Title

On Risk and Safety Communication in Amazonian Mixed-Species Bird Flocks: Species-Level Vocal Structure, Community States, and Agroforestry Disturbance

ISBN

9798190925296

Author

Parra, Eliseo Rafael

Publication Date

2026-09-09

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

On Risk and Safety Communication in Amazonian Mixed-Species Bird Flocks: Species-Level Vocal Structure, Community States, and Agroforestry Disturbance

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biology

by

Eliseo Rafael Parra

2026

ABSTRACT OF THE DISSERTATION

On Risk and Safety Communication in Amazonian Mixed-Species Bird Flocks:

Species-Level Vocal Structure, Community States, and Agroforestry Disturbance

by

Eliseo Rafael Parra

Doctor of Philosophy in Biology

University of California, Los Angeles, 2026

Professor Daniel T. Blumstein, Chair

Predation requires animals to balance immediate safety against the need to resume foraging, yet most studies emphasize alarm calls rather than the full transition from danger to safety. This dissertation examines how risk-related information is organized within Amazonian mixed-species bird flocks, from the vocal behavior of a sentinel species to community-level communication and its sensitivity to forest management. Standardized presentations of trained Bicolored Hawks *Accipiter bicolor* were combined with hand annotation, sequence analysis, hidden Markov models, functional-trait comparisons, and cross-site tests. Chapter 1 analyzed 3,739 vocalizations of the Dusky-throated Antshrike, *Thamnomanes ardesiacus*, before and after hawk exposure. Repertoire composition shifted most strongly during the first post-exposure minute, with Jensen–Shannon divergence declining from 0.365 to 0.038 by 540–600 s. No interval from 300 s onward was displaced more than pre-exposure variation. Early calls were longer, bout size declined during recovery, and the observed call-order predicted subsequent calls better than shuffled sequences. Chapter 2 expanded the analysis to the mixed-species flock and identified four ordered states corresponding to acute threat, heightened vigilance, recovery, and baseline safety. Early states were characterized by alarm calls, lower-foraging species, and birds farther from cover; smaller species contributed disproportionately during heightened vigilance. Several temporally restricted vocalizations emerged as candidate safety signals. Chapter 3 compared these dynamics across intact and Brazil nut–managed forests. Managed sites retained four-state responses but differed in state occupancy, duration, posterior entropy, call composition, and

species-specific repertoires. All eight state-composition contrasts and all fourteen valid species-repertoire comparisons were significant, whereas species-level overall call timing did not differ after correction. The strongest divergence occurred in later recovery and safety states. Together, these findings show that interspecific communication is a measurable functional dimension of biodiversity and may reveal ecological degradation before species loss becomes evident. It therefore offers a highly sensitive, mechanistic, non-invasive indicator of changing ecosystem function.

The dissertation of Eliseo Rafael Parra is approved.

Morgan Tingley

Nathan Kraft

Felipe Zapata

Daniel T. Blumstein, Committee Chair

University of California, Los Angeles

2026

CONTENTS

ABSTRACT OF THE DISSERTATION ii
LIST OF FIGURES vii
LIST OF TABLES viii
ACKNOWLEDGMENTS ix
VITA xi
GENERAL INTRODUCTION 1
CHAPTER 1 Trained Hawk Flight Reorganizes Sentinel Antshrike Vocal Event Streams in an Amazonian Forest 4
Abstract 4
Introduction 4
Methods 8
Study system 8
Experimental Design 9
Acoustic Data Codification 10
Token Data Preparation 11
Temporal Windows and Call Rates 11
Statistical Analyses 12
Repertoire Composition and Recovery to Baseline 12
Early Token Enrichment 14
Duration, Intervals, and Bouts 14
Context-dependent Token Role 15
Ordered Sequence Tests 15
Results 16
Event Set and Baseline Repertoire 16
Repertoire Displacement and Recovery 17
Call Rate and Early Token Enrichment 18
Duration, Intervals, and Bouts 19
Order-sensitive Sequence Structure 21
Results Summary 24
Discussion 25
Supplemental Tables and Figures 29
References 29
CHAPTER 2 A Symphony of Fear and Safety: Distinct Vocalizations Indicate Relative Risk in Mixed-Species Groups 33
Abstract 33
Introduction 33
Materials and Methods 38
Overview 38
Study Site 38
Focal Species Selection 38
Experimental design 38
Vocalization reference codes and audio processing 39
Identifying potential safety signals 40
Hidden Markov Model analysis 40
Estimating state number and bin size 42
Estimating vocalization state associations and risk periods 43
Permutation analysis of functional traits 44

Machine learning performance comparison 45
Control data performance comparison 45
Results 46
Hidden Markov Model state number and bin size: results 46
Estimating vocalization state associations and risk periods: results 46
Identifying potential safety signals: results 49
Permutation analysis of functional traits: results 50
Machine learning performance comparison: results 51
Control data HMM comparison: results 52
Discussion 52
Risk periods: overview 53
Period 1: Acute threat 53
Period 2: Heightened vigilance 54
Period 3: Recovery and foraging resumption 54
Period 4: Baseline safety and foraging 55
General synthesis 55
Acknowledgements 56
Data Availability Statement 56
References 56
CHAPTER 3 Agroforestry Degrades Multi-species Risk-Communication in Amazonian Forests 65
Chapter Abstract 65
Introduction 65
Results 69
Discussion 74
Methods 77
Study system and experimental design 77
Acoustic recording and annotation 77
Hidden Markov model screening and state estimation 78
Cross-site HMM summary metrics 79
State composition and species-level permutation tests 80
Focal species point-count 81
Species trait data and forest structure assessments 82
Supplementary Materials 83
References 84
GENERAL DISCUSSION AND CONCLUSIONS 90
GENERAL REFERENCES 96

LIST OF FIGURES

Figure 1-1. Field site location and sampling design 9
Figure 1-2. Dusky-throated Antshrike vocal-event stream 17
Figure 1-3. Vocal-token composition before and after hawk exposure 19
Figure 1-4. Repertoire recovery relative to baseline variation. 24
Supplemental Figure 1-S1. Trial-level recovery, bout length estimates, and call-duration distributions 29
Figure 2-1. Predictions and analytical workflow. 37
Figure 2-2. Vocalizations, foraging heights, and inferred risk periods. 47
Figure 2-3. Risk-period estimates and functional-trait comparisons 51
Figure 3-1. Trial locations and forest-structure context 68
Figure 3-2. Cross-site call timing and state composition 70
Figure 3-3. Species- and state-level call-count contrasts 72
Figure 3-4. Cross-site hidden-state summary metrics 73
Supplemental Figure 3-S1. Focal-species foraging-height distributions 83

LIST OF TABLES

Table 1-1. Recovery and planned trial-level contrasts. 20
Table 1-2. Tests of local order in vocal-token sequences. 22
Table 2-1. Temporal and distributional characteristics by K-means cluster. 48
Table 2-2. Functional-trait comparisons among risk periods. 49
Table 3-1. Summary of cross-site tests of post-predator vocal communication. 69
Supplemental Table 3-S1. Point-count comparison between forest classes. 83
Supplemental Table 3-S2. GEDI lidar vegetation-metric summaries. 84

ACKNOWLEDGMENTS

I would like to acknowledge the support provided by a team of mentors. The guidance of my advisor Professor Dan Blumstein has been invaluable and exhaustive; his experience has animated yet another empirical journey that will inspire many future ecologists. Special thanks to collaborator Ari Martinez, whose inclusion forged many unforgettable lessons in mentorship. I'd like to thank my committee; Morgan Tingley, Felipe Zapata and Nathan Kraft. Their reassurance and support have always been perfectly timed and gratefully accepted. To the many landowners and property managers who eagerly welcomed our novel, and sometimes risky, experimental designs onto their lands, this project would not have been possible without your support. Thank you to Don Victor, Jorge Novoa, and Edwin Velasquez. Lastly, thank you to Lola and Cartavio, our Bicolored Hawk research collaborators, whose elegant and precise flights helped reveal a new understanding of communication in the rainforest.

The following chapters incorporate work completed in collaboration with my coauthors. Chapter 1 is "Trained Hawk Flight Reorganizes Sentinel Antshrike Vocal Event Streams in an Amazonian Forest". This chapter was authored by Eliseo Parra, Ari E. Martínez, and Daniel T. Blumstein, and is in preparation for publication. EP developed the ideas; EP and AEM collected the data; all authors contributed to data analysis. The dissertation author was the primary investigator and author of this paper.

Chapter 2 is a version of "A Symphony of Fear and Safety: Distinct Vocalizations Indicate Relative Risk in Mixed-Species Groups" by Eliseo Parra, Benjamin Hoffman, José Miguel Ponciano, Ari E. Martínez, and Daniel T. Blumstein. This manuscript was submitted to *Ecology Letters* and is currently in 2nd review. EP, AEM, and DTB developed the ideas; EP and AEM collected the data; all authors contributed to data analysis; EP led the first draft; and all authors contributed substantially to revisions. The dissertation author was the primary investigator and author of this paper.

Chapter 3, “Agroforestry Degrades Multi-species Risk-Communication in Amazonian forests”, by Eliseo Parra, Ari E. Martínez, and Daniel T. Blumstein, is in preparation for publication. EP, AEM, and DTB developed the ideas; EP and AEM collected the data; all authors contributed to data analysis; EP led the first draft; and all authors contributed to revisions. The dissertation author was the primary investigator and author of this paper.

During the preparation of this manuscript, the authors utilized the GPT-4o large language model (OpenAI, August 2026) to assist with [1] grammar review tasks including formatting, typo identification, citation formatting [2] code review and refinement, and [3] the development of software applications for the production of graphic components. Following the use of this tool, the authors critically reviewed, revised, and verified all outputs, and they accept full accountability for the accuracy and integrity of the final text.

Financial support was provided by National Science Foundation awards IOS-2032478 (Daniel T. Blumstein) and IOS-2032469 (Ari E. Martínez). In addition, support was provided by the Eugene V. Cota-Robles fellowship to the dissertation author from 2021 to 2023.

VITA

EDUCATION

2015 — Bachelor of Science in Ecology and Evolution, San Francisco State University

2017 — Master of Science in Ecology, Evolution, and Conservation Biology, San Francisco State University

ACADEMIC AND PROFESSIONAL EMPLOYMENT

2023 — Adjunct Professor of Ecology and Evolution: *Los Angeles Mission College*

2023 — Adjunct Professor of Biology: *Pierce College of Los Angeles*

2019 — Adjunct Professor of Biology: *Diablo Valley College*

2018 — Adjunct Professor of Biology: *Cañada College*

2017 — Laboratory Instructor, Biology I & II: *San Francisco State University*

PUBLICATIONS

2026. Camerlenghi, E., **Parra, E**., et al. Small canopy species drive information highways about predators in an Amazonian rainforest. *Current Biology* 36(8), R318–R319.

2026. Figreros-Andrade, A. F., Gulman, M., Novoa, J., **Parra, E**., et al. Evaluating the effectiveness of model predators in simulating real predators in tropical bird communities. *Current Zoology*, zoag042.

2023. Blumstein, D. T., Johnson, N. A., Katz, S., Kharpatin, S., Ortiz-Ross, X., **Parra, E**., et al. Biological lessons for strategic resistance management. *Evolutionary Applications* 16(12), 1861–1871.

2023. Martínez, A. E., Ponciano, J. M., Gómez, J. P., et al., **Parra, E**., et al. The structure and organization of an Amazonian bird community remains little changed after nearly four decades in Manu National Park. *Ecology Letters* 26(2), 335–346.

2023. Ghose, S. L., Yap, T. A., Byrne, A. Q., et al., **Parra, E**., et al. Continent-wide recent emergence of a global pathogen in African amphibians. *Frontiers in Conservation Science* 4, 1069490.

2022. Martínez, A. E., **Parra, E**., Gómez, J. P., & Vredenburg, V. T. Shared predators between primate groups and mixed species bird flocks: the potential for forest-wide eavesdropping networks. *Oikos* 2022(10), e08274.

2019. **Parra, E**., Riegner, M., Novoa, J., & Martínez, A. E. Observations of Rufous-vented Ground Cuckoo *Neomorphus geoffroyi* associating with mixed-species flocks. *Cotinga*, 94–97.

2019. Hunts, C., Heather, M., Martínez, A. E., & **Parra, E**. Experimental evidence of alarm eavesdropping on Amazonian mixed-flock followers by two social sentinel species. *Ornitología Neotropical* 30, 224–228.

2018. Martínez, A. E., **Parra, E**., Muellerklein, O., & Vredenburg, V. T. Fear-based niche shifts in neotropical birds. *Ecology* 99(6), 1338–1346.

2017. Martínez, A. E., **Parra, E**., Collado, L. F., & Vredenburg, V. T. Deconstructing the landscape of fear in stable multi-species societies. *Ecology* 98(9), 2447–2455.

2012. Micheletti, S., **Parra, E**., & Routman, E. J. Adaptive color polymorphism and unusually high local genetic diversity in the side-blotched lizard, *Uta stansburiana*. *PLoS ONE* 7(10), e47694.

HONORS, AWARDS, AND FELLOWSHIPS

2021 **-** 2023 Eugene V. Cota-Robles Fellowship, University of California, Los Angeles

GENERAL INTRODUCTION

Predation forces animals to continually balance two competing demands: avoiding danger and acquiring the resources needed to survive. Social information can reduce the uncertainty surrounding this trade-off by allowing individuals to use the behavior and vocalizations of others when deciding whether to hide, remain vigilant, move, or resume foraging (Lima & Dill, 1990; Magrath et al., 2015). Most research on acoustic responses to predators has emphasized conspicuous alarm calls produced during immediate danger. Far less is known about the broader vocal process through which animals move from acute threat back toward safer foraging conditions. In this dissertation, we define risk communication as vocal information associated with elevated perceived danger, while safety communication includes cues or signals associated with declining danger and the resumption of normal flocking and foraging. We treat changes in call identity, timing, sequence, and species participation as components of a dynamic information environment. In systems with high species diversity and niche specialization, the richness of communication interplay may be used to ecologically interpolate other ecosystem parameters, such as vegetation-type or heterogeneity between species' preferred microhabitats. In doing so, this framework may be applied to compliment and inform current advances in remote censusing whereby species report their perceived conditions through patterns of interspecific communication.

Amazonian mixed-species understory bird flocks provide an unusually strong system for examining these dynamics. These flocks contain persistent associations among species that differ in body size, foraging position, vigilance, and dependence on heterospecific information (Munn & Terborgh, 1979; Goodale et al., 2010). Unique to Neotropical mixed-species flocks, these groups contain a high representation of social tracheophone suboscines—including antshrikes, antwrens, woodcreepers, and their allies—whose vocal biology is characterized by

largely innate development, limited song plasticity, and broadly stereotyped vocalizations across geographic populations. Collectively, these traits make the group well suited for field studies of communication, social behavior, and their evolution (Tobias et al. 2012). Sentinel antshrikes contribute disproportionately to flock cohesion and antipredator behavior, while smaller flock members may experience different levels of vulnerability and contribute information during intermediate stages of recovery. Because these assemblages depend on repeated interspecific associations, their communication may also represent a functional ecological process that is altered before species and associations disappear completely from a habitat.

The following chapters ask three linked questions in order of ecological complexity. First, does predator exposure reorganize the vocal behavior of an obligate mixed-flock sentinel species, and can recovery be measured as a return toward its ordinary pre-disturbance repertoire? Second, do the combined vocalizations of the mixed-flock form ordered community-level states corresponding to acute threat, heightened vigilance, recovery, and baseline safety, and are those states associated with particular vocalizations and species traits? Our last question approaches the utility of such ecological study. Third, to test the sensitivity of these communications to perturbation, we ask are these risk-to-safety dynamics altered in forests subject to a gradation of agroforestry use? These questions are addressed through standardized presentations of trained Bicolored Hawks, intensive hand annotation of vocal events, sequence and compositional analyses, hidden Markov models, functional-trait comparisons, and cross-site tests.

Chapter 1 establishes the single-species foundation by describing how the vocal event stream of the Dusky-throated Antshrike *Thamnomanes ardesiacus* changes after hawk exposure and recovers toward its pre-hawk structure. Chapter 2 expands the analysis to the mixed-species community, using hidden Markov models to infer latent risk periods and linking those periods to call types, candidate safety signals, and the traits of nine vocalizing species.

Chapter 3 then tests the ecological sensitivity of this framework across intact and Brazil-nut–managed forests, asking whether state duration, occupancy, uncertainty, call composition, and species contributions change with human use. Together, the chapters expand their inquiry from individual vocal organization to community-level risk states, and lastly, to landscape-scale comparisons and conservation application.

CHAPTER 1
Trained Hawk Flight Reorganizes Sentinel Antshrike Vocal Event Streams in an Amazonian Forest

Abstract

Exposure to predators elicits alarm calls, yet most studies focus on call rates or single call types rather than complete vocal event-stream structure. We analyzed annotated vocalizations of the dusky-throated antshrike (*Thamnomanes ardesiacus*) before and after a trained bicolored hawk (*Astur bicolor*), a natural predator, was flown through foraging flocks. We treated each call type as an arbitrary vocal call-token and tested whether hawk exposure changed repertoire composition, recovery toward pre-hawk baseline, call duration, inter-call intervals, bout structure, and call-token order. The event table contained 3,739 antshrike vocalizations across 16 trials (1,855 PRE and 1,884 POST). Across ten 60-s POST bins, aggregate composition was most displaced in the first minute (Jensen-Shannon divergence = 0.365) and declined to 0.038 by POST 540-600 s. A repeated-measures mixed model showed that trial-level divergence declined with time after exposure. Relative to each trial's ordinary PRE minute-to-minute variation, composition remained detectably displaced during POST 0-60, 60-120, 120-180, and 240-300 s after correction for multiple comparisons; no interval from 300 s onward was detectably displaced. POST 0-60 s calls were longer than PRE calls and ordered token sequences predicted next tokens better than shuffled sequences in post recovery windows. These results suggest that predator exposure temporarily reorganizes antshrike vocal behavior across token identity, duration, bout structure, and local order, followed by a noisy recovery toward baseline vocal production.

Introduction

Predation risk shapes animal behavior at many scales, from momentary decisions about vigilance to long-term habitat use and social organization (Lima and Dill 1990). Many species emit alarm calls upon detecting a predator and these calls can transmit information about

predator presence and/or risk intensity to receivers (Magrath et al. 2015). In many systems where animals have complex vocal repertoires, information about risk may be transmitted by the entire repertoire of vocalizations rather than a single call type (Rauber and Manser 2017). Animals may change how much they call, which elements of a repertoire are used, how calls are timed, and how calls are arranged into sequences or repeated bouts. These vocalizations contribute to public information which can strongly influence animal communities (Goodale et al. 2010), and alarm-call eavesdropping is widespread across taxa (Magrath et al. 2015). It is therefore profitable to understand a response to a predator as a dynamic event stream, rather than as a simple increase in the production of an alarm call type.

Amazonian mixed-species understory flocks provide a particularly strong system in which to study this question. These flocks are spatially cohesive, contain recurring species associations, and often form around nuclear or sentinel-like species that help structure flock movement and anti-predator behavior (Munn and Terborgh 1979; Munn 1986; Jullien and Thiollay 1998; Thiollay and Jullien 1998). In Amazonian understory flocks, dusky-throated antshrikes *Thamnomanes ardesiacus* (THAR) are especially important because they provide alarm information and maintain cohesion among flock members. Experimental work has shown that removing antshrikes (and therefore antshrike risk information) results in flocks moving to protected areas and reduces overall flock occurrence (Martínez et al. 2018). Other work shows that antshrikes pair vocal behavior with repositioning and perching cues used by flock mates, suggesting that antshrike signals help organize interspecific movement (Williams and Lindell 2021). Further, Antshrike alarm calls have been shown to elicit strong predator-avoidance behaviors in many co-occurring species (including primates), while also encoding information regarding predator threat (Martínez et al. 2017, Martínez & Zenil 2012, Martínez et al. 2022). While these studies begin to unpack the broader importance of particular vocalizations, complex patterns and event streams remain poorly understood.

A central challenge in capturing the details of risk-related flock communication is its dependence on rare events like predator attacks. One solution is to implement trained predators, such as hawks, who can be taught to fly through mixed flocks. Other challenges include isolating the cause of the vocalization, which may be a response to another species' vocalization, a response to a predator, or an instance of intraspecific communication. These factors limit inference to interspecific communication from adjacency of signals alone. By presenting a predator and identifying all the subsequent response vocalizations, we can characterize both the repertoire and event structure of putative information-producing species to better identify how other species' signals interact with relevant communication structure, as opposed to isolated calls. This logic follows the broader eavesdropping literature which emphasizes that heterospecific alarm information can dynamically influence receivers as the signaler's own behavior is also changing with context (Goodale et al. 2010; Magrath et al. 2015). Lastly, these context-derived patterns may shed light on previously understudied vocalization structure yet should be seen as a first step. Experimental work in birds has shown that call combinations and ordering can affect receiver behavior, but strong claims about syntax or semantics require direct playback or receiver-response tests (Suzuki et al. 2016, 2017) and compositional claims require careful demonstration that units and arrangements alter function (Engesser et al. 2015).

Recent work on other animal communication systems provides useful methodological analogies. Sperm whale *Physeter macrocephalus* coda studies show that vocal units can have contextual and combinatorial structure, with information distributed across rhythm, tempo, duration drift, and sequence position rather than only across flat labels (Sharma et al. 2024). A related spectral analysis showed that traditional coda labels can miss additional acoustic dimensions, emphasizing that annotated categories may be useful but incomplete descriptions of vocal structure (Beguš et al. 2025). Bird communication-network models likewise show that

the temporal ordering and waiting times between calls can reveal structure that is invisible in pooled call counts (Stowell et al. 2016). These studies justify applying a simple, annotation-based event-stream framework to antshrike communication that fits the practical constraints of tropical field bioacoustics. Dense forest recordings often contain faint or obstacle-degraded signals, overlapping species with similar calls, variable distance to the receiving microphone, and background noise. We therefore use the most reliable, consistently interpretable, and biologically justifiable information in the dataset: annotated call identity tokens (discrete vocalizations), begin time, end time, duration, time between calls, trial, and stage. We use *token* to refer to an individual classified vocal event and *call type* to refer to the category assigned to that event. Hand annotation allows exact token order to be analyzed efficiently and consistently, even in the absence of raw acoustic signals sufficiently clean for reliable automated acoustic feature extraction.

Here we directly test whether a predatory hawk's flight through a flock significantly reorganizes the antshrike vocal event stream. The primary data used are individually identified vocalizations, pre- and post-presentation event times, call durations, and trial identities. This token based strategy follows a meaningful level of organization, since the timing and arrangement of risk-related calls may be more important than their fine acoustic details. This is especially relevant in time-critical predation event contexts where receivers experience a rapid sequence of consequential vocalizations.

We ask four questions: 1) Does antshrike repertoire composition change immediately after hawk exposure relative to pre-hawk baseline? 2) Does the post-hawk repertoire return toward pre-hawk structure over time, and when is it no longer detectably more variable than ordinary pre-hawk fluctuation? 3) Do call duration, inter-call intervals, and bout organization change after hawk exposure? 4) Does observed token order predict future tokens better than shuffled order? By measuring recovery as return to levels of pre-disturbance state variability

(Shade et al. 2012a) we test whether predator exposure changes the pattern of calls produced after exposure and when that pattern subsequently recovers.

Methods

Study system

The focal species for this study is the dusky-throated Antshrike (*Thamnomanes ardesiacus*), an insectivorous tracheophone suboscine bird. This species is a common alarm-producing species in Amazonian mixed-species understory flocks and is known to influence flock cohesion, habitat use, and responses to predation risk (Martínez et al. 2018; Williams and Lindell 2021). They are found throughout western Amazonia and in forest north of the Amazon River, nearly always found in mixed-species flocks and considered a mixed-flock obligate species. Our study was conducted at the Los Amigos Biological Station (EBLA), Madre de Dios, Peru (-12.556371, -70.106327) (figure 1-1, A), where several flock territories had been identified in previous years' research. EBLA terra firme forest represents a lowland Amazonian ecosystem that remains unflooded by seasonal river inundation and is characterized by a closed-canopy forest with stands of Brazil-nut Tree (*Bertholletia* excelsa) and a structurally complex understory. The habitat consists of heterogeneous topography, including ridges, slopes, and small drainage features, tracts of bamboo, and supports high plant and animal diversity typical of southwestern Amazonian terra firme forests.

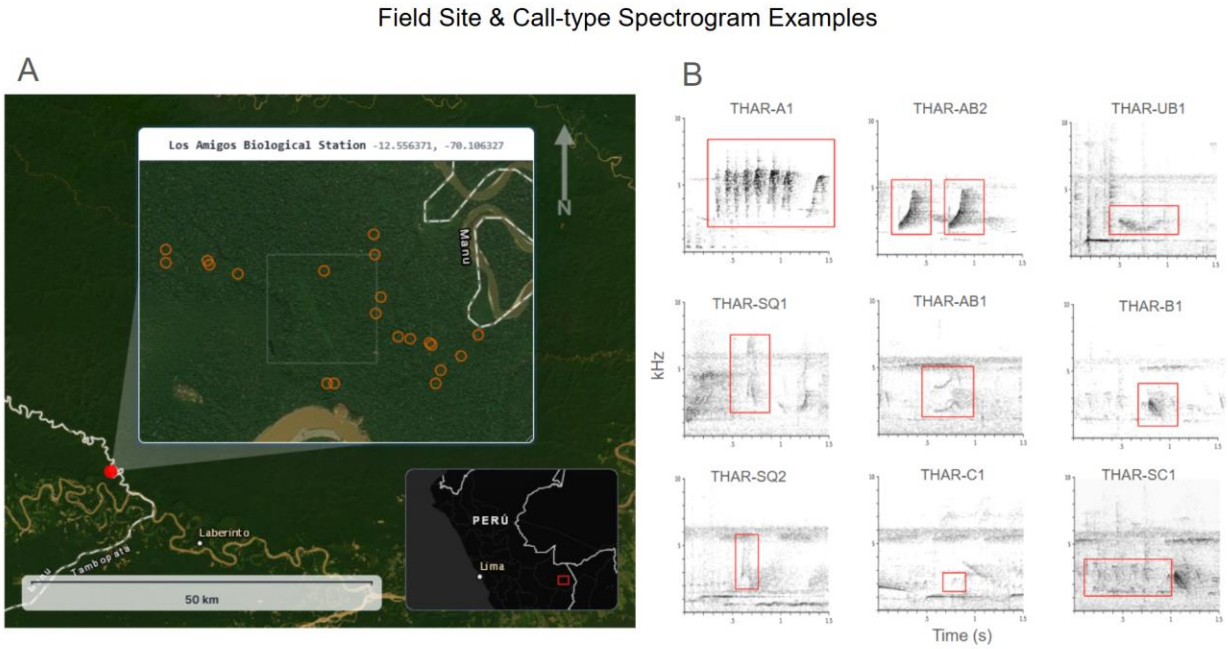


Figure 1-1. Field site location and sampling design

(A) Field site Location. Los Amigos Biological Station in Madre de Dios, Peru, shown with orange circles indicating sampling locations within the study area and the inset showing the station's regional position. Intra-sampling polygon represents a 1 km square for scale. (B) Representative spectrograms of nine Dusky-throated Antshrike (*Thamnomanes ardesiacus*; THAR) call annotations. THAR token categories; red rectangles show the time–frequency selections used to assign token identity and extract event timing and duration, with token labels treated as arbitrary categorical units rather than assumed functional call types.

Experimental Design

To elicit antshrike alarm vocalizations, we conducted a fly-through of the flock using a trained bicolored hawk. Hawks were trained Colka Raptors, based in Piura, Peru, following protocols established in prior research in rainforest field conditions (Martinez et al 2018). The hawk was trained to fly to a perch positioned 30-40 meters towards the flock, 15 meters above ground, and return immediately.

