3.8$ playback experiments population$^{-1}$; range: 4–23; see Supplemental Material Table S2). Song discrimination in these reciprocal cases was highly correlated ($r = 0.88$, $t = 8.4$, df $= 21$, $P < 0.0001$; Figure 1), and we therefore assume that unidirectional data accurately describe song discrimination within taxon pairs in our database.

Birdsong functions in both mate choice and territorial defense. Our playback experiments measure territorial defense but are likely a conservative proxy for inferring the strength of a behavioral barrier to reproduction. This

**TABLE 2.** Twenty-one taxon pairs currently classified as subspecies that show strong behavioral discrimination against allopatric song. We suggest that these populations merit recognition as distinct biological species. Song discrimination is the proportion of territories of the first-listed subspecies that ignored song from the second, allopatric subspecies (sample size of territories tested is given in parentheses). We conducted reciprocal playback experiments for 5 pairs; in each of these cases, song discrimination was high in both directions (population A typically ignored song from population B, and population B typically ignored song from population A).

| Taxon pair | | | | | |
|---|---|---|---|---|---|
| Species | Population A subspecies | Population B subspecies | Clade | Song discrimination | Reciprocal song discrimination |
| Buffy Tuftedcheek (*Pseudocolaptes lawrencii*) | *lawrencii* (Central America) | *johnsoni* (NW Andes) | Suboscine | 0.9 (10) | – |
| Striped Woodhaunter (*Automolus subulatus*) | *virgatus* (W of Andes) | *subulatus/lemae* (Amazon basin) | Suboscine | 0.92 (12) | – |
| Plain Xenops (*Xenops minutus*) | *littoralis* (W Ecuador) | *obsoletus* (W Amazon basin) | Suboscine | 0.75 (8) | – |
| Buff-throated Foliage-gleaner (*Automolus ochrolaemus*) | *exsertus* (Pacific slope Central America) | *hypophaeus* (Caribbean slope Central America) | Suboscine | 0.87 (15) | 0.79 (14) |
| Olivaceous Woodcreeper (*Sittasomus griseicapillus*) | *aequatorialis* (W Ecuador, NW Peru) | *amazonus* (W Amazon basin) | Suboscine | 1 (15) | 1 (2) |
| Wedge-billed Woodcreeper (*Glyphorynchus spirurus*) | *sublestus* (W Ecuador, W Colombia) | *castelnaudii* (W Amazon basin) | Suboscine | 0.78 (9) | – |
| Dusky Antbird (*Cercomacroides tyrannina*) | *crepera* (Central America) | *saturatior* (E Venezuela, Guiana Shield) | Suboscine | 0.67 (18) | – |
| Tawny Antpitta (*Grallaria quitensis*) | *quitensis* (Northern Andes) | *alticola* (E Andes Colombia) | Suboscine | 0.67 (9) | – |
| White-crowned Tapaculo (*Scytalopus atratus*) | *atratus* (E slope Andes, N of Marañon Gap) | *atratus* (E slope Andes, S of Marañon Gap) | Suboscine | 0.86 (6) | – |
| Blackish Tapaculo (*Scytalopus latrans*) | *latrans* (E slope Andes, N of Marañon Gap) | *intermedius* (N Peru, S of Marañon Gap) | Suboscine | 1 (8) | – |
| Yellow-olive Flycatcher (*Tolmomyias sulphurescens*) | *aequatorialis* (W Ecuador, NW Peru) | *peruvianus* (foothills E Ecaudor, E Peru) | Suboscine | 0.64 (14) | 1 (3) |
| Slaty-capped Flycatcher (*Leptopogon superciliaris*) | *transandinus* (W Ecuador, W Colombia) | *superciliaris* (E Ecuador) | Suboscine | 0.77 (13) | 0.8 (5) |
| Slaty-backed Chat-Tyrant (*Ochthoeca cinnamomeiventris*) | *cinnamomeiventris* (Andes N of Marañon Gap) | *thoracica* (Andes, primarily S of Marañon Gap) | Suboscine | 1 (5) | – |
| Andean Solitaire (*Myadestes ralloides*) | *venezuelens*is (E slope Andes, N of Marañon Gap) | *ralloides* (Andes S of Marañon Gap) | Oscine | 0.83 (12) | – |
| Plain-tailed Wren (*Pheugopedius euophrys*) | *longipes* (S Ecuador, N Peru, N of Marañon Gap) | *schulenbergi* (Peru, S of Marañon Gap) | Oscine | 0.63 (19) | – |
| Tawny-crowned Greenlet (*Tunchiornis ochraceiceps*) | *pallidipectus* (Central America) | *ferrugineifrons* (W Amazon basin) | Oscine | 0.8 (10) | – |
| Golden-crowned Warbler (*Basileuterus culicivorus*) | *godmani* (Central America) | *cabanisi* (N South America) | Oscine | 0.76 (17) | – |
| Golden-bellied Warbler (*Myiothlypis chrysogaster*) | *chlorophrys* (NW Ecuador, W Colombia) | *chrysogaster* (E Peru, N Bolivia) | Oscine | 0.91 (11) | – |
| Citrine Warbler (*Myiothlypis luteoviridis*) | *luteoviridis* (Andes N of Marañon Gap) | *striaticeps* (Andes S of Marañon Gap) | Oscine | 0.73 (11) | – |
| Tricolored Brushfinch (*Atlapetes tricolor*) | *crassus* (NW Ecuador, W Colombia) | *tricolor* (E Peru) | Oscine | 0.83 (6) | – |
| Orange-billed Sparrow (*Arremon aurantiirostris*) | *occidentalis* (W Ecuador) | *spectabilis* (W Amazon basin) | Oscine | 0.7 (23) | 0.67 (12) |

