TITLE

Evaluating the effectiveness of model predators in simulating real predators in tropical bird communities

AUTHORS AND AFFILIATIONS

Andrés Felipe Tigreros-Andrade[a] , Michael Gulman[b], Jorge Novoa[c], Eliseo Parra[d], Daniel T. Blumstein[d], Ari E. Martínez[a]

[a]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA

[b]Colka Raptors SAC, Piura, Perú

[c]CORBIDI, Lima, Perú

[d]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA

***Corresponding Author**

***Address correspondence to Andrés Felipe Tigreros-Andrade. E-mail: andres_tigreros94@hotmail.com**

Handling editor: Chunlin Li

Received on 11 February 2026; accepted on 11 June 2026

ABSTRACT

Predatory attacks are difficult to observe in nature, and thus, researchers use model predators to simulate real predators. However, few studies have systematically compared the response of prey to models versus living predators. We flew a trained live raptor across mixed species flocks of Amazonian birds and compared the acoustic and behavioural response of individual species and the entire flock to that following a flight of a painted model glider. We found no statistically significant differences between treatments in alarm call bandwidth 50%, duration 90%, latency for most species, or the proportion of species producing alarm calls. A significant overall treatment effect was detected for alarm call frequency 75%, although this was not driven by any specific species and was small in magnitude. The number of urgent notes showed a marginal trend toward fewer notes in model trials. Latency to call differed between treatments overall, driven by White-flanked Antwren, which called more rapidly to the live hawk than to the model. About one-third of flock species responded with alarm calls, a proportion that did not differ between treatments. Models can indeed be an effective method to study anti-predator behaviour, but more research on how and why species may or may not discriminate between models and live predators is warranted.

Keywords: predator-evoked vocalizations, alarm calls, propensity to alarm, methods to study alarm communication

Predation is a key mechanism that modulates many aspects of ecosystems, including the interactions among individuals and species that affect prey populations and the structure of communities (Smee 2010, Mogali et al. 2011, Batabyal 2023). Animals have evolved a variety of

behavioral and morphological adaptations to manage predation risks (Lima and Dill 1990, Mogali et al. 2011, Rosier 2011, Palmer and Packer 2021). Among them, the use of alarm calls is especially important (Price et al. 2015, Potvin et al. 2018). Through alarm call signaling, individuals can warn both conspecifics and heterospecifics of approaching predation threats, reducing their own risk by promoting collective vigilance and evasive actions (Ellard and Byers 2005, Podos and Webster 2022). Understanding the contexts under which alarm calls are produced and the variation in their structure is crucial for understanding a variety of questions regarding predator-prey behavior, social communication, and evaluating aspects of vocal complexity and the evolution of signals.

Given the rarity with which predator attacks are observed in many systems, answering questions related to predator-driven alarm calling in prey communities requires simulating predator encounters using model predators. The use of predator models has been applied to different ecological contexts to investigate multiple aspects of predator-prey interactions and has a considerable history in ethology. Dating back to the 1930s, Tinbergen and Lorenz conducted classic experiments using a model bird silhouette flying over the head of domestic turkeys, ducks, and geese (Tinbergen 1989). Later, Curio et al. (1978) used stuffed animals to test the Cultural Transmission Hypothesis in captive Common Blackbirds (*Turdus merula*), and similarly Hegner (1985) used a wooden model of a Eurasian Sparrowhawk (*Accipiter nisus*) to examine whether dominance hierarchies within bird flocks predicted the sequence of return to foraging following predator exposure. Also, such models have been applied across a range of taxa. They have been used to study fish foraging decisions in response to the risk of Black Groupers (*Mycteroperca bonaci*) predation (Catano et al. 2016), to simulate threats of an approaching spear fisherman to study social feeding behavior of coral reef fishes (Gil and Hein 2017), via the

use of taxidermic 'robobadgers' to simulate ground predators and life sized model gliders to simulate raptors (Blumstein and Armitage 1997, life sized photographs to simulate terrestrial predators (Blumstein et al. 2009), and more recently, by the use of taxidermic badgers to test potential eavesdropping between long-billed curlews and prairie dogs (Dreelin et al. 2025). In avian systems, predator models such as taxidermy mounts of raptors or painted Styrofoam models are used to study alarm calling responses (see Griesser 2008, Soard and Ritchison 2009, Magrath et al. 2015b) and to test antipredator behavior in mixed flocks (Browning 2015, Jiang et al. 2020, Frazier et al. 2025). These models offer a non-invasive and standardized method for presenting predator stimuli. However, it is assumed implicitly that these predator models elicit a realistic alarm-calling response from animals that is similar to what is evoked by a real predator.

Despite the use of models in behavioral studies and on target prey species, few studies have validated whether these models evoke equivalent responses to those obtained from using real predators (Blumstein and Armitage 1997). This gap limits our ability to assess the reliability of model-based experiments for testing alarm communication and social coordination in a natural setting. Mixed-species bird flocks (hereafter flocks) are an ideal system for studying alarm-calling in response to sharing information about predation threats between different species (e.g., Goodale and Kotagama 2005, Vander Meiden 2023). These flocks are composed of at least three or more individuals from at least two species that move and forage together in a coordinated manner (Mangini et al. 2023). Flock members have specialized roles, and some species act as sentinels who detect predators and emit alarm calls that function to warn others, while different species primarily respond to this information (Goodale and Kotagama 2005, Martínez and Zenil 2012). Thus, mixed species flocks function as dynamic information-sharing networks, enabling

synergistic responses to predators and other disturbances (Goodale and Kotagama 2005, Griesser 2008, Magrath et al. 2015a).

While models can be used to determine how the information networks work, there is a need to compare and contrast the response to live raptors with that produced by models. In this study, we ask whether the mixed flocks in the southern Peruvian Amazon respond similarly to exposure of a predation event using models and real raptors. By comparing the acoustic response to simulated or real predator presentation, we assessed the extent to which models evoke "realistic" alarming behavior. We evaluated the proportion of species producing alarm calls, the latency to produce alarm calls, and the acoustic structure of alarm signals elicited by live raptors and models used to simulate predation threat. We hypothesized that the vocal characteristics of the alarm responses would be similar for species responding between the two treatments, but that the overall number of species responding would differ. We predicted that alarm call frequency 75%, bandwidth 50%, duration 90%, latency to first alarm, the number of urgent notes and the proportion of species in the flock alarming would be similar between real and simulated predators for individual species. We also predicted that the number of mixed species flock members responding would be greater for the real predator compared to the model. This study aims to evaluate the effectiveness of using models to simulate natural predation events to accurately assess prey responses in natural settings.

Materials and METHODS

Study Site and Species

Experiments were conducted in a tropical lowland forest in the Madre de Dios region of southeastern Peru at Villaana's Biological Station between 3 June and 29 June 2024. Villaana's is

a Brazilian Nut (*Bertholletia excelsa*) concession located at S12° 11.026' W69° 04.822' at around 250 m above sea level, approximately 60 km north of Puerto Maldonado, consisting of more than 1 000 ha of mature Amazonian Terra Firma Forest.

We identified 24 flocks within the concession boundaries**.** Mixed flocks in this area are stable year-round and are often composed of a leader species, Dusky-throated Antshrike (*Thamnomanes ardesiacus*), which is known for its high vigilance and alarm-calling role (Munn and Terborgh 1979). They are also accompanied by obligate follower species such as White-flanked Antwren (*Myrmotherula axillaris*), Red-crowned Ant-Tanager (*Habia rubica*), and Long-winged Antwren (*M. longipennis*), as well as facultative follower species such as the Gray Antwren (*M. menetriesii*), Buff-throated Woodcreeper (*Xiphorhynchus guttatus*), Elegant Woodcreeper (*X. elegans*), among others.

Experimental design

We conducted field experiments by flying a trained Bicolored Hawk (*Accipiter bicolor,* hawk hereafter, Figure S1A) over mixed-species flocks or by flying a model predator glider painted to resemble the hawk, which was designed by one of the authors (Figure S1B). The hawk was trained to be fed only from the trainer's hand, and thus, no birds were physically attacked by the raptor. We recognize that we are presenting one single individual predator (model or live raptor) to the flocks. All trials used the same trained individual raptor, meaning that individual variation in flight dynamics or predator-specific individual differences cannot be assessed. However, prior results have shown that in our system, predator size and not individual size variation within a species is what is relevant in risk perception in our system. Specifically, similarly sized but very different raptors (*Falco sp.* and *Accipiter sp.*) elicited similar responses in these species (Martínez et al. 2017). This is consistent with predator responses in broadly different systems

(see Templeton et al. 2005, Preisser and Orrock 2012). Additionally, alarm calls elicited by different replicas of our predator models have generated similar responses from these target species (Tigreros-Andrade unpublished data).