Eight different mixed-species flocks were used to gather recordings of flock-vocal response behavior, visited in random order between the hours of 07:00 and 14:00 h until a total of 16 trials were completed. Flocks engaged in non-foraging behavior (e.g., territory disputes) were skipped and re-visited at a later time or day. Flocks were given an initial acclimation time

of a minimum of 10 minutes prior to beginning a pre-recording of 10 minutes (PRE). After this time, the falconer was silently summoned, and the hawk was flown and retrieved. Post recording then began. The falconer then immediately left the area to a distance of > 100 m. Recording continued from hawk flight until 10 minutes post-flight (POST). Two researchers remained in place to record the flock response, and take notes on species composition, vegetation parameters, and other important details of each trial.

Acoustic Data Codification

Acoustic data were recorded using a ZOOM H5 multichannel digital audio recorder and saved as WAV audio files (44.1 kHz, 16 bit). Vocalizations were first categorized and catalogued by identifying each unique sound and establishing a nomenclature structure, consisting first of a species four-letter species code (e.g., THAR) followed by the vocalization id (e.g., A1) to make a unique code for each vocalization (e.g., THAR-A1) (Figure 1-1B). These labels were used then to annotate full trial recordings. Acoustic recording annotations were made by creating time-length and frequency-range selections for each vocalization in each trial using Raven Pro bioacoustic analysis software (K. Lisa Yang Center for Conservation Bioacoustics, 2026), resulting in a hand-annotated event table where each row represented one annotated vocal event with trial identity, stage, begin time, end time, duration, species code, and call suffix. PRE and POST segments were treated as separate sampling periods because the time gap between the end of PRE recording and the beginning of POST recording varied slightly among trials. Therefore, timing, waiting-time, bout, and sequence calculations were made within trial and stage.

The primary analysis is focused on the sentinel antshrike species THAR in order to establish whether a single focal species shows a structured predator-response event stream. The analysis used token labels as arbitrary categorical units. Tokens with fewer than 5 total THAR occurrences were excluded from token-based composition, enrichment, and sequence

analyses. All valid THAR events were retained for call-rate, duration, wait-time, and bout analyses.

Token Data Preparation

Although some token nomenclature was descriptive (e.g., A1 = alarm), we did not use the appearance of the label as a biological predictor. Any apparent relationship between a prefix and a putative call function was ignored in the statistical design. All analyses were based on observed frequencies, timing, duration, and sequential order to derive a de novo catalogue of call use directly from the behavioral data rather than from a priori functional classifications.

The selection table file contains begin and end times for each call. We used the provided duration field when available and otherwise interpreted duration as the difference between end and begin time of a call. Rare labels were removed because token-based analyses become unstable with sparse categories. The final event table contained 3,739 THAR vocalizations across 16 trials (1,855 PRE and 1,884 POST).

Temporal Windows and Call Rates

PRE and POST recordings were treated as separate 600-s sampling periods. PRE was retained as a single full-baseline layer for primary baseline comparisons. For the primary temporal display, POST was divided into ten consecutive, 60-s bins from POST 0-60 s through POST 540-600 s. The 60-s scale was chosen a priori because antipredator and playback studies frequently summarize rapid post-stimulus changes over 1-min intervals, and because the first minute isolates the clearest acute-response period (Altmann 1974; Griffin and Galef 2005; Griffin et al. 2005).

Call rates were expressed as calls per minute to match the 60-s temporal bins. For planned acute-response inference, POST 0-60 s was compared with the full PRE baseline. For sequence-order inference, adjacent 60-s bins were pooled into POST 0-60 s, POST 60-180 s,

POST 180-420 s, and POST 420-600 s to maintain sufficient event counts for two-token (k = 2) contexts. A six-bin 100-s sequence analysis was retained as an equal-duration sensitivity check.

Statistical Analyses

All analyses were implemented in R (R Core Team 2024), and trial was the unit of inference wherever possible. Planned acute-response contrasts for call rate, duration, inter-call interval, and bout size used paired Wilcoxon signed-rank tests because the number of trials was modest and not normally distributed (Wilcoxon 1945). Repertoire recovery was analyzed with linear mixed-effects models because each trial contributed repeated, temporally ordered observations; these models allow within-trial covariance to be represented directly rather than requiring the sphericity assumption of conventional repeated-measures ANOVA (Gueorguieva and Krystal 2004). The statistical threshold was alpha = 0.05. Holm family-wise-error correction was applied across the ten bin-wise recovery contrasts (Holm 1979), whereas Benjamini-Hochberg false-discovery-rate correction was applied across token-specific enrichment tests, which compare the production of tokens post with rates recorded in pre(Benjamini and Hochberg 1995). One trial lacked usable PRE events, so PRE-based analyses used 15 trials; post-only contrasts used 16 trials when both POST windows were available.

Repertoire Composition and Recovery to Baseline

For each 60-s interval, we calculated the proportion of THAR events represented by each retained vocal token. Let P = (p1,..., pK) denote the vector of proportions across the K retained token categories in a POST interval and Q = (q1,..., qK) the corresponding vector for the reference PRE repertoire. Thus, P and Q describe two complete repertoire distributions, not two individual call types. We quantified their dissimilarity with Jensen-Shannon divergence (JSD; Shannon 1948; Lin 1991):

$$JSD(P,Q) = H((P + Q)/2) - (1/2)H(P) - (1/2)H(Q)$$

where H is Shannon entropy in bits:

$$H(P) = -\sum_i \pi_i \log_2(\pi_i)$$

With base-2 logarithms, JSD ranges from 0 to 1: a value of 0 indicates identical token distributions, and larger values indicate greater repertoire dissimilarity. Aggregate JSD, calculated after pooling events across trials, was retained for descriptive figures. Formal inference used trial-level JSD, calculated for each trial by comparing each POST bin's complete token distribution with that trial's full PRE distribution.

To estimate ordinary baseline variability independently of the focal minute, we divided each trial's PRE period into ten 60-s bins. For PRE bin i in trial j, we calculated JSD(PREji, PREj,-i), comparing the focal minute with the remaining nine PRE minutes:

$$\mathrm{JSD}_{\mathrm{PRE-LOO}}(j,i) = \mathrm{JSD}\left(\mathrm{PRE}_{j,i}, \mathrm{PRE}_{j,-i}\right)$$

Excluding the focal PRE minute from its reference distribution avoids a partially self-referential comparison. We summarized each trial's usable leave-one-bin-out (LOO) PRE values by their mean, B_j, and calculated excess post-hawk displacement as:

$$\Delta\mathrm{JSD}_{j,t} = \mathrm{JSD}\left(\mathrm{POST}_{j,t}, \mathrm{PRE}_j\right) - B_j$$

A value of ΔJSD = 0 therefore represents the trial's typical minute-to-minute PRE fluctuation; positive values indicate greater dissimilarity than ordinary PRE variation. This definition follows disturbance-ecology treatments of resilience as return toward a pre-disturbance state relative to its intrinsic variability or normal operating range, rather than exact identity with a baseline sample (Shade et al. 2012a), and is analogous to distance-to-baseline analyses of multivariate community recovery after controlled disturbance (Shade et al. 2012b).

We fit a linear mixed-effects model with 60-s POST bin as a categorical fixed effect, a random intercept for trial, and a first-order autoregressive [AR(1)] residual correlation across successive bins. This is an ANOVA-style repeated-measures comparison that accounts for the non-independence of multiple bins from the same trial (Gueorguieva and Krystal 2004). A no-intercept parameterization estimated the mean delta JSD for each bin, and each estimate was tested against zero; two-sided p-values were Holm-adjusted across the ten bins (Holm 1979).

We used this model rather than a pooled Spearman correlation for bin-wise inference because a correlation tests only an overall monotonic association and does not compare each interval with baseline while accounting for within-trial dependence. We also fit raw trial-level JSD against POST-bin midpoint as a continuous fixed effect, with the same random and correlation structure, to test the overall recovery trajectory. The pooled PRE percentile envelope was retained only as a descriptive visualization. Because failure to reject a difference is not evidence of equivalence, nonsignificant bins are described as not detectably above ordinary PRE variation, not as identical to baseline.

Early Token Enrichment

To identify which THAR call types contributed to the first-minute post-hawk shift, we conducted the enrichment analysis separately for every retained THAR vocal token. For each token, we calculated its proportion within each trial and stage, then tested whether the paired trial-level proportion was higher in POST 0-60 s than in PRE using a one-sided Wilcoxon signed-rank test. We also tested the reverse direction to identify tokens depleted in the first minute. P-values were adjusted across all retained tokens using the Benjamini-Hochberg false-discovery-rate procedure. Pooled proportions are reported as descriptive summaries; trial-level tests were used for inference to avoid treating repeated calls within a trial as independent replicates.

Duration, Intervals, and Bouts

Call duration was taken directly from the corrected annotation table as end time minus begin time. We calculated the median THAR call duration within each trial and 60-s interval and used planned paired Wilcoxon tests to compare POST 0-60 s with PRE and with POST 540-600 s.

For inter-call intervals, we calculated the interval from the previous THAR event within the same trial and stage as the current call's begin time minus the previous call's end time.

Negative values caused by overlapping annotations were set to zero. The first event in each trial-stage sequence had undefined wait time and was excluded from wait-time tests. Before assigning calls to bouts, we estimated the interval separating within-bout call spacing from longer pauses between bouts. We fit a three-component Gaussian mixture model to intervals between consecutive calls, representing short, intermediate, and long gaps. We interpreted long-gaps as pauses between bouts and used the intersection of the fitted intermediate- and long-gap distributions as the bout-separation threshold. This yielded a threshold of 2.85 s. Consecutive calls separated by no more than 2.85 s were therefore assigned to the same bout, whereas a longer interval initiated a new bout.

Context-dependent Token Role

Some vocal tokens occurred in both PRE and POST contexts. For each B1 event, we recorded the previous token, next token, bout position, stage, 60-s layer, duration, and wait time to compare tokens embedded in different local contexts before, immediately after, and later after hawk exposure. This logic follows the general principle that vocal labels may be incomplete descriptions unless timing and sequence context are considered (Sharma et al. 2024; Beguš et al. 2025).

Ordered Sequence Tests

We tested whether local THAR token order contained predictive sequential structure beyond marginal token abundance using count-based n-gram models evaluated with leave-one-trial-out cross-validation (Kershenbaum et al. 2016; Hedley 2016). Context length k was set to 0, 1, 2, 4, or 8 previous THAR tokens. The k = 0 model is the no-order repertoire baseline; it predicts from overall token frequency only. For k > 0, the model predicts the next token from the preceding k-token context. Additive smoothing with alpha = 0.5 was used to avoid zero probabilities for unseen contexts. Model performance was summarized with top-1 accuracy and held-out perplexity:

$$\text{Perplexity} = \exp\left[-\frac{1}{N}\sum_{i=1}^{N}\log p(x_i \mid \text{history}_i)\right]$$

where N is the number of predicted tokens and p(x_i | history_i) is the probability assigned to the observed next token. Lower held-out perplexity indicates that the model assigned greater probability, on average, to the observed next tokens and therefore provided better probabilistic prediction (Chen and Goodman 1999). To test whether order mattered, we compared ordered sequences with shuffled null sequences ($N_{shuffles}$=200) created by randomly permuting token order within each trial and analysis period while preserving the same token counts. This null retains repertoire composition but destroys local order. We calculated one-sided empirical p-values as:

$$p = \frac{1 + n(P_{\text{shuffled}} \leq P_{\text{ordered}})}{1 + N_{\text{shuffles}}}$$

Ordered models were considered stronger than shuffled models when ordered perplexity was lower than the shuffled distribution. Because individual 60-s bins become sparse for k = 2 contexts, the primary order analysis used pooled adjacent 60-s recovery windows: PRE, POST 0-60 s, POST 60-180 s, POST 180-420 s, and POST 420-600 s. A 100-s equal-bin analysis was used only as a sensitivity check for whether order effects persisted under longer equal-duration windows with larger sequence samples.

Results

Event Set and Baseline Repertoire

After filtering to THAR, retaining valid PRE/POST records, and excluding suspected typographic errors, the analysis contained 3,739 THAR events from 16 trials. These included 1,855 PRE events and 1,884 POST events. Fifteen trials contained usable PRE and POST data for PRE-paired analyses; post-only contrasts used all 16 trials when both post-hawk bins being compared were available. Token-based analyses retained 3,720 common-token events (Figure 1-2A) after excluding 19 events with fewer than 5 total THAR occurrences.

The retained PRE repertoire was dominated by C1 = 555 events (30.1%), SQ2 = 524 events (28.4%), B1 = 298 events (16.2%), and AB1 = 281 events (15.3%). In the first 60 s after hawk exposure across all 16 trials, the retained-token repertoire shifted strongly toward B1 = 226 events (56.6%), SQ1 = 47 events (11.8%), A1 = 43 events (10.8%), C1 = 28 events (7.0%), and UB1 = 27 events (6.8%) (Figure 1-2B: 0-60; Figure 1-3A/B). In the final POST minute, composition was dominated by SQ2 = 55 events (40.4%), C1 = 42 events (30.9%), AB1 = 19 events (14.0%), and B1 = 12 events (8.8%) (Figure 1-2B: 540-600; Figure 1-3A).

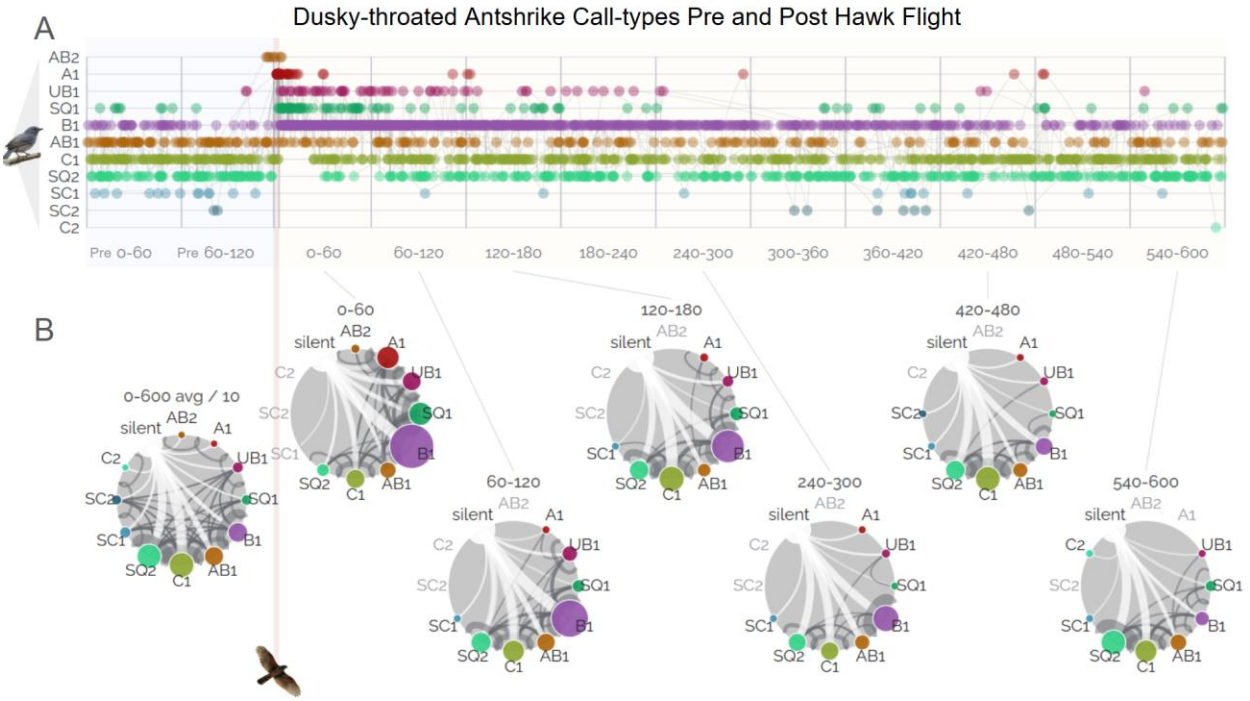


Figure 1-2. Dusky-throated Antshrike vocal-event stream

(A) Pooled Dusky-throated Antshrike (THAR) vocal-token events. Each point represents an annotated THAR event, colored by its token label, across two displayed 60-s PRE intervals and ten consecutive 60-s POST intervals; the red vertical divider marks hawk exposure time and division of PRE/POST. (B) Transition structure before and after a trained Bicolored Hawk flight. Circular networks summarize the full PRE period as a mean 60-s network and selected POST intervals, with node size representing token frequency and ribbon width representing transition weight, including transitions involving the silent state (with a gap of >2.853 s); the acute response was dominated by B1 and associated A1/SQ1 activity, followed by progressive recovery toward the C1–SQ2–AB1-dominated PRE configuration.

Repertoire Displacement and Recovery

Aggregate JSD from PRE was highest immediately after hawk exposure and generally declined across the ten 60-s POST bins. Aggregate values were 0.365, 0.154, 0.133, 0.069,

0.101, 0.040, 0.076, 0.025, 0.059, and 0.038 from POST 0-60 s through POST 540-600 s, respectively (Figure 1-4A). The repeated-measures trend model confirmed an overall decline in trial-level JSD with time after exposure (beta = -0.000518 JSD s-1, SE = 0.000106, t = -4.88, df = 128, p = 3.10 x 10-6), equivalent to an average decline of 0.0311 JSD units per minute (Table 1-1).

Relative to each trial's mean leave-one-minute-out PRE JSD, excess dissimilarity was detectably greater than zero during POST 0-60 s (delta JSD = 0.405, Holm-adjusted $p < 0.001$), 60-120 s (0.227, $p < 0.001$), 120-180 s (0.264, $p < 0.001$), and 240-300 s (0.172, $p = 0.0186$) (Table 1-1). The 180-240-s interval was not significant after correction (0.130, $p = 0.132$), followed by renewed detectable displacement at 240-300 s, indicating a non-monotonic recovery trajectory. No interval from POST 300-360 s onward was detectably above ordinary PRE variation after Holm correction (all adjusted $p >= 0.197$). Thus, sustained absence of detectable excess repertoire dissimilarity began at approximately 300 s. All estimated differences remained positive, so these nonsignificant late intervals indicate recovery toward baseline rather than proof of statistical equivalence. The pooled PRE envelope in Figure 1-4A was retained as a descriptive comparison and showed the same four elevated aggregate intervals.

Call Rate and Early Token Enrichment

THAR call rate increased in the first minute after hawk exposure. Among the 15 PRE-paired trials, median trial-level call rate was 11.60 calls min-1 in PRE and 22.00 calls min-1 in POST 0-60 s. POST 0-60 s rate was higher than PRE ($n = 15$, $V = 95$, $p = 0.025$) and higher than POST 540-600 s in the post-only contrast (median 24.00 vs. 5.50 calls min-1, $n = 16$, $V = 122.5$, $p = 0.0026$) (Table 1-1).

Trial-level token-enrichment tests identified A1 ($q = 0.009$) and B1 ($q = 0.0434$) as significantly enriched in POST 0-60 s after false-discovery-rate correction. SQ1 increased

strongly but did not meet the FDR threshold (q = 0.0702) (Figure 1-3B). Tokens SQ2 (q = 0.003996), SC1 (q = 0.0163), C1 (q = 0.0217), and AB1 (q = 0.0217) were significantly depleted in the first post-hawk minute (Figure 1-3B). These results identify which token labels contributed most to the first-minute repertoire displacement.

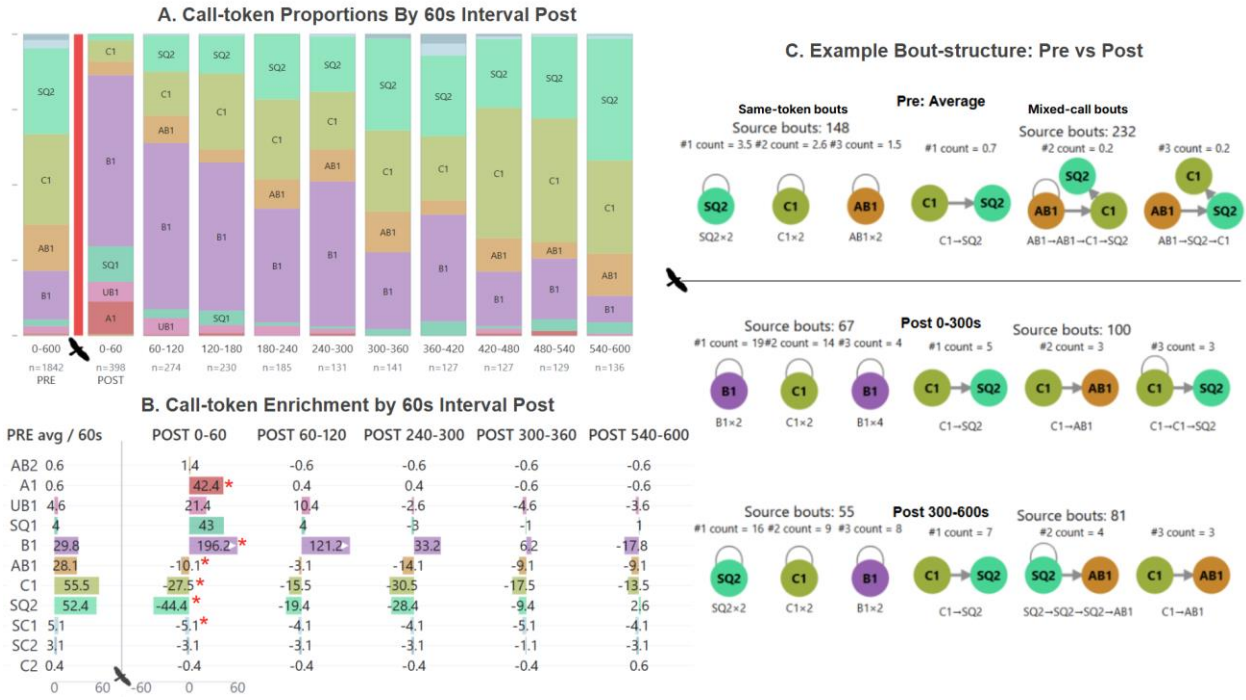


Figure 1-3. Vocal-token composition before and after hawk exposure

(A) THAR vocal-token composition before and after trained hawk exposure. Stacked bars show the proportion of THAR annotated tokens in the full 600-s PRE baseline and in ten 60-s POST bins after the Bicolored Hawk flight. Vertical red bar and hawk icon designate transition from PRE/POST. (B) Enrichment diagram of THAR vocal tokens. Bars show trial-level differences in token proportion between POST 60 s bins and PRE. A1 (q = 0.00914) and B1 (q = 0.0434) were significantly enriched in POST 0-60 s, whereas SQ2 (q = 0.003996), SC1 (q = 0.0163), C1 (q = 0.0217), and AB1 (q = 0.0217) were significantly depleted, (red asterisk: Benjamini-Hochberg false-discovery-rate outcomes from one-sided paired Wilcoxon tests). SQ1 increased strongly but did not meet the FDR threshold (q = 0.0702). (C) Common THAR bout motifs before and after trained-hawk exposure. Bouts were defined within trial and stage using the estimated 2.85-s inter-call gap. Repeated same-token bouts and mixed-token bouts are shown for PRE, early POST recovery (0–300 s), and late POST recovery (300–600 s). Source-bout counts give the number of non-singleton bouts contributing to each panel, and callouts show the most frequent bout examples.

Duration, Intervals, and Bouts

Call duration changed with stage. For the PRE-paired contrast, median trial-level duration was 0.316 s in PRE and 0.445 s in POST 0-60 s. POST 0-60 s calls were longer than PRE calls (n = 15, V = 95, p = 0.025). For the post-only first-versus-final contrast, median

duration was 0.452 s in POST 0-60 s and 0.329 s in POST 540-600 s (n = 16, V = 91, p = 0.122) (Table 1-1. Supplemental Figure 1-S1C).

Median inter-call intervals were 0.990 s in PRE and 1.429 s in POST 0-60 s for the PRE-paired contrast. In the post-only first-versus-final contrast, median interval was 1.431 s in POST 0-60 s and 3.088 s in POST 540-600 s. The a-priori interval contrasts were not significantly different (POST 0-60 s less than PRE: n = 15, V = 84, p = 0.918; POST 540-600 s greater than POST 0-60 s: n = 16, V = 92, p = 0.112) (Table 1-1).

The estimated bout gap from the log-interval mixture was 2.85 s (Supplemental Figure 1-S1B). Using this gap, median trial-level mean bout size was 2.68 calls in PRE and 3.14 calls in POST 0-60 s for the PRE-paired contrast; this contrast was not significant (n = 15, V = 81, p = 0.122). In the post-only first-versus-final contrast, median trial-level mean bout size was 3.26 calls in POST 0-60 s and 2.00 calls in POST 540-600 s. POST 0-60 s bouts were larger than final-minute bouts (n = 16, V = 100, p = 0.0124) (Table 1-1). Thus, bout size declined during recovery (Figure 1-3C).

B1 remained a context-dependent token. In PRE, B1 was usually preceded by B1 (71.6%), with smaller contributions from C1 (8.4%) and SQ2 (6.4%). In POST 0-60 s, B1 was preceded by B1 (76.4%) and SQ1 (9.3%), whereas final-minute B1 contexts included more C1 and SQ2 adjacency, although final-minute B1 counts were lower (Figure 1-3). Thus, B1 appears to be a context-dependent token rather than stage-specific.

Table 1-1. Recovery and planned trial-level contrasts.

For each trial and POST bin, delta JSD is the trial-level POST-to-PRE JSD minus that trial's mean leave-one-PRE-minute-out JSD. Positive estimates therefore indicate more repertoire dissimilarity than ordinary PRE minute-to-minute variation. Values are model estimates +/- SE; bin-wise tests use df = 120 and two-sided Holm-adjusted (HA) p-values across ten POST intervals. The trend row reports the continuous-time mixed-model slope for raw JSD. Both mixed models include a trial random intercept and AR(1) residual correlation across successive bins. Nonsignificant delta JSD contrasts indicate that excess dissimilarity was not detected; they do not establish equivalence to PRE. Remaining rows report trial medians and prespecified one-sided paired Wilcoxon signed-rank tests (n, paired trials; V, test statistic).

Period and comparison	Estimate or median	Test statistic and p-value

Period and comparison	Estimate or median	Test statistic and p-value
POST 0-60 s vs. ordinary PRE variation; excess JSD	0.405 +/- 0.055	t = 7.41, df = 120; HA p < 0.001
POST 60-120 s vs. ordinary PRE variation; excess JSD	0.227 +/- 0.056	t = 4.07, df = 120; HA p < 0.001
POST 120-180 s vs. ordinary PRE variation; excess JSD	0.264 +/- 0.055	t = 4.82, df = 120; HA p < 0.001
POST 180-240 s vs. ordinary PRE variation; excess JSD	0.130 +/- 0.056	t = 2.32, df = 120; HA p = 0.132
POST 240-300 s vs. ordinary PRE variation; excess JSD	0.172 +/- 0.056	t = 3.07, df = 120; HA p = 0.0186
POST 300-360 s vs. ordinary PRE variation; excess JSD	0.069 +/- 0.055	t = 1.27, df = 120; HA p = 0.438
POST 360-420 s vs. ordinary PRE variation; excess JSD	0.117 +/- 0.056	t = 2.08, df = 120; HA p = 0.197
POST 420-480 s vs. ordinary PRE variation; excess JSD	0.082 +/- 0.056	t = 1.46, df = 120; HA p= 0.438
POST 480-540 s vs. ordinary PRE variation; excess JSD	0.033 +/- 0.056	t = 0.59, df = 120; HA p = 0.558
POST 540-600 s vs. ordinary PRE variation; excess JSD	0.113 +/- 0.055	t = 2.06, df = 120; HA p = 0.197
POST 0-600 s; JSD trend with time	beta = -0.000518 +/- 0.000106 JSD s-1	t = -4.88, df = 128; p < 0.001
POST 0-60 s vs. PRE; median rate (calls min-1)	22.00 vs. 11.60; n = 15	V = 95; p = 0.025
POST 0-60 s vs. POST 540-600 s; median rate (calls min-1)	24.00 vs. 5.50; n = 16	V = 122.5; p = 0.0026
POST 0-60 s vs. PRE; median duration (s)	0.445 vs. 0.316; n = 15	V = 95; p = 0.025
POST 0-60 s vs. POST 540-600 s; median duration (s)	0.452 vs. 0.329; n = 16	V = 91; p = 0.122
POST 0-60 s vs. PRE; median inter-call interval (s)	1.429 vs. 0.990; n = 15	V = 84; p = 0.918
POST 540-600 s vs. POST 0-60 s; median inter-call interval (s)	3.088 vs. 1.431; n = 16	V = 92; p = 0.112
POST 0-60 s vs. PRE; median bout size	3.14 vs. 2.68; n = 15	V = 81; p = 0.122
POST 0-60 s vs. POST 540-600 s; median bout size	3.26 vs. 2.00; n = 16	V = 100; p = 0.0124

Order-sensitive Sequence Structure

Next-token prediction showed that short-range order improved predictive performance. In the pooled PRE/POST stage analysis, the PRE no-context model had perplexity 6.26 and accuracy 23.7%, whereas the previous-token model had perplexity 4.30 and accuracy 51.1%. In POST, the no-context model had perplexity 5.28 and accuracy 40.7%, whereas the previous-token model had perplexity 3.66 and accuracy 59.0%. Longer histories generally performed

worse, consistent with sparse longer contexts. Therefore, primary order inference emphasized k = 1 and k = 2 models (Figure 1-4B; Table 1-2).