**TABLE 3.** Pairs of taxa currently classified as subspecies that show strong behavioral discrimination against allopatric song and that likely meet in a contact zone. Song discrimination is the proportion of territories of the first-listed subspecies that ignored song from the second, allopatric subspecies (sample size of territories tested is given in parentheses). We conducted reciprocal playback experiments for 3 taxon pairs (but not near the contact zone); in each of these cases, song discrimination was high in both directions (population A typically ignored song from population B, and population B typically ignored song from population A). Field studies documenting genetic, vocal, and morphological variation in the presumed contact zones offer an opportunity to test our assumption that song discrimination in the allopatric portion of a population's range is associated with mate choice in the contact zone.

| Taxon pair | | | | | | |
|---|---|---|---|---|---|---|
| Species | Population A subspecies | Population B subspecies | Contact zone | Clade | Song discrimination | Reciprocal song discrimination |
| Spotted Woodcreeper (*Xiphorhynchus erythropygius*) | *aequatorialis* (NW Andes) | *punctigula* (Central America) | E Panama/NW Colombia | Suboscine | 0.64 (14) | 1 (12) |
| Black-faced Antthrush (*Formicarius analis*) | *hoffmanni* (Central America lowlands) | *moniliger* (Mesoamerica) | Nicaragua/Honduras | Suboscine | 0.67 (12) | – |
| Black-faced Antthrush (*Formicarius analis*) | *hoffmanni* (Central America lowlands) | *griseoventris* (Amazon Basin) | Venezuela | Suboscine | 0.67 (9) | – |
| Paltry Tyrannulet (*Zimmerius vilissimus*) | *parvus* (Central America) | *villissimus* (Mesoamerica) | Nicaragua/Honduras | Suboscine | 0.8 (15) | – |
| Buff-throated Foliage-gleaner (*Automolus ochrolaemus*) | *hypophaeus* (Caribbean slope Central America) | *auricularis* (C & E Panama) | C Panama | Suboscine | 0.9 (10) | – |
| White-breasted Wood-Wren (*Henicorhina leucosticta*) | *costaricensis* (Central America) | *darienensis* (E Panama) | C Panama | Oscine | 0.85 (13) | – |
| Scaly-breasted Wren (*Microcerculus marginatus*) | *luscinia* (Central America) | *occidentalis* (NW Ecuador) | E Panama/NW Colombia | Oscine | 0.88 (8) | 0.67 (9) |
| Orange-billed Sparrow (*Arremon aurantiirostris*) | *rufidorsalis* (Caribbean slope Central America) | *aurantiirostris* (Pacific slope Central America) | C Panama | Oscine | 0.6 (10) | 0.75 (20) |

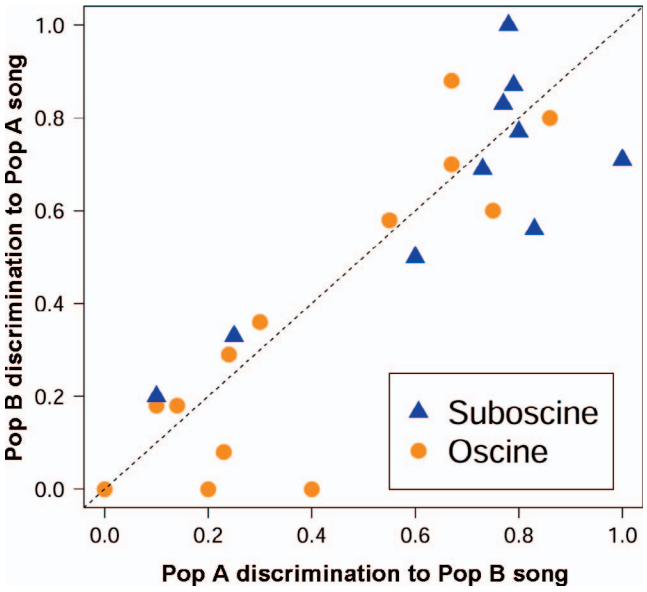


**FIGURE 1.** Behavioral discrimination against allopatric song was symmetric within taxon pairs. Song discrimination within a taxon pair was highly correlated ($r = 0.87$, $t = 7.95$, df = 21, $P < 0.0001$) for the 23 taxon pairs (13 oscines and 10 suboscines; see Supplemental Material Table S2) for which we measured both how population A discriminated against song from population B and the reverse. The dotted line shows the 1:1 line.

is because selection on females choosing mates is stronger than selection on birds (typically considered to be males) engaging in territorial defense. As a result, the response function to song is typically broader for birds engaging in territorial defense (again, typically considered to be males) and narrower for females choosing mates (Searcy and Brenowitz 1988, Seddon and Tobias 2010, Danner et al. 2011, Curé et al. 2012). That is, if a territorial bird ignores a song in a territorial context, it is likely that a female would also discriminate against that song in a mate choice context. We highlight that territorial defense is not the exclusive purview of males, and that female territorial defense is common in tropical birds (Odom et al. 2014). Indeed, we observed multiple individuals responding to playback in around half the cases when we observed aggressive responses to playback treatments; for sexually dichromatic taxa (e.g., antbirds) for which we could identify sex, multiple individuals typically consisted of a male and a female, presumably a mated pair; and we assume that female territorial defense was likewise common in the sexually monomorphic taxa we studied (B. G. Freeman et al. personal observation). In sum, we assume that taxon pairs for which a majority of territory-defending individuals (either males alone or mated pairs) fail to approach the speaker have evolved a substantial degree of premating reproductive isolation based on song.