Trials were conducted between 06:00 and 12:00 h to minimize variation in flock activity, weather conditions were stable across the whole field season and were recorded at the start of each trial. Specifically, a single researcher located Dusky-throated Antshrike flocks. Then, we started following them carefully so as not to reveal our presence and alter their behavior. Next, the designated observer made notes on flock composition, recorded their GPS location, and quantified baseline calling behavior for 10 minutes prior to starting an experiment. After 10 minutes of baseline observation (pre-trial), the predator stimulus (real or model) was deployed in a predetermined random order. The reaction elicited by the treatment was also recorded for 10 minutes (post-trial). Live hawk and model presentations were discarded if the flight distance travelled from the release point to the landing/perch was shorter than 15 m for either the model or the hawk. We waited at least 48 hours between consecutive trials on a flock to minimize habituation. Most flocks received one to two trials per treatment; however, two flocks were tested across four trials each (two per treatment), and one flock received five trials (three hawk, two model) over the course of the study period. This remains conservative with respect to habituation, given prior experiments conducted in this system (Martínez et al. 2017, Camerlenghi et al. 2019). Several flocks contributed trials of only one treatment due to logistical constraints during fieldwork, including difficulty relocating certain flocks across the study period.

Statistical analyses

All recorded alarm calls were processed using Raven Pro 1.6.5 (K. Lisa Yang Center for Conservation Bioacoustics 2026). We focused on the first minute following the predator or

model presentation. To ask whether the experimental treatment modified vocal characteristics or the community response, we fitted a series of linear mixed-effects models. We focused on five continuous acoustic variables: frequency 75%, bandwidth 50%, alarm duration 90%, latency from the flight onset to the first alarm call, and the number of 'urgent' notes in each alarm call.

We chose these measurements since some studies have documented that some alarm calls are high frequency signals (see Templeton et al. 2005), that frequency bandwidth and alarm duration vary with calls emitted in response to the presence of predators (Templeton et al. 2005, Sieving et al. 2010), and that the number of urgent notes in the alarm calls encodes the magnitude of threat (Templeton et al. 2005, Martinez et al. 2017). Particularly, we used these more robust measurements from Raven Pro, which take into account the energy of the signal within the selection box based on percentiles (frequency 75%, bandwidth 50% and alarm duration 90%) because they provide a more reliable estimation of acoustic parameters when recording distance from the microphone to the vocalizing birds is not standardized and by the bias that may be caused using the by-eye method when creating the boxes in the software (Ríos-Chelén et al. 2017, Wierucka et al. 2021, Dharmasiri et al. 2022). For example, frequency 75% is the frequency below which 75% of the signal's energy occurs within the selection; bandwidth 50% is the frequency range that contains the central 50% of energy; and duration 90% is the time window containing 90% of the signal energy. The researcher extracting the acoustic measurements in Raven Pro (A. Tigreros) was not blind to treatment conditions, as the identity of the stimuli was registered alongside each audio file. While this could introduce subtle observer bias in selection-box placement, we mitigated this risk by using the above-mentioned energy-percentile measurements, which are more objective and less susceptible to observer-driven variation than by-eye measurement methods.

All but one of our focal species are sub-oscines, which are known to have limited variation in frequency range within their call structures; nonetheless, given the importance of frequency metrics in oscines and the fact that some variation in frequency structure exists within calls within sub-oscines, we compare frequency metrics to evaluate whether they may merit further exploration in the future.

For each model, fixed effects were predator type and species, adding an interaction term between species and predator type, and flock identity was a random effect. We used lmerTest (Kuznetsova et al. 2017) in R to estimate *P*-values using Satterthwaite's approximation for degrees of freedom. To analyse the proportion of species within each flock that produced alarm calls, we used a beta regression model with a logit link function implemented via the glmmTMB package. The model included the same fixed and random effects as the LMMs. Due to some flocks contributing multiple trials to the dataset, observations were not fully independent at the trial level. We accounted for this non-independence by including Flock ID as a random intercept in all mixed models, which accounted for between-flock variance and allowed within-flock repeated measures to be modelled appropriately. Trial-level independence within flocks was supported by the minimum 48-h interval between consecutive trials on the same flock.

For the Latency and the Community Proportion models, the species variable represents the identity of the first species to produce an alarm call for each trial, included as a fixed effect to control for potential differences in community responses depending on which species initiated the alarm responses. This metric reflects participation at the species level. Specifically, the proportion of core species (not individual birds) present in the flock during the pre-trial period that produced at least one alarm call during the post-trial period. It therefore captures the

participation of species-level engagement rather than individual call rate. For the acoustic models, species refers to the identity of the focal vocalizing individual.

Furthermore, to assess whether the model predator elicited similar responses to the real raptor across species, we performed Type III ANOVA tests on each of the above-mentioned linear mixed models, excluding the one analysing the proportion of the community emitting alarm calls. This approach allowed us to test for overall treatment effects and interactions while controlling for known species-level variation in alarm behavior. Type III sums of squares were chosen because it evaluates each effect after accounting for all others in the model, which is appropriate when interaction terms are present, and cell sizes are unequal across species-by-treatment combinations (Langsrud 2003, Shaw and Mitchell-Olds 1993). We also compared whether the flight distance between the two treatments was similar using a Welch's *t*-test. All analyses were performed using the lme4 (Bates et al. 2015) and glmmTMB (Brooks et al. 2017) packages in R (R Core Team 2025). We report model coefficient estimates ± SE alongside the 95% confidence intervals (CIs) and *P*-values for all fixed effects. Effect sizes for all models are reported as marginal R^2 (R^2m, variance explained by fixed effects only) and conditional R^2 (R^2c, variance explained by fixed and random effects combined), calculated following Nakagawa and Schielzeth (2013) using the MuMIn package for LMMs (Bartoń 2024) and the performance package for the beta regression model (Lüdecke et al. 2021). All response variables were approximately normally distributed with no flagrant violations of homogeneity of variance assumptions, as verified by visual inspection of DHARMa residual diagnostics and confirmed by dispersion and outlier tests (Hartig 2024). The sole exception was Alarm Duration 90%, which showed left-skewed residuals. To meet model assumptions, this variable was first log-transformed and then subjected to a Box-Cox transformation using the optimal lambda ($\lambda = 0.29$)

estimated via maximum likelihood using the EnvStats package (Millard, 2013). Values were shifted to positive numbers before transformation. The resulting values were scaled and centered (z-transformation) before model fitting. All other response variables were entered into models without transformation.

RESULTS

In total, our final dataset comprised 45 trials across 24 flocks, with 20 trials using the painted model predator and 25 trials using the real predator. One trial could not be assigned to a specific flock due to the absence of a GPS record and was excluded from all analyses. Flock ID was included as a random intercept in all models to account for repeated measures, and variation in the number of trials contributed per flock. We also compared the flight distances (or approach distances) of both treatments, defined as the distance that the predator (live hawk or model) travelled from its release point to its landing or perch location, using a t-test in which we found that there is no significant difference between the hawk and the model (Welch's $t = 0.51$, $df = 21.27$, $P = 0.613$) with averages of 23.42 and 22.36 m respectively.

Overall, there were no substantial differences in how the four key species responded to the hawk vs. the model (Tables 1 and 2). Significant effects for each Species were found across different parameters in the models for frequency 75%, bandwidth 50%, and alarm duration 90%, which indicates that the baseline alarm call structure differs across species regardless of predator type, a pattern consistent with species-specific acoustic call structure.

Table 1. Fixed-effect coefficient estimates from linear mixed models (LMMs) and a beta regression model evaluating the effects of predator treatment, species identity, and their

interaction on alarm response metrics in mixed-species bird flocks. Flock ID was included as a random intercept in all models. Bold *P*-values indicate statistical significance ($P < 0.05$). R^2m = marginal R^2 (fixed effects only); R^2c = conditional R^2 (fixed + random effects). † Alarm Duration 90% coefficients are in standardized units following log + Box-Cox transformation ($\lambda = 0.29$) and z-scoring. THAR = *Thamnomanes ardesiacus* (reference); MYME = *Myrmotherula menetriesii*; MYAX = *M. axillaris*; HARU = *Habia rubica*.