For primary sequence-order inference, we used pooled adjacent 60-s recovery windows rather than individual 60-s bins. Ordered-vs-shuffled null tests showed that real order was more predictive than token abundance alone. For k = 1, ordered versus shuffled mean perplexity was 4.30 vs. 5.91 in PRE, 3.46 vs. 4.45 in POST 0-60 s, 3.39 vs. 4.09 in POST 60-180 s, 3.77 vs. 4.62 in POST 180-420 s, and 4.18 vs. 4.98 in POST 420-600 s. For k = 2, ordered versus shuffled mean perplexity was 4.56 vs. 5.94 in PRE, 3.52 vs. 4.57 in POST 0-60 s, 3.45 vs. 4.11 in POST 60-180 s, 3.84 vs. 4.67 in POST 180-420 s, and 4.82 vs. 5.53 in POST 420-600 s (all pooled-window p = 0.005) (Figure 1-4C, Table 1-2).

Pooled data were used because calls became less common in post-exposure recordings as the flock moved away. However, we performed an additional 100 s bin analysis as a sensitivity analysis. Ordered 100 s sequences generally had lower perplexity than shuffled sequences, but support was not completely uniform in every bin: k = 1 was not supported in POST 400-500 s (p = 0.0597), and k = 2 was not supported in POST 500-600 s (p = 0.0746) (Figure 1-4D; Table 1-2).

Table 1-2. Tests of local order in vocal-token sequences.

Models using the preceding one (*k* = 1) or two (*k* = 2) tokens were compared with 200 within-trial, within-period shuffles that preserved token abundance while destroying local order; *n* is the number of held-out next-token predictions, "bits" is $\log_2$ (shuffled mean perplexity/observed perplexity), and prediction gain is the percentage reduction in perplexity relative to the shuffled null.

Period	Comparison	Analysis	Value	P value
pre	shuffled	order k=1	n=1827; bits=0.459; prediction gain=27.27%	*P*=.005*
0-60 s	shuffled	order k=1	n=383; bits=0.363; prediction gain=22.26%	*P*=.005*
60-180 s	shuffled	order k=1	n=487; bits=0.273; prediction gain=17.22%	*P*=.005*
180-420 s	shuffled	order k=1	n=568; bits=0.293; prediction gain=18.38%	*P*=.005*
420-600 s	shuffled	order k=1	n=376; bits=0.254; prediction gain=16.11%	*P*=.005*
0-60 s	shuffled	order k=2	n=367; bits=0.378; prediction gain=23.03%	*P*=.005*
60-180 s	shuffled	order k=2	n=471; bits=0.253; prediction gain=16.11%	*P*=.005*
180-420 s	shuffled	order k=2	n=552; bits=0.283; prediction gain=17.81%	*P*=.005*
420-600 s	shuffled	order k=2	n=360; bits=0.199; prediction gain=12.91%	*P*=.005*
0-100 s	shuffled	order k=1	n=585; bits=0.379; prediction gain=23.14%	*P*=.005*

Period	Comparison	Analysis	Value	P value
100-200 s	shuffled	order k=1	n=358; bits=0.263; prediction gain=16.69%	*P*=.005*
200-300 s	shuffled	order k=1	n=227; bits=0.366; prediction gain=22.40%	*P*=.005*
300-400 s	shuffled	order k=1	n=202; bits=0.211; prediction gain=13.57%	*P*=.005*
400-500 s	shuffled	order k=1	n=199; bits=0.133; prediction gain=8.79%	*P*=.059
500-600 s	shuffled	order k=1	n=212; bits=0.205; prediction gain=13.26%	*P*=.005*
0-100 s	shuffled	order k=2	n=569; bits=0.362; prediction gain=22.16%	*P*=.005*
100-200 s	shuffled	order k=2	n=342; bits=0.226; prediction gain=14.52%	*P*=.005*
200-300 s	shuffled	order k=2	n=211; bits=0.222; prediction gain=14.25%	*P*=.005*
300-400 s	shuffled	order k=2	n=189; bits=0.205; prediction gain=13.22%	*P*=.005*
400-500 s	shuffled	order k=2	n=183; bits=0.213; prediction gain=13.75%	*P*=.005*
500-600 s	shuffled	order k=2	n=197; bits=0.102; prediction gain=6.84%	*P*=.075

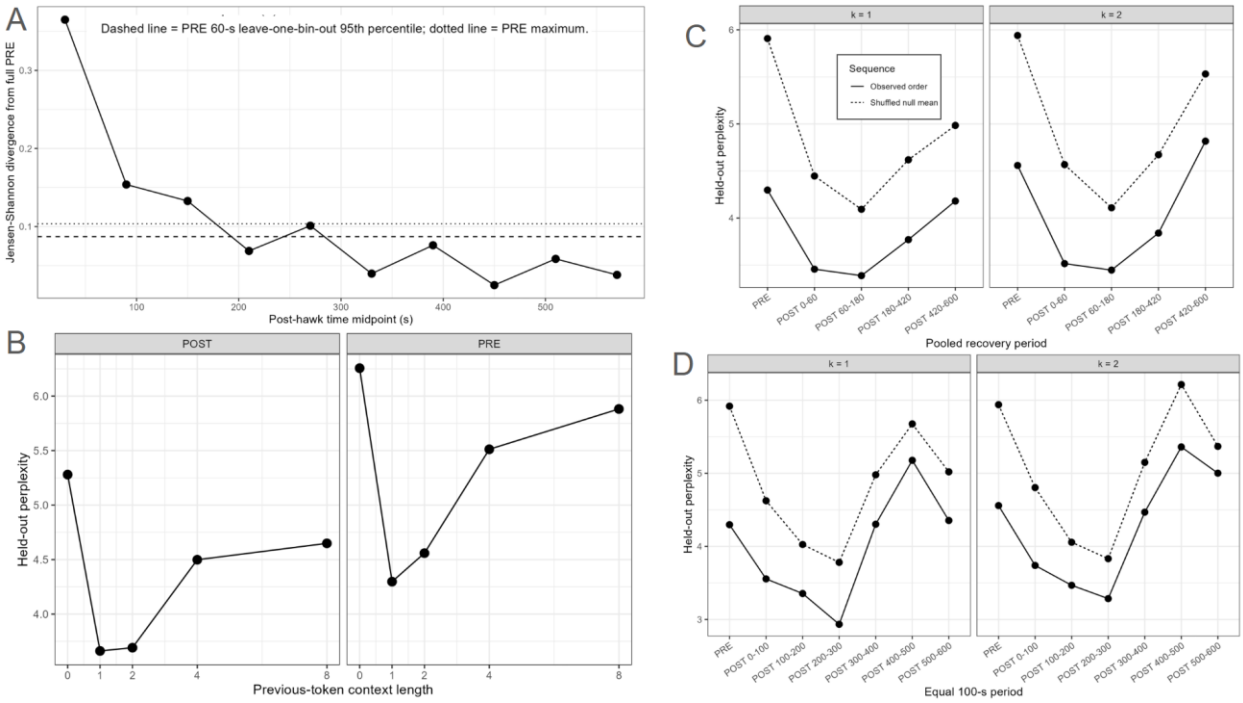


Figure 1-4. Repertoire recovery relative to baseline variation.

(A) Aggregate THAR repertoire recovery relative to natural PRE 60-s variation. Points show aggregate Jensen-Shannon divergence (JSD) between each 60-s POST bin and the full PRE token distribution. The dashed line marks the 95th percentile of the pooled leave-one-PRE-minute-out JSD envelope (0.087), and the dotted line marks its maximum (0.104). Aggregate JSD exceeded the 95% line during POST 0-60, 60-120, 120-180, and 240-300 s. Trial-level mixed-model contrasts likewise identified detectable excess JSD in those four intervals after Holm correction, whereas no interval from 300 s onward was detectably above ordinary PRE variation (Table 1-1). (B) Previous-token context length and held-out next-token prediction in PRE and POST sequences. Leave-one-trial-out n-gram models predicted each retained common THAR token from the preceding k tokens; k = 0 is the no-order repertoire baseline and lower perplexity indicates better prediction. In PRE, perplexity declined from 6.26 at k = 0 to 4.30 at k = 1, and in POST it declined from 5.28 at k = 0 to 3.66 at k = 1. Longer histories generally did not improve performance. (C) Observed versus shuffled THAR token order across pooled recovery periods. For each pooled period, leave-one-trial-out n-gram models with k = 1 or k = 2 previous-token contexts were scored on held-out perplexity. Solid lines show observed token order; dashed lines show the mean shuffled null generated by permuting tokens within trial and period while preserving token counts. Observed sequences had lower perplexity than shuffled sequences in PRE and all pooled POST periods for both k values (all empirical p = 0.005), indicating local order beyond token abundance. (D) Sensitivity analysis using six equal 100-s POST bins. Observed order generally produced lower perplexity than shuffled null sequences.

Results Summary

Across analyses, the most consistent statistical pattern was recovery of token composition, with a large first-minute displacement followed by a noisy but significant decline in JSD. Trial-level excess JSD was detectable in the first three POST minutes and again at 240-300 s; no bin from 300 s onward was detectably above ordinary PRE variation after Holm

correction. Call rate and duration increased in POST 0-60 s, and bout size was larger in the first minute than in the final minute but not clearly larger than PRE under the estimated bout gap. A1 and B1 were the only FDR-significant enriched THAR tokens; SQ1 increased but did not pass the FDR threshold. Inter-call intervals were descriptively longer in later POST bins, but planned interval contrasts did not produce strong paired-test evidence. Sequence analyses supported persistent local order, with k = 2 inference most reliable in pooled adjacent recovery windows rather than sparse individual bins.

Discussion

Hawk exposure reorganized the dusky-throated antshrike vocal event stream across several dimensions. The strongest result was repertoire displacement followed by recovery: first-minute POST tokens differed sharply from PRE, and trial-level JSD declined significantly with time. Excess JSD remained detectable during the first three POST minutes and again during 240-300 s. Because the 180-240 s contrast was not significant but the 240-300 s constrast was, recovery was not strictly monotonic. From approximately 300 s onward, however, no interval was detectably more dissimilar than ordinary PRE variation after correction for multiple comparisons. This pattern supports a temporary reconfiguration of vocal behavior followed by a noisy, graded recovery rather than a single abrupt return point. Antshrikes are sentinel species in Amazonian mixed-species flocks (Munn and Terborgh 1979; Martínez et al. 2018; Williams and Lindell 2021), and such reconfiguration may matter not only to conspecifics but to the broader flock information environment. Previous work shows that flock members benefit from antshrikes anti-predator information and respond to both their movement and calling behavior (Martínez et al. 2018; Williams and Lindell 2021). The present study adds that the sentinel's own vocal structure changes predictably through the course of predator exposure and recovery.

The recovery result is especially important because it changes how late post-hawk calls should be interpreted. Without PRE data, late calls might be classified as part of a sustained alarm response simply because they followed predator presentation. With PRE data, many late events instead appear to form part of a reassembling baseline repertoire. Disturbance ecology similarly defines resilience relative to the intrinsic variability or normal operating range of the pre-disturbance state (Shade et al. 2012a), and empirical community studies have used multivariate distance from pre-disturbance composition to track recovery through time (Shade et al. 2012b). Our late-bin results should nevertheless be interpreted as an absence of detectable excess dissimilarity, not proof that POST and PRE repertoires were identical. This interpretation is also consistent with the risk-allocation view that antipredator behavior changes as danger varies through time (Lima and Bednekoff 1999).

The findings also clarify the role of call labels. We did not treat A1, B1, or other labels as fixed semantic categories. Instead, the same label was allowed to have different event contexts. B1 is the clearest example. It was common in PRE, dominant immediately after hawk presentation as a same-token bout, and present again during recovery. Its local context changed across stages, suggesting that its biological role cannot be inferred from the label alone. This is consistent with broader animal communication work showing that formal labels may miss context-sensitive dimensions of a signal (Sharma et al. 2024; Beguš et al. 2025). In THAR, the relevant additional dimensions are not automatically extracted spectral features, but event-level properties: duration, inter-call waits, bout position, and neighboring tokens.

The sequence results show that order matters, and period-specific analyses clarify the interpretation. Predator exposure did not create local order de novo. Antshrike vocal behavior was already sequence-structured in PRE, remained structured during the acute first-minute response, and continued to show order during pooled recovery windows. The predator response therefore changed the composition and dominant local arrangements of the event stream rather

than converting an unordered baseline into an ordered alarm sequence. This does not demonstrate syntax or semantic compositionality. It does suggest that future vocalization-based experimental designs should account for persistent local organization, with stronger claims about meaning requiring well-planned playback experiments, receiver responses, or field manipulations linking token sequences to flock behavior.

Like call sequence order, bout occurrences appear to be uniformly distributed. Under the estimated 2.85-s gap, first-minute bouts were larger than final-minute bouts, but the first-minute versus PRE contrast was not clearly significant. This shifts the bout result toward a more specific claim that bout organization generally declined across POST recovery, however the later stages of both pre and post are long enough to confer some effects of flock movement away from the recordist.

The analysis provides a disciplined route toward future interspecific communication network work. A mixed-species network should not begin by pooling every species-token transition into one graph. Instead, the THAR event stream can provide a focal reference: which heterospecific calls follow the first-minute THAR configuration, which follow the recovery configuration, and which occur inside the same bouts or perhaps more interestingly, between. Because this paper establishes the focal THAR trajectory, a later study can now ask whether other species respond to specific THAR tokens, to THAR bout onsets, or simply to the shared predator context. That would be a stronger test of interspecific communication than a single pooled adjacency network.

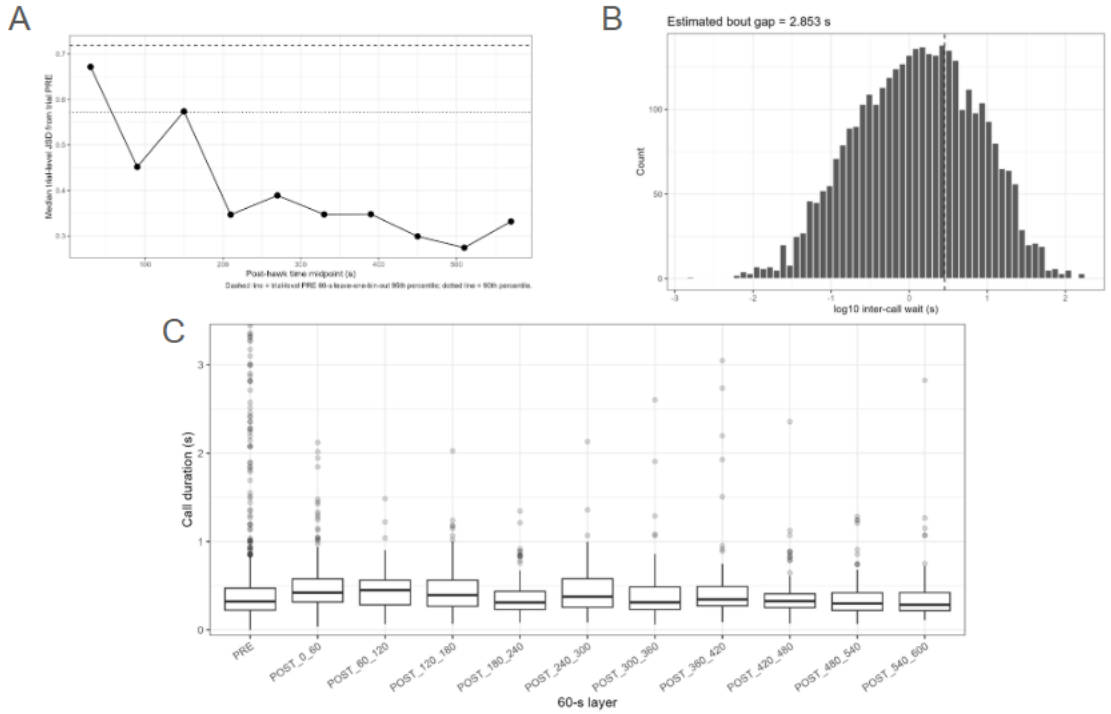
Duration was a useful addition to token identity. POST 0-60 s calls were longer than PRE calls, indicating that the predator response altered event form as well as token composition. Future work with higher signal-to-noise recordings could test whether tokens such as B1 differ acoustically across PRE, first-minute POST, and recovery contexts, paralleling

recent cetacean work showing that traditional labels can hide additional spectral structure (Beguš et al. 2025).

Several limitations are important. First, the analysis focuses on THAR alone. Because this species is seldom found in isolation, future work should build communication patterns with species that form close relationships, such as the white-flanked antwren *Myrmotherula axillaris*. Second, individual 60-s bins are useful for displaying repertoire recovery but can be sparse for token-enrichment and k = 2 sequence-order inference; future sequence analyses could use biologically estimated state periods, including hidden Markov models. Third, the mixed-model recovery analysis tests whether POST bins remain detectably more dissimilar than ordinary PRE variation. Nonsignificant late bins are not a formal equivalence result; establishing statistical similarity would require a prespecified biological equivalence margin and an equivalence test.

Despite these limitations, several observations are clear. A predator did not merely increase THAR use of an alarm call; it shifted the THAR event stream into a short-lived predator-response configuration involving token identity, call duration, bout organization, and local order. As post-hawk time increased, repertoire dissimilarity declined, and from approximately 300 s onward no interval was detectably above ordinary PRE minute-to-minute variation after multiplicity correction. This provides a reproducible framework for studying recovery from perceived danger as recovery of communication structure. Since these trials were conducted in protected forest, it remains to be shown how anthropogenic disturbance affects patterns in antshrike communication. In mixed-species flocks, where public information can affect community movement and habitat use, vocal-event recovery may be an important but undermeasured component of antipredator behavior.

Supplemental Tables and Figures



Supplemental Figure 1-S1. Trial-level recovery, bout length estimates, and call-duration distributions.

(A) Trial-level diagnostic of THAR repertoire recovery relative to PRE variation. Points show the median across trials of trial-level JSD between each 60-s POST bin and that trial's full PRE token distribution; reference lines show pooled trial-level PRE leave-one-bin-out percentiles and are descriptive. Trial-level JSD declined significantly with time in the repeated-measures mixed model (beta = -0.000518 JSD s-1, SE = 0.000106, t = -4.88, df = 128, p = 3.10 x 10-6). (B) Estimated inter-call gap criterion used to define THAR bouts. The histogram shows positive within-trial, within-stage inter-call intervals on a log10 scale. The dashed vertical line marks the 2.853-s bout-separation criterion, calculated as the intersection between the medium- and long-gap components of a three-component Gaussian mixture. A new bout was defined when the interval from the preceding THAR call exceeded this threshold. (C) Distribution of THAR call durations across PRE and 60-s POST layers. Boxplots show event-level call-duration distributions for all THAR events in the full PRE baseline and each 60-s POST bin; the display is truncated to the 1st-99th percentile.

References

Altmann J. 1974. Observational study of behavior: sampling methods. Behaviour. 49:227-267. doi:10.1163/156853974X00534.

Beguš G, Sprouse RL, Leban A, Silva M, Gero S. 2025. Vowel- and diphthong-like spectral patterns in sperm whale codas. Open Mind. 9:1849-1874. doi:10.1162/OPMI.a.252.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 57:289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.

Chen SF, Goodman J. 1999. An empirical study of smoothing techniques for language modeling. *Comput Speech Lang.* 13(4):359–394. doi:10.1006/csla.1999.0128.

Engesser S, Crane JMS, Savage JL, Russell AF, Townsend SW. 2015. Experimental evidence for phonemic contrasts in a nonhuman vocal system. PLoS Biol. 13:e1002171. doi:10.1371/journal.pbio.1002171.

Goodale E, Beauchamp G, Magrath RD, Nieh JC, Ruxton GD. 2010. Interspecific information transfer influences animal community structure. Trends Ecol Evol. 25:354-361. doi:10.1016/j.tree.2010.01.002.

Gueorguieva R, Krystal JH. 2004. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry. 61:310-317. doi:10.1001/archpsyc.61.3.310.

Griffin AS, Galef BG Jr. 2005. Social learning about predators: does timing matter? Anim Behav. 69:669-678. doi:10.1016/j.anbehav.2004.05.020.

Griffin AS, Savani RS, Hausmanis K, Lefebvre L. 2005. Mixed-species aggregations in birds: zenaida doves, Zenaida aurita, respond to the alarm calls of Carib grackles, Quiscalus lugubris. Anim Behav. 70:507-515. doi:10.1016/j.anbehav.2004.11.023.

Hedley RW. 2016. Complexity, predictability and time homogeneity of syntax in the songs of Cassin's Vireo (*Vireo cassinii*). *PLoS One.* 11(4):e0150822. doi:10.1371/journal.pone.0150822.

Holm S. 1979. A simple sequentially rejective multiple test procedure. Scand J Stat. 6:65-70.

Jullien M, Thiollay JM. 1998. Multi-species territoriality and dynamic of neotropical forest understorey bird flocks. J Anim Ecol. 67:227-252.

K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. 2026. Raven Pro: Interactive Sound Analysis Software. Version 1.6.5 [computer software]. Ithaca (NY): Cornell Lab of Ornithology.

Kershenbaum A, Blumstein DT, Roch MA, Akçay Ç, Backus G, Bee MA, et al. 2016. Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol Rev.* 91(1):13–52. doi:10.1111/brv.12160.

Lima SL, Bednekoff PA. 1999. Temporal variation in danger drives antipredator behavior: the predation risk allocation hypothesis. Am Nat. 153:649-659. doi:10.1086/303202.

Lima SL, Dill LM. 1990. Behavioral decisions made under the risk of predation: a review and prospectus. Can J Zool. 68:619-640. doi:10.1139/z90-092.

Lin J. 1991. Divergence measures based on the Shannon entropy. IEEE Trans Inf Theory. 37:145-151. doi:10.1109/18.61115.

Magrath RD, Haff TM, Fallow PM, Radford AN. 2015. Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. Biol Rev. 90:560-586. doi:10.1111/brv.12122.

Martínez AE, Gomez JP. 2013. Are mixed-species bird flocks stable through two decades? Am Nat. 181:E53-E59. doi: 10.1086/669152.

Martínez AE, Parra E, Collado LF, Vredenburg VT. 2017. Deconstructing the landscape of fear in stable multi-species societies. Ecology. 98:2447–2455. doi:10.1002/ecy.1935

Martínez AE, Parra E, Muellerklein OC, Vredenburg VT. 2018. Fear-based niche shifts in neotropical birds. Ecology. 99:1338-1346. doi:10.1002/ecy.2217.

Martínez AE, Parra E, Gomez JP, Vredenburg VT. 2022. Shared predators between primate groups and mixed species bird flocks: the potential for forest-wide eavesdropping networks. Oikos. 2022:e08274. doi:10.1111/oik.08274.

Martínez AE, Zenil RT. 2012. Foraging guild influences dependence on heterospecific alarm calls in Amazonian bird flocks. Behav Ecol. 23:544–550. doi:10.1093/beheco/arr222

Mokross K, Ryder TB, Côrtes MC, Wolfe JD, Stouffer PC. 2014. Decay of interspecific avian flock networks along a disturbance gradient in Amazonia. Proc R Soc B. 281:20132599. doi:10.1098/rspb.2013.2599.

Munn CA. 1986. Birds that cry wolf. Nature. 319:143-145. doi:10.1038/319143a0.

Munn CA, Terborgh JW. 1979. Multi-species territoriality in neotropical foraging flocks. Condor. 81:338-347.

Powell GVN. 1989. On the possible contribution of mixed species flocks to species richness in neotropical avifaunas. Behav Ecol Sociobiol. 24:387-393.

Rauber R, Manser MB. 2017. Discrete call types referring to predation risk enhance the efficiency of the meerkat sentinel system. *Sci Rep.* 7:44436. doi:10.1038/srep44436.

R Core Team. 2024. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Shannon CE. 1948. A mathematical theory of communication. Bell Syst Tech J. 27:379-423, 623-656. doi:10.1002/j.1538-7305.1948.tb01338.x.

Shade A, Peter H, Allison SD, Baho DL, Berga M, Bürgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JBH, Matulich KL, Schmidt TM, Handelsman J. 2012a. Fundamentals of microbial community resistance and resilience. Front Microbiol. 3:417. doi:10.3389/fmicb.2012.00417.

Shade A, Read JS, Youngblut ND, Fierer N, Knight R, Kratz TK, Lottig NR, Roden EE, Stanley EH, Stombaugh J, Whitaker RJ, Wu CH, McMahon KD. 2012b. Lake microbial communities are resilient after a whole-ecosystem disturbance. ISME J. 6:2153-2167. doi:10.1038/ismej.2012.56.

Sharma P, Gero S, Payne R, Gruber DF, Rus D, Torralba A, Andreas J. 2024. Contextual and combinatorial structure in sperm whale vocalisations. Nat Commun. 15:3617. doi:10.1038/s41467-024-47221-8.

Slater PJB, Lester NP. 1982. Minimising errors in splitting behaviour into bouts. Behaviour. 79:153-161. doi:10.1163/156853982X00229.

Stowell D, Gill LF, Clayton DF. 2016. Detailed temporal structure of communication networks in groups of songbirds. J R Soc Interface. 13:20160296. doi:10.1098/rsif.2016.0296.

Suzuki TN, Wheatcroft D, Griesser M. 2016. Experimental evidence for compositional syntax in bird calls. Nat Commun. 7:10986. doi:10.1038/ncomms10986.

Suzuki TN, Wheatcroft D, Griesser M. 2017. Wild birds use an ordering rule to decode novel call sequences. Curr Biol. 27:2331-2336.e3. doi:10.1016/j.cub.2017.06.031.

Templeton CN, Greene E, Davis K. 2005. Allometry of alarm calls: black-capped chickadees encode information about predator size. Science. 308:1934-1937. doi:10.1126/science.1108841.

Templeton CN, Greene E. 2007. Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. Proc Natl Acad Sci USA. 104:5479-5482. doi:10.1073/pnas.0605183104.

Thiollay JM, Jullien M. 1998. Flocking behaviour of foraging birds in a neotropical rain forest and the antipredator defence hypothesis. Ibis. 140:382-394.

Tolkamp BJ, Kyriazakis I. 1999. To split behaviour into bouts, log-transform the intervals. Anim Behav. 57:807-817. doi:10.1006/anbe.1998.1022.

Wilcoxon F. 1945. Individual comparisons by ranking methods. Biometrics Bull. 1:80-83. doi:10.2307/3001968.

Williams HM, Lindell CA. 2021. Context-specific behavior serves as a mechanism of interspecific cohesion in mixed-species flocks. Ornithology. 138:ukab046. doi:10.1093/ornithology/ukab046.

Yeates MP, Tolkamp BJ, Allcroft DJ, Kyriazakis I. 2001. The use of mixed distribution models to determine bout criteria for analysis of animal behaviour. J Theor Biol. 213:413-425. doi:10.1006/jtbi.2001.2425.