Finally, we note that we did not expect our song discrimination scores to equal either 0 or 1 even if populations' species recognition capacities were lacking or complete, respectively. This is because measurement error in our field experiments can only bias our estimate of song discrimination up from 0 (if "true" song discrimination equals 0) or down from 1 (if "true" song discrimination equals 1). In particular, we note that we did not include a negative control in our playback experiments. Thus, we are unable to define the background rate at which territorial individuals approach the speaker during the allopatric treatment simply by happenstance, because they are curious about a novel sound. As a consequence, we expect song discrimination scores to be <1, even if song were, in fact, a complete behavioral barrier to reproduction. For example, we previously quantified the relationship between song discrimination and genetic distance using Michaelis-Menten models, finding that song discrimination asymptotes were ~0.78 when taxon pairs have been isolated for ~5 million yr (B. G. Freeman et al. personal observation). This suggests that over evolutionary time, for the genetic distances in our dataset, song discrimination peaks not at 1 but at ~0.78. As a consequence, we interpret taxon pairs with song discrimination scores around this value to represent cases where song discrimination is essentially complete.

## Acoustic Trait Analysis

We analyzed 1,087 songs from 72 taxon pairs (8.0 ± 1.8 songs population$^{-1}$; all >5 songs population$^{-1}$) using the software Raven Pro 1.5 (Bioacoustics Research Program 2014; see Supplemental Material Table S3). We measured acoustic traits for the same recordings used in playback experiments, supplementing these with additional recordings downloaded from xeno-canto (http://xeno-canto.org) and the Macaulay Library of Natural Sounds (http://macaulaylibrary.org) to boost sample sizes. For a representative song from each recording, we measured 7 song variables (total note count, mean note rate, mean note length, peak frequency, low frequency, mean note frequency range, and total song frequency range), following Mason et al. (2014). Within Raven, we used a Hann spectrogram window with 512 samples, a time grid with an overlap of 50% and a hop size of 256 samples, and a frequency grid with discrete Fourier transform set at 512 and grid spacing of 86.1 Hz.

We analyzed patterns of variation in acoustic traits for each taxon pair by log-transforming total note count and running a principal component analysis (PCA), using the correlation matrix based on the centered and scaled dataset wherein all variables were set to mean = 0 and SD = 1. This method produces distinct PCAs for each of the 72 taxon pairs. For these 72 comparisons, PC1 explained, on average, 48.2% of variation in multidimen-

sional acoustic space (range: 30.8–73.8%). To compare taxon pairs' divergence in acoustic space using a common currency, we quantified standardized acoustic divergence between populations within a taxon pair as the distance between population means along PC1, measured in units of pooled standard deviations.

### Statistical Analysis

All statistical analyses were carried out in R (R Development Core Team 2014). We first used the "cor.test" function to quantify the correlation between standardized acoustic divergence and behavioral song discrimination. Because standardized acoustic divergence was nonlinearly related to song discrimination, we used nonlinear regressions to investigate this relationship. Specifically, we used Michaelis-Menten curves, fit using the "nls" function, to investigate whether this nonlinear relationship differed depending on taxonomic rank (intraspecific vs. interspecific) or clade identity (oscine vs. suboscine). Our null model fit the formula $y = ax/(b + x)$ to the full dataset, where $a$ is the asymptote and $b$ is a measure of the rate of increase. Our alternative model fit the formula $y = ax/(b + c\delta + x)$, where $\delta$ is an indicator variable of taxonomic status or clade identity and $c$ is the difference in rate between the 2 groups. We then used the "anova" function to compare the relative fit of null and alternate models. Finally, we used $t$-tests to test whether species and subspecies differed in their song discrimination for both oscines and suboscines.

## RESULTS

We found a strong positive correlation between standardized acoustic divergence and song discrimination ($r = 0.57$, $P << 0.001$; Figure 2). However, this relationship was nonlinear. There was substantial variation in song discrimination at low levels of standardized acoustic divergence (0 to ~3) and nearly uniformly high song discrimination when standardized acoustic divergence was greater than ~3 (i.e. when populations' mean positions along PC1 within multivariate acoustic space were greater than ~3 standard deviations apart; Figure 2). We used Michaelis-Menten curves to model the relationship between standardized acoustic divergence and song discrimination; including taxonomic status (intraspecific vs. interspecific; $F = 0.48$, df = 1, $P = 0.49$) or clade identity (suboscine vs. oscine; $F = 0.95$, df = 1, $P = 0.33$; Figure 2) in models did not improve model fit. Thus, the nonlinear relationship between standardized acoustic divergence and song discrimination document here is a general relationship that applies to taxon pairs of Neotropical passerine birds across clades and taxonomic ranks.