Response Variable	Predictor	Estimate	SE	[2.5, 97.5%] CI	*P*	R^2m	R^2c
Frequency 75%	Intercept	5705.73	148.11	[5415.4, 5996.1]	**<0.001**	0.65	0.78
	MYME	−506.80	167.29	[−834.7, −178.9]	**0.003**		
	MYAX	−758.97	237.17	[−1223.8, −294.1]	**0.002**		
	HARU	−2258.16	193.2	[−2636.8, −1879.5]	**<0.001**		
	Model	2.58	164.13	[−319.1, 324.3]	0.987		
	MYME × Model	−614.39	320.35	[−1242.3, 13.5]	0.057		
	MYAX × Model	−180.17	386.74	[−938.2, 577.8]	0.642		
	HARU × Model	−304.40	284.87	[−862.7, 254.9]	0.287		
Bandwidth 50%	Intercept	990.8	94.69	[805.2, 1176.4]	**<0.001**	0.11	0.39
	MYME	333.87	109.78	[118.7, 549.0]	**0.003**		
	MYAX	195.64	156.16	[−110.4, 501.7]	0.212		
	HARU	57.28	126.41	[−190.5, 305.1]	0.651		

	Model	181.32	107.34	[−29.1, 391.7]	0.094		
	MYME × Model	−185.11	211.14	[−598.9, 228.7]	0.382		
	MYAX × Model	−149.69	254.84	[−649.2, 349.8]	0.558		
	HARU × Model	−372.84	187.29	[−740.0, −5.7]	**0.048**		
Duration 90% †	Intercept	-0.58	0.153	[-0.879, -0.278]	**<0.001**	0.58	0.73
	MYME	1.09	0.174	[0.752, 1.434]	**<0.001**		
	MYAX	1.67	0.247	[1.186, 2.153]	**<0.001**		
	HARU	1.55	0.201	[1.154, 1.942]	**<0.001**		
	Model	-0.42	0.171	[−0.759, -0.091]	**0.014**		
	MYME × Model	0.87	0.333	[0.222, 1.529]	**0.009**		
	MYAX × Model	0.09	0.402	[−0.701, 0.877]	0.827		
	HARU × Model	0.27	0.296	[−0.313, 0.848]	0.368		
Latency	Intercept	0.58	0.134	[0.317, 0.843]	**<0.001**	0.42	0.65
	MYAX	−0.342	0.255	[−0.842, 0.158]	0.194		
	HARU	−0.064	0.172	[−0.401, 0.273]	0.716		
	Model	−0.091	0.165	[−0.414, 0.232]	0.591		
	MYAX × Model	1.123	0.305	[0.525, 1.721]	**0.002**		
	HARU × Model	−0.062	0.234	[−0.521, 0.397]	0.795		
Number of Urgent Notes	Intercept	7.818	1.579	[4.723, 10.913]	**<0.001**	0.21	0.21

	MYME	3.432	3.059	[−2.564, 9.428]	0.267		
	MYAX	5.182	2.826	[−0.357, 10.721]	0.072		
	HARU	−2.735	2.187	[−7.021, 1.551]	0.216		
	Model	−4.318	2.289	[−8.804, 0.168]	0.052		
	MYME × Model	−0.932	6.289	[−13.258, 11.394]	0.883		
	MYAX × Model	−1.253	3.828	[−8.756, 6.250]	0.745		
	HARU × Model	4.901	3.133	[−1.240, 11.042]	0.123		
Proportion of Flock Species Alarming	Intercept	−0.494	0.272	[−1.027, 0.039]	0.069	0.26	0.65
	MYAX	0.487	0.5	[−0.493, 1.467]	0.33		
	HARU	−0.256	0.335	[−0.913, 0.401]	0.445		
	Model	−0.022	0.339	[−0.686, 0.642]	0.949		
	MYAX × Model	−1.191	0.672	[−2.508, 0.126]	0.076		
	HARU × Model	0.246	0.466	[−0.667, 1.159]	0.599		

None of the four species showed any differences with respect to alarm call frequency 75% between the model and the live predator independently when comparing the interactions across species with treatments (Figure 1, Table 1). Nonetheless, when comparing treatments based on the ANOVA results (Table 2, P <0.05), we observed a small but statistically significant overall difference between both treatments.

Table 2. Type III ANOVA results testing effects of species, treatment, and their interaction on alarm response metrics. Results were obtained from running this test on the executed linear mixed models with Species and Predator Treatment (*Experiment*) as fixed effects and the Flock ID

as a random effect. Species evaluated in these ANOVAs are the same four core species from the linear models: THAR: *Thamnomanes ardesiacus* (Dusky-throated Antshrike), MYME: *Myrmotherula menetriesii* (Gray Antwren), MYAX: *M. axillaris* (White-flanked Antwren) and HARU: *Habia rubica* (Red-crowned Ant-Tanager).

Response Variable	**Effect**	**NumDF**	**DenDF**	***P*-value**
Frequency 75%	Species	3	155.03	**<0.001**
	Treatment	1	159.81	**0.047**
	Species*Treatment	3	158.07	0.262
Bandwidth 50%	Species	3	151.76	**0.014**
	Treatment	1	159.12	0.961
	Species*Treatment	3	156.89	0.254
Duration 90%	Species	3	153.97	**<0.001**
	Treatment	1	159.71	0.408
	Species*Treatment	3	157.68	0.071
Latency to Alarm	Species	2	19.643	0.247
	Treatment	1	12.94	**0.037**
	Species*Treatment	2	17.379	**0.003**
Number of Urgent Notes	Species	3	69	0.056
	Treatment	1	69	0.052
	Species*Treatment	3	69	0.282

We also examined whether the bandwidth 50% in alarm calls differed between treatments; it did not differ between treatments (Figure 2; Table 2, $P = 0.961$) and the Species*Treatment interaction was non-significant ($P = 0.254$). For Red-crowned Ant-tanager, White-flanked Antwren and Gray Antwren, frequency range values were tightly distributed and did not differ between treatments, reflecting consistent alarm structure regardless of predator type. For Dusky-

throated Antshrike, although the values for bandwidth were broader in the real predator treatment, the medians and overall distributions remained comparable.

While alarm call duration 90% differed markedly among the four species, it did not differ between treatments (Figure 3; Table 2, P = 0.408). However, the latency to first call following the presentation differed for one species; White-flanked Antwrens called significantly more rapidly following the presentation of the live hawk compared to the model (Figure 4; Table 2, P = 0.003). The significant effect of Treatment reflected a general tendency for faster alarm responses to the live hawk overall, but this pattern was not uniform across species. The significant Species * Treatment interaction indicated that this trend was driven primarily by this species, which called substantially more rapidly following the Hawk presentation (see Table 1, *Myrmotherula axillaris*: Model estimate = 1.123). For the other species, latency did not differ significantly between treatments. Additionally, we observed no statistically significant difference between the two treatments for the interaction with the different species when evaluating the number of urgent notes emitted per alarm call for each treatment and species, although the main effect of treatment approached a conventional threshold, suggesting a possible marginal tendency toward fewer urgent notes in model trials (Figure 5, Table 2, P = 0.052).

To evaluate whether the presence of a real versus simulated predator influenced overall alarm activity within flocks, we calculated the proportion of the flock species that vocalized in each trial overall (Figure 6). Community size was defined as the number of core species detected during the pre-trial period (a species-level metric, not individual count). The beta regression model of the proportion of the flock community that joined the alarm response showed no significant treatment effect ($\beta = -0.022 \pm 0.339$, P = 0.949).

On average, about one-third of the flocking species responded with alarm calls, and this proportion did not differ between treatments. This result suggests that flock-level engagement in alarm behavior is similar whether a real raptor or a model is used. These results indicate that the likelihood of community-wide vocal engagement was not influenced by the predator treatment or by which species first vocalized.

DISCUSSION

Our study aimed to evaluate whether simulated predator threats using painted raptor models can elicit "realistic" alarm-calling behavior in mixed-species Amazonian bird flocks, compared to the response elicited by a real predatory bird. Our findings largely support the hypothesis that acoustic characteristics were similar, but contradicted our hypothesis for flock-wide responses. Both treatments, the live-trained raptor and the model, elicited consistent alarm responses with no statistically significant differences across major acoustic and behavioral metrics, including frequency structure, alarm duration, latency for most species (the only statistically significant difference was for White-flanked Antwrens), number of urgent notes, and the proportion of community response. In addition, and contrary to what we expected, even the overall level of response of the flock community was similar between treatments. Overall, these results provide empirical support, within the scope of the acoustic and immediate behavioral responses measured here, for the continued use of predator models as a cost-effective and replicable tool in behavioral ecology research, particularly when investigating predator-prey dynamics and alarm communication in Amazonian understory mixed-species bird flock patterns as those dimensioned in this study.