CHAPTER 2
A Symphony of Fear and Safety: Distinct Vocalizations Indicate Relative Risk in Mixed-Species Groups

This chapter is based on a manuscript by Eliseo Parra, Benjamin Hoffman, José Miguel Ponciano, Ari E. Martínez, and Daniel T. Blumstein.

Abstract

Animals in mixed-species groups are known to share social information regarding predators, often through alarm-calls linked to immediate threats. Far less is known regarding communication in the time between alarm-calling and active foraging, including the call types and species associated with transitions through periods of *intermediate risk*. We employed hidden Markov models (HMMs) to classify vocalizations from interspecific assemblages following a flyby from a trained hawk. We then compared foraging traits of species vocalizing in each period to understand the relationship between foraging niche and information. Our results identified four periods comprised of differing species characteristics: lower-stratum foraging species alarming in period 1, smaller species vocalizing in period 2, and exposed-position foragers in periods 1-3. By comparing candidate safety vocalization features we describe several calls associated with transitions to safer periods that may function as safety cues or signals. This novel framework highlights unexplored pathways for future ecological studies.

Introduction

Animals living in groups rely extensively on social information and communication to coordinate their behaviors, enhance survival, and drive group dynamics that may couple biodiversity and ecosystem functioning (Bauer & Hoye, 2014; Pringle, 2008). In many systems, animal groups are directly influenced by shared information regarding predators (Laundre et al., 2010). Prey often respond to increases in perceived risk by producing distinct vocalizations, engaging in evasive actions (such as freezing or diving), and making changes to their niche-space to coincide with information-producing species or refugia (Creel & Christianson, 2008; Schmitz et al., 2000). While studies have begun to provide mechanistic examples of how animal

group relationships are maintained (Goodale & Beauchamp, 2010), understanding how social information directly mediates collective behavior is a critical step to assess ecological effects of perturbations between species (Gil et al., 2017; Estes et al., 2011; Schmitz et al., 2010).

Considerable research on animal group communication has focused on the transmission of foraging information and the use of alarm-calls to warn of threats (Magrath et al., 2015). Additionally, many alarm-producing species have been shown to respond to alarm-calls of other species, suggesting complex communication dynamics (Goodale & Kotagama, 2005; Magrath et al., 2009; Hunts et al., 2019; Meise et al., 2020). By focusing on alarm-calling, these studies tend to treat risk communication in binary terms—presence or absence of predators—without exploring the nuances of how varying levels of threat are signaled and perceived among group members. This may overlook the broader repertoire of vocalizations that indicate levels of risk and influence group behaviors (Estes et al., 2011). Understanding graded communication is crucial as it may affect mixed-group cohesion, decision-making, and overall fitness.

Amazonian mixed-species flocks of insectivorous birds provide an ideal system to study these complex behavioral dynamics (Jullien & Thiollay, 1998; Tobias et al., 2012). These flocks are identifiable by their two most common genera of suboscine antbirds (family: Thamnophilidae): ant-shrikes of the genus Thamnomanes and ant-wrens of the genus Myrmotherula (Munn & Terborgh, 1979). Within flocks, Antshrikes represent a larger-bodied genus (17.7 g) who typically forage by fly-catching from a stationary perch, making frequent sallying flights to capture insects flying or falling from above. Antshrikes produce alarm calls that elicit predator-avoidance behaviors in eavesdropping heterospecifics. Antwrens are represented by several, typically smaller species (8.4–9.4 g) within mixed flocks, who forage by closely inspecting substrate for arthropods (Martínez & Zenil, 2012). Territories shared between these species can span many decades, reflecting multiple generations (Martínez et al., 2023). Flock formation centers around the dual challenges of avoiding predators and maximizing foraging

efficiency (Goodale et al., 2020; Mangini et al., 2022). Individuals mitigate these challenges by associating with species that flush insect prey or produce alarms when predators are detected (Munn & Terborgh, 1979; Munn, 1986), optimizing the trade-offs between foraging effort and predator awareness through sociality and acoustic information (Goodale et al., 2020; Martínez & Zenil, 2012).

Foraging behavioral traits such as hunting insects in leaf clusters (i.e., Epinecrophylla antwrens) reduce visibility of predators, increasing dependence on alarm-calling from other individuals (Martínez & Zenil, 2012). Other traits, like body mass, may drive more complex flock patterns (Sridhar et al., 2012). Previous studies found that birds produced lower-quality alarms as predator size increased in response to an increasing size mismatch between predator and prey (Templeton et al., 2005; Martínez et al., 2017). This is consistent with observations of small birds maintaining proximity to large predators (Greeney et al., 2015; Cerpa & Medrano, 2016; Kim et al., 2023) and optimal foraging theory which predicts larger raptors typically consume larger prey (Bozinovic & Medel, 1988; Fargallo et al., 2020). Behavioral research on Accipiter hawks has shown prey vulnerability increases with prey size (Selås, 1993; Cresswell, 1995; Rytkönen et al., 1998). Since Accipiter spp. and other bird-hunting raptors are primary predators of mixed-flock birds, smaller flock species (<15 g) may exist in a size-based refugia while larger flock species experience relatively higher risk, an example of a trait-mediated trophic cascade (Greeney et al., 2015). This may allow smaller species to forage further from protective cover (Whitfield, 2003) and lower foraging heights (Post & Götmark, 1996).

If smaller species in mixed flocks i) experience increased foraging benefits from mixed groups, ii) experience less predation pressure relative to larger flock species, and iii) vocalize information regarding risk, they may be more likely to provide information during times of predation stress, including putative safety cues or “all clear” signals. While recent models and experimental results found that safety cue information may be beneficial in conditions of high

predation pressures (Luttbeg et al., 2020), there have been few studies on the role of safety cues in mixed flocks.

Although prior studies have explored how individuals and groups communicate threats through alarm-calls (Munn & Terborgh, 1979; Munn, 1985; Goodale & Kotagama, 2005; Flower, 2011), the identification of communication associated with decreasing-risk conditions remains underexplored, including those that may facilitate transitions to foraging. Putative safety signals that communicate safe foraging conditions also remain largely unidentified in most ecological contexts. An important step towards identifying safety signals is to define transitional call patterns from periods of alarm through foraging. A simple two-step process may suffice to accurately describe these transitions, whereby alarm-calls precede foraging sounds (Figure 2-1). A more complex pattern would include several stages, with vocalizations from different species when compared across stages (Figure 2-1B).

Given the heterogeneity of flock composition and traits, we predicted that flocks would respond to predation threats with a minimum of three states. The first state contains the well-studied alarm-calls that are emitted following the detection of a predator (Figure 2-1, Prediction B, p1, calls 1-3), hereafter the “acute threat state”. The second state may be characterized by repetitive distress vocalizations (Prediction B, calls 4-6, p2), hereafter the “heightened vigilance state”. We predicted a third safe-state characterized by foraging and its associated sounds (Figure 2-1, Prediction B, p1-p3) hereafter “baseline safety and foraging”. Last, a composite of unique signals and surrounding period sounds may contribute to 4 or more states.

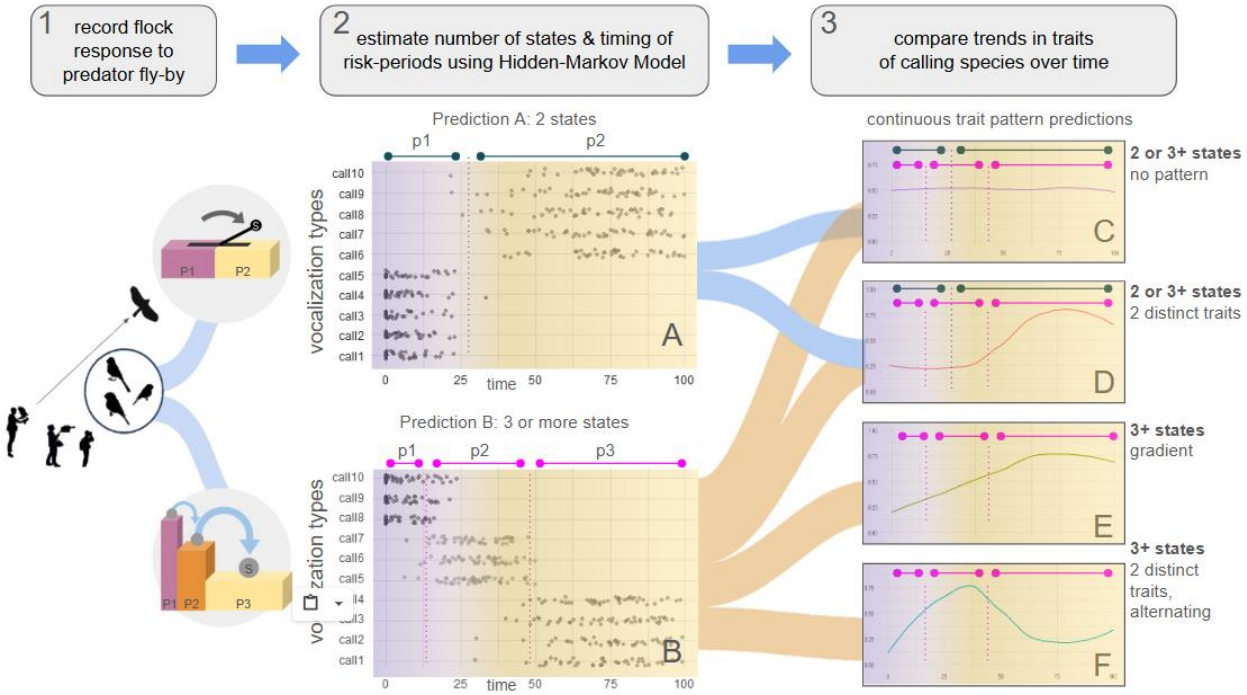


Figure 2-1. Predictions and analytical workflow.

Visual guide to predictions tested with Hidden Markov Model. Step 1 illustrates the collection and recording of mixed-flock vocal responses to predator fly-by. Step 2 shows 2 example predictions of vocalization type responses to a predator using simulated data, with Prediction A showing a "switch" prediction with call distributions in 2 states with 2 risk periods (p1 and p2), and Prediction B showing a quantized "step" prediction with call distributions in 3 states (p1-p3). Step 3 demonstrates several possible patterns using simulated data of species traits over time, with the line representing an average trait value of species calling in 5 s bins over the 600 s time window following the detection of a predator.

Considering the broad range of body masses among species in mixed flocks, we predicted that species which vocalize in intermediate high-risk periods (p2) would be smaller in size (lower risk exposure) and there would be a return to average mass over time (Figure 2-1, Part 3, Prediction F). Alarms may be produced by more exposed and lower-foraging species, followed by calls from higher and protected foragers (Figure 2-1, prediction E).

By mapping vocalizations to risk periods, we compared temporal patterns of alarm-calls, intermediate-risk vocalizations, and putative safety signals. We further compared foraging and morphological characteristics of species vocalizing in each risk period to evaluate changes in species traits between periods (Figure 2-1, Step 3).

Materials and Methods

Overview

To obtain vocalizations from flocks, we experimentally elicited alarm-calls using a trained predator (Bicolored Hawk, *Accipiter bicolor*) and recorded the flock responses as they resumed foraging (Figure 2-1, Step 1). We then classified and timestamped each unique vocalization (alarm, song, or other vocalizations) for 600 s after the hawk fly-by. We used Hidden Markov Models (HMMs), a statistical tool that allows researchers to understand state patterns, to describe the gradient of risk inferred from changes in vocalizations after the hawk flyover (Figure 2-1, Step 2).

Study Site

We conducted our study at the Los Amigos Biological Station (EBLA) 12°34′08.0″ S, 70°06′02.2″ W, located in the Madre de Dios department of Peru. The site features extensive upland forest terrain, marked by the presence of Castaña trees (*Bertholletia excelsa*).

Focal Species Selection

To account for small differences among flocks, comparisons were restricted to species previously identified as mixed-flock participants (Munn & Terborgh, 1979) that occurred in more than four of the eight focal flocks. This criterion excluded rare species while retaining the repeated core of the assemblage, which showed broadly consistent flock diversity and generally one territorial pair per focal species.

Experimental design

The bi-colored hawks were managed by a professional falconry service (Colka Raptors, Peru) and required training prior to work in the field. The training emphasized the hawk leaving the falconer's glove, flying to a perch position of 30-40 m in distance and 15 m height, and returning immediately (Martínez et al., 2017).

Eight different mixed-species flocks were used to gather recordings of flock-vocal response behavior, visited in random order until a total of 20 trials were completed. No flock was tested more than once per day. All trials were conducted between 07:00 and 14:00 h, after the dawn chorus. Flocks in territory disputes were skipped and revisited.

We acclimated the flock to our presence by waiting >10 min before all trials (Martinez 2017). Recording began at the hawk's flight and continued for >600 s after, allowing time for flocks to resume normal foraging vocal behavior. Further attempts at recording past these times disturbed the flocks, causing them to vocalize in response to researcher movement.

To prevent disturbance, three researchers maintained specific distances while searching for flocks. The lead searched ahead, spotting flocks 30-40 m away. The second, with recording equipment, stayed 30 m behind. The falconer and hawk followed 30-40 m further back, ensuring at least 90 m between the falconer and flocks, beyond range of detection. Researchers collected trait data (foraging height, coverage density, distance to cover) during pre and post periods. We used hand signals and a green laser to silently direct the falconer to release the raptor. The hawk was flown to a perch ~30 m away and returned to the falconer's glove. The falconer then retreated to a 100 m distance.

Vocalization reference codes and audio processing

We recorded 200 min of audio after 20 simulated predation events across eight flocks, capturing alarms, songs, and other vocalizations as flocks transitioned from predator detection to foraging. We identified and tagged each flock-species vocalization starting at the hawk flight and extending to 600 s following the flight using Raven Pro 1.6.

Call codes consist of two parts. First, the four-letter species codes (e.g., HARU for *Habia rubica*). Second, each vocalization *type* was coded (e.g., C1, A1) and combined with species codes (e.g., HARU-C1) to create a dataset of vocalizations and timestamps following the predator presentation (Heckman et al. 2016). Rare vocalizations (<5 occurrences, n = 12) and

repetitive calls (>300 in a trial, n = 4) were excluded, leaving 2,304 vocalizations of 23 types from 9 species. Some call-codes identify well studied vocalizations (such as the alarm of the Dusky-throated Antshrike, THAR-A1) but most represent calls that do not have a well-understood use or documented function.

Identifying potential safety signals

To identify the potential use of putative safety signals, we used k-means clustering to classify vocalization types based on their temporal occurrence patterns (skewness, kurtosis, and interquartile range), comparing their characteristics to known alarm vocalizations like those from THAR-A, MYAX-A, and MYME-A (Figure 2-2) (Hunts et al. 2019; Martínez et al. 2017; Munn & Terborgh 1979). Since alarm-calls occur in concentrated bursts following threats, they have distinct distributional behaviors when compared to non-alarm vocalizations. We defined potential safety signals as vocalizations sharing alarm-like distribution patterns, with small interquartile ranges and high kurtosis values, occurring in later periods while transitioning to foraging behavior. Without experimental validation (i.e. playback), we interpret these vocalizations as candidate safety signals rather than confirmed signals.

K-means clustering was implemented in R (version 4.2.3; R Core Team, 2023). Clustering used the Hartigan-Wong algorithm with a maximum of 10 iterations and 25 random initializations. Prior to clustering, all variables (mean time, skewness, kurtosis, and interquartile range) were centered and scaled using the scale() function so that each variable contributed equally to Euclidean distance. Vocalizations were binned in 5 s intervals.

Hidden Markov Model analysis

Since mixed flocks represent temporally stable, multigenerational flocking specialists (Munn & Terborgh,1979; Martínez et al. 2023) with known eavesdropping and cross-communication behavior (Hunts et al., 2019; Camerlenghi et al., 2026), we pooled all species and focused the analysis on call types, retrieving species and trait data after fitting the models.

We employed a multivariate Hidden Markov Model (HMM) (Rabiner, 1989; Putland et al., 2018; Glennie et al., 2023) to infer the level of perceived risk as a hidden "state", relating to the production of call types in response to simulated predation events. This approach transformed the level of risk perceived on a community level into a series of ordered, discrete "state" classifications. The estimation of total state-number was assessed using response data from multiple flocks, reflecting patterns shared by multiple, neighboring groups. Since a combination of call types indicates a hidden state, the unit of observation consisted of a temporal window ("bin"), capturing a collection of calls within its boundaries. Arranging the data in these bins over time makes this a discrete-time model, where each time step (bin) will be denoted with the letter. Their "emissions" over time were used by the model to infer a state or transition from one state to the next. Since the predator was quickly flown and removed from the area, it is reasonable to assume that all subsequent vocal responses following the removal of the predator reflected the current risk condition. Formally, we define the following terms.

States: Let S denote the set of K hidden states of perceived risk:

$$S = \{s_1, s_2, \dots, s_K\}$$

where K denotes the total number of hidden states.

Observations: Let O_t denote the vector of call-type counts observed in temporal bin t:

$$O_t = (o_{1t}, o_{2t}, \dots, o_{bt})$$

where O_{1t} is the number of calls of type i observed in temporal bin t, and b is the number of call types.

State transition probabilities: Let A = a_{ij} be the K × K transition matrix, where each a_{ij} is the conditional probability of moving from state i at time t − 1 to state j at time t:

$$a_{ij} = P\left(S_t = s_j \mid S_{t-1} = s_i\right)$$

This definition expresses the first-order Markov assumption that the next hidden state depends on the immediately preceding hidden state.

Emission probabilities: Let B = b_{jk} denote the probability of observing call-count vector O_k while the hidden process is in state j:

$$b_{jk} = P(O_t = o_k \mid S_t = s_j)$$

These probabilities link the observed call composition in each time bin to the underlying risk state.

Initial state probabilities: Let π be the vector of probabilities that a trial begins in each hidden state:

$$\pi_j = P(S_1 = s_j)$$

Together, the initial-state, transition, and emission probabilities define the likelihood of the observed sequence and permit estimation of the most probable latent-state sequence.

We estimated two key probabilities: emission probabilities (probability of a vocalization occurring), which reflect the likelihood of observing specific vocalizations from each state, and transition probabilities, which describe the likelihood of moving between risk states. Emission probabilities classified temporal bins to discrete risk states, whose transitions to different risk states defined boundaries of risk-periods (a period of several consecutive bins of a similar type) (Glennie et al., 2023). Posterior probabilities were estimated using the forward-backward algorithm. Transition and emission probabilities were estimated via the Baum-Welch algorithm, and the Viterbi algorithm (Visser & Speekenbrink, 2010) identified the most probable sequence of hidden states. The models, built using the depmixS4 package (Visser & Speekenbrink, 2010) in R (R Core Team, 2023), used multinomial response densities to analyze vocalization transitions relative to temporal covariates.

Estimating state number and bin size

To assess the optimal bin length and total number of possible states for our model, we compared HMMs using Integrated Completed Likelihood (ICL) tests. ICL, known for conservative estimates and reduced overfitting, was prioritized due to its reliability in

distinguishing non-overlapping distributions of state-emission probabilities, particularly in datasets with infrequent state changes between observations (Pohle et al., 2017). A large bin size may overlook details of state-changes by the inclusion of too many vocalization types in each frame, while a small bin size may experience over-estimation of state number due to the relatively large differences between adjacent bins. State number was evaluated between two and six states, with more than six states experiencing model instability or oversensitivity (i.e., “splitting” larger states into many, smaller state-switching events), due to differences of call timing between flocks and general tendencies of HMMs to over-estimate state numbers (Pohle et al., 2017). We thus determined that six states were sufficient to establish a reduction in model improvement. We evaluated HMMs with 2-6 states across bin sizes of 5, 10, 20, 30, 40, and 50s, performing 10 iterations per combination, resulting in 300 model comparisons. These analyses guided the selection of state numbers and bin sizes for our final HMM configuration. Technically, this approach is akin to profile likelihood optimization of one of the unknown parameters with the final HMM configuration being such parameter. This model structure specification approach has been advocated before in Taper & Lele (2004) and Taper et al. (2021).

Estimating vocalization state associations and risk periods

We used HMMs to identify transition points between risk-period regimes by recording the transition point for sequential states in a single HMM analysis and repeating the analysis with a random seed 1000 times. The average transition time for each state transition was used for comparing traits and vocalization types across periods.

Incorporating rare signals like alarm-calls posed challenges due to their low occurrence. To address this, vocalization data were bootstrapped 1,000 times, ensuring consistent representation across types. To reflect the relative timing of vocalizations, temporal tags were estimated for each call type. These tags are used to capture the order in which language

naturally unfolds, often applied to linguistic models (Chiche & Yitagesu, 2022). Temporal tags were created by using survival analysis to generate an estimation of the area under a distribution curve (AUC) of a call type over time. This allows for estimation of “censored” observations that continue long past a measurement period (Damuzzo et al., 2019; Messori et al., 1997). This value was established by extracting survival times for each vocalization from their Kaplan-Meier curves using the R package “survival” (Therneau, 2001), averaging adjacent survival probabilities, multiplying by difference between time points and summing the total area under each survival curve.

Posterior probabilities were used to quantify the strength of each vocalization-state association. For each vocalization, the posterior probability was averaged across all successful model iterations and represents the average support for associating that vocalization with a state.

Permutation analysis of functional traits

We compared species trait patterns across four risk periods, using vocalizations recorded during the final 200 s (400-600 s) as the lowest-risk reference period. Preliminary test trials and researcher observations indicated that t > 300 s was sufficient for re-establishing foraging behavior for most flocks, while further recording/following t > 600s would induce flock disturbance vocalizations. Therefore, we determined that 400-600s was optimal for establishing a lowest-risk period. We compared body mass, foraging height, and distance to protective cover, selected *a priori* because of their expected associations with risk. Body-mass data were obtained from AVONET (Tobias et al., 2022).

Because trait values were non-normally distributed and the number of vocalizations varied among periods, we used two-sided studentized permutation tests to compare each risk period r with reference period 5. Studentization incorporates each period’s sample size and variance into the estimated standard error; thus, periods with fewer vocalizations or greater

within-period variability must show a larger difference in means to produce a significant statistic. Effect sizes were estimated by permuting comparison period (CP) and risk periods (n = 10,000) for each of the four comparisons and three functional traits, producing 12 distributions. Test statistics were compared to these distributions to calculate probabilities for differences in functional traits by period. P-values reflect the proportion of permuted test statistics exceeding the observed value. To minimize Type I errors from multiple comparisons, we used the conservative Bonferroni correction, adjusting the threshold to 0.004 for 12 tests (4 periods x 3 traits).

Machine learning performance comparison

We trained a deep neural network model to simultaneously detect and classify vocalizations, using the open-source software Voxaboxen (Mahon et al., 2025). This was done to ensure that vocalizations in this study were annotated consistently across- and within-recordings, and to develop a detection procedure that could be used to analyze other datasets. Using the vocalizations detected by this model, we replicated HMM and trait pattern analysis from the previous steps and compared the results with those based on manually annotated data. Details are in Supplement 1.

Control data performance comparison

To further assess the robustness of our results and to ensure that our findings were not an artifact of our HMM, a control dataset was made by randomizing vocalization types and assigning random timestamps between 0 and 600 s. This dataset, comprising 1,000 samples of each vocalization type, matched the bootstrapped data used in the HMM. We applied identical statistical methods to analyze risk periods and trait patterns of the control data.

Results

Hidden Markov Model state number and bin size: results

Our analysis identified a minimum of four states following a predator flyover. Integrated Completed Likelihood (ICL) comparisons (from 300 HMM iterations with varying state numbers) revealed similar performance for models with four to six states. Bin size comparisons favored 20-50 s bins over smaller intervals, with 30 s and 40 s bins performing best overall. For final analysis, we selected a four-state model with 20 s bins. This choice balances detailed temporal resolution with state clarity. Although 30 s bins performed slightly better over 600 s, 20 s bins better captured rapid vocalization changes immediately after predator detection, such as THAR-A alarms produced by Antshrikes, reflecting a known biologically meaningful pattern.

Estimating vocalization state associations and risk periods: results

In the analysis with 1,000 random starts, 921 fits were successful. The estimated State 1→2 boundary was 42.50 s (SD = 22.70 s; empirical 95% run interval = 19.83–97.83 s; recovered in 813/921 fits), the State 2→3 boundary was 87.47 s (SD = 50.71 s; interval = 39.68–239.08 s; 814/921 fits), and the State 3→4 boundary was 202.42 s (SD = 68.31 s; interval = 87.60–345.60 s; 912/921 fits) (Figure 2-3, A). 21 vocalizations were associated with a specific state with mean posterior probabilities of at least 0.5 (Table 2-1).

State 1 vocalizations, including THAR-A, MYAX-A, HARU-A, and MYLO-A (figure 2-2) are previously identified alarm-calls and have been studied in the context of mixed-species communication (Martínez et al. 2017; Hunts et al. 2019). For example, the White-flanked Antwrens MYAX-A call elicits fleeing responses in heterospecifics, and are strongly associated with predation threats (Hunts et al., 2019).

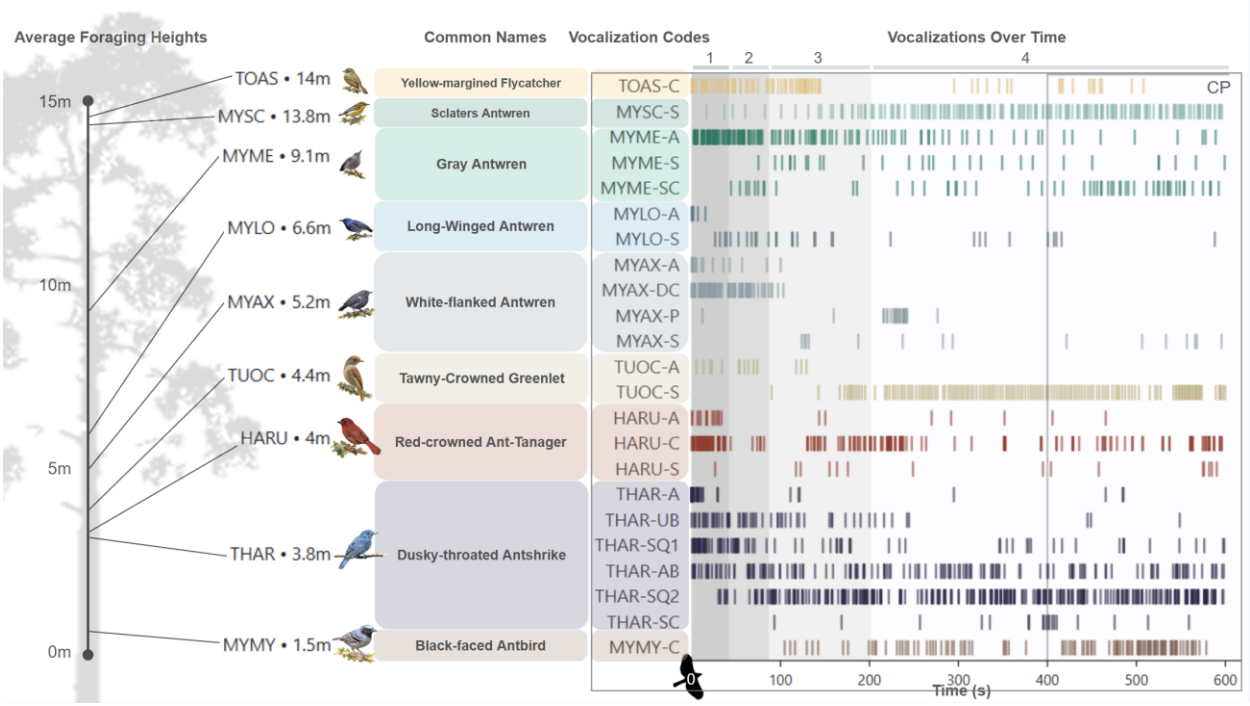


Figure 2-2. Vocalizations, foraging heights, and inferred risk periods.