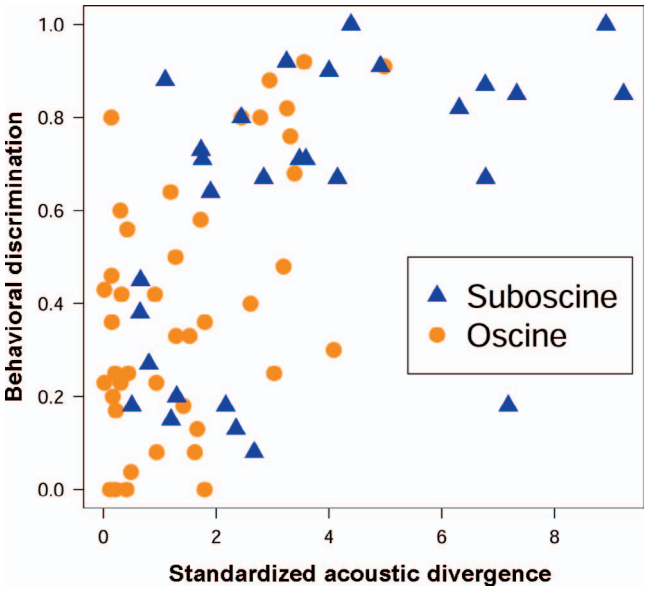


**FIGURE 2.** Standardized acoustic divergence is nonlinearly related to behavioral song discrimination in a dataset of 72 taxon pairs. Suboscines tend to have greater standardized acoustic divergences and behavioral discrimination values than oscines, but the 2 clades have the same nonlinear relationship between standardized acoustic divergence and behavioral song discrimination—a Michaelis-Menten model without clade identify is a better fit than a model that includes clade identity ($F = 0.95$, df = 1, $P = 0.33$). In both clades, song discrimination is nearly uniformly high when standardized acoustic divergence is greater than ~3, and highly variable at low levels of standardized acoustic divergence. Song discrimination is the percentage of territories in a population that failed to approach the speaker in response to playback of its allopatric taxon pair (i.e. song discrimination scores >0.5 indicate taxon pairs in which the majority of territories discriminated against allopatric song), and standardized acoustic divergence is the distance between population means along PC1 within a taxon pair, expressed in pooled standard deviations. The outlier taxon pair in the bottom right (high acoustic divergence but low discrimination) is Yellowish Flycatcher (*Empidonax flavescens*)–Cordilleran Flycatcher (*E. occidentalis*).

### Taxonomic Implications

Our results are relevant to the taxonomic classification of allopatric populations under the BSC. We focus on our song discrimination data, because our finding that song discrimination and standardized acoustic divergence are not linearly related suggests that song discrimination is a more direct proxy for whether song constitutes a behavioral barrier to reproduction. Surprisingly, for suboscines, song discrimination was significantly *greater* in intraspecific taxon pairs (mean song discrimination = 0.70, $n = 18$) than in interspecific taxon pairs (mean song discrimination = 0.47, $n = 11$; $t = -2.1$, df = 23.4, $P = 0.047$). In oscines, song discrimination was unrelated to taxonomic rank ($t = 0.91$, df = 29.2, $P = 0.37$). These mismatches between song discrimination and taxonomic

rank, particularly for suboscines, indicate that current taxonomy does not reflect the role of song divergence as a barrier to reproduction between allopatric populations of Neotropical passerines. To partially address this mismatch, we suggest that geographically isolated taxon pairs that are currently classified as conspecific, but that show strong song discrimination to allopatric song, merit classification as distinct biological species. We apply this logic to evaluate species limits for all taxon pairs for which we have conducted playback experiments (i.e. including taxon pairs for which we have not measured divergence in acoustic traits).

As described above, we do not expect our methodology to produce song discrimination scores of 1 (100% of territories ignore allopatric song), even if song in fact constituted a complete behavioral barrier to reproduction between 2 isolated populations. To provide a yardstick of how much song discrimination in our dataset is "enough" to merit classifying allopatric populations as distinct biological species, we consider song discrimination scores for the 9 interspecific taxon pairs in our dataset that have been split into distinct species within the past 2 decades in part on the basis of their divergent vocalizations (Table 1). Song discrimination in these cases averaged 0.59 (6 of 10 territorial birds ignored allopatric song). We use this average as a rough benchmark and suggest that the 21 taxon pairs currently classified as conspecific that show song discrimination greater than ~0.6 merit classification as distinct biological species (Table 2).

In Table 3, we highlight 8 additional taxon pairs in which populations showed strong behavioral discrimination (>0.6) toward allopatric song, but in which it is likely that the 2 populations meet in a contact zone. Our playback experiments in these cases suggest that song constitutes a behavioral barrier to reproduction. But because our experiment took place far from the presumed contact zone, further work in the putative contact zones is needed to determine whether this is correct. We note that these cases thus represent tests of our assumption that high song discrimination between allopatric populations indicates that populations would also show high song discrimination and fail to interbreed when given the opportunity to actually interact (Hudson and Price 2014).

## DISCUSSION

Geographically isolated populations of birds often differ in song, which may indicate that they have evolved behavioral barriers to reproduction. We measured song divergence using both acoustic trait analyses and song playback experiments in 72 taxon pairs of Neotropical passerines. We found that divergence in acoustic traits is positively correlated with behavioral discrimination against allopatric song, but that this relationship is nonlinear for both suboscines and oscines (Figure 2). Territorial birds consistently discriminated against song from allopatric populations when populations differed in their mean position in multivariate acoustic space by more than ~3 standard deviations (Figure 2). However, this relationship is variable at low to moderate levels of acoustic divergence, such that quantifying acoustic divergence (at lower levels of acoustic divergence) provides little insight into whether a population would behaviorally discriminate against allopatric song. Moreover, we found that song discrimination is not greater in taxon pairs ranked as species than in taxon pairs ranked as subspecies. We assume that strong song discrimination indicates a substantial behavioral barrier to reproduction. As such, we propose that 21 taxon pairs whose current classification as subspecies largely ignores playback of allopatric song merit recognition as distinct biological species (Table 2). More broadly, our results support the use of playback experiments as the preferred tool to assess whether song divergence between isolated populations constitutes a substantial premating barrier to reproduction.

### Acoustic Trait Analyses

Our conclusion that acoustic divergence is weakly related to behavioral discrimination when acoustic divergence is low (Figure 2), and thus that playback experiments provide better inference of reproductive isolation under these conditions, relies on our methodology for measuring and analyzing acoustic traits. We quantified 7 acoustic traits, measured from an average of 8 recordings per population, and used multivariate statistics to measure acoustic divergence between populations. By contrast, studies of acoustic variation among closely related populations often consider many dozens of traits, some of which are customized to the specific vocalization of the taxa under consideration, measure hundreds of recordings, and also conduct univariate comparisons for each individual acoustic trait (e.g., Ng et al. 2016). Because the present study measured acoustic divergence across a diverse set of Neotropical passerines that differ dramatically in song, we focused on a small number of acoustic traits that are likely to be broadly relevant across taxa (and also likely to have low collinearity). Our sample sizes of recordings analyzed per population per taxon pair were not large enough to permit meaningful univariate analyses; we note that conclusions from multivariate acoustic analyses are typically similar to those from univariate analyses (e.g., Ng et al. 2016).