Alarm signaling is a critical component in the collective predator detection in animal communities and also in the information sharing in mixed flocks (Magrath et al. 2015a), and to study these dynamics, it's often essential to simulate predatory attacks. Our findings showed that both treatments triggered similar vocal responses in the four most common species in the Amazonian understory mixed-species flocks. These results are consistent with prior reports that showed that predator models may elicit reliable and natural anti-predator behavior in birds (Griesser 2008, Soard and Ritchison 2009). The lack of differences in latency to call and the proportion of flock members responding vocally further support the idea that models can be used to simulate realistic predation events. These findings align with validation efforts in other taxa, including mammalian decoy-camera setups (Boone et al. 2023) and less powerful model predator validations that elicit mammalian alarm calls (Blumstein and Armitage 1997).

Moreover, the similarities between the responses of the four focal alarming species and the proportion of the community alarming also suggest that heterospecific communication and networked vigilance may remain functionally similar even when a model was used. This supports previous work showing that species in mixed flocks, including both sentinels and followers, coordinate their behavior in response to shared threats (Goodale and Kotagama 2005) and that these alarm networks are robust to variation in the nature of threat cues (Vander Meiden et al. 2023). The lack of statistical difference in bandwidth, a metric thought to reflect arousal or urgency in alarm contexts (Morton 1977), duration, and in the latency to first call (for most of the species), which can indicate either perception delay or signal decision-making (Caro 2005), suggests model exposure effectively simulates the behavioral cues necessary for predator detection and response. For urgent notes, although a marginal trend toward fewer notes in model trials was observed, this did not translate into species-specific differences in overall response

intensity. The consistent species-level differences in alarm call structure (frequency, bandwidth, duration) likely reflect ecological differentiation in acoustic adaptations.

While our results validate the use of models to study the behavioral response by mixed flocks to predation events, we caution their use to understand large-scale effects on the community. Our experiments only evaluated the responses of the birds in close proximity to the focal species in the mixed flocks. Thus, we were unable to assess differences between the model and the live raptor in the larger community. Future work could deploy autonomous audio recorders to assess whether differences in the calls elicited by models and live raptors are present at larger spatial scales. Longitudinal studies could explore habituation patterns over time, whether birds respond less to models with repeated exposure compared to live predators, and how these dynamics vary across different forested ecosystems. Expanding these studies globally could also assess the generality of our findings across diverse predator-prey systems.

In our model, we included Flock ID as a random intercept to account for the repeated measures taken within each flock. However, it's important to note that all hawk trials involved the same trained individual. The individual-level variation in a real predator's behavior may significantly influence prey responses, presenting a limitation in our study. However, responses to predators are very likely highly conserved, such that similar responses are likely to be produced regardless of individual variation. In our system, birds have similarly responded to large pigeons and woodpeckers flying suddenly through the flock (pers. obs.), and in previous studies, we have shown that very different but similarly sized predators (*Accipiter sp.* versus *Falco sp.*) generate statistically similar responses. Nevertheless, this study was primarily focused on short-term and immediate behavioral responses occurring within the first minute following predator presentation. We did not assess long-term vigilance adjustments, alterations in foraging

patterns, spatial redistribution of flock members, or other relevant factors. Furthermore, model predators may differ from live predators in their ability to elicit these long-term responses, and future studies should explicitly explore these dimensions.

Whether other taxa beyond birds, such as mammals, amphibians, or fish, respond equivalently to predator models remains an open question, as discrimination ability may depend on cognitive complexity and specific sensory cues available in each system. While evidence in mammals (Blumstein and Armitage 1997, Blumstein et al. 2009) and fish (Catano et al. 2016) suggests that models can simulate real threats, the ability to discriminate between models and live threats may depend on cognitive complexity or evolutionary pressures within each group. Our study contributes to the validation of a widely used methodology in behavioral ecology, supporting the use of standardized, cost-effective tools in research on communication about predators. Precisely, by confirming the reliability of predator models in eliciting natural responses, we open avenues for exploring communication networks around predators, predator recognition, and information flow within mixed-species flock communities, specifically for Neotropical flocks, advancing both theoretical and applied research on alarm communication and anti-predator strategies.

DATA AVAILABILITY

All the data is accessible at https://doi.org/10.5281/zenodo.17373175

ACKNOWLEDGEMENTS

We thank Danny Muñoz for his valuable comments and recommendations on how to analyse acoustic data in RavenPro, and Justin Baldwin for his thoughtful comments on this manuscript's structure, figures, and overall presentation. We are also deeply grateful to Marita Vargas and David Camargo Vargas

from Villaana for their kind support and assistance during fieldwork on their property. To Noah Arthur and Seleona Hunter – Seletino for the illustrations included in the graphical abstract. And finally, we extend our appreciation to Colka Raptors SAC and to "Lola" (our trained raptor) for her dedication and effort in her flights during our experimental trials.

FUNDING

This work was supported by funding to DTB NSF IOS-2032478 and AEM NSF IOS-NSF-2032469.

ETHICS STATEMENT

This research was conducted under the research permit SERFOR No. AUT-IFS-2022-051, issued by the Servicio Nacional Forestal y de Fauna Silvestre (SERFOR) of the Peruvian Ministry of the Environment (Ministerio del Ambiente, MINAM). The study followed all local regulations for non-invasive wildlife research and complied with the ASAB/ABS Guidelines for the ethical treatment of nonhuman animals in behavioral research and teaching and with all relevant institutional requirements. No birds were captured, handled, or harmed during the experiments. Observations were conducted on free-ranging wild birds using a trained raptor and a painted model predator, both of which posed no physical risk to the subjects. The research design, therefore, had no welfare or environmental implications for the study species or their habitat.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGIES

The authors used AI-assisted tools (Claude) for statistical code review, and figure formatting suggestions. All scientific interpretation, data analysis decisions, and final text are the responsibility of the authors.

AUTHOR'S CONTRIBUTIONS

Andrés Felipe Tigreros-Andrade and Ari E. Martínez conceptualized the study (Equal) and contributed to methodology development (Equal); Andrés Felipe Tigreros-Andrade, Michael Gulman, Jorge Novoa and Ari E. Martínez conducted fieldwork and data collection; Andrés Felipe Tigreros-Andrade led data curation (Lead), formal analysis (Lead), and writing of the original draft (Lead); Eliseo Parra contributed to methodology (Equal), formal analysis (Supporting), review and editing (Supporting); Daniel T. Blumstein contributed to conceptualization (Supporting), funding acquisition (Equal), supervision (Supporting), and writing, review and editing (Equal); Ari Martínez contributed to conceptualization (Equal), methodology (Equal), supervision (Lead), funding acquisition (Equal), and writing, review and editing (Equal).

REFERENCES

Bartoń K, 2024. MuMIn: multi-model inference. R package version 1.48.4. Available from: https://CRAN.R-project.org/package=MuMIn.

Batabyal A, 2023. Predator–prey systems as models for integrative research in biology: the value of a non-consumptive effects framework. *J Exp Biol* 226(19):jeb245851.

Bates D, Mächler M, Bolker B, Walker S, 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1).

Blumstein DT, Armitage KB, 1997. Alarm calling in yellow-bellied marmots: I. The meaning of situationally variable alarm calls. *Anim Behav* 53:143–171.

Blumstein DT, Ferando E, Stankowich T, 2009. A test of the multipredator hypothesis: yellow-bellied marmots respond fearfully to the sight of novel and extinct predators. *Anim Behav* 78(4):873–878.

Boone HM, Pacifici K, Moorman CE, Kays R, 2023. Using models and camera traps to estimate depredation rates and neonate survival. *PLoS One* 18(10):e0293328.

Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW et al., 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal* 9(2): 378–400.

Browning SA, 2015. Mixed-species flock members' reactions to novel and predator stimuli. PhD diss. University of Tennessee. Available from: https://trace.tennessee.edu/utk_graddiss/3327.

Camerlenghi E, Tellaroli P, Griggio M, Martínez AE, 2019. Information about predators varies across an Amazonian rain forest as a result of sentinel species distribution. *Am Nat* 194(5):E134–E139.

Caro TM, 2005. Antipredator defenses in birds and mammals. Chicago (IL): University of Chicago Press.

Catano LB, Rojas MC, Malossi RJ, Peters JR, Heithaus MR, Fourqurean JW, Burkepile DE, 2016. Reefscapes of fear: predation risk and reef heterogeneity interact to shape herbivore foraging behaviour. *J Anim Ecol* 85(1):146–156.