Foraging heights, vocalizations codes, and bird vocalizations over time following hawk fly-by (hawk is presented at 0 s). Separate shaded regions behind vocalization graph indicate times separating "risk periods" 1-4 identified by Hidden Markov Model. Lowest risk comparison period is indicated by grey box and labeled "CP", comparison period. Call data is pooled from all trials. Bird illustrations were digitally adapted from photographs by Hector Bottai (*Tolmomyias assimilis*, 2014; *Myrmotherula sclateri*, 2023; *Myrmotherula menetriesii*, 2015b; *Myrmotherula longipennis*, 2015a; *Myrmotherula axillaris*, 2022; *Thamnomanes ardesiacus*, 2021), Dominic Sherony (*Tunchiornis ochraceiceps*, photographed under the name *Hylophilus ochraceiceps*, 2012), Instituto Últimos Refúgios (*Habia rubica*, 2012), and João Quental (*Myrmoborus myotherinus*, 2014) via Wikimedia Commons. The corresponding adapted illustrations are licensed under CC BY-SA 4.0 (Bottai and Instituto Últimos Refúgios), CC BY-SA 2.0 (Sherony and Timm) and CC BY 2.0 (Quental).

Seven vocalizations associated with Risk Period 2 have posterior probabilities > 0.5, including: TUOC-A, THAR-UB, MYME-A, THAR-SQ1, TOAS-C, HARU-S, and MYLO-S. Combined, Risk Periods 1 and 2 accounted for less time than any subsequent risk period and feature a majority of the calls unique to the time immediately following the detection of a predator.

Risk Periods 3 and 4 featured state calls with broader interquartile ranges (IQR, Table 2-1), many categorized as "songs" in acoustic reference catalogs. Only two calls exceeded a 0.5

posterior probability in State 3 (MYMY-C, MYAX-P), with MYAX-P showing a probability of 1.0 (Table 2-2). Risk Period 4 had reduced heterogeneity and included vocalizations with the largest IQR values. Some have been previously associated with flock-movement contexts, such as THAR-SQ1, THAR-SQ2, and THAR-AB, and have been described as repositioning calls responded to by heterospecific antwrens (Williams and Lindell 2021). Others are territorial songs or lack defined function, such as MYAX-S, MYSC-S, TUOC-S, and MYME-S.

Table 2-1. Temporal and distributional characteristics by K-means cluster.

Mean time and measurements of distribution characters for all vocalization types used in the analysis. Cluster assignments reflect the grouping results of the unsupervised K-means analysis using kurtosis, skewness and IQR. Reported average distribution characteristics are the mean values of all vocalizations within the cluster. Bold and (*) indicate vocalizations previously studied or described as an alarm.

Vocalization	Cluster	Mean time (s)	Risk period	Kurtosis	Skewness	IQR (s)	Posterior state	Posterior probability	Cluster mean kurtosis	Cluster mean skewness	Cluster mean IQR (s)
MYLO-A	1	5.28	1	66.94	7.77	6.24	1	**1.00**	37.28	5.21	31.88
MYAX-A*	1	24.42	1	52.84	6.57	35.85	1	**1.00**	37.28	5.21	31.88
MYAX-DC	1	36.58	1	10.65	2.82	40.37	1	**0.63**	37.28	5.21	31.88
THAR-A*	1	45.08	2	85.52	8.97	8.26	1	**1.00**	37.28	5.21	31.88
TUOC-A	1	60.19	2	11.44	2.96	47.25	2	**0.78**	37.28	5.21	31.88
MYAX-P	1	221.47	4	18.67	4.02	14.99	3	**1.00**	37.28	5.21	31.88
THAR-SC	1	375.62	4	14.91	3.39	70.18	4	**0.98**	37.28	5.21	31.88
THAR-UB	2	102.96	3	6.76	1.95	97.62	2	**0.60**	8.44	2.02	141.08
MYME-A*	2	122.52	3	6.47	1.96	130.77	2	**0.55**	8.44	2.02	141.08
HARU-A*	2	110.86	3	13.65	3.27	165.24	1	**0.75**	8.44	2.02	141.08
THAR-SQ1	2	147.91	3	14.61	3.18	155.27	2	**0.50**	8.44	2.02	141.08
TOAS-C	2	156.42	3	4.02	1.34	99.47	2	**0.69**	8.44	2.02	141.08
MYAX-S	2	350.67	4	17.45	3.74	153.62	4	**0.93**	8.44	2.02	141.08
MYSC-S	2	382.35	4	2.92	0.74	171.03	4	**0.60**	8.44	2.02	141.08
TUOC-S	2	388.03	4	1.62	0.00	155.60	4	**0.60**	8.44	2.02	141.08
HARU-C	3	221.57	3	8.87	2.43	214.94	1	0.42	5.18	1.56	214.75
THAR-AB	3	268.43	4	3.31	0.97	232.78	4	**0.52**	5.18	1.56	214.75
HARU-S	3	344.98	4	8.11	2.67	227.65	2	**0.59**	5.18	1.56	214.75
MYME-S	3	285.48	4	4.20	1.54	190.35	4	**0.90**	5.18	1.56	214.75
THAR-SQ2	3	340.70	4	2.40	0.49	218.87	4	**0.68**	5.18	1.56	214.75
MYMY-C	3	412.26	4	4.17	1.26	203.92	3	**0.64**	5.18	1.56	214.75
MYLO-S	4	183.33	3	5.29	1.84	250.12	2	**0.68**	4.55	1.63	263.22
MYME-SC	4	382.03	4	3.82	1.42	276.31	3	0.46	4.55	1.63	263.22

Identifying potential safety signals: results

The k-means cluster analysis identified four distinct clusters reflecting risk-period patterns. Cluster 1 contained vocalizations with heavy-trailed distributions (kurtosis mean = 37.28, species n = 4, calls n = 7) and small interquartile ranges (mean = 31.88 s), including many previously-studied alarm calls which occur immediately following predator flight. These calls include the most well-studied signals in the cohort of vocalizations, and calls MYAX-P and THAR-SC appear the most likely signals of safety periods (figure 2-3, C. cluster 1). Cluster 2 featured calls with lower kurtosis (mean = 8.44, species n = 7, calls n = 8) and larger IQRs (mean = 141.08), many skewed to the 100 s period following predator detection, such as HARU-A, MYME-A and THAR-UB. Notable exceptions to this pattern include MYSC-S and TUOC-S (figure 2-3, C. cluster 2) which may be cues or signals of safe foraging conditions. Clusters 3 (kurtosis mean = 5.18, species n = 4, calls n = 6) and 4 (kurtosis mean 4.55, species n = 2, calls n = 2) consisted of low kurtosis calls and wide IQRs (IQR cluster 3 mean = 214.75 s, 4 = 263.22 s), representing calls prevalent across the measurement period, except for pauses in risk-periods 1 and 2 (figure 2-3, C. Clusters 3 & 4).

Table 2-2. Functional-trait comparisons among risk periods.
Values comparing several functional trait parameters between risk periods. Mass is bird weight in grams (g). Bird foraging height is the average observed height of a species over the field experiment. Distance to cover is the average estimated distance a species was found in relation to protective cover. Bonferroni-corrected significant values (p < 0.004) are in bold.

Trait	Period	Hand mean	Hand difference	Hand p value	Machine-model mean	Machine-model difference	Machine-model p value	Null mean	Null difference	Null p value
Mass (g)	**1**	17.51	1.47	0.008	14.87	1.54	0.011	16.01	-0.07	0.858
Mass (g)	**2**	**13.42**	**-2.62**	**<0.001**	11.42	-1.92	0.018	15.71	-0.37	0.681
Mass (g)	**3**	16.84	0.80	0.112	12.90	-0.43	0.444	17.29	1.21	0.271
Mass (g)	4	15.47	-0.697	0.028	13.6	-0.157	0.494	15.0	-0.124	0.675
Avg. foraging height (m)	**1**	**6.37**	**-2.97**	**<0.001**	**7.72**	**-9.38**	**<0.001**	8.35	-.21	0.667
Avg. foraging height (m)	**2**	7.85	-1.50	0.014	**10.60**	**-6.50**	**<0.001**	8.72	0.16	0.899
Avg. foraging height (m)	**3**	7.96	-1.38	0.024	14.09	-3.01	0.019	7.58	-0.99	0.313

Trait	Period	Hand mean	Hand difference	Hand p value	Machine-model mean	Machine-model difference	Machine-model p value	Null mean	Null difference	Null p value
Avg. foraging height (m)	4	9.44	0.09	0.893	18.98	1.88	0.107	8.52	-0.04	0.949
Distance to cover (m)	**1**	**4.25**	**1.29**	**<0.001**	**3.51**	**1.42**	**<0.001**	3.08	-0.62	0.042
Distance to cover (m)	**2**	**4.90**	**1.94**	**<0.001**	**3.51**	**1.42**	**<0.001**	3.08	-0.62	0.043
Distance to cover (m)	**3**	**4.54**	**1.58**	**<0.001**	**3.00**	**0.90**	**0.001**	3.86	0.16	0.674
Distance to cover (m)	4	3.11	0.15	0.314	1.63	-0.47	0.011	3.75	0.05	0.812

Permutation analysis of functional traits: results

The results of the permutation analysis suggest that, over the response period, the average traits of vocalizing birds differed significantly in five comparisons ($p < 0.004$, Table 2-2). Compared to the reference period, Period 1 contained calls from birds who forage lower in the forest column (-2.97 m, $p < 0.001$). Period 2 was composed of calls from smaller, lower mass birds (-2.62 g, $p < 0.001$). Periods 1,2, and 3 contained calls from birds who forage further from cover, in less protected locations (+1.29 m, +1.94 m, +1.58 m, $p < 0.001$) (figure 2-3, B).

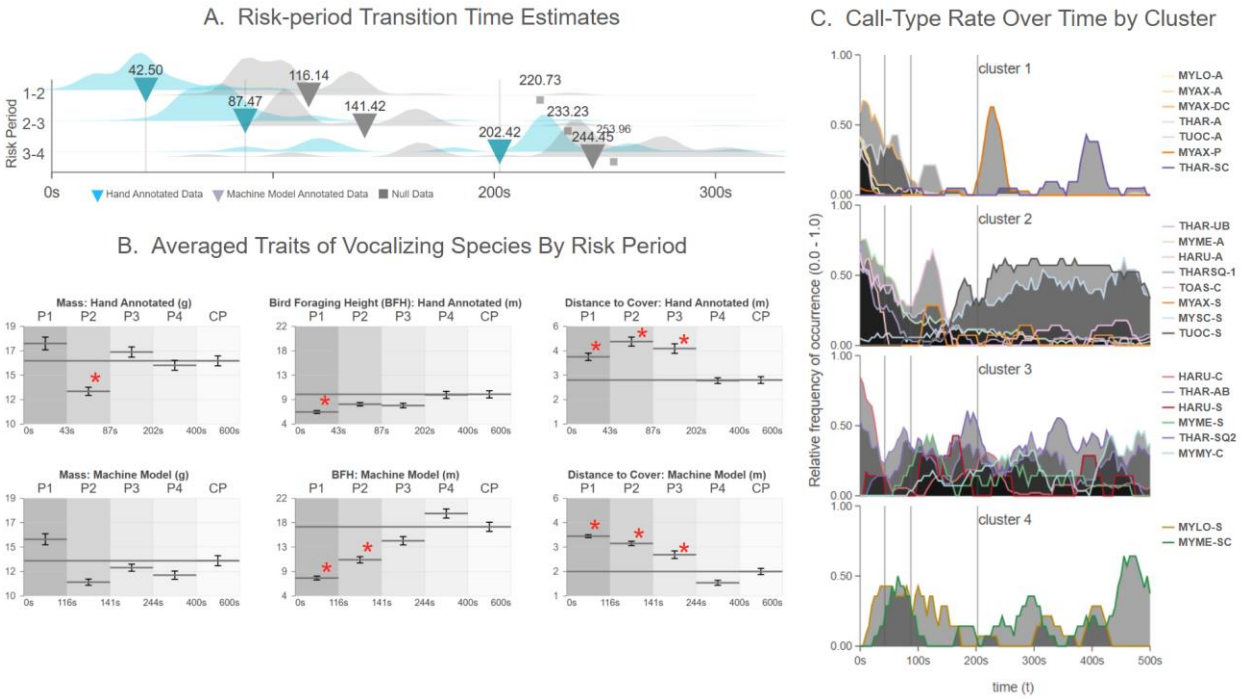


Figure 2-3. Risk-period estimates and functional-trait comparisons

(A) Risk-period estimates and average from multiple Hidden Markov Model iterations (n = 1000). Mean transition times for hand-annotated data are 1-2 = 42.50 s, 2-3 = 87.47 s and 3-4 = 202.42 s, indicated by blue triangles. Grey triangles indicate transition times estimated from machine-annotated data, 1-2 = 116.14 s, 2-3 = 141.42 s, and 3-4 = 244.45 s. Grey squares indicate transition estimates from null data, 1-2 = 220.73 s, 2-3 = 233.23 s, and 3-4 = 253.96 s. (B) Patterns of trait value means from calling species in 5 s bins following detection of predator at time 0. Hand and machine values are scaled equally for visual comparison. Risk periods 1-4 and comparison period are shown as grey shaded regions, significant contrasts have a red asterisk. (C) Vocalization types over time by cluster and scaled by proportion of 1. Each cluster (1-4) is grouped by the output of k-means results. The distribution of the MYAX-P and THAR-SC can be visually inspected in the cluster 1 call frequencies graph.

Machine learning performance comparison: results

In the analysis using machine learning annotated data with 1,000 random starts, 918 fits were successful. The estimated State 1→2 boundary was 116.14 s (SD = 54.45 s; empirical 95% run interval = 87.54–234.04 s; recovered in 793/918 fits), the State 2→3 boundary was 141.42 s (SD = 60.21 s; interval = 104.43–302.97 s; 761/918 fits), and the State 3→4 boundary was 244.45 s (SD = 69.23 s; interval = 104.43–458.57 s; 891/918 fits). Permutation analysis using machine learning-based detections resulted in five significant trait differences. No differences were found significant in mass comparisons, while another foraging height was found significantly lower in periods 1 (-9.38, $p < 0.001$) and 2 (-6.5m, $p < 0.001$). Values for

average distance-to-cover, a value associated with species-level boldness, did not return to baseline similarity until period 4, with states 1, 2, and 3 being significantly higher9 (period 1 +1.42, $p < 0.001$, period 2 +1.42, $p < 0.001$, period 3 + 0.90, $p = 0.001$) (figure 2-2). While machine annotation may be insensitive to the most subtle calls of smaller species, overall patterns appeared similar between the hand annotated dataset and machine-annotated (figure 2-3, B).

Control data HMM comparison: results

The control model using randomized vocalization types and timestamps displayed greater model instability when compared to hand and machine-annotated data. The model retained 299 out of 1,000 random starts, estimated the State 1→2 boundary at 220.73 s (SD = 18.08 s; empirical 95% run interval = 175.77–235.40 s; recovered in 299/299 retained fits), the State 2→3 boundary at 233.23 s (SD = 6.37 s; interval = 224.51–241.63 s; 299/299 retained fits), and the State 3→4 boundary at 253.96 s (SD = 12.46 s; interval = 237.11–282.58 s; 299/299 retained fits). No trait differences from control data state boundaries were found significant.

Discussion

Animals may emit alarm-calls following predator exposure but eventually must resume foraging (Sridhar et al., 2009; Sridhar & Shanker, 2014). Understanding these transitions in mixed-species groups is challenging due to the complexity and scarcity of annotated datasets. Our experimental design used integrated repeated, in-situ presentations and meticulously hand-annotated vocalization datasets to describe novel interspecific communication patterns, train machine learning models, and identify calls associated with distinct risk periods. We interpreted the four HMM-derived periods as an ordinal risk–safety continuum comprising 1) acute threat, 2) heightened vigilance, 3) recovery, and 4) baseline safety.

Risk periods: overview

HMMs and k-means clustering revealed temporal patterns that align with seminal work on the graded progression of threat assessment (Lima & Dill, 1990). In trait-mediated trophic cascades, a species may alter their behavior around predators when considering the effects of a third species (Greeney et al., 2015; Cerpa & Medrano, 2016; Kim et al., 2023). For example, very large or small species may experience "size refugia". Hummingbirds often appear fearless in approaching hawks or other large predators (Greeney et al., 2015). Other small birds have been observed nesting within a hawk's nest (Cerpa & Medrano, 2016). These patterns align with other taxonomic systems: models of social bee behavior predict smaller bees experience lower predation risk (Rodríguez-Gironés & Bosch, 2012), while African lions (*Panthera leo*) are shown to hunt fewer smaller Impala (*Aepyceros melampus*) (Owen-Smith & Mills, 2008). These effects may best explain the mass-trait patterns we observed in the progression of multiple risk periods following the experimental predation event.

Period 1: Acute threat

Acute threat periods featured alarms from species like the Dusky-throated Antshrike *Thamnomanes ardesiacus*, White-flanked Antwren *Myrmotherula axillaris*, Long-winged Antwren *M. longipennis* and Red-crowned Ant-tanager *Habia rubica* (Figure 2-2: THAR-A, HARU-A, MYLO-A, MYAX-A). HMM estimates assigned these calls with posterior probabilities > 0.50 (table 2-1). While model predictions extend this period to about 42 s, many alarm vocalizations have low kurtosis, high skewness, and smaller interquartile ranges, suggesting calls are compressed to the earliest moments in the trial. When compared to safety periods, the species vocalizing in this period are those who forage lower in the forest column, further from cover (figure 2-3, B).

Period 2: Heightened vigilance

Trait comparisons showed that smaller species, like antwrens and flycatchers, dominated the vocalizations of Period 2. Smaller species may more safely signal reduced threats from these riskier locations, supported by a size-refugia (Greeney et al., 2015; Post & Götmark, 1996; Rodríguez-Gironés & Bosch, 2012; Martínez et al., 2018; Fargallo et al., 2020). Previous studies support this: antshrikes produce weaker alarms in response to larger raptors, following a prey-predator size mismatch (Martínez et al., 2017). White-flanked antwrens (5-10.6 g) may form mixed flocks away from alarm-producing species, showing that antwrens are not dependent on alarm information (Botero, 2002). Our findings are consistent with previous analysis of body size and predation risk which reported consistent relationships between body mass and risk-taking behavior across taxa (Beauchamp, 2023).

Species vocalizing in this period also forage further from cover, and the foraging height of calling species returns to baseline (figure 2-3, B). Notable calls with posterior probabilities >.5 for the period of heightened vigilance include MYME-A, TUOC-A, TOAS-C, and THAR-UB (figure 2-2). Smaller species may provide safety information to maintain flock cohesion, influencing intermediate-risk signaling and safety dynamics. This could help explain reduced flock-recruitment in fragmented forests, potentially de-coupling associations between species who utilize niches of very different strata (Mokross et al., 2014, Rutt & Stouffer, 2021).

Period 3: Recovery and foraging resumption

Period 3 is characterized by overlapping vocalizations from the end of preceding higher-risk stages and the beginning of Period 4 vocalizations. A single call, MYAX-P, had a posterior probability of 1.0 for recovery Period 3 (Table 2-1; Figure 2-2). The MYAX-P call's high kurtosis, Cluster 1 designation, and apparent temporal concentration at the intersection of high risk and foraging are consistent with a potential safety-signal function; however, its use in heterospecific contexts should be tested with an in situ playback study.

Period 4: Baseline safety and foraging

Period 4, estimated to begin around 200 s post hawk-flyby, has no significant patterns in vocalizing species composition when compared to the final 200 s of the census period. Period 4 includes repetitive songs and calls from species like Sclater's antwren *Myrmotherula sclateri*, Greenlets, and semi-terrestrial antbirds. Many calls in this period are songs, e.g. MYME-S and MYAX-S. In baseline periods, some vocalizations shift to quieter calls, including "soft calls" that combine multiple song elements (e.g., THAR-SC).

General synthesis

We have shown that when a predator is detected by a mixed-species flock, clear temporal patterns emerge in the types of vocalizations produced and that these patterns contain information about relative predation risk. Such patterns may extend to other taxa, such as in alarm-eavesdropping relationships described in Antshrike and Tamarin systems (Martinez et al., 2022), or other systems outside of the Neotropics. Resolving behavioral complexity at the intersection of risk, diversity, and vocal communication within mixed-species animal groups provides a more complete understanding of how diverse communities function. By extending the characterization of ecological systems beyond conventional alpha- and beta-diversity metrics, our inferential approach captures often-overlooked variation in behavioral processes that should be treated as a fundamental axis of ecological assessment. Ground-truthed through direct observation and hand-annotated recordings, this approach also provides a foundation for integrating contemporary large-scale data-processing tools to investigate understudied social complexity. This framework translates interspecific vocal repertoires into functional units linked to species traits and habitat parameters and may thereby contribute to future assessments of ecological function through measurements of acoustic interactions among species that respond directly to changing environmental conditions.

Acknowledgements

We thank local logistical stewards and the EBLA staff for facilitating conservation-compatible research. We are grateful to Colka Raptors (Piura, Peru) for professional falconry support and to the trained Bicolored Hawks “Lola” and “Cartavio”. This work was supported by the U.S. National Science Foundation (NSF IOS-2032478; NSF IOS-2032469). Special thanks to ornithologist Jorge Novoa and falconer Edwin Velasquez, whose expertise and knowledge are irreplaceable.

Data Availability Statement

Additional supporting information and all code and data supporting the results have been archived and may be accessed at https://doi.org/10.17605/OSF.IO/54URY

References

Acevedo, M.A., Corrada-Bravo, C.J., Corrada-Bravo, H., Villanueva-Rivera, L.J. & Aide, T.M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, **4**, 206–214. https://doi.org/10.1016/j.ecoinf.2009.06.005

Amador, A. & Mindlin, G.B. (2023). The dynamics behind diversity in suboscine songs. *Journal of Experimental Biology*, **226**, jeb227975. https://doi.org/10.1242/jeb.227975

Ames, P.L. (1971). *The morphology of the syrinx in passerine birds*. New Haven, CT: Peabody Museum of Natural History, Yale University.

Bauer, S., & Hoye, B. J. (2014). Migratory animals couple biodiversity and ecosystem functioning worldwide. *Science, 344*(6179), 1242552. https://doi.org/10.1126/science.1242552

Beauchamp, G. (2023). Susceptibility to predation varies with body mass, foraging niche, and anti-predator responses among bird species. *Birds*, **4**, 73–84. doi: 10.3390/birds4010006.

Botero, C.A. (2002). Is the white-flanked antwren (Formicariidae: *Myrmotherula axillaris*) a nuclear species in mixed-species flocks? A field experiment. *Journal of Field Ornithology*, **73**, 74–81. https://doi.org/10.1648/0273-8570-73.1.74

Bottai, H. (2014a). Myrmotherula brachyura – Pigmy antwren [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_brachyura_-_Pigmy_antwren.JPG.

Bottai, H. (2014b). Tolmomyias assimilis – Zimmer's flatbill [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Tolmomyias_assimilis_-_Zimmer%27s_flatbill.JPG.

Bottai, H. (2015a). Myrmotherula longipennis – Long-winged antwren (male) [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_longipennis_-_Long-winged_antwren_%28male%29.jpg.

Bottai, H. (2015b). Myrmotherula menetriesii – Grey antwren (male) (cropped) [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_menetriesii_-_Grey_antwren_%28male%29_%28cropped%29.jpg.

Bottai, H. (2022). Myrmotherula axillaris White-flanked Antwren (male); Ramal do Noca, Rio Branco, Acre, Brazil [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_axillaris_White-flanked_Antwren_%28male%29;_Ramal_do_Noca,_Rio_Branco,_Acre_,_Brazil.jpg.

Bozinovic, F. & Medel, R.G. (1988). Body size, energetic and foraging mode of raptors in central Chile. *Oecologia*, **75**, 456–458. https://doi.org/10.1007/BF00376951

Bradbury, J.W. & Vehrencamp, S.L. (2011). *Principles of animal communication* (2nd edn). Sunderland, MA: Sinauer.

Camerlenghi, E., Parra, E., Novoa, J., Blumstein, D.T., Ponciano, J.M. & Martínez, A.E. (2026). Small canopy species drive information highways about predators in an Amazonian rainforest. *Current Biology*, **36**, R318–R319. doi: 10.1016/j.cub.2026.02.006.

Cerpa, P. & Medrano, F. (2016). Breeding of the plain-mantled tit-spinetail (*Leptasthenura aegithaloides*) in a variable hawk (*Geranoaetus polyosoma*) nest in the Atacama Desert, Chile. *Biodiversity and Natural History*, **2**, 6–9.

Chiche, A. & Yitagesu, B. (2022). Part of speech tagging: A systematic review of deep learning and machine learning approaches. *Journal of Big Data*, **9**, 117. https://doi.org/10.1186/s40537-022-00561-y

Creel, S., & Christianson, D. (2008). Relationships between direct predation and risk effects. *Trends in Ecology & Evolution, 23*(4), 194–201. https://doi.org/10.1016/j.tree.2007.12.004

Cresswell, W. (1995). Selection of avian prey by wintering sparrowhawks *Accipiter nisus* in southern Scotland. *Ardea*, **83**, 381–389.

Damuzzo, V., Agnoletto, L., Leonardi, L., Chiumente, M., Mengato, D. & Messori, A. (2019). Analysis of survival curves: statistical methods accounting for the presence of long-term survivors. *Frontiers in Oncology*, **9**, 453. https://doi.org/10.3389/fonc.2019.00453

Dalziell, A.H., Welbergen, J.A., Igic, B. & Magrath, R.D. (2015). Avian vocal mimicry: A unified conceptual framework. *Biological Reviews*, **90**, 643–668. https://doi.org/10.1111/brv.12129

Dawkins, M.S. (1989). *Unravelling animal behaviour*. Oxford: Wiley.

Denton, T., Wisdom, S. & Hershey, J.R. (2022). Improving bird classification with unsupervised sound separation. In: *ICASSP 2022 – IEEE International Conference on Acoustics, Speech and Signal Processing*, 826–830. https://doi.org/10.1109/ICASSP43922.2022.9747202

Dolby, A.S. & Grubb, T.C. Jr (1998). Benefits to satellite members in mixed-species foraging groups: An experimental analysis. *Animal Behaviour*, **56**, 501–509. https://doi.org/10.1006/anbe.1998.0808

Estes, J.A., Terborgh, J., Brashares, J.S., Power, M.E., Berger, J., Bond, W.J. *et al.* (2011). Trophic downgrading of planet Earth. *Science*, **333**, 301–306. https://doi.org/10.1126/science.1205106

Fallow, P.M. & Magrath, R.D. (2010). Eavesdropping on other species: Mutual interspecific understanding of urgency information in avian alarm calls. *Animal Behaviour*, **79**, 411–417. https://doi.org/10.1016/j.anbehav.2009.11.018

Fargallo, J.A., Navarro-López, J., Palma-Granados, P. & Nieto, R.M. (2020). Foraging strategy of a carnivorous–insectivorous raptor species based on prey size, capturability and nutritional components. *Scientific Reports*, **10**, 10287. https://doi.org/10.1038/s41598-020-64504-4

Flower, T. (2011). Fork-tailed drongos use deceptive mimicked alarm calls to steal food. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1548–1555. doi: 10.1098/rspb.2010.1932.