It is possible that a different methodology for measuring acoustic divergence could recover a tighter correspondence between acoustic divergence and behavioral discrimination, but we highlight that populations may be statistically different in their acoustic traits but fail to discriminate against allopatric song in playback experi-

ments (e.g., Nelson 1998, Soha et al. 2016). If our findings are valid, we suggest that acoustic trait analyses are best employed (1) as an initial exploration of song divergence (i.e. a stopgap measure pending experimental data), especially when acoustic divergence is weak or moderate, and (2) in concert with playback experiments; for example, the shortcomings associated with conducting playback experiments in a small number of geographic locations can be overcome by comparing vocal characters from across a broader geographic spread. Future research can assess the validity of our results by focusing on single clades and by combining playback data with acoustic analyses (with acoustic traits tailored to the clade in question) to quantify the correlation between these 2 methods for measuring song divergence.

### Implications for Species Limits

Avian taxonomy typically follows the BSC, which defines species as reproductively isolated populations. Applying the BSC to geographically isolated populations that do not have the opportunity to interbreed is challenging. In practice, ornithologists use a comparative framework to assess whether isolated populations are likely to interbreed (or not) should they come into sympatry (Isler et al. 1998, Tobias et al. 2010b). Here, we focus on divergence in song—a single, behavioral barrier. Although a variety of ecological, morphological, behavioral, and genetic factors can all constitute barriers to reproduction, including divergence in calls and other vocalizations (reviewed in Price 2008), song divergence is an important barrier to reproduction in birds, perhaps particularly so in suboscines (Edwards et al. 2005, Tobias et al. 2012).

We use song discrimination scores for taxon pairs that have recently been split, in part on the basis of their divergent vocalizations, as a yardstick for when to split taxa on the basis of our song discrimination data (Table 1). The average song discrimination scores for these taxa are ~0.6, which means that 6 of 10 territorial birds ignored allopatric song. We found 21 intraspecific taxon pairs that cleared this threshold, and we suggest that these populations merit status as distinct biological species (Table 2). In many cases, these taxon pairs have been previously considered to represent distinct species on the basis of obvious vocal differentiation. For example, Striped Woodhaunter (*Automolus subulatus*) populations east and west of the Andes differ in voice and have thus been considered to represent distinct species by some authorities (e.g., Ridgely and Greenfield 2001). It will thus surprise few ornithologists familiar with Neotropical birds to learn that territorial woodhaunters from west of the Andes typically ignore playback of Amazonian song. However, we also document strong song discrimination between many populations that are known to differ in song but that have not been previously considered to represent distinct species—for example, Andean Solitaire (*Myadestes ralloides*) populations north and south of the Marañon Gap in Peru. Finally, we document high song discrimination in 8 cases of allopatric populations that likely meet in a contact zone (Table 3). Such examples offer an opportunity to test our assumption that song discrimination in the allopatric portion of the range is indeed associated with mate choice in the contact zone.

We suggest elevating the taxa listed in Table 2 to species rank. One complicating factor to carrying out this suggestion in practice is that there may be more than 2 taxa involved in many of these species. That is, we argue that allopatric populations A and B (for which we conducted playback experiments) should be considered distinct biological species, but where does allopatric population C (for which we did not conduct experiments) fit in? In most cases, major geographic barriers can provide a reasonable basis for placing additional populations. For example, the spine of the Andes divides Striped Woodhaunter populations in two (east and west), while the Marañon Gap divides Andean Solitaire populations in two (north and south). In other cases, geographic barriers are not complete—for example, distributions extend across the northern reaches of the Andes, as in the Yellow-olive Flycatcher (*Tolmomyias sulphurescens*)—and the question of affinities of additional allopatric populations is more difficult to address. It may be that such cases require a more comprehensive analysis, for example by conducting playback experiments between each of dozens of populations. Nevertheless, our data suggest that multiple biological species lurk within these complicated species complexes; if published information on song variation can be used to group additional allopatric populations not considered in our playback experiments, then the data we present here may be sufficient to redefine species limits in these complexes.

Finally, we refrain from interpreting low song discrimination values as indicating an absence of premating reproductive isolation that might constitute evidence for "lumping" populations currently considered distinct species into a single taxon. We take this cautious approach for 2 reasons. First, 2 allopatric populations may be nearly identical in song, with low song discrimination scores, but divergent in other signaling traits (e.g., plumage or call notes) that generate substantial premating reproductive isolation. Second, we argue that if a bird ignores a song in a territorial context, it is likely that a female would also discriminate against that song in a mate choice context. However, whether the reverse is true is less clear. Indeed, there are cases where males respond to a song in a territorial context but females discriminate against this same song in a mate choice context (note that, in this example, females may cue to male calls more then male songs; Seddon and Tobias 2010). In sum, we interpret a

lack of response to allopatric song as indicating a behavioral barrier to reproduction, but we do not interpret a response to allopatric song as indicating the absence of a behavioral barrier to reproduction.