Curio E, Ernst U, Vieth W, 1978. The adaptive significance of avian mobbing: II. Cultural transmission of enemy recognition in blackbirds: effectiveness and some constraints. *Z Tierpsychol* 48(2):184–202.

Dharmasiri ME, Barber CA, Horn AG, 2022. Nestling European starlings (*Sturnus vulgaris*) adjust their begging calls in noise. *Bioacoustics* 31(5):594–613.

Dreelin RA, Jones HP, Gutiérrez-Carrillo DA, Piper C, Goodman BD et al., 2025. Interspecific eavesdropping by birds on a keystone mammal improves antipredator behaviour. *Anim Behav* 225: 123219.

Ellard CG, Byers RD, 2005. The influence of the behaviour of conspecifics on responses to threat in the Mongolian gerbil (*Meriones unguiculatus*). *Anim Behav* 70(1):49–58.

Frazier EK, Selman ZA, Price CA, Papeş M, Freeberg TM, 2025. Mixed-species flock diversity and habitat density are associated with antipredator behavior in songbirds. *Diversity* 17(5):363.

Gil MA, Hein AM, 2017. Social interactions among grazing reef fish drive material flux in a coral reef ecosystem. *Proc Natl Acad Sci* USA 114(18):4703–4708.

Goodale E, Kotagama SW, 2005. Alarm calling in Sri Lankan mixed-species bird flocks. *Auk* 122(1):108–120.

Griesser M, 2008. Referential calls signal predator behavior in a group-living bird species. *Curr Biol* 18(1):69–73.

Hartig F, 2024. *DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models*. The R Project for Statistical Computing: Vienna, Austria.

Hegner RE, 1985. Dominance and anti-predator behaviour in blue tits (*Parus caeruleus*). *Anim Behav* 33(3):762–768.

Jiang D, Sieving KE, Meaux E, Goodale E, 2020. Seasonal changes in mixed-species bird flocks and antipredator information. *Ecol Evol* 10(12):5368–5382.

K. Lisa Yang Center for Conservation Bioacoustics, 2026. *Raven Pro: Interactive Sound Analysis Software (Version 1.6.5)*. Ithaca (NY): Cornell Lab of Ornithology. Available from: https://www.ravensoundsoftware.com/.

Kuznetsova A, Brockhoff PB, Christensen RHB, 2017. lmerTest package: tests in linear mixed effects models. *J Stat Softw* 82(13):1–26.

Langsrud Ø, 2003. ANOVA for unbalanced data: use Type II instead of Type III sums of squares. *Stat Comput* 13(2):163–167.

Lima SL, Dill LM, 1990. Behavioral decisions made under the risk of predation: a review and prospectus. *Can J Zool* 68(4):619–640.

Lüdecke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D, 2021. performance: an R package for assessment, comparison and testing of statistical models. *J Open Source Softw* 6(60):3139.

Magrath RD, Haff TM, McLachlan JR, Igic B, 2015a. Wild birds learn to eavesdrop on heterospecific alarm calls. *Curr Biol* 25(15):2047–2050.

Magrath RD, Haff TM, Fallow PM, Radford AN, 2015b. Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. *Biol Rev* 90(2):560–586.

Mangini GG, Rutt CL, Sridhar H, Buitron G, Muñoz J et al., 2023. A classification scheme for mixed-species bird flocks. *Philos Trans R Soc B Biol Sci* 378(1878): 20220100.

Martínez AE, Parra E, Collado LF, Vredenburg VT, 2017. Deconstructing the landscape of fear in stable multi-species societies. *Ecology* 98(9):2447–2455.

Martínez AE, Zenil RT, 2012. Foraging guild influences dependence on heterospecific alarm calls in Amazonian bird flocks. *Behav Ecol* 23(3):544–550.

Millard, SP, 2013. *EnvStats: An R Package for Environmental Statistics*. Springer, New York.

Mogali SM, Saidapur SK, Shanbhag BA, 2011. Levels of predation modulate antipredator defense behavior and metamorphic traits in the toad *Bufo melanostictus*. *J Herpetol* 45: 428–431.

Morton ES, 1977. On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am Nat* 111(981):855–869.

Munn CA, Terborgh JW, 1979. Multi-species territoriality in neotropical foraging flocks. *Condor* 81(4):338.

Nakagawa S, Schielzeth H, 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol* 4(2):133–142.

Palmer MS, Packer C, 2021. Reactive anti-predator behavioral strategy shaped by predator characteristics. *PLoS One* 16(8):e0256147.

Podos J, Webster MS, 2022. Ecology and evolution of bird sounds. *Curr Biol* 32(20):R1100–R1104.

Potvin DA, Ratnayake CP, Radford AN, Magrath RD, 2018. Birds learn socially to recognize heterospecific alarm calls by acoustic association. *Curr Biol* 28(16):2632–2637.

Preisser EL, Orrock JL, 2012. The allometry of fear: interspecific relationships between body size and response to predation risk. *Ecosphere* 3(9):1–27.

Price T, Wadewitz P, Cheney D, Seyfarth R, Hammerschmidt K, Fischer J, 2015. Vervets revisited: a quantitative analysis of alarm call structure and context specificity. *Sci Rep* 5(1):13220.

R Core Team, 2025. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. https://www.R-project.org/.

Ríos-Chelén AA, McDonald AN, Berger A, Perry AC, Krakauer AH et al., 2017. Do birds vocalize at higher pitch in noise, or is it a matter of measurement? *Behav Ecol Sociobiol* 71(1): 29.

Rosier RL, Langkilde T, 2011. Behavior under risk: how animals avoid becoming dinner. *Nat Educ Knowl* 2(11):8.

Shaw RG, Mitchell-Olds T, 1993. ANOVA for unbalanced data: an overview. *Ecology* 74(6):1638–1645.

Sieving KE, Hetrick SA, Avery ML, 2010. The versatility of graded acoustic measures in classification of predation threats by the tufted titmouse *Baeolophus bicolor*: exploring a mixed framework for threat communication. *Oikos* 119(2):264–276.

Smee D, 2010. Species with a large impact on community structure. *Nat Educ Knowl* 3(10):40.

Soard CM, Ritchison G, 2009. 'Chick-a-dee' calls of Carolina chickadees convey information about degree of threat posed by avian predators. *Anim Behav* 78:1447–1453.

Templeton CN, Greene E, Davis K, 2005. Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science* 308(5730):1934–1937.

Tinbergen N, 1989. The study of instinct. Oxford (UK): Clarendon Press.

Vander Meiden LN, Shizuka D, Johnson AE, 2023. Studying individual-level interactions can transform our understanding of avian mixed-species flocks. *Ornithology* 140(2):ukad007.

Wierucka K, Henley MD, Mumby HS, 2021. Acoustic cues to individuality in wild male adult African savannah elephants (*Loxodonta africana*). *PeerJ* 9:e10736.

FIGURE LEGENDS AND ALT TEXT

Figure 1. Frequency 75% (Hz) of alarm calls across the four core alarm-calling species in mixed-species flocks, by treatment. Red violins represent responses to the live raptor (Hawk); blue violins represent responses to the painted model (Model). Violin plots display the full distribution of values with embedded boxplots showing the median and interquartile range. THAR: *Thamnomanes ardesiacus* (Dusky-throated Antshrike); MYME: *Myrmotherula menetriesii* (Gray Antwren); MYAX: *M. axillaris* (White-flanked Antwren); HARU: *Habia rubica* (Red-crowned Ant-Tanager).

Alt text: Four-panel violin plot showing alarm call frequency 75% (Hz) of alarm calls on the y-axis for each of our core flocking bird species (THAR, MYME, MYAX, HARU).

Figure 2. Bandwidth 50% (Hz) of alarm calls across four core alarm-calling species in mixed-species flocks, by treatment. Red violins represent responses to the live raptor (Hawk); blue violins represent responses to the painted model (Model). Violin plots display the full distribution of values with embedded boxplots showing the median and interquartile range. THAR: *Thamnomanes ardesiacus* (Dusky-throated Antshrike); MYME: *Myrmotherula menetriesii* (Gray Antwren); MYAX: *M. axillaris* (White-flanked Antwren); HARU: *Habia rubica* (Red-crowned Ant-Tanager).

Alt text: Four-panel violin plot showing alarm call bandwidth 50% (Hz) of alarm calls on the y-axis for each of our core flocking bird species (THAR, MYME, MYAX, HARU).