Geer, T. (1981). Factors affecting the delivery of prey to nestling sparrowhawks (*Accipiter nisus*). *Journal of Zoology*, **195**, 71–80. https://doi.org/10.1111/j.1469-7998.1981.tb01894.x

Gil, M.A., Zill, J. & Ponciano, J.M. (2017). Context-dependent landscape of fear: Algal density elicits risky herbivory in a coral reef. *Ecology*, **98**, 534–544. https://doi.org/10.1002/ecy.1668

Glennie, R., Adam, T., Leos-Barajas, V., Michelot, T., Photopoulou, T. & McClintock, B.T. (2023). Hidden Markov models: Pitfalls and opportunities in ecology. *Methods in Ecology and Evolution*, **14**, 43–56. https://doi.org/10.1111/2041-210X.13801

Goodale, E. & Beauchamp, G. (2010). The relationship between leadership and gregariousness in mixed-species bird flocks. *Journal of Avian Biology*, **41**, 99–103. https://doi.org/10.1111/j.1600-048X.2009.04828.x

Goodale, E. & Kotagama, S.W. (2005). Alarm calling in Sri Lankan mixed-species bird flocks. *The Auk*, **122**, 108–120. https://doi.org/10.1093/auk/122.1.108

Goodale, E., Beauchamp, G., Magrath, R. D., Nieh, J. C., & Ruxton, G. D. (2020). Mixed company: A framework for understanding the composition and organisation of mixed-species animal groups. *Biological Reviews, 95*(5), 1114–1130. https://doi.org/10.1111/brv.12683

Greeney, H.F., Meneses, M.R., Hamilton, C.E., Lichter-Marck, E., Mannan, R.W., Snyder, N. *et al.* (2015). Trait-mediated trophic cascade creates enemy-free space for nesting hummingbirds. *Science Advances*, **1**, e1500310. https://doi.org/10.1126/sciadv.1500310

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **28**, 100–108.

Heckman, J., McGuinness, B., Celikel, T. & Englitz, B. (2016). Determinants of the mouse ultrasonic vocal structure and repertoire. *Neuroscience & Biobehavioral Reviews*, **65**, 313–325. https://doi.org/10.1016/j.neubiorev.2016.03.029

Hunts, C., Heather, M., Martínez, A.E. & Parra, E. (2019). Experimental evidence of alarm eavesdropping on Amazonian mixed-flock followers by two social sentinel species: the dusky-throated antshrike (*Thamnomanes ardesiacus*) and the cinereous antshrike (*T. schistogynus*). *Ornitología Neotropical*, **30**, 224–228. https://doi.org/10.58843/ornneo.v30i0.452

Instituto Últimos Refúgios (2012). Tiê-do-mato-grosso (Habia rubica) – Red-crowned Ant-Tanager [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Ti%C3%AA-do-mato-grosso_%28Habia_rubica%29_-_Red-crowned_Ant-Tanager.jpg.

Jullien, M. & Thiollay, J. (1998). Multi-species territoriality and dynamics of Neotropical forest understorey bird flocks. *Journal of Animal Ecology*, **67**, 227–252. https://doi.org/10.1046/j.1365-2656.1998.00171.x

Kershenbaum, A., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A. *et al.* (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, **91**, 13–52. https://doi.org/10.1111/brv.12160

Kim, M., Chung, O.-S. & Lee, J.K. (2023). The relationship between nest location selection of barn swallows (*Hirundo rustica*) and human activity and residence. *Scientific Reports*, **13**, 21442. https://doi.org/10.1038/s41598-023-50149-6

Krebs, J.R. (1986). *Behavioral ecology: An evolutionary approach* (3rd edn). Oxford: Blackwell.

Laundré, J.W., Hernández, L. & Ripple, W.J. (2010). The landscape of fear: Ecological implications of being afraid. *The Open Ecology Journal*, **3**, 1–7. https://doi.org/10.2174/1874213001003030001

Leavesley, A.J. & Magrath, R.D. (2005). Communicating about danger: Urgency alarm calling in a bird. *Animal Behaviour*, **70**, 365–373. https://doi.org/10.1016/j.anbehav.2004.10.017

Lilly, M.V., Lucore, E.C. & Tarvin, K.A. (2019). Eavesdropping grey squirrels infer safety from bird chatter. *PLOS ONE*, **14**, e0221279. https://doi.org/10.1371/journal.pone.0221279

Lima, S.L. & Dill, L.M. (1990). Behavioral decisions made under the risk of predation: A review and prospectus. *Canadian Journal of Zoology*, **68**, 619–640. doi: 10.1139/z90-092.

Luttbeg, B., Ferrari, M.C., Blumstein, D.T. & Chivers, D.P. (2020). Safety cues can give prey more valuable information than danger cues. *The American Naturalist*, **195**, 636–648. https://doi.org/10.1086/707544

Madin, E.M., Madin, J.S. & Booth, D.J. (2011). Landscape of fear visible from space. *Scientific Reports*, **1**, 14. https://doi.org/10.1038/srep00014

Magrath, R.D., Haff, T.M., Fallow, P.M. & Radford, A.N. (2015). Eavesdropping on heterospecific alarm calls: From mechanisms to consequences. *Biological Reviews*, **90**, 560–586. https://doi.org/10.1111/brv.12122

Magrath, R.D., Pitcher, B.J. & Gardner, J.L. (2009). Recognition of other species' aerial alarm calls: Speaking the same language or learning another? *Proceedings of the Royal Society B: Biological Sciences*, **276**, 769–774. https://doi.org/10.1098/rspb.2008.1368

Mangini, G., Mokross, K., Gandoy, F. & Areta, J.I. (2022). Mixed-species flocking is associated with low arthropod detectability and increased foraging efficiency by Yungas forest birds in Argentina. *Ornithology*, **139**, ukab087. https://doi.org/10.1093/ornithology/ukab087

Mahon, L., Hoffman, B., James, L. S., Cusimano, M., Hagiwara, M., Woolley, S. C.,... & Pietquin, O. (2025) Robust detection of overlapping bioacoustic sound events. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems workshop: AI for non-human animal communication*.

Martínez, A.E., Parra, E., Collado, L.F. & Vredenburg, V.T. (2017). Deconstructing the landscape of fear in stable multi-species societies. *Ecology*, **98**, 2447–2455. https://doi.org/10.1002/ecy.1935

Martínez, A.E., Parra, E., Gomez, J.P. & Vredenburg, V.T. (2022). Shared predators between primate groups and mixed species bird flocks: The potential for forest-wide eavesdropping networks. *Oikos*, **2022**(10), e08274. doi: 10.1111/oik.08274

Martínez, A. E., Parra, E., Muellerklein, O., Vredenberg, V.T. (2018). Fear-based niche shifts in neotropical birds. *Ecology*, **99**, 1338-1346. https://doi.org/10.1002/ecy.2217

Martínez, A.E. & Zenil, R.T. (2012). Foraging guild influences dependence on heterospecific alarm calls in Amazonian bird flocks. *Behavioral Ecology*, **23**, 544–550. https://doi.org/10.1093/beheco/arr222

Martínez, A.E., Parra, E., Gil, M.A., Zill, J., Ponciano, J.M. & Vredenburg, V.T. (2023). The structure and organisation of an Amazonian bird community remains little changed after nearly four decades in Manu National Park. *Ecology Letters*, **26**, e14159. https://doi.org/10.1111/ele.14159

Meise, K., Franks, D.W. & Bro-Jørgensen, J. (2020). Alarm communication networks as a driver of community structure in African savannah herbivores. *Ecology Letters*, **23**, 293–304. https://doi.org/10.1111/ele.13432

Messori, A., Becagli, P. & Trippoli, S. (1997). Median versus mean lifetime survival in the analysis of survival data. *Haematologica*, **82**, 730.

Mokross, K., Ryder, T.B., Côrtes, M.C., Wolfe, J.D. & Stouffer, P.C. (2014). Decay of interspecific avian flock networks along a disturbance gradient in Amazonia. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20132599. https://doi.org/10.1098/rspb.2013.2599

Munn, C.A. (1985). Permanent canopy and understory flocks in Amazonia: Species composition and population density. *Ornithological Monographs*, **36**, 683–712. https://doi.org/10.2307/40168278

Munn, C.A. (1986). Birds that 'cry wolf'. *Nature*, **319**, 143–145. https://doi.org/10.1038/319143a0

Munn, C.A. & Terborgh, J.W. (1979). Multi-species territoriality in Neotropical foraging flocks. *The Condor*, **81**, 338–347. https://doi.org/10.2307/1366956

Owen-Smith, N. & Mills, M.G.L. (2008). Predator–prey size relationships in an African large-mammal food web. *Journal of Animal Ecology*, **77**, 173–183. doi: 10.1111/j.1365-2656.2007.01314.x.

Pohle, J., Langrock, R., van Beest, F.M. & Schmidt, N.M. (2017). Selecting the number of states in Hidden Markov models: Pragmatic solutions illustrated using animal movement. *Journal of Agricultural, Biological and Environmental Statistics*, **22**, 270–293. https://doi.org/10.1007/s13253-017-0283-8

Post, P. & Götmark, F. (1996). Prey selection by sparrowhawks, *Accipiter nisus*: Relative predation risk for breeding passerine birds in relation to their size, ecology and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **351**, 1559–1577. https://doi.org/10.1098/rstb.1996.0141

Pringle, R.M. (2008). Elephants as agents of habitat creation for small vertebrates at the patch scale. *Ecology*, **89**, 26–33. https://doi.org/10.1890/07-0776.1

Putland, R.L., Ranjard, L., Constantine, R. & Radford, C.A. (2018). A Hidden Markov model approach to indicate Bryde's whale presence from ocean glider passive acoustic data. *Ecological Indicators*, **84**, 479–487. https://doi.org/10.1016/j.ecolind.2017.09.025

Quental, J. (2014). Myrmoborus myotherinus ardesiacus – Black-faced Antbird (male) [photograph]. Wikimedia Commons. CC BY 2.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmoborus_myotherinus_ardesiacus_-_Black-faced_Antbird_%28male%29.jpg.

Rabiner, L.R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**, 257–286. https://doi.org/10.1109/5.18626

R Core Team (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rodríguez-Gironés, M.A. & Bosch, J. (2012). Effects of body size and sociality on the anti-predator behaviour of foraging bees. *Oikos*, **121**, 1473–1482. https://doi.org/10.1111/j.1600-0706.2011.19473.x

Rutt, C.L., Mokross, K., Kaller, M.D. & Stouffer, P.C. (2020). Experimental forest fragmentation alters Amazonian mixed-species flocks. *Biological Conservation*, **242**, 108415. https://doi.org/10.1016/j.biocon.2020.108415

Rutt, C. L. & Stouffer, P. C. (2021). Seasonal dynamics of flock interaction networks across a human-modified landscape in lowland Amazonian rain forest. Ecological Applications, 31, e02235. https://doi.org/10.1002/eap.2235

Rytkönen, S., Kuokkanen, P., Hukkanen, M. & Huhtala, K. (1998). Prey selection by Sparrowhawks *Accipiter nisus* and characteristics of vulnerable prey. *Ornis Fennica*, **75**, 77–87.

Schmitz, O.J., Hambäck, P.A. & Beckerman, A.P. (2000). Trophic cascades in terrestrial systems: A review of the effects of carnivore removals on plants. *The American Naturalist*, **155**, 141–153. https://doi.org/10.1086/303311

Schmitz, O.J., Hawlena, D. & Trussell, G.C. (2010). Predator control of ecosystem nutrient dynamics. *Ecology Letters*, **13**, 1199–1209. https://doi.org/10.1111/j.1461-0248.2010.01511.x

Selås, V. (1993). Selection of avian prey by breeding Sparrowhawks *Accipiter nisus* in southern Norway: The importance of size and foraging behaviour of prey. *Ornis Fennica*, **70**, 144–154.

Sherony, D. (2012). Tawny-crowned Greenlet (Hylophilus ochraceiceps) (7222935148) [photograph]. Wikimedia Commons. CC BY-SA 2.0. Available at: https://commons.wikimedia.org/wiki/File:Tawny-crowned_Greenlet_%28Hylophilus_ochraceiceps%29_%287222935148%29.jpg.

Smith, M.J. & Harper, D.G.C. (1995). Animal signals: Models and terminology. *Journal of Theoretical Biology*, **177**, 305–311. https://doi.org/10.1006/jtbi.1995.0248

Sridhar, H. & Shanker, K. (2014). Using intra-flock association patterns to understand why birds participate in mixed-species foraging flocks in terrestrial habitats. *Behavioral Ecology and Sociobiology*, **68**, 185–196. https://doi.org/10.1007/s00265-013-1633-3

Sridhar, H., Beauchamp, G. & Shanker, K. (2009). Why do birds participate in mixed-species foraging flocks? A large-scale synthesis. *Animal Behaviour*, **78**, 337–347. https://doi.org/10.1016/j.anbehav.2009.05.008

Sridhar, H., Srinivasan, U., Askins, R.A., Canales-Delgadillo, J.C., Chen, C.-C., Ewert, D.N. *et al.* (2012). Positive relationships between association strength and phenotypic similarity characterize the assembly of mixed-species bird flocks worldwide. *The American Naturalist*, **180**, 777–790. doi: 10.1086/668012.

Taper, Mark L., and Subhash R. Lele. (2004). Dynamical models as paths to evidence in ecology. *The nature of scientific evidence: Statistical, philosophical and empirical considerations*, 275-297.

Taper, Mark L., Subhash R. Lele, José M. Ponciano, Brian Dennis, and Christopher L. Jerde. (2021). Assessing the global and local uncertainty of scientific evidence in the presence of model misspecification. *Frontiers in Ecology and Evolution,* **9,** 679155.

Templeton, C.N., Greene, E. & Davis, K. (2005). Allometry of alarm calls: Black-capped chickadees encode information about predator size. *Science*, **308**, 1934–1937. https://doi.org/10.1126/science.1108841

Therneau, T.M. (2001). *A package for survival analysis in S* (version updates in R). https://CRAN.R-project.org/package=survival

Thiollay, J.-M. & Jullien, M. (1998). Flocking behaviour of foraging birds in a Neotropical rain forest and the Antipredator Defence Hypothesis. *Ibis*, **140**, 382–394. https://doi.org/10.1111/j.1474-919X.1998.tb04599.x

Timm, C.D. (2010). Xiphorhynchus elegans [photograph]. Wikimedia Commons. CC BY-SA 2.0. Available at: https://commons.wikimedia.org/wiki/File:Xiphorhynchus_elegans.jpg. Last accessed 5 September 2026.

Tobias, J.A., Brawn, J.D., Brumfield, R.T., Derryberry, E.P., Kirschel, A.N.G. & Seddon, N. (2012). The importance of suboscine birds as study systems in ecology and evolution. *Ornitología Neotropical*, **23**, 261–274.

Tobias, J.A., Sheard, C., Pigot, A.L., Devenish, A.J., Yang, J., Sayol, F. *et al.* (2022). AVONET: Morphological, ecological and geographical data for all birds. *Ecology Letters*, **25**, 581–597. https://doi.org/10.1111/ele.13898

Touchton, J.M., Seddon, N. & Tobias, J.A. (2014). Captive rearing experiments confirm song development without learning in a tracheophone suboscine bird. *PLOS ONE*, **9**, e95746. https://doi.org/10.1371/journal.pone.0095746

Tornberg, R. (1997). Prey selection of the goshawk *Accipiter gentilis* during the breeding season: The role of prey profitability and vulnerability. *Ornis Fennica*, **74**, 15–28.

Visser, I. & Speekenbrink, M. (2010). depmixS4: An R package for Hidden Markov Models. *Journal of Statistical Software*, **36**, 1–21. https://doi.org/10.18637/jss.v036.i07

Whitfield, D.P. (2003). Predation by Eurasian sparrowhawks produces density-dependent mortality of wintering redshanks. *Journal of Animal Ecology*, **72**, 27–35. https://doi.org/10.1046/j.1365-2656.2003.00672.x

Williams, S.M. & Lindell, C.A. (2021). Context-specific behavior serves as a mechanism of interspecific cohesion in mixed-species flocks. *Ornithology*, **138**, 1-10. https://doi.org/10.1093/ornithology/ukab046

Yan, K.M., Pinto, S.P., Vartany, C. & Blumstein, D.T. (2019). Shift down, look up: A test of the non-linearity and fear hypothesis in a non-vocal skink. *Ethology*, **125**, 153–158. https://doi.org/10.1111/eth.12839

CHAPTER 3
Agroforestry Degrades Multi-Species Risk-Communication in Amazonian Forests

Chapter Abstract

Animals use social information to manage the trade-offs between feeding and predation. In Neotropical forests, mixed-species bird flocks are a canonical example: flock members monitor one another's alarms, contact calls, and movements while collectively returning to routine foraging after danger has passed (Munn & Terborgh 1979; Goodale et al. 2010; Magrath et al. 2015). Recent work suggests that vocal exchanges encode not only alarm, but also an ordered acoustic transition from acute danger toward relative safety (Parra et al. in 2nd ; Camerlenghi et al. 2026). Whether that structure persists in human-used forests is unclear. Using standardized live presentations of a trained Bicolored Hawk (*Accipiter bicolor*), exhaustive annotation of audible post-predator calls, and hidden Markov models across a Brazil-nut harvest gradient in western Amazonia, we show that managed forests retain discrete post-predator states but differ from the reference forest in recovery-state characteristics: state occupancy, state duration, posterior entropy, and state-level call composition. Species repertoires shifted broadly, call output shifted selectively, and median timing did not. The clearest divergence occurred in later states associated with safety rather than initial alarm. Mixed-flock communication therefore appears to be a sensitive, mechanistic, non-invasive early indicator of ecosystem change in human-used tropical forests.

Introduction

Mixed-species flocks are among the clearest examples of interspecific information use in tropical forests. In the Neotropics, they are persistent social assemblages in which species differ in leadership, vigilance, foraging position, and responsiveness to shared cues (Munn & Terborgh 1979; Jullien & Thiollay 1998; Goodale & Beauchamp 2010; Goodale et al. 2020).

Communication about predators is central to these assemblages. Birds routinely eavesdrop on heterospecific alarm calls, and those cues can alter vigilance, movement, flock cohesion, and patch use (Goodale et al. 2010; Magrath et al. 2015). In Amazonian flocks, a subset of nuclear or sentinel species contributes disproportionately to the spread of risk information across strata and among flock members (Camerlenghi et al. 2019). Martínez et al. (2018) further showed that fear can reshape habitat use itself: when key sentinel species were removed, flock members shifted into denser, lower-risk microhabitats. Mixed flocks therefore organize not only who moves together, but how birds occupy space under risk.

Recent work suggests that communication in these flocks is richer than a simple alarm-versus-no-alarm distinction. In the southeastern Peruvian Amazon, standardized predator presentations showed that post-predator calling can be resolved into ordered hidden states, with different call types, taxa, and foraging traits characterizing early, intermediate, and late phases of the response (Parra et al. in revision). A complementary study of predator-information flow in the same system showed that information can propagate rapidly through the wider community and that small canopy species play a disproportionate role in that process (Camerlenghi et al. 2026). Taken together, these results imply that mixed flocks encode a graded and complex progression from danger to safety. Revealing that structure, however, requires unique methodological resolution: predator encounters must be standardized in the field, low-amplitude and non-song vocalizations must be coded across species, and the resulting time series must be analyzed with models that can infer latent state changes rather than simple counts or call rates alone.

There are several reasons to expect that mixed flocks will be especially sensitive to human disturbance, even where canopy forest remains and many constituent species persist (Mokross et al. 2014). Flocks depend on repeated associations among behaviorally compatible species, predictable vertical stratification, and reliable access to shared information. Logging,

fragmentation, edge creation, and compositional turnover can erode those properties before local extinction becomes obvious in inventories (Zou et al. 2018; Goodale et al. 2015). Empirical work supports that expectation. Habitat alteration and deforestation reduce flock richness and restructure flock composition in the Andes (Colorado Zuluaga & Rodewald 2015; Vásquez-Ávila et al. 2021). In Amazonia, flock networks decay across disturbance gradients and collapse under experimental fragmentation (Mokross et al. 2014; Rutt et al. 2020). Selective logging likewise alters interspecific associations in mixed flocks elsewhere in the tropics (Borah et al. 2018). These studies imply that flock organization may be a signal that identifies ecological degradation earlier, and more mechanistically, than coarse occurrence data alone.

This suggestion is consistent with a broader literature showing that human modification reorganizes interactions and acoustic communities before all constituent species disappear. Deforestation alters species interactions across ecological systems (Howes et al. 2023), and heavily deforested tropical forests lose both species and functions (Fuzessy et al. 2024). At broader acoustic scales, Amazonian soundscapes retain a signature of disturbances such as logging and fire (Rappaport et al. 2022). Soundscape analyses are powerful, but they generally integrate across many taxa and unspecified signal classes. Mixed-species risk communication provides a more mechanistic complement because it isolates a specific social process, identifies the species and call types involved, and identifies disruption of ecologically important traits–alarm and safety communication.

Brazil nut *Bertholletia excelsa* landscapes in Madre de Dios, Peru are a useful setting in which to study the communication consequences of forest exploitation (figure 3-1, A). Peru's Brazil nut concession system was formalized in 2000 and now spans over 1123 concessions of an average of 860 hectares, totaling about one million hectares of closed-canopy forest: roughly 10.5% of the Madre De Dios region (Nunes et al. 2012; Willem et al. 2019). These concessions are often framed as a comparatively sustainable form of non-timber forest use, yet they vary in

access, extraction history, associated timber use, and degradation (Rockwell et al. 2015; Ramírez & Belcher 2020; Jansen et al. 2021) (figure 3-1, B). Brazil nut forests retain sufficient canopy structure and avifauna to support natural bird communities including mixed flocks, but they still can be characterized along a gradient of human use (figure 3-1, B). They therefore provide an ideal system to test whether the degradation of interspecific communication is sensitive to lighter, policy-relevant forms of forest management.

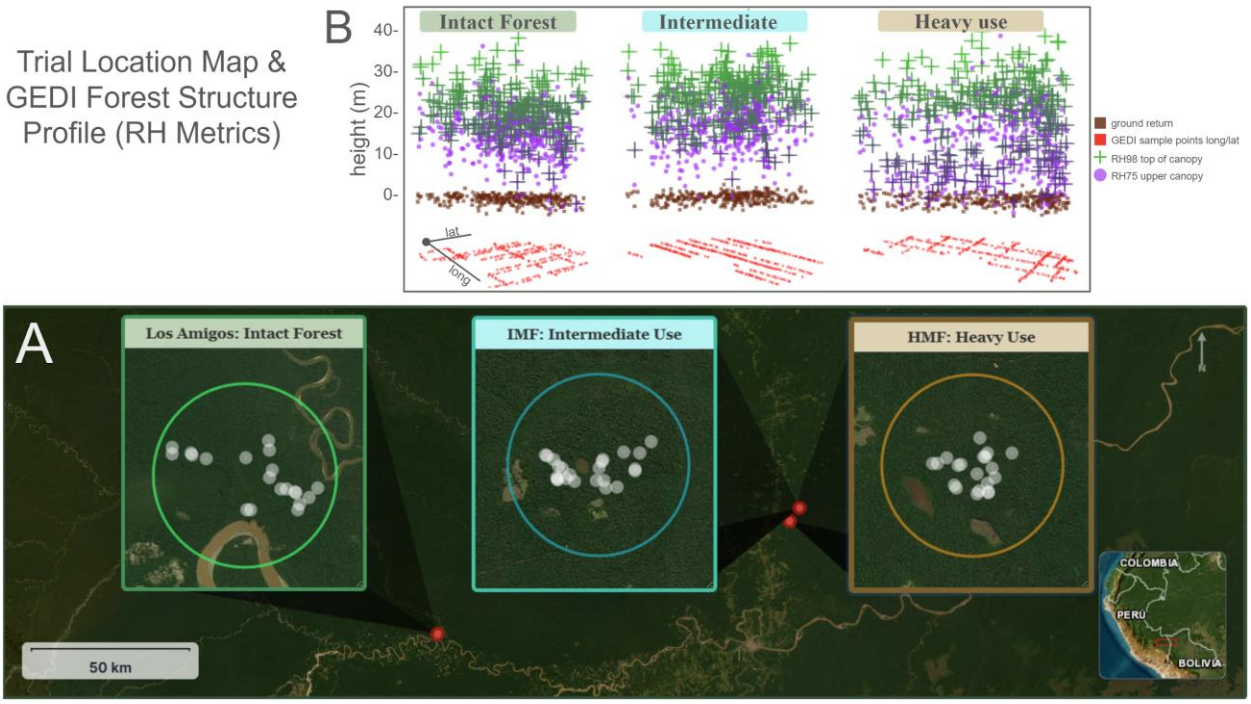


Figure 3-1. Trial locations and forest-structure context

A. Map of Trial locations. Red points indicate trial locations along the Madre de Dios river in SE Peru. White points on inset maps are locations of individual trials. Colored circles indicate sample region of GEDI points used to extract forest profile metrics. B. GEDI forest profile metrics. Green crosses indicate canopy point estimates. Purple circles indicate upper 75% of forest biomass. Brown squares indicate ground points. Red points indicate sampling locations across angled latitude and longitude axis. Note: space-born LiDAR points are not always evenly distributed.

Here we test this framework across an important agroforestry gradient. We presented a trained Bicolored Hawk to mixed-species flocks, after which all audible calls were manually coded, including alarms, quiet contact notes, mobbing calls, and other non-song vocalizations (Parra et al. in revision). We then used hidden Markov models to infer post-predator states and compared those states among sites with respect to proportional occupancy, temporal duration,

posterior entropy, and state-level call composition. To separate state-level disruption from species-level changes, we also tested whether focal species differed among sites in call-type repertoires, total vocal output, and median call timing. To our knowledge, no previous field study has combined live raptor presentations, exhaustive cross-species annotation, and state-based time-series modelling to test how interspecific risk communication changes across a gradient of tropical forest use. We expected managed forests to retain recognizable response states but to show longer early alarm phases, higher entropy, later safe-foraging states, and reduced fidelity in the transition toward safety (Mokross et al. 2014; Rutt et al. 2020; Rappaport et al. 2022).

Results

Overall, managed forests retained an ordered post-predator response sequence, but the states differed from Los Amigos reference forest in how much they were used, how sharply they were defined, how long they lasted, and which calls composed them (figure 3-2, table 3-1). In managed forests, early alarm-associated states tended to occupy more time or persist longer, whereas later recovery-associated states were broader, less cleanly separated from adjacent states, and compositionally different from the reference forest (figure 3-2, B).

Table 3-1. Summary of cross-site tests of post-predator vocal communication.
Los Amigos was treated as the reference forest and compared separately with intermediate-use managed forest (IMF) and highly disturbed managed forest (HMF). P values were adjusted within each analysis family using the Benjamini-Hochberg false discovery rate procedure. BC, Bray-Curtis dissimilarity; HARU, Habia rubica; MYAX, Myrmotherula axillaris; MYLO, Myrmotherula longipennis; MYMY, Myrmoborus myotherinus; MYSC, Myrmotherula sclateri; THAR, Thamnomanes ardesiacus.

Focal Analysis	Test statistic	Valid tests	FDR-significant	Effect summary	Interpretation
State call type composition	Bray-Curtis distance between site mean call-composition vectors within matched HMM states	8	8	BC = 0.449-0.575; strongest state 4, IMF - intact: BC = 0.575, adjusted P < 0.001	All states differed compositionally between intact and managed forests.
Within species call repertoire	Monte Carlo chi-square homogeneity tests on species-by-site call-type tables	14 valid; 8 skipped	14	Cramer's V = 0.061-0.629; strongest MYLO HMF - intact, MYAX HMF - intact, MYMY IMF - intact	Every tested species changed its call-type distribution.

Focal Analysis	Test statistic	Valid tests	FDR-significant	Effect summary	Interpretation
Species total call output	Permutation tests of mean calls per trial, including zero-call trials	22	4	HARU higher in IMF; MYSC lower in both managed forests; THAR lower in HMF	Output shifts were taxon-specific, not universal.
Species median call timing	Permutation tests of mean trial-level median call time among calling trials	15 valid; 7 skipped	0	Largest raw trend: MYLO HMF - intact = -125.41 s, adjusted P = 0.475	No species-level median timing differences.