### Playback Experiments Can Help Address the Latitudinal Gradient in Taxonomy

It is widely acknowledged that species are more narrowly defined in the temperate zone than in tropical latitudes (for birds, see Tobias et al. 2008, Weir 2009, Milá et al. 2012). In birds, this bias arises for both biological and nonbiological reasons. First, temperate zone species tend to attain secondary contact more quickly than tropical species (Martin et al. 2010, Weir and Price 2011), with the consequence that genetically divergent, allopatric populations are more common in the tropics than in the temperate zone (Martin and Tewksbury 2008). Second, temperate zone scientists carry out the bulk of avian systematic research, and many tropical taxa are poorly studied. This geographic bias in taxonomy has practical consequences for comparative studies and our understanding of avian diversity. Studies of diversification along latitudinal gradients are hampered by variation in taxonomy (e.g., Tobias et al. 2008), as are investigations of the evolutionary dynamics of particular clades. For example, analyses of the tempo and drivers of diversification in *Scytalopus* would presumably have reached different conclusions if they were conducted in 1990, when 13 species were recognized (Sibley and Monroe 1990), compared with at present, when 42 species are recognized (Remsen et al. 2017).

There is widespread sentiment that the latitudinal gradient in taxonomy ought not to exist, but little agreement on how to implement the widespread taxonomic changes necessary to address this bias. The status quo is that authors publish taxonomic studies that assess species limits within a complex, and taxonomic committees apply the BSC to pass (or reject) these studies' recommendations. This results in a slow but steady flattening of the latitudinal gradient in taxonomy; for example, the South American Classification Committee of the American Ornithologists' Union recognized 52 additional species of continental landbirds found in South America in the 5 yr period between 2012 and 2016 (Remsen et al. 2017), far more than the 2 additional species of continental landbirds found in the United States and Canada that the North American Classification Committee recognized during the same period (Chesser et al. 2017). However, others advocate greatly accelerating the push to rank tropical taxa at the species level (e.g., Gill 2014); for example, the *Handbook of the Birds of the World* recently applied species delimitation criteria (Tobias et al. 2010b) to elevate >1,000 populations of birds (primarily in the tropics) to species status (see http://www.lynxeds.com/product/hbw-and-birdlife-international-illustrated-checklist-birds-world-0). Needless to say, this proposal to increase extant species-level diversity of birds by ~10% in one fell swoop is controversial (Remsen 2015, Collar et al. 2016).

The approach we describe here—conducting playback experiments to measure species recognition between allopatric populations—offers a way forward to addressing the latitudinal gradient in taxonomy, for 4 reasons. First, it maintains the primacy of the BSC, which has many advantages—primarily, that a focus on reproductive isolation means that "species" is the only taxonomic rank that has a biological basis (e.g., Toews 2015). Second, it focuses on a trait directly relevant to reproductive isolation—song. Tropical birds often exhibit minimal plumage variation, and, because genetic divergence is only loosely correlated with reproductive barriers (Price and Bouvier 2002), focusing on song divergence is an especially profitable approach to assessment of species limits in allopatric tropical taxa. Third, it directly assesses whether song divergence is relevant to reproductive isolation. While the *Handbook of the Birds of the World* approach considers vocal differentiation as evidence for species status (Collar et al. 2016), not all vocal differences are created equal (see Figure 2); playback experiments "ask the birds themselves" if differences between populations matter to species recognition. Fourth, playback experiments are easier than ever to carry out. Field guides and online resources provide descriptions of geographic variation in song that suggest populations to target for playback studies; large repositories of high-quality, publicly available recordings provide recordings to use in playback experiments; and affordable and lightweight wireless speakers make conducting fieldwork relatively straightforward. We envision that citizen scientists and ornithologists, particularly those residing in tropical nations, will organize and carry out playback studies to measure species recognition between allopatric populations of tropical birds on an increasingly broad scale.

## Conclusion

The use of field playbacks to assess whether vocal differences between isolated populations constitute barriers to reproduction has a long history (e.g., Lanyon 1978). Because song is an important barrier to reproduction in birds, perhaps particularly so in clades with innate song, evaluating song divergence is an important component of assessing the taxonomic rank of allopatric populations. Our results show that playback experiments and acoustic trait analyses do not provide interchangeable inference, and they demonstrate that it is feasible to conduct playback experiments on a broad scale. We argue that this approach provides valuable data for assessing species limits using the BSC, and we advocate for the continued use of

playback experiments on tropical birds to help flatten the latitudinal gradient in taxonomy. In sum, we encourage the further use of field playback experiments to measure species recognition between allopatric songbird populations, an approach that has been termed "behavioral systematics" (Pegan et al. 2015).

## ACKNOWLEDGMENTS

We are indebted to the many recordists who made their song recordings available through xeno-canto.org and the Macaulay Library, as well as the administrators who maintain these collections—our work would not have been possible without their efforts. We thank B. Van Doren and the spring 2015 Advanced Tropical Field Ornithology course for assistance with data collection, and N. Mason for providing a particularly useful R script. N. Buttner, M. Ellis, L. D. Gonzalez, P. O'Donnell, R. Parsons, A. Rodriguez, and the people of the town of Utuana provided valuable logistical assistance in the field. Comments from three anonymous reviewers greatly improved the manuscript.

**Funding statement:** This research was supported by an AOU Research Grant and by a National Science Foundation Postdoctoral Fellowship in Biology (award no. 1523695) to B.G.F. None of our funders had any influence on the content of the submitted manuscript, and none of our funders required approval of the final manuscript to be published.

**Ethics statement:** This research was conducted in compliance with the Guidelines to the Use of Wild Birds in Research.

**Author contributions:** B.G.F. formulated the question and analyzed the data. G.A.M. and B.G.F. collected the data. B.G.F. and G.A.M. wrote the paper.

**Data deposits:** Our dataset is included in Supplemental Material Tables S1 and S3.

## LITERATURE CITED

Alström, P., and R. Ranft (2003). The use of sounds in avian systematics and the importance of bird sound archives. Bulletin of the British Ornithologists' Club 123:114–135.

Araya-Salas, M., and G. Smith-Vidaurre (2017). warbleR: An R package to streamline analysis of animal acoustic signals. Methods in Ecology and Evolution 8:184–191.