Figure 3. Alarm call duration 90% (s) across four core alarm-calling species in mixed-species flocks, by treatment. Red violins represent responses to the live raptor (Hawk); blue violins represent responses to the painted model (Model). Violin plots display the full distribution of values with embedded boxplots showing the median and interquartile range. Note that values shown are on the raw scale. THAR: *Thamnomanes ardesiacus* (Dusky-throated Antshrike); MYME: *Myrmotherula menetriesii* (Gray Antwren); MYAX: *M. axillaris* (White-flanked Antwren); HARU: *Habia rubica* (Red-crowned Ant-Tanager).

Alt text: Four-panel violin plot showing alarm call duration 90% (s) of alarm calls on the y-axis for each of our core flocking bird species (THAR, MYME, MYAX, HARU).

Figure 4. Latency (s) to first alarm call across three core alarm-calling species in mixed-species flocks, by treatment. Time is measured from the onset of the treatment to the emission of the first alarm call. Red violins represent responses to the live raptor (Hawk); blue violins represent responses to the painted model (Model). Violin plots display the full distribution of values with

embedded boxplots showing the median and interquartile range. THAR: *Thamnomanes ardesiacus* (Dusky-throated Antshrike); MYAX: *Myrmotherula axillaris* (White-flanked Antwren); HARU: *Habia rubica* (Red-crowned Ant-Tanager). MYME is not displayed because it did not initiate an alarm call in any of the trials.

Alt text: Three-panel violin plot showing latency (s) to first alarm call on the y-axis for each of our core flocking bird species (THAR, MYAX, HARU).

Figure 5. Number of urgent notes per alarm call across the four core alarm-calling species in mixed-species flocks, by treatment. Red violins represent responses to the live raptor (Hawk); blue violins represent responses to the painted model (Model). Violin plots display the full distribution of values with embedded boxplots showing the median and interquartile range. THAR: *Thamnomanes ardesiacus* (Dusky-throated Antshrike); MYME: *Myrmotherula menetriesii* (Gray Antwren); MYAX: *M. axillaris* (White-flanked Antwren); HARU: *Habia rubica* (Red-crowned Ant-Tanager).

Alt text: Four-panel violin plot showing number of urgent notes per alarm call on the y-axis for each of our core flocking bird species (HARU, MYAX, MYME, THAR).

Figure 6. Proportion of flock species producing alarm calls in response to the predator stimulus, by treatment. Each point represents a single experimental trial. Red violins represent responses to the live raptor (Hawk); blue violins represent responses to the painted model (Model). Violin plots display the full distribution of values with embedded boxplots showing the median and interquartile range.

Alt text: Single-panel violin plot showing the proportion of species producing alarm calls on the y-axis, with two violins: red for live hawk treatment and blue for model treatment.

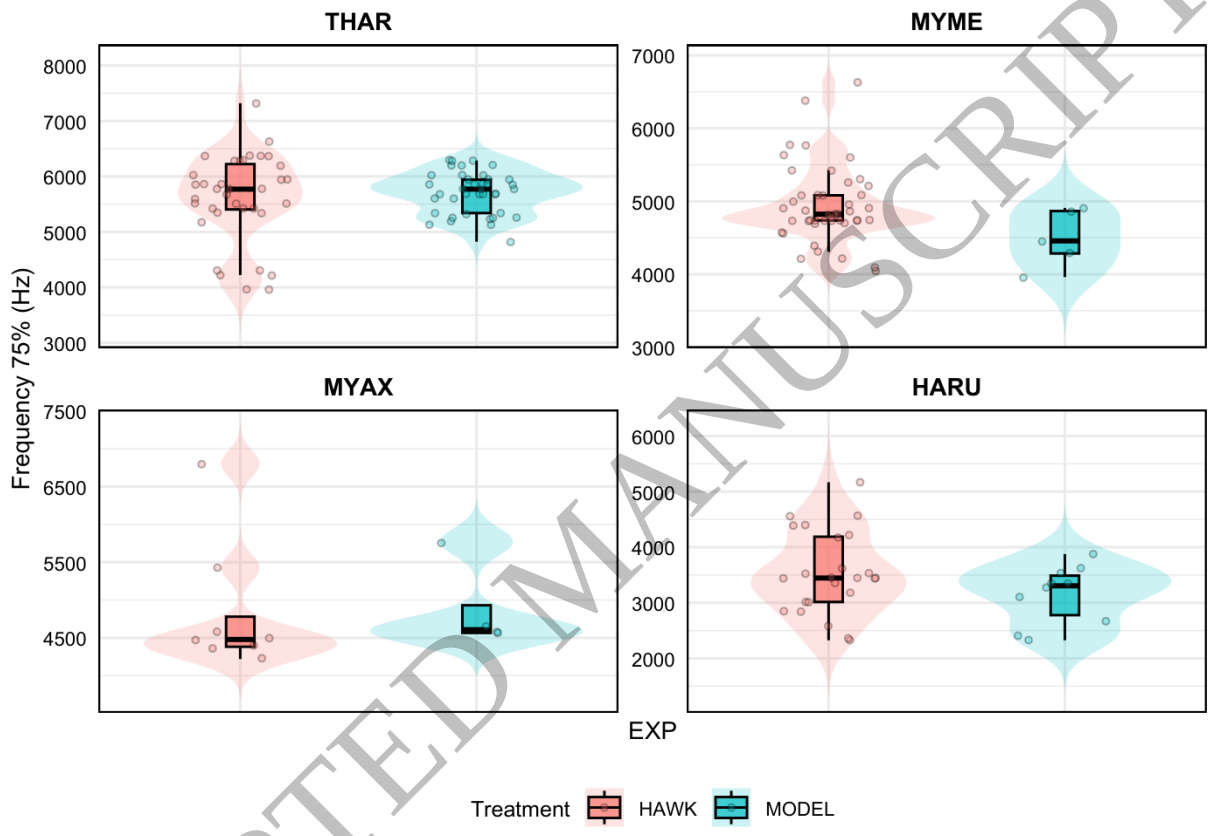


Figure 1
178x127 mm (x DPI)

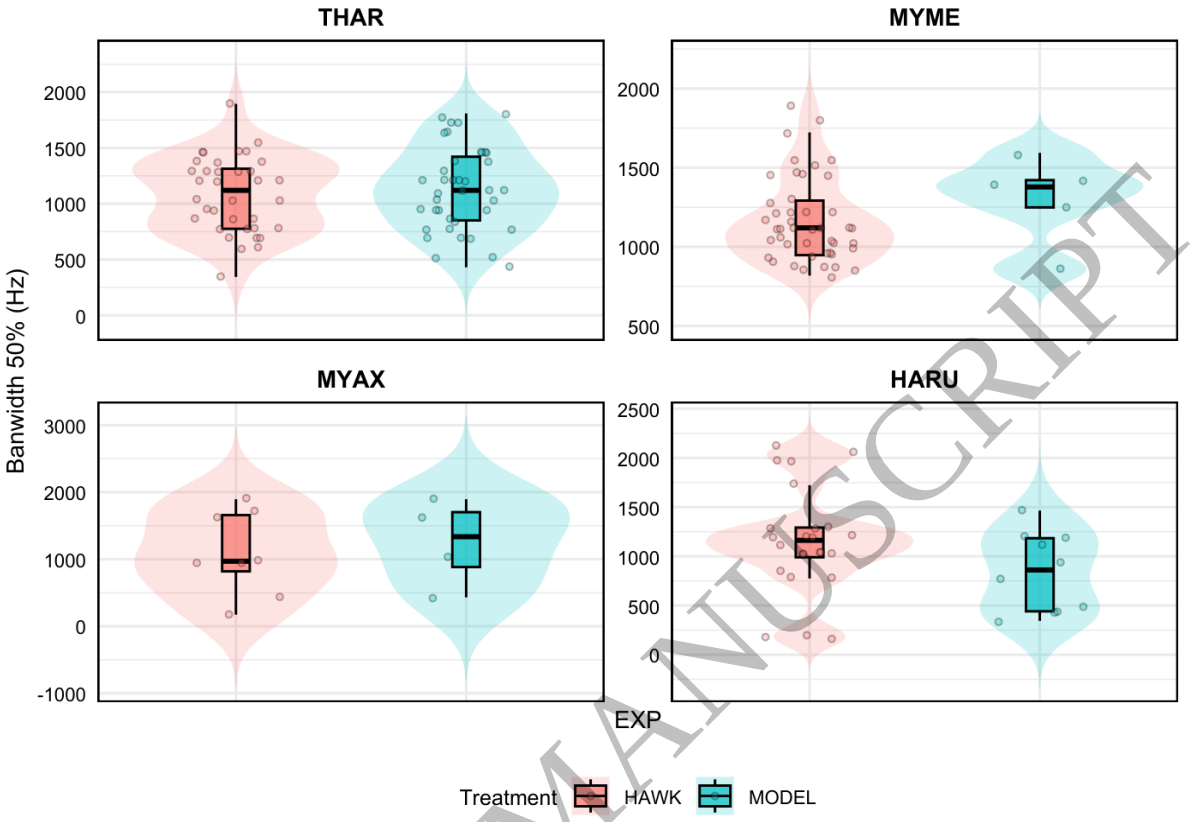


Figure 2
178x127 mm (x DPI)