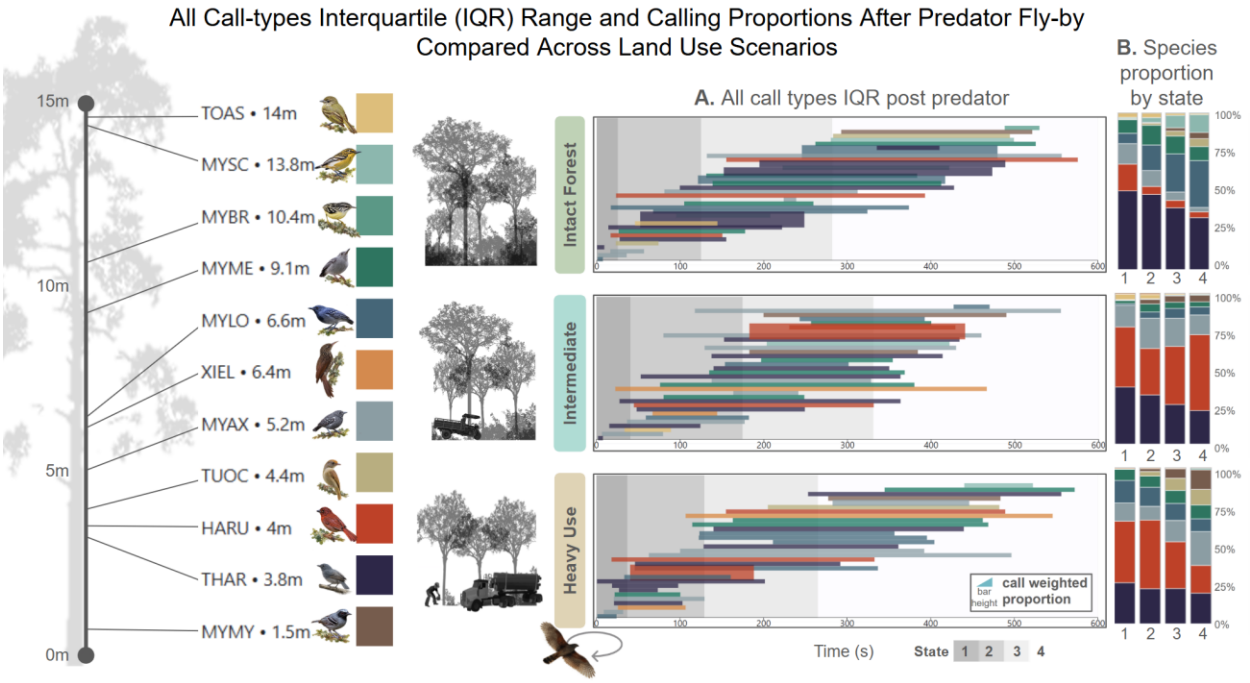


Figure 3-2. Cross-site call timing and state composition

(A) Interquartile range of call type production over time at 3 sites. (B) Composition metrics are established within states defined by mean HMM Viterbi assignments. Reported estimated foraging heights are based on observed heights across terra firma sites in the department of Madre De Dios, Peru (see: Supplemental figure 3-S1). Bird illustrations were digitally adapted from photographs by Hector Bottai (Tolmomyias assimilis, 2014b; Myrmotherula sclateri, 2023; Myrmotherula menetriesii, 2015b; Myrmotherula longipennis, 2015a; Myrmotherula axillaris, 2022; Myrmotherula brachyura, 2014a), Dominic Sherony (Tunchiornis ochraceiceps, photographed under the name Hylophilus ochraceiceps, 2012), Instituto Últimos Refúgios (Habia rubica, 2012), João Quental (Myrmoborus myotherinus, 2014) and Cláudio Dias Timm (Xiphorhynchus elegans, 2010), via Wikimedia Commons. Logging-truck and forestry-worker illustrations adapted from Tracey Saxby (2012) and Tracey Saxby (2013). The illustrations are licensed under CC BY-SA 4.0 (Bottai and Instituto Últimos Refúgios), CC BY-SA 2.0 (Sherony and Timm) and CC BY 2.0 (Quental).

Relative to the control site, overall call activity was higher for HARU at the intermediate-impact site and lower for THAR and MYSC at the high-impact site (Figure 3-3A). Relative to the

control site, highly disturbed managed forest showed more call activity in State 1 (difference = +0.091, z = 2.80, adjusted P = 0.041), whereas intermediate-use managed forest showed reduced call activity in State 4 (difference = −0.093, z = −2.83, adjusted P = 0.041) (Figure 3-3B). Posterior entropy was higher in highly disturbed managed forest for State 2 (difference = +0.020, t = 7.01, adjusted P < 0.001) and higher in States 3 and 4 for both managed forests (State 3: intermediate-use, +0.011, adjusted P = 0.022; highly disturbed, +0.013, adjusted P = 0.005; State 4: intermediate-use, +0.008, adjusted P < 0.001; highly disturbed, +0.005, adjusted P = 0.022) (Figure 3-4D). Higher posterior entropy indicates that calls assigned to a state were less confidently separated from other states. Thus, the later risk-to-safety phases in managed forests were not simply occupied differently; they were also less sharply defined than at the control site. State duration showed the same pattern of disruption in the risk-to-safety sequence. The alarm-calling stage (State 1) was longer in both intermediate-use managed forest (difference = +196.950 s, adjusted P = 0.012) and highly disturbed managed forest (+27.660 s, adjusted P = 0.012). A transitional risk stage (State 3) was compressed in highly disturbed managed forest (−42.500 s, adjusted P = 0.010), whereas the later recovery stage (State 4) was longer in both managed forests (intermediate-use: +34.440 s, adjusted P = 0.011; highly disturbed: +36.760 s, adjusted P = 0.022) (Figure 3-4A).

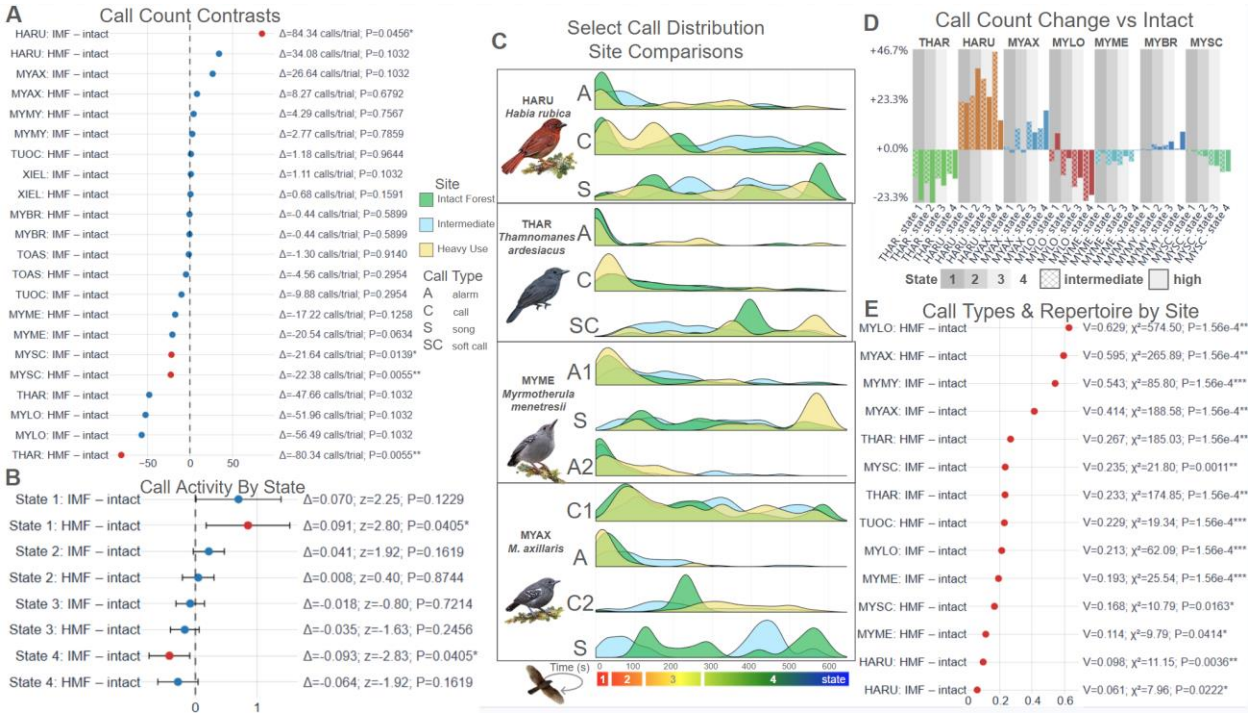


Figure 3-3. Species- and state-level call-count contrasts

(A) Species-level call-count contrasts between managed and reference forest. (B) State-level call-activity contrasts between managed and reference forest. (C) Site-specific call-type distributions for selected species. (D) Percent change in call counts by species, call type, and state relative to intact forest. (E) Species-level call-repertoire contrasts between managed and intact sites. Statistically significant contrasts after correction are indicated in red. Bird illustrations were digitally adapted from photographs by Hector Bottai (Myrmotherula menetriesii, 2015b; Myrmotherula axillaris, 2022), Instituto Últimos Refúgios (Habia rubica, 2012), João Quental (Myrmoborus myotherinus, 2014). The adapted illustrations are licensed under CC BY-SA 4.0 (Bottai and Instituto Últimos Refúgios), CC BY-SA 2.0 (Sherony and Timm) and CC BY 2.0 (Quental).

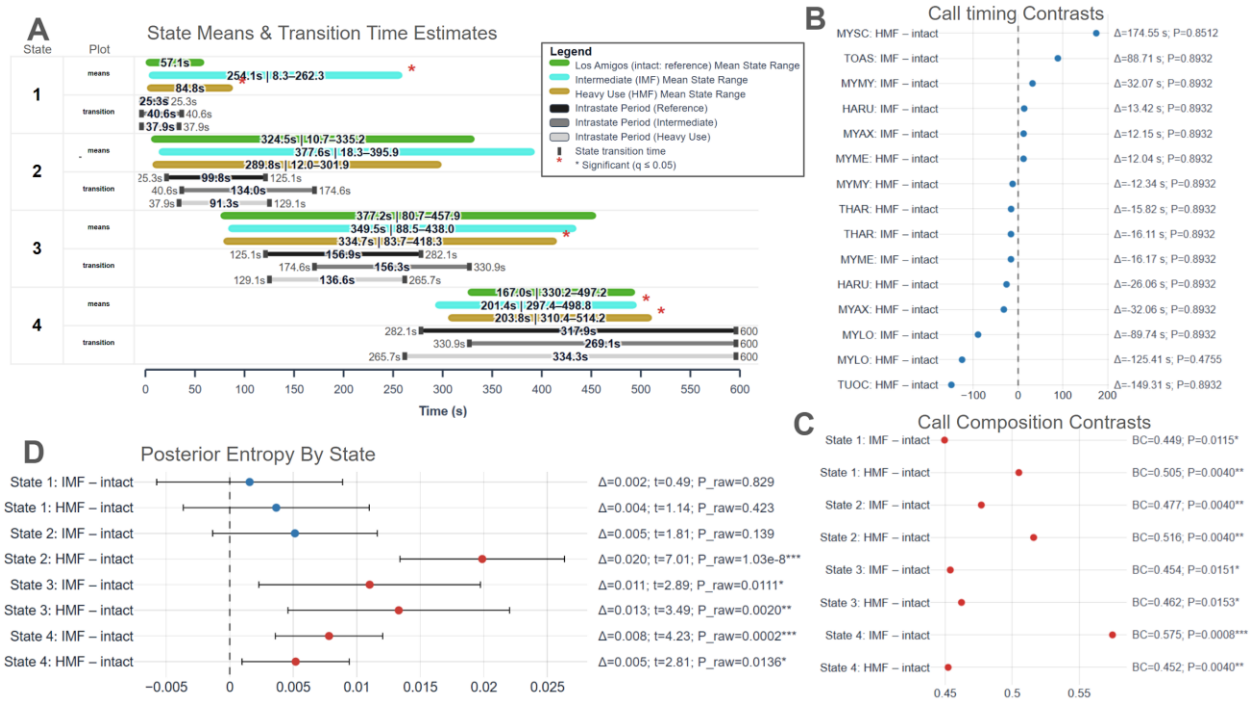


Figure 3-4. Cross-site hidden-state summary metrics

(A) All models state mean-time ranges separated by site (colored bars). Transition time estimates represent the time at which the total accumulated posterior probability of previous and proximate states are equivalent (grey-scale lines). (B) Call-timing of individual call types at managed sites compared to intact forest. (C) call-composition of each state compared to intact forest. (D) Model posterior entropy for each state compared to intact forest. All points significant after correction in red.

Call composition was the most consistent cross-site difference. State-level call composition differed between the control site and both managed forests in all eight state-by-site comparisons (Bray-Curtis dissimilarity = 0.449-0.575; all FDR-adjusted P < 0.05) (Figure 3-4, C). The largest state-level contrast occurred in state 4 between intermediate-use managed forest and the control site (Bray-Curtis = 0.575, adjusted P < 0.001), consistent with other changes occurring in later recovery-associated states. Species-level call-type repertoires showed the same broad pattern: all 14 valid species-by-site comparisons differed after FDR correction (Cramer's V = 0.061-0.629) (figure 3-3, E). The strongest repertoire shifts were in *Myrmotherula longipennis* between highly disturbed managed forest and Los Amigos (Cramer's V = 0.629), *Myrmotherula axillaris* between highly disturbed managed forest and Los Amigos (V = 0.595), and *Myrmoborus myotherinus* between intermediate-use managed forest and Los

Amigos (V = 0.543). Eight species-by-site comparisons could not be tested because one site had no calls or a single-call-type repertoire.

Species-level total call output shifted more selectively. *Habia rubica* called more in intermediate-use managed forest than in Los Amigos (98.84 versus 14.50 calls per trial; difference = +84.34; adjusted P = 0.046) (figure 3-3, A). There were fewer *Myrmotherula sclateri* calls in both managed forests than in Los Amigos (intermediate-use: 1.42 versus 23.06 calls per trial; difference = -21.64; adjusted P = 0.014; highly disturbed: 0.68 versus 23.06; difference = -22.38; adjusted P = 0.005). *Thamnomanes ardesiacus* also had fewer recorded calls in highly disturbed managed forest than in Los Amigos (37.16 versus 117.50 calls per trial; difference = -80.34; adjusted P = 0.005) (figure 3-3, D). In contrast, species-level median call timing did not differ after FDR correction in any valid comparison (15 valid tests, seven skipped for insufficient calling trials) (figure 3-4, B). The largest uncorrected timing trend was earlier *Myrmotherula longipennis* calling in highly disturbed managed forest than in Los Amigos (difference = -125.41 s), but this did not survive correction (adjusted P = 0.475).

All sites were compared using the same HMM model. Multiple criteria were used to evaluate which model parameters made for optimal interpretability. Taking site variability and biological context into consideration, site-specific ICL scores supported an ordered four-state model (k=4), a 10-s early-weighting window, and bin sizes approaching 20 s. These parameters permitted a direct state-matched comparisons across all sites.

Discussion

These results indicate that mixed flocks are even more sensitive indicators of ecological change than previously suggested (Zou et al. 2018), even in forests that remain structurally intact and still contain flock species and a largely intact bird community. Managed forests retained discrete states, yet those states were occupied differently, became less sharply

defined, and differed consistently in call composition. One of the strongest signals of degradation in our dataset was the reduced fidelity in the progression from early alarm toward later recovery. That pattern extends previous work showing that disturbance alters flock richness, composition, pairwise associations, and network structure (Colorado Zuluaga & Rodewald 2015; Mokross et al. 2014; Borah et al. 2018; Rutt et al. 2020). Here the disrupted unit is the temporal reliability of the information exchange that typically follows a predator encounter.

The Brazil-nut concession landscape is especially informative because it tests sensitivity under comparatively low-impact, management-relevant forest use. Forest cover persists, mixed flocks remain, and these concessions are often treated as compatible with biodiversity conservation (Willem et al. 2019; Rockwell et al. 2015; Jansen et al. 2021). Nevertheless, a vital risk- and safety-communication network was modified in significant and potentially consequential ways. This interpretation is consistent with Martínez et al. (2018), who showed that changes in perceived safety can reorganize flock habitat use, and with Rappaport et al. (2022), who demonstrated that acoustic communities can index forest degradation not evident from canopy cover alone. Our approach provides a finer, mechanistic level of resolution. Because the stimulus was standardized, the vocal repertoires were exhaustively coded, and state structure was estimated rather than assumed, the resulting signals are behaviorally interpretable.

The species-level results indicate that disruption was not limited to a change in total calling rate. Whenever a species had enough calls and call types to test its repertoire, the distribution of call types differed between Los Amigos and the managed forests. This suggests that managed sites changed not only how much some species called, but also which calls they contributed to the post-predator exchange. The strongest repertoire shifts occurred in *M. longipennis*, *M. axillaris*, and *M. myotherinus*, whereas the strongest total-output shifts involved

H. rubica, *M. sclateri*, and *T. ardesiacus*. These taxa span different positions in mixed-flock communication, suggesting that degradation is expressed across multiple contributors rather than through the disappearance of a single acoustic role.

The practical implication is that mixed-flock communication may be useful as a monitoring tool for mechanisms of change. Broad soundscape metrics remain valuable because they scale well, but they do not identify which interspecific pathways fail first (Rappaport et al. 2022). By contrast, the present framework identifies when degradation appears in the sequence from alarm to recovery, which states change most, and which species drive those changes. The absence of corrected species-level median-timing differences, together with strong state-composition and selective call-output effects, suggests that communication degrades through changes in which species contribute and which calls they use, rather than through a wholesale shift in when each species calls. This distinction is important because state duration and state ambiguity did change: the temporal organization of the response sequence was altered, but individual species did not show consistent corrected shifts in their median call times. Future studies integrating vegetation preferences and microhabitat data with interspecific call activity may allow considerable gain in understanding the impacts of small-scale forest change.

In following the development in the study of the interspecific communication ecology in mixed flock systems across sustained, small changes in habitat structure, we have shown that graded risk-to-safety communication is sensitive enough to detect changes in relationships that occur before the loss of species within a system. The implications of this study place interspecific communication frameworks as a viable candidate in the pursuit of advancing future remote censusing and bioacoustic monitoring tools, connecting ecosystem health management directly with the behavior of organisms whom themselves experience the direct consequences of degraded forest systems.

Methods

Study system and experimental design

Fieldwork was conducted at three lowland terra firme rainforest sites in Madre de Dios, Peru, spanning a gradient of human use in Brazil nut landscapes. Los Amigos served as the reference forest (12°33'22.9"S, 70°06'22.8"W: mean canopy height [GEDI rh98] 27.0 m, height IQR 23.39–30.46 m). Two additional sites, North PM 1 (12°12'03.2"S, 69°03'12.0"W: mean canopy height [GEDI rh98] 29.5 m, height IQR 24.81–33.74 m) and North PM 2 (12°14'18.2"S, 69°04'52.8"W: mean canopy height [GEDI rh98] 25.3 m, height IQR 17.00–32.04 m), represented more intensively worked forest. Focal flocks were located visually during active foraging and challenged with a standardized predator presentation based on the Los Amigos protocol (Parra et al. in revision). A trained Bicolored Hawk (*Accipiter bicolor*), handled by a licensed falconer, was flown from concealment to a perch near the focal flock and then retrieved immediately, using a previously established methodology (Martinez et al. 2017). Trials were conducted between 07:00 and 14:00 h, avoiding dawn chorus and obvious non-foraging periods. Flocks were followed until conditions were suitable and allowed at least 10 min to settle before presentation. Each trial was recorded for 600 s after the fly-by. To minimize observer effects, the flock observer, recordist, and the falconer maintained distance of 30 meters between each other when detecting flocks, with the falconer at the furthest distance. When a flock was found, researchers would advance to 15-20 meters from the focal flock, and the falconer remained ca 90 meters away until signaled to advance and present the hawk.

Acoustic recording and annotation

Recordings were made with a Zoom H5 recorder using a Zoom XYH-5 X/Y condenser pair and a Sennheiser ME66/K6 microphone. Audio was recorded at 44.1 kHz and 16-bit resolution. Vocal events were annotated manually using the Los Amigos coding framework and reduced call dictionary (http://www.neotropicalscience.com/atac.html). We deliberately retained

the full post-predator repertoire rather than only canonical alarms: soft contact notes, mobbing calls, quiet calls, and other non-song vocalizations were coded where possible, because the aim was to recover the full progression from danger toward routine flocking rather than alarm alone. Analyses were restricted to post-presentation events within 0-600 s. Explicit silence codes were excluded, no empty time bins were inserted, and rare call types with fewer than five occurrences across the filtered dataset were removed before modelling to avoid unstable state estimates. Following these guidelines, initial datasets included 7,412 post-predator presentation calls from intact forests, 5,007 calls from intermediate use managed forest, and 3,405 calls from heavy-use managed forest for a total of n = 15,824 initial hand-annotated time-stamped vocalizations.

Hidden Markov model screening and state estimation

Hidden Markov models (HMMs) were fitted separately for each site with depmixS4 (Rabiner 1989; Visser & Speekenbrink 2010; McClintock et al. 2020; Glennie et al. 2023). Each trial was treated as an independent sequence through ntimes, so transitions were estimated within trials only and never across trial boundaries. The observed data for each event consisted of two categorical channels: reduced call identity and a discretized post-stimulus time bin. Transitions were specified with an intercept-only structure so that hidden states reflected latent phases of site-level calling rather than covariate-conditioned transition regimes. During screening, candidate models varied in time-bin width (10, 20, and 30 s), number of hidden states (2-6), and early-window weighting (0, 10, 20, and 30 s). Early weighting was implemented by oversampling events within the selected window, which increased the influence of brief alarm-rich intervals without inserting synthetic silent observations. Each candidate specification was fitted from multiple random starts and ranked by the integrated completed likelihood normalized by the number of original events, which penalizes both overparameterization and diffuse state assignment (Biernacki et al. 2000; Schwarz 1978). After screening, the selected site models were refitted repeatedly from random starts, posterior probabilities were collapsed

back to the original unreplicated events, and states within each fit were re-labelled by posterior-weighted mean event times so that lower-numbered states always denoted earlier phases of the response. The matched four-state solution recovered in all sites was used for cross-site comparison.

Cross-site HMM summary metrics

Site comparisons were based on trial-level summaries derived from the repeated-fit HMM outputs rather than on any single decoded sequence. Three primary state metrics were analyzed. Occupancy was the proportion of total trial state-time assigned to a given state. Entropy was the posterior-weighted mean uncertainty of event assignment within a state and therefore measured how sharply defined that state was. For event e, posterior assignment uncertainty was calculated as:

$$H_e = -\sum_{s=1}^{K} \gamma_{es} log(\gamma_{es})$$

where K is the number of hidden states and γ_{es} is the posterior probability that event e belonged to state s. State-specific entropy was then the posterior-weighted mean event uncertainty:

$$\underline{H_s} = \frac{\sum_{e=1}^{E} \gamma_{es} H_e}{\sum_{e=1}^{E} \gamma_{es}}$$

State-time spread was calculated across repeated fitted models as:

$$Spread_s = \max_f(\overline{t}_{sf}) - \min_f(\overline{t}_{sf})$$

where the over barred *t* term is the posterior-weighted mean event time for states in repeated fit f.

State-time spread was calculated from the repeated-fit state summaries as the range of mean event times assigned to a state. For occupancy, trial-level means were analyzed with fractional-response models using Los Amigos as the reference site and Dunnett-adjusted contrasts for IMF -LA and HMF-LA (Papke & Wooldridge 1996; Dunnett 1955). Entropy was analyzed analogously with linear models and the same planned contrasts. Spread was

evaluated with permutation tests based on 4,999 label shuffles. False discovery rate was controlled within metric family with the Benjamini-Hochberg procedure (Benjamini & Hochberg 1995).

State composition and species-level permutation tests

Additional analyses were conducted on the original filtered events after projecting them into site-specific state windows defined by the saved transition summaries. Los Amigos was treated as the reference forest and compared separately with intermediate-use managed forest and highly-disturbed managed forest. For each trial and state, call composition was summarized as a proportional vector across reduced call types. For each matched state, we calculated the mean call-composition vector for each site and quantified between-site dissimilarity with Bray-Curtis distance:

$$BC_{AB} = \frac{\sum_{k=1}^{C} |p_{Ak} - p_{Bk}|}{\sum_{k=1}^{C} (p_{Ak} + p_{Bk})}$$

where pAk and pBk are the mean proportional contributions of call type k in sites A and B, and C is the number of reduced call types in the comparison. Because these were proportional composition vectors, the associated L1 distance was:

$$L1_{AB} = \sum_{k=1}^{C} |p_{Ak} - p_{Bk}| = 2BC_{AB}$$

Significance was assessed with 10,000 trial-label permutations that preserved the observed number of trials per site.

Species-level analyses used species identities derived from the first four letters of each reduced call code. We first tested whether each species' call-type repertoire differed between Los Amigos and each managed site. For each species-by-site comparison, we constructed a 2 x C contingency table of call counts, where C was the number of reduced call types observed for that species in the comparison. Call-type distributions were compared with chi-square tests of

homogeneity using Monte Carlo *P* values based on 10,000 simulations. Cramér's *V* was reported as an effect size:

$$V = \sqrt{\frac{\chi^2}{Nmin(r-1, c-1)}}$$

where N is the total number of calls and r and c are the numbers of rows and columns in the contingency table. Repertoire tests were skipped when either site had zero calls or when fewer than two nonzero call types were available.

We then tested two trial-level species summaries. Total call output was the number of calls by a species in each trial, including zero-call trials; site differences were assessed with 10,000 trial-label permutations of the difference in mean calls per trial. Species-level call timing was calculated only for trials in which the species called. For each such trial, we calculated the median post-presentation call time and then tested the difference between sites in the mean of those trial-level medians using the same 10,000-permutation framework. Timing tests were skipped when either site had fewer than two calling trials. *P* values were adjusted within each analysis family using the Benjamini-Hochberg false discovery rate procedure: state-level call composition, species call-type repertoire, species total call output, and species median timing.

Focal species point-count

Point-count census was used to confirm the presence of flocking species at both sites. Since IMF and HMF are neighboring concessions, a single transect, equivalent in census points to the intact site, was used for comparison (figure 3-1,A) (Los Amigos n = 30, IMF n = 33). Repeated visits to a point were collapsed to presence if a species was detected at least once and absence otherwise. We then calculated naive occupancy as occupied points divided by surveyed points for each species and site, compared Los Amigos and IMF with two-sided Fisher exact tests, and adjusted the resulting P-values with the Benjamini-Hochberg procedure. Because detectability was not estimated explicitly, these summaries are interpreted as

detection-unadjusted occupancy indices (MacKenzie et al. 2002; Joseph et al. 2006; Hayes & Monfils 2015). We report occupied points, total points, naive occupancy, odds ratios, raw p-values, and Benjamini–Hochberg adjusted p-values (supplemental table 3-S1). This comparison follows standard point-count occupancy practice while acknowledging assumptions about closure and detectability (Hayes & Monfils 2015).

Naive occupancy was calculated as:

$$\psi_i = \frac{x_i}{n_i}$$

where ψ_i is naive occupancy for species i, x_i is the number of points at which species i was detected at least once, and n_i is the number of surveyed points at that site.