Baker, M. C., and A. E. Miller Baker (1990). Reproductive behavior of female buntings: Isolating mechanisms in a hybridizing pair of species. Evolution 44:332–338.

Barraclough, T. G., and A. P. Vogler (2000). Detecting the geographical pattern of speciation from species-level phylogenies. The American Naturalist 155:419–434.

Bioacoustics Research Program (2014). Raven Pro: Interactive Sound Analysis Software. Cornell Lab of Ornithology, Ithaca, NY, USA. http://www.birds.cornell.edu/raven

Cadena, C. D., and A. M. Cuervo (2010). Molecules, ecology, morphology, and songs in concert: How many species is *Arremon torquatus* (Aves: Emberizidae)? Biological Journal of the Linnean Society 99:152–176.

Chaves, J. C., A. M. Cuervo, M. J. Miller, and C. D. Cadena (2010). Revising species limits in a group of *Myrmeciza* antbirds reveals a cryptic species within *M. laemosticta* (Thamnophilidae). The Condor 112:718–730.

Chesser, R. T., K. Burns, C. Cicero, J. L. Dunn, A. W. Kratter, I. J. Lovette, P. C. Rasmussen, J. V. Remsen, Jr., J. D. Rising, D. F. Stotz, and K. Winker (2017). Checklist of North and Middle American Birds. American Ornithological Society. http://checklist.aou.org

Collar, N. J., L. D. C. Fishpool, J. del Hoyo, J. D. Pilgrim, N. Seddon, C. N. Spottiswoode, and J. A. Tobias (2016). Toward a scoring system for species delimitation: A response to Remsen. Journal of Field Ornithology 87:104–115.

Coyne, J. A., and T. D. Price (2000). Little evidence for sympatric speciation in island birds. Evolution 54:2166–2171.

Curé, C., N. Mathevon, R. Mundry, and T. Aubin (2012). Acoustic cues used for species recognition can differ between sexes and sibling species: Evidence in shearwaters. Animal Behaviour 84:239–250.

Danner, J. E., R. M. Danner, F. Bonier, P. R. Martin, T. W. Small, and I. T. Moore (2011). Female, but not male, tropical sparrows respond more strongly to the local song dialect: Implications for population divergence. The American Naturalist 178:53–63.

Edwards, S. V, S. B. Kingan, J. D. Calkins, C. N. Balakrishnan, W. B. Jennings, W. J. Swanson, and M. D. Sorenson (2005). Speciation in birds: genes, geography, and sexual selection. Proceedings of the National Academy of Sciences USA 102 (Supplement 1):6550–6557.

Freeman, B. G., A. M. Class Freeman, and W. M. Hochachka (2016). Asymmetric interspecific aggression in New Guinean songbirds that replace one another along an elevational gradient. Ibis 158:726–737.

Freeman, B. G., and G. Montgomery (2016). Interspecific aggression by the Swainson's Thrush (*Catharus ustulatus*) may limit the distribution of the threatened Bicknell's Thrush (*Catharus bicknelli*) in the Adirondack Mountains. The Condor: Ornithological Applications 118:169–178.

Gill, F. B. (2014). Species taxonomy of birds: Which null hypothesis? The Auk: Ornithological Advances 131:150–161.

Hudson, E. J., and T. D. Price (2014). Pervasive reinforcement and the role of sexual selection in biological speciation. Journal of Heredity 105 (Supplement 1):821–833.

Irwin, D. E., and T. Price (1999). Sexual imprinting, learning and speciation. Heredity 82:347–354.

Isler, M. L., P. R. Isler, and B. M. Whitney (1998). Use of vocalizations to establish species limits in antbirds (Passeriformes: Thamnophilidae). The Auk 115:577–590.

Jankowski, J. E., S. K. Robinson, and D. J. Levey (2010). Squeezed at the top: Interspecific aggression may constrain elevational ranges in tropical birds. Ecology 91:1877–1884.

Lanyon, W. E. (1978). Revision of the *Myiarchus* flycatchers of South America. Bulletin of the American Museum of Natural History 161:article 4.

Martin, P. R., and T. E. Martin (2001). Behavioral interactions between coexisting species: Song playback experiments with wood warblers. Ecology 82:207–218.

Martin, P. R., and J. J. Tewksbury (2008). Latitudinal variation in subspecific diversification of birds. Evolution 62:2775–2788.

Martin, P. R., R. Montgomerie, and S. C. Lougheed (2010). Rapid sympatry explains greater color pattern divergence in high latitude birds. Evolution 64:336–347.

Mason, N. A., K. J. Burns, J. A. Tobias, S. Claramunt, N. Seddon, and E. P. Derryberry (2017). Song evolution, speciation, and vocal learning in passerine birds. Evolution 71:786–796.

Mason, N. A., A. J. Shultz, and K. J. Burns (2014). Elaborate visual and acoustic signals evolve independently in a large, phenotypically diverse radiation of songbirds. Proceedings of the Royal Society B 281:20140967.

McEntee, J. P., J. V. Peñalba, C. Werema, E. Mulungu, M. Mbilinyi, D. Moyer, L. Hansen, J. Fjeldså, and R. C. K. Bowie (2016). Social selection parapatry in Afrotropical sunbirds. Evolution 70:1307–1321.

McGregor, P. K., C. K. Catchpole, T. Dabelsteen, J. B. Falls, L. Fusani, H. C. Gerhardt, F. Gilbert, A. G. Horn, G. M. Klump, D. E. Kroodsma, M. M. Lambrechts, et al. (1992). Design of playback experiments: The Thornbridge Hall NATO ARW consensus. In Playback and Studies of Animal Communication (P. K. McGregor, Editor). Springer, Boston, MA, USA. pp. 1–9.