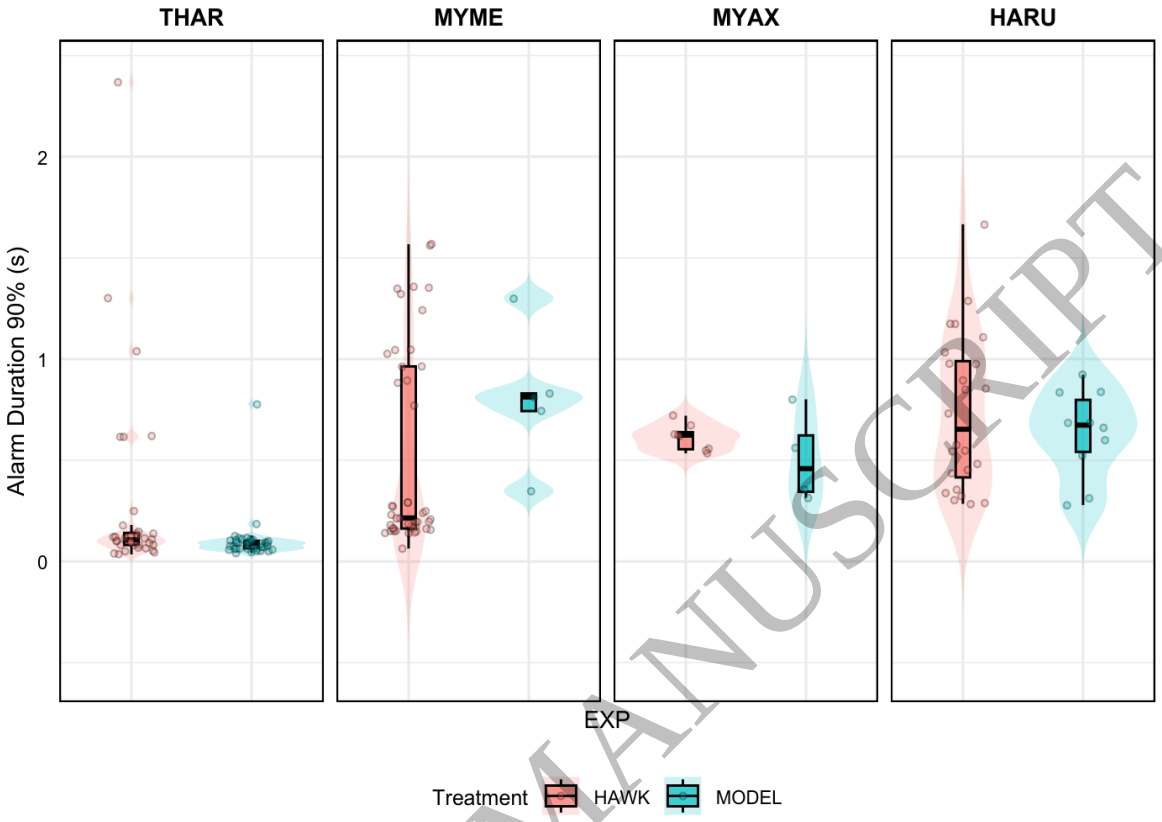


Figure 3
178x127 mm (x DPI)

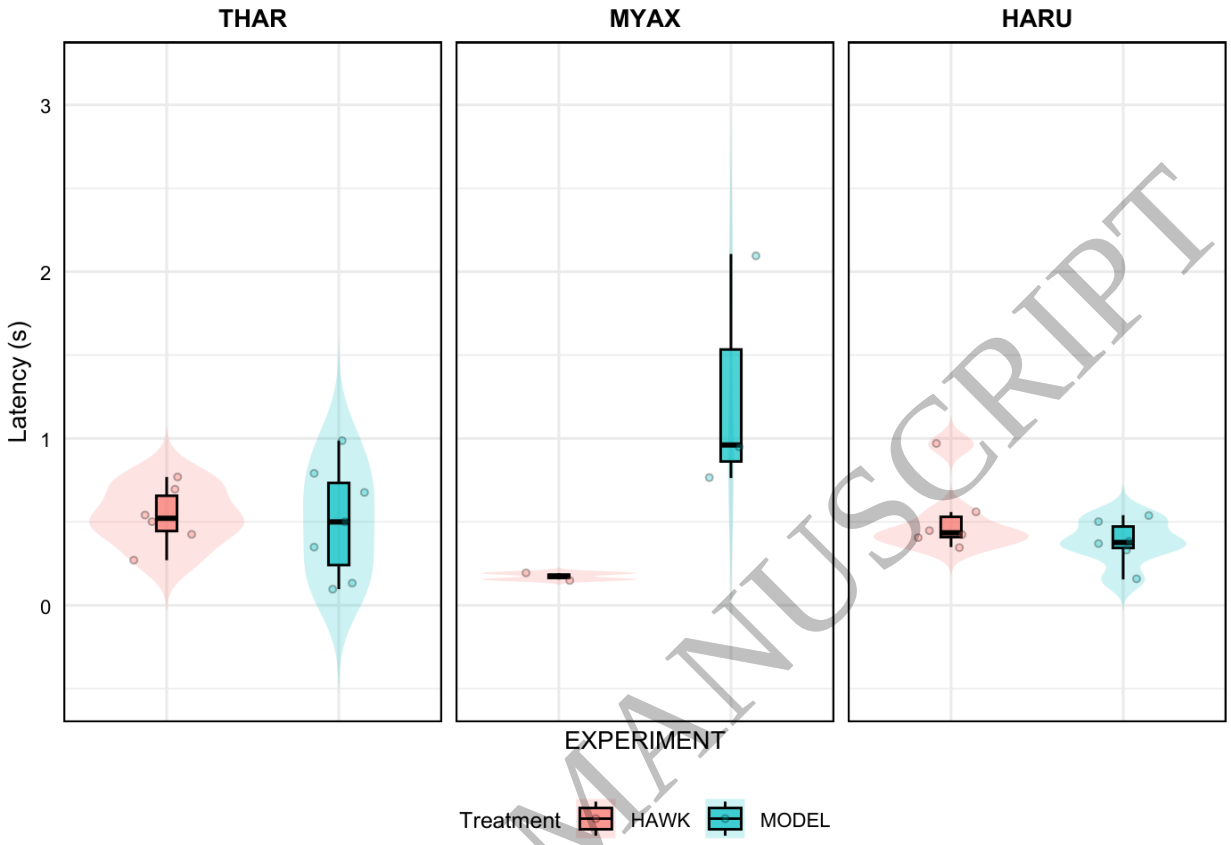


Figure 4
178x127 mm (x DPI)

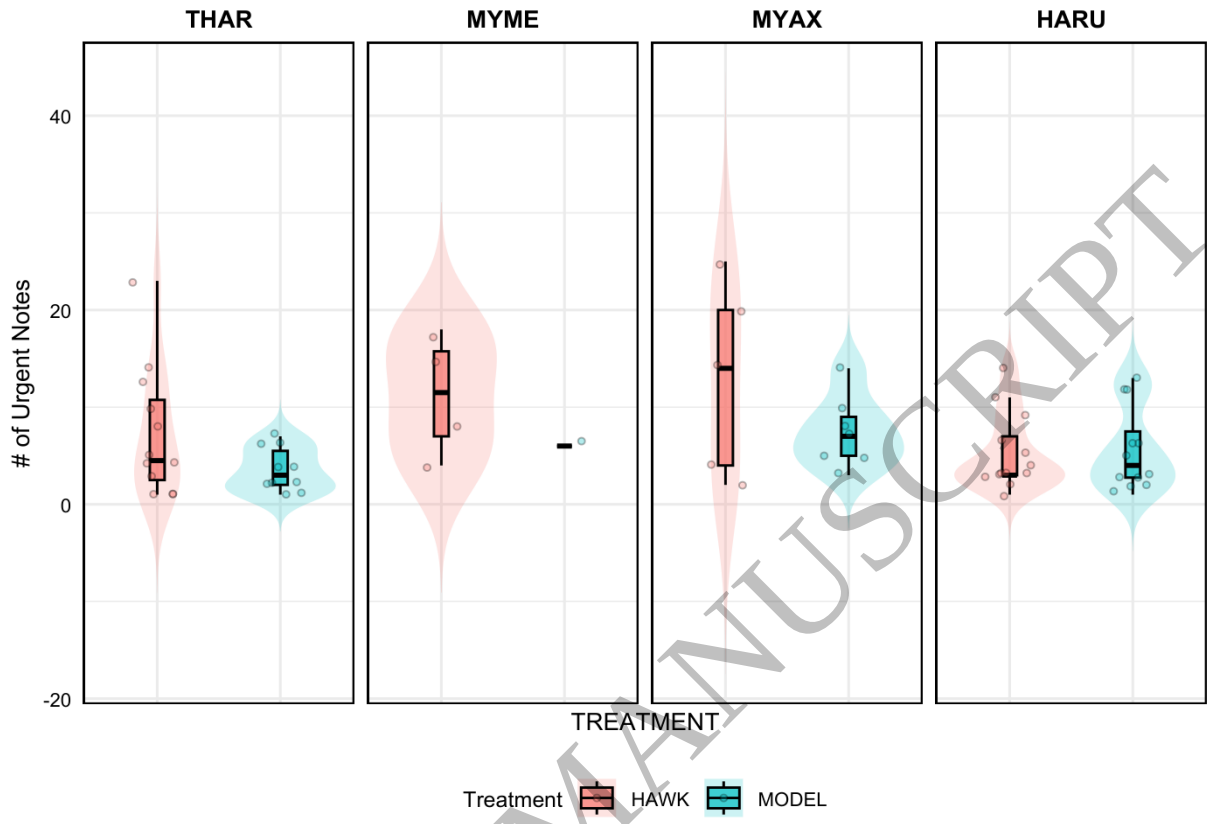


Figure 5
178x127 mm (x DPI)

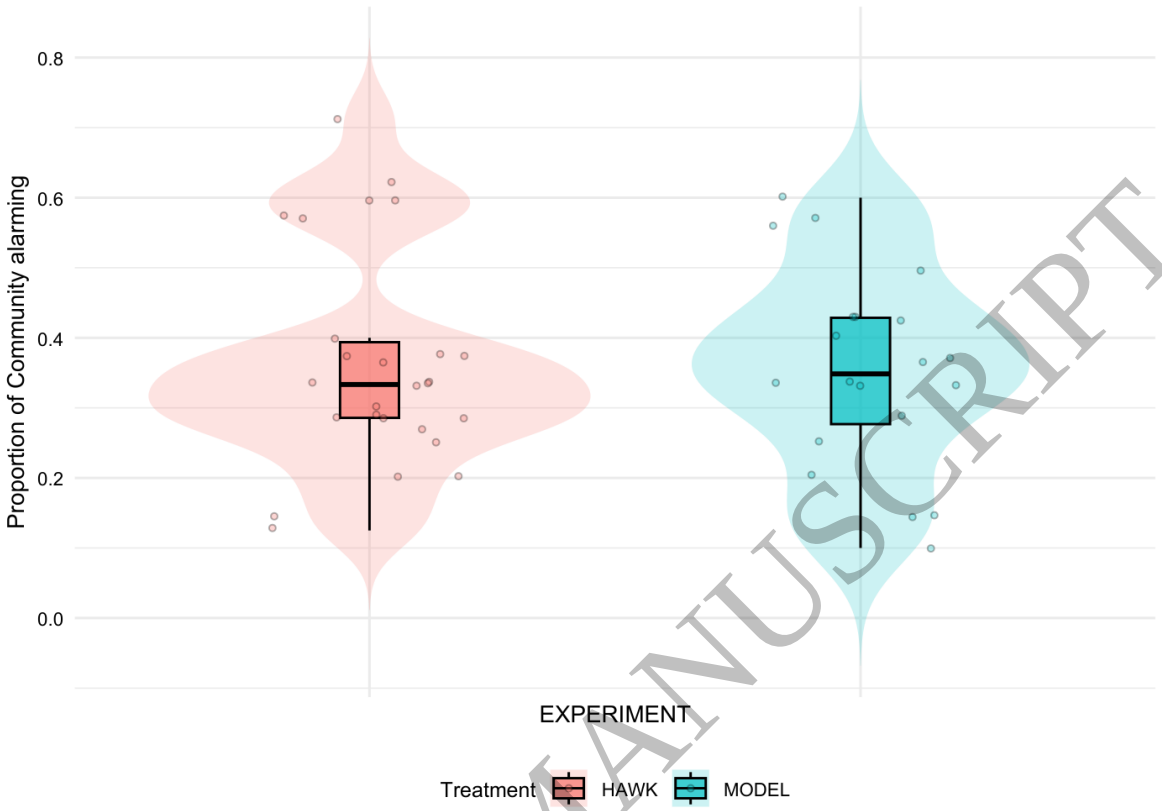
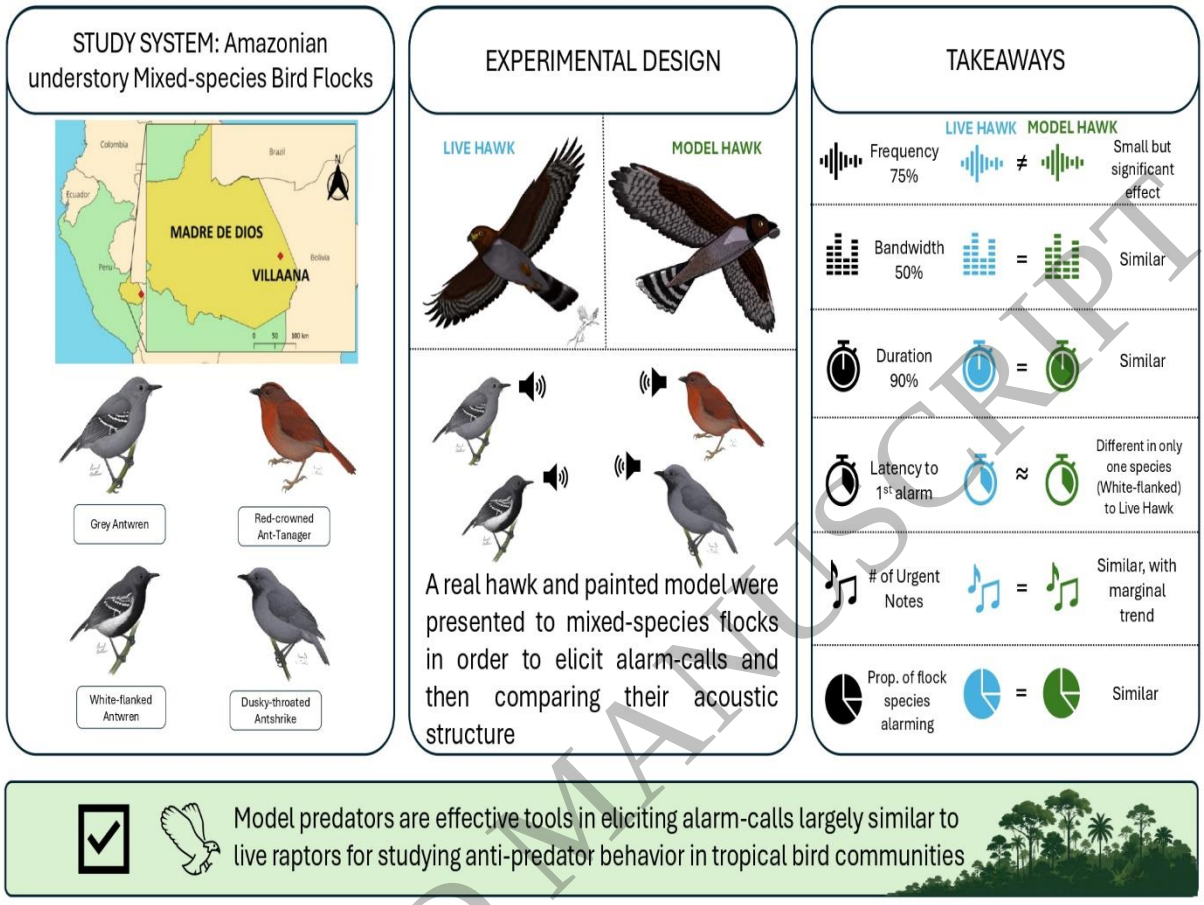


Figure 6
178x127 mm (x DPI)



Graphical Abstract
339x190 mm (x DPI)