Milá, B., E. S. Tavares, A. Muñoz Saldaña, J. Karubian, T. B. Smith, and A. J. Baker (2012). A Trans-Amazonian screening of mtDNA reveals deep intraspecific divergence in forest birds and suggests a vast underestimation of species diversity. PLoS ONE 7:e40541.

Nelson, D. A. (1998). Geographic variation in song of Gambel's White-crowned Sparrow. Behaviour 135:321–342.

Ng, E. Y. X., J. A. Eaton, P. Verbelen, R. O. Hutchinson, and F. E. Rheindt (2016). Using bioacoustic data to test species limits in an Indo-Pacific radiation of *Macropygia* cuckoo doves. Biological Journal of the Linnean Society 118:786–812.

Odom, K. J., M. L. Hall, K. Riebel, K. E. Omland, and N. E. Langmore (2014). Female song is widespread and ancestral in songbirds. Nature Communications 5:article 3379.

Payne, R. B. (1986). Bird songs and avian systematics. Current Ornithology 3:87–126.

Pegan, T., R. B. Rumelt, S. Dzielski, M. M. Ferraro, L. E. Flesher, N. Young, A. Class Freeman, and B. G. Freeman (2015). Asymmetric response of Costa Rican White-breasted Wood-Wrens (*Henicorhina leucosticta*) to vocalizations from allopatric populations. PLoS ONE 10:e0144949.

Peters, S. S., W. A. Searcy, and P. Marler (1980). Species song discrimination in choice experiments with territorial male swamp and song sparrows. Animal Behaviour 28:393–404.

Price, T. D. (2008). Speciation in Birds. Roberts, Greenwood Village, CO, USA.

Price, T. D., and M. M. Bouvier (2002). The evolution of F1 postzygotic incompatibilities in birds. Evolution 56:2083–2089.

R Development Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org

Remsen, J. V., Jr. (2005). Pattern, process, and rigor meet classification. The Auk 122:403–413.

Remsen, J. V., Jr. (2015). Review of *HBW and Birdlife International Illustrated Checklist of the Birds of the World, vol. 1: Non-passerines* (N. J. Collar and J. del Hoyo, Editors). Journal of Field Ornithology 86:182–187.

Remsen, J. V., Jr., J. I. Areta, C. D. Cadena, S. Claramunt, A. Jaramillo, J. F. Pacheco, J. Pérez-Emán, M. B. Robbins, F. G. Stiles, D. F. Stotz, and K. J. Zimmer (2017). A classification of the bird species of South America. American Ornithologists' Union. http://www.museum.lsu.edu/~Remsen/SACCBaseline.htm

Rheindt, F. E., J. A. Norman, and L. Christidis (2008). DNA evidence shows vocalizations to be a better indicator of taxonomic limits than plumage patterns in *Zimmerius* tyrant-flycatchers. Molecular Phylogenetics and Evolution 48:150–156.

Ridgely, R. S., and P. J. Greenfield (2001). The Birds of Ecuador, vol. 1. Christopher Helm, London, UK.

Searcy, W. A., and E. A. Brenowitz (1988). Sexual differences in species recognition of avian song. Nature 332:152–154.

Seddon, N., and J. A. Tobias (2010). Character displacement from the receiver's perspective: Species and mate recognition despite convergent signals in suboscine birds. Proceedings of the Royal Society B 277:2475–2483.

Sibley, C. G., and B. L. Monroe (1990). Distribution and Taxonomy of Birds of the World. Yale University Press, New Haven, CT, USA.

Slabbekoorn, H., and T. B. Smith (2002). Bird song, ecology and speciation. Philosophical Transactions of the Royal Society of London, Series B 357:493–503.

Soha, J. A., and P. Marler (2000). A species-specific acoustic cue for selective song learning in the white-crowned sparrow. Animal Behaviour 60:297–306.

Soha, J. A., A. Poesel, D. A. Nelson, and B. Lohr (2016). Non-salient geographic variation in birdsong in a species that learns by improvisation. Ethology 122:343–353.

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

Tobias, J. A., J. Aben, R. T. Brumfield, E. P. Derryberry, W. Halfwerk, H. Slabbekoorn, and N. Seddon (2010a). Song divergence by sensory drive in Amazonian birds. Evolution 64:2820–2839.

Tobias, J. A., J. M. Bates, S. J. Hackett, and N. Seddon (2008). Comment on "The latitudinal gradient in recent speciation and extinction rates of birds and mammals." Science 319: 901.

Tobias, J. A., J. D. Brawn, R. T. Brumfield, E. P. Derryberry, A. N. G. Kirschel, and N. Seddon (2012). The importance of Neotropical suboscine birds as study systems in ecology and evolution. Ornithologia Neotropical 23:259–272.

Tobias, J. A., N. Seddon, C. N. Spottiswoode, J. D. Pilgrim, L. D. C. Fishpool, and N. J. Collar (2010b). Quantitative criteria for species delimitation. Ibis 152:724–746.

Toews, D. P. L. (2015). Biological species and taxonomic species: Will a new null hypothesis help? (A comment on Gill 2014). The Auk: Ornithological Advances 132:78–81.

Touchton, J. M., N. Seddon, and J. A. Tobias (2014). Captive rearing experiments confirm song development without learning in a tracheophone suboscine bird. PLoS ONE 9: e95746.

Uy, J. A. C., R. G. Moyle, and C. E. Filardi (2009). Plumage and song differences mediate species recognition between

incipient flycatcher species of the Solomon Islands. Evolution 63:153–164.

Weir, J. T. (2009). Implications of genetic differentiation in Neotropical montane forest birds. Annals of the Missouri Botanical Garden 96:410–433.

Weir, J. T., and T. D. Price (2011). Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. The American Naturalist 177:462–469.