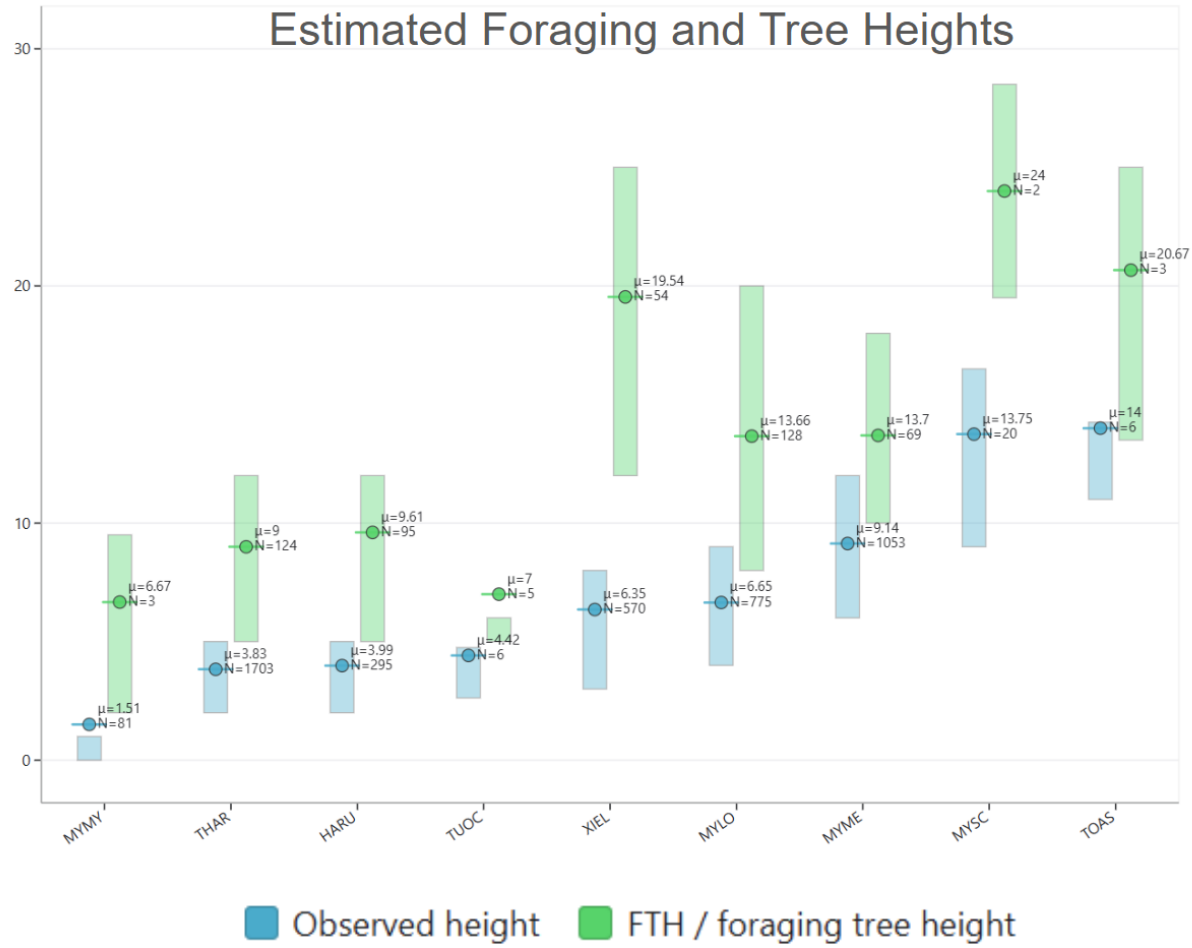
Species trait data and forest structure assessments

Species body mass was taken from AVONET (Tobias et al. 2022). Foraging height, foraging tree height (FTH), cover density, perch diameter estimates, and distance to protective cover were drawn from opportunistic field observations collected during the pre-trial acclimation period and during previous work (specifically foraging height estimates) in the Madre De Dios Region (see: supplemental figure 3-S1). These values are used as explanatory and not in model assessments or statistical work.

GEDI space-borne LiDAR products were used to contextualize site differences in habitat condition. Files were accessed through Google Earth Engine using rgee and reticulate, and monthly Level 2A, Level 2B, and Level 4A products were summarized from 1 January 2020 through 30 November 2024 after standard quality filtering (Gorelick et al. 2017; Dubayah et al. 2020, 2021a, 2021b, 2022; Duncanson et al. 2022). After filtering, the dataset contained 1,259 GEDI height samples across the three site buffers in UTM Zone 19S: 482 samples at Los Amigos, 416 at North PM 1, and 361 at North PM 2 (supplemental table 3-S2). Site-level height summaries were calculated from rh98_m, with the mean and interquartile range reported in supplemental table 3-S2; rh75_m measurements were retained as supplementary relative-

height metrics (figure 3-1, B). Sample area was chosen to allow for more points in low-sample density areas because GEDI footprints occur non-uniformly across the landscape and must be averaged.

Supplementary Materials



Supplemental Figure 3-S1. Focal-species foraging-height distributions

Interquartile values of estimated foraging heights and foraging tree heights (FTH) of focal mixed-flock species. Mean value and number of observations (n) are reported at mean centers.

Supplemental Table 3-S1. Point-count comparison between forest classes.

Species four letter codes and corresponding species names. Point count census comparing intact species detections (intact) with detections in managed forests (concession). Significant differences are indicated by a star *.

Code	Species	Intact	Concession	Difference	P	BH-adjusted P
THAR	*Thamnomanes ardesiacus**	0.700	0.273	0.427	0.001	0.016*
THSY	*Thamnomanes schistogynus*	0.133	0.030	0.103	0.183	0.616
MYAX	*Myrmotherula axillaris*	0.433	0.333	0.100	0.447	0.802

Code	Species	Intact	Concession	Difference	P	BH-adjusted P
MYME	*Myrmotherula menetriesii*	0.433	0.212	0.221	0.103	0.456
MYLO	*Myrmotherula longipennis*	0.400	0.242	0.158	0.278	0.689
MYSC	*Myrmotherula sclateri*	0.467	0.121	0.345	0.004	0.064
MYBR	*Myrmotherula brachyura*	0.133	0.182	-0.048	0.735	0.973
HARU	*Habia rubica*	0.333	0.515	-0.182	0.203	0.616
TUOC	*Tunchiornis ochraceiceps*	0.067	0.030	0.036	0.601	0.883
LEAM	*Leptopogon amaurocephalus*	0.200	0.000	0.200	0.009	0.093
ISHX	*Isleria hauxwelli*	0.067	0.061	0.006	1.000	1.000
MYMY	*Myrmoborus myotherinus*	0.867	0.788	0.079	0.515	0.818
TOAS	*Tolmomyias assimilis*	0.200	0.121	0.079	0.498	0.803
AUIN	*Automolus infuscatus*	0.233	0.121	0.112	0.325	0.760
DELO	*Deconychura longicauda*	0.333	0.182	0.152	0.247	0.650
XIEL	*Xiphorhynchus elegans*	0.767	0.788	-0.021	1.000	1.000
XIGU	*Xiphorhynchus guttatus*	0.733	0.818	-0.085	0.547	0.850

Supplemental Table 3-S2. GEDI lidar vegetation-metric summaries.

Note. IQR is the 25th–75th percentile range. Height estimates are derived from the rh98 canopy measurements provided by NASA's GEDI spaceborne lidar samples.

Site	n samples	Mean height (m)	Height IQR (m)
Los Amigos (intact)	482	27.00	23.39–30.46
Intermediate managed forest (IMF)	416	29.54	24.81–33.74
Heavy-use managed forest (HMF)	361	25.28	17.00–32.04

References

Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57, 289-300 (1995). https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Biernacki, C., Celeux, G. & Govaert, G. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Trans. Pattern Anal. Mach. Intell. 22, 719-725 (2000). https://doi.org/10.1109/34.865189

Borah, B., Quader, S. & Srinivasan, U. Responses of interspecific associations in mixed-species bird flocks to selective logging. J. Appl. Ecol. 55, 1637-1646 (2018). https://doi.org/10.1111/1365-2664.13097

Bottai, H. (2014a). Myrmotherula brachyura – Pigmy antwren [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_brachyura_-_Pigmy_antwren.JPG.

Bottai, H. (2014b). Tolmomyias assimilis – Zimmer's flatbill [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Tolmomyias_assimilis_-_Zimmer%27s_flatbill.JPG.

Bottai, H. (2015a). Myrmotherula longipennis – Long-winged antwren (male) [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_longipennis_-_Long-winged_antwren_%28male%29.jpg.

Bottai, H. (2015b). Myrmotherula menetriesii – Grey antwren (male) (cropped) [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_menetriesii_-_Grey_antwren_%28male%29_%28cropped%29.jpg.

Bottai, H. (2022). Myrmotherula axillaris White-flanked Antwren (male); Ramal do Noca, Rio Branco, Acre, Brazil [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_axillaris_White-flanked_Antwren_%28male%29;_Ramal_do_Noca,_Rio_Branco,_Acre_,_Brazil.jpg.

Bottai, H. (2023). Myrmotherula sclateri Sclater's Antwren; Porto Velho, Rondônia, Brazil [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmotherula_sclateri_Sclater%27s_Antwren;_Porto_Velho,_Rond%C3%B4nia,_Brazil.jpg.

Camerlenghi, E., Tellaroli, P., Griggio, M. & Martínez, A. E. Information about predators varies across an Amazonian rain forest as a result of sentinel species distribution. Am. Nat. 194, E134-E139 (2019). https://doi.org/10.1086/705242

Camerlenghi, E. et al. Small canopy species drive information highways about predators in an Amazonian rainforest. Curr. Biol. 36, R318-R319 (2026). https://doi.org/10.1016/j.cub.2026.02.006

Colorado Zuluaga, G. J. & Rodewald, A. D. Response of mixed-species flocks to habitat alteration and deforestation in the Andes. Biol. Conserv. 188, 72-81 (2015). https://doi.org/10.1016/j.biocon.2015.02.008

Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B 39, 1-22 (1977). https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Dubayah, R. et al. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. Sci. Remote Sens. 1, 100002 (2020). https://doi.org/10.1016/j.srs.2020.100002

Dubayah, R., Hofton, M., Blair, J. B., Armston, J., Tang, H. & Luthcke, S. GEDI L2A elevation and height metrics data global footprint level V002. NASA EOSDIS Land Processes DAAC (2021). https://doi.org/10.5067/GEDI/GEDI02_A.002

Dubayah, R., Tang, H., Armston, J., Luthcke, S., Hofton, M. & Blair, J. B. GEDI L2B canopy cover and vertical profile metrics data global footprint level V002. NASA EOSDIS Land Processes DAAC (2021). https://doi.org/10.5067/GEDI/GEDI02_B.002

Dubayah, R. O. et al. GEDI L4A footprint level aboveground biomass density, version 2.1. ORNL DAAC (2022). https://doi.org/10.3334/ORNLDAAC/2056

Duncanson, L. et al. Aboveground biomass density models for NASA's Global Ecosystem Dynamics Investigation lidar mission. Remote Sens. Environ. 270, 112845 (2022). https://doi.org/10.1016/j.rse.2021.112845

Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. J. Am. Stat. Assoc. 50, 1096-1121 (1955). https://doi.org/10.1080/01621459.1955.10501294

Ernst, M. D. Permutation methods: a basis for exact inference. Stat. Sci. 19, 676-685 (2004). https://doi.org/10.1214/088342304000000396

Faith, D. P., Minchin, P. R. & Belbin, L. Compositional dissimilarity as a robust measure of ecological distance. Vegetatio 69, 57-68 (1987). https://doi.org/10.1007/BF00038687

Fuzessy, L., Pavoine, S., Cardador, L., Maspons, J. & Sol, D. Loss of species and functions in a deforested megadiverse tropical forest. Conserv. Biol. 38, e14250 (2024). https://doi.org/10.1111/cobi.14250

Glennie, R. et al. Hidden Markov models: pitfalls and opportunities in ecology. Methods Ecol. Evol. 14, 43-56 (2023). https://doi.org/10.1111/2041-210X.13801

Goodale, E. & Beauchamp, G. The relationship between leadership and gregariousness in mixed-species bird flocks. J. Avian Biol. 41, 99-103 (2010). https://doi.org/10.1111/j.1600-048X.2009.04828.x

Goodale, E., Beauchamp, G., Magrath, R. D., Nieh, J. C. & Ruxton, G. D. Interspecific information transfer influences animal community structure. Trends Ecol. Evol. 25, 354-361 (2010). https://doi.org/10.1016/j.tree.2010.01.002

Goodale, E. et al. The structure of mixed-species bird flocks, and their response to anthropogenic disturbance, with special reference to East Asia. Avian Res. 6, 14 (2015). https://doi.org/10.1186/s40657-015-0023-0

Goodale, E. et al. Mixed company: a framework for understanding the composition and organization of mixed-species animal groups. Biol. Rev. 95, 889-910 (2020). https://doi.org/10.1111/brv.12591

Gorelick, N. et al. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18-27 (2017). https://doi.org/10.1016/j.rse.2017.06.031

Hayes, D. B. & Monfils, M. J. Occupancy modeling of bird point counts: implications of mobile animals. J. Wildl. Manag. 79, 1361-1368 (2015). https://doi.org/10.1002/jwmg.943

Howes, B. et al. Deforestation alters species interactions. Nat. Sci. 3, e20220027 (2023). https://doi.org/10.1002/ntls.20220027

Instituto Últimos Refúgios (2012). Tiê-do-mato-grosso (Habia rubica) – Red-crowned Ant-Tanager [photograph]. Wikimedia Commons. CC BY-SA 4.0. Available at: https://commons.wikimedia.org/wiki/File:Ti%C3%AA-do-mato-grosso_%28Habia_rubica%29_-_Red-crowned_Ant-Tanager.jpg.

Jansen, M. et al. Forest degradation and inter-annual tree level Brazil nut production in the Peruvian Amazon. Front. For. Glob. Change 3, 525533 (2021). https://doi.org/10.3389/ffgc.2020.525533

Joseph, L. N., Field, S. A., Wilcox, C. & Possingham, H. P. Presence-absence versus abundance data for monitoring threatened species. Conserv. Biol. 20, 1679-1687 (2006). https://doi.org/10.1111/j.1523-1739.2006.00529.x

Jullien, M. & Thiollay, J.-M. Multi-species territoriality and dynamic of neotropical forest understory bird flocks. J. Anim. Ecol. 67, 227-252 (1998). https://doi.org/10.1046/j.1365-2656.1998.00171.x

MacKenzie, D. I. et al. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248-2255 (2002). https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

Magrath, R. D., Haff, T. M., Fallow, P. M. & Radford, A. N. Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. Biol. Rev. 90, 560-586 (2015). https://doi.org/10.1111/brv.12122

McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., Patterson, T. A. & Coulson, T. (2020). Uncovering ecological state dynamics with hidden Markov models. Ecology Letters, 23, 1878–1903. https://doi.org/10.1111/ele.13610

Martínez, A. E., Parra, E., Muellerklein, O. & Vredenburg, V. T. Fear-based niche shifts in neotropical birds. Ecology 99, 1338-1346 (2018). https://doi.org/10.1002/ecy.2217

Mokross, K. et al. Decay of interspecific avian flock networks along a disturbance gradient in Amazonia. Proc. R. Soc. B 281, 20132599 (2014). https://doi.org/10.1098/rspb.2013.2599

Munn, C. A. & Terborgh, J. W. Multi-species territoriality in Neotropical foraging flocks. Condor 81, 338-347 (1979). https://doi.org/10.2307/1366956

Nunes, F. et al. Economic benefits of forest conservation: assessing the potential rents from Brazil nut concessions in Madre de Dios, Peru, to channel REDD+ investments. Environ. Conserv. 39, 132-143 (2012). https://doi.org/10.1017/S0376892911000671

Papke, L. E. & Wooldridge, J. M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. J. Appl. Econom. 11, 619-632 (1996). https://doi.org/10.1002/(SICI)1099-1255(199611)11:6<619::AID-JAE418>3.0.CO;2-1

Parra, E., Hoffman, B., Ponciano, J. M., Martínez, A. E. & Blumstein, D. T. A symphony of fear and safety: distinct vocalizations indicate relative risk in mixed-species groups. Manuscript under revision at Ecology Letters (2026).

Quental, J. (2014). Myrmoborus myotherinus ardesiacus – Black-faced Antbird (male) [photograph]. Wikimedia Commons. CC BY 2.0. Available at: https://commons.wikimedia.org/wiki/File:Myrmoborus_myotherinus_ardesiacus_-_Black-faced_Antbird_%28male%29.jpg.

Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE 77, 257-286 (1989). https://doi.org/10.1109/5.18626

Ramírez, L. F. & Belcher, B. M. Crossing the science-policy interface: lessons from a research project on Brazil nut management in Peru. For. Policy Econ. 114, 101789 (2020). https://doi.org/10.1016/j.forpol.2018.07.018

Rappaport, D. I. et al. Animal soundscapes reveal key markers of Amazon forest degradation from fire and logging. Proc. Natl Acad. Sci. USA 119, e2102878119 (2022). https://doi.org/10.1073/pnas.2102878119

Rockwell, C. A. et al. Nut production in *Bertholletia excelsa* across a logged forest mosaic: implications for multiple forest use. PLOS One 10, e0135464 (2015). https://doi.org/10.1371/journal.pone.0135464

Rutt, C. L., Mokross, K., Kaller, M. D. & Stouffer, P. C. Experimental forest fragmentation alters Amazonian mixed-species flocks. Biol. Conserv. 242, 108415 (2020). https://doi.org/10.1016/j.biocon.2020.108415

Saxby, T. (2012). Logging truck [illustration]. Integration and Application Network, University of Maryland Center for Environmental Science. CC BY-SA 4.0. Available at: source illustration. Last accessed 5 September 2026.

Saxby, T. (2013). Forestry worker [illustration]. Integration and Application Network, University of Maryland Center for Environmental Science. CC BY-SA 4.0. Available at: source illustration. Last accessed 5 September 2026.

Schwarz, G. Estimating the dimension of a model. Ann. Stat. 6, 461-464 (1978). https://doi.org/10.1214/aos/1176344136

Sherony, D. (2012). Tawny-crowned Greenlet (Hylophilus ochraceiceps) (7222935148) [photograph]. Wikimedia Commons. CC BY-SA 2.0. Available at: https://commons.wikimedia.org/wiki/File:Tawny-crowned_Greenlet_%28Hylophilus_ochraceiceps%29_%287222935148%29.jpg.

Timm, C.D. (2010). Xiphorhynchus elegans [photograph]. Wikimedia Commons. CC BY-SA 2.0. Available at: https://commons.wikimedia.org/wiki/File:Xiphorhynchus_elegans.jpg. Last accessed 5 September 2026.

Tobias, J. A. et al. AVONET: morphological, ecological and geographical data for all birds. Ecol. Lett. 25, 581-597 (2022). https://doi.org/10.1111/ele.13898

Vásquez-Ávila, B., Knowlton, J. L., Espinosa, C. I. & Tinoco, B. A. Habitat alteration modifies the structure and function of mixed-species flocks in an Andean landscape. Biotropica 53, 1153-1162 (2021). https://doi.org/10.1111/btp.12957

Visser, I. & Speekenbrink, M. depmixS4: an R package for hidden Markov models. J. Stat. Softw. 36, 1-21 (2010). https://doi.org/10.18637/jss.v036.i07

Willem, H. V., Ingram, V. J. & Guariguata, M. R. Brazil nut forest concessions in the Peruvian Amazon: success or failure? Int. For. Rev. 21, 254-265 (2019). https://doi.org/10.1505/146554819826606540

Zou, F. et al. The conservation implications of mixed-species flocking in terrestrial birds, a globally distributed species interaction network. Biol. Conserv. 224, 267-276 (2018). https://doi.org/10.1016/j.biocon.2018.06.004

GENERAL DISCUSSION AND CONCLUSIONS

The long-term stability of Amazonian mixed-species flocks provides a tractable foundation for addressing ecological questions that may also apply, in modified form, to less persistent multispecies assemblages. This dissertation asked two linked questions: (1) whether recovery from acute predation risk to routine foraging proceeds through multiple discrete stages and, if so, (2) whether the structure of those stages is measurably altered by fine-scale habitat disturbance. We addressed these questions at progressively broader levels of biological organization. Chapter 1 examined the vocal event stream of a single sentinel species, the Dusky-throated Antshrike, Chapter 2 expanded the analysis to the repertoires of multiple associated species, and Chapter 3 compared the organization of those repertoires across a gradient of forest use. The single-species analysis showed that antshrike vocalizations are not isolated events; rather, they occur in reproducible combinations and sequences whose composition, duration, bout structure, and local order change following predator exposure and subsequently recover toward baseline. This result identifies a pathway to signaling complexity in tracheophone suboscines that does not depend on large or highly plastic call repertoires. Instead, complexity can emerge through the context-dependent deployment and temporal organization of relatively stereotyped vocal types. Characterizing this focal event stream also provides a foundation for identifying how heterospecific vocalizations align with, respond to, or modify the communication sequence of a sentinel species.

At the community level, the mixed-flock repertoire provided the basis for the next major result: the identification of intermediate stages in the transition from acute risk to routine foraging. In intact forest, hidden Markov models resolved four ordered states—acute threat, heightened vigilance, recovery, and baseline safety—thereby identifying two transitional stages between immediate alarm and ordinary foraging. These states were characterized by distinct combinations of vocalization types and species traits rather than by a uniform decline in call

rate. Several calls occurred at restricted positions within the transition and therefore represent candidate cues or signals of declining risk. MYAX-P, for example, was strongly associated with the recovery state and exhibited a temporally concentrated distribution consistent with a potential safety function (Figure 2-3C), although playback experiments will be required to determine whether receivers use it as an “all-clear” signal. These results show that relatively constrained antbird repertoires do not preclude rich, context-dependent communication. Rather, intra- and interspecific combinations of calls can be organized in time into reproducible patterns associated with inferred levels of risk.

This functional interpretation builds on an established Amazonian literature showing that mixed-species flock networks respond to habitat modification before the underlying bird assemblage disappears entirely. Across primary forest, fragments, and secondary forest, species-level association frequency and weighted degree decline with increasing disturbance, while flock attendance and clustering are positively associated with vegetation height (Mokross et al. 2014). Spatial analyses further show that immature secondary forest redirects flock movement, expands home-range requirements, and remains suboptimal until vegetation approaches the structure of primary forest (Mokross et al. 2018). Experimental re-isolation of forest fragments produced even stronger effects, including reduced attendance by obligate flock members, movement constrained by newly created edges, and the disappearance of an entire flock (Rutt et al. 2020). Network complexity and cohesion also vary seasonally, with the strongest seasonal differences occurring in fragments and secondary forest, while fine-scale understory and subcanopy structure predict flock species richness and functional diversity (Rutt & Stouffer 2021; Coddington et al. 2023). Collectively, these studies establish a repeated pattern in which human modification erodes the architecture, cohesion, and spatial stability of Amazonian mixed-flock networks.

Association-network studies are essential for identifying this structural decay, but their edges are generally constructed from repeated co-attendance, proximity, or shared space use. They therefore reveal whether species continue to associate and how frequently they do so, but not necessarily what ecological dependency each association supports or what information moves through it. This distinction is important because network links are unlikely to be functionally interchangeable. The loss of a connection to a facultative participant may have different consequences from the weakening of a connection to a species that detects predators, maintains cohesion, or relays information between otherwise weakly connected strata. Network topology can therefore document that connectivity has declined without fully explaining why that decline changes flock behavior or compromises the collective benefits of flocking.

A richer functional characterization of network nodes and edges is therefore needed before tools such as multiplex networks and percolation analysis can be used to predict the effects of disturbance. Experimental work in Amazonian flocks provides an ideal system for developing a framework to assess these asymmetric, role-specific dependencies. Species from different foraging guilds vary in their reliance on heterospecific alarm calls, indicating that their dependence on public information is partly determined by how they search for food and how much private information about predators they can obtain (Martínez & Zenil 2012). Receiver species also respond more strongly to alarms from the sentinel species associated with their own habitat, showing that the reliability and value of predator information vary spatially rather than being equivalent among possible informants (Camerlenghi et al. 2019). At the group level, flock space use can be disproportionately associated with the vegetation conditions used by species such as the Dusky-throated Antshrike, demonstrating how the ecological requirements of a single sentinel species can influence the movement of the wider assemblage (Williams & Lindell 2019). More broadly, small canopy species can occupy central positions in predator-information pathways and propagate alarms across forest strata and taxonomic boundaries

(Camerlenghi et al. 2026). Taken together, these studies suggest that flock connections integrate observations from species occupying different sensory, spatial, and foraging niches. The weakening of particular nodes or pathways may therefore reduce the coverage, redundancy, and temporal resolution of collective risk information even when many of the network's species remain locally present.

The contribution of this dissertation is consequently not the first demonstration that Amazonian mixed-flock networks degrade with habitat modification. Rather, Chapters 2 and 3 extend this literature from network topology to information-processing mechanism. Chapter 2 shows that species contributions are organized across an ordered transition from acute threat through heightened vigilance and recovery to baseline safety. Species with different body sizes, foraging positions, and exposure to protective cover contribute distinct vocalizations during successive stages, indicating that the flock pools information from ecologically differentiated participants rather than responding through a single generalized alarm channel. Chapter 3 then shows that managed forests alter the duration, clarity, and vocal composition of these states even though mixed flocks and an ordered response sequence remain detectable. In this sense, network degradation may involve not only fewer associations, but also a reduced ability to transform distributed observations into a timely and reliable collective estimate of changing danger.

Although the present results do not by themselves demonstrate that altered vocal-state dynamics cause flock dissolution, they identify a measurable mechanism through which previously documented network decay could become functionally consequential. If the identity, timing, or reliability of species contributions is altered, flock members may receive less complete information about when danger has declined. This could prolong vigilance, delay efficient foraging, or weaken cohesion during movement and recovery. These outcomes provide a plausible functional bridge between fine-scale changes in habitat, the disruption of role-specific

connective dependencies, and the reduced attendance or collapse observed under stronger fragmentation.

Several limitations should be considered before extending these inferences broadly. Amazonian understory mixed-species flocks have evolutionary histories in habitats with relatively continuous structure, high degrees of sociality, and largely innate vocal-development systems. This combination may make their interspecific vocal behavior particularly responsive to environmental change and therefore especially suitable for detecting alterations in habitat condition. In other regions, including temperate and Paleotropical systems, different ecological conditions may complicate similarly fine-scale inference about levels of disturbance from communication dynamics. For example, along Andean elevational gradients, mixed flocks occupy habitats with increasing natural discontinuity and physical disturbance, including frequent ravines, extreme geographic features, rockslides, and altitude-driven environmental changes. Research suggests that these conditions reduce the tendency for birds to form the persistent mixed flocks characteristic of Amazonian lowlands. Flocks may instead form more open-membership, less structured aggregations that frequently divide and move through the landscape independently (Montaño-Centellas et al. 2023). A similar loosening of interspecific associations occurs along Central American flyways and in temperate regions, where seasonal and opportunistic mixed flocks contain many migratory species and may be driven primarily by foraging efficiencies and risk reduction through dilution effects. Interspecific communication complexity in these systems may therefore be constrained or expressed differently by their natural ecological conditions, reducing the utility of directly comparing cross-species vocalization dynamics between intact and disturbed sites.

Additional limitations to applying this framework across regions may arise from differences in the biogeographic and evolutionary histories of mixed-species flock systems. Paleotropical mixed-flock systems can contain numerous oscine passerines, including drongos,

babblers, bulbuls, starlings, and related taxa. In contrast to the largely innate vocal development and limited vocal-production learning characteristic of tracheophone suboscines, many oscines acquire vocalizations through social learning from adult tutors. This learning can generate substantial individual and geographic variation, including local dialects and context-dependent vocal mimicry, making signal repertoires more dynamic across space and time. In Sri Lankan mixed-species flocks, Greater Racket-tailed Drongos use context-dependent mimicry of predator and heterospecific alarm-associated calls and may use song and contact-call mimicry to attract other flock participants (Goodale & Kotagama 2006a,b). In a separate African system, Fork-tailed Drongos use both species-specific and mimicked false alarms to steal food and switch alarm-call types when repeated signals lose their effectiveness (Flower 2011; Flower et al. 2014). By contrast, tracheophone suboscines show more developmentally constrained vocal production, although the timing, rate, and sequential deployment of their vocal forms may remain context-dependent. Whether these differences in vocal-developmental mode influence mixed-flock stability remains unresolved. Nevertheless, they may complicate acoustic annotation and machine-learning comparisons by weakening the assumption that a vocal form maps consistently onto a species, call type, or behavioral function across sites.

Within these limitations, this dissertation demonstrates that communication can reveal how multispecies assemblages process ecological information, rather than simply whether their constituent species continue to co-occur. By linking species traits and call repertoires to ordered transitions in perceived risk, and by showing that the duration, clarity, and composition of those transitions change under forest management, the dissertation provides a mechanistic bridge between network structure and ecosystem function. Communication dynamics may therefore offer a measurable means of detecting functional degradation before it becomes evident through local species loss or complete network collapse.

GENERAL REFERENCES

Coddington, C. P. J., Cooper, W. J., Mokross, K. & Luther, D. A. (2023). Forest structure predicts species richness and functional diversity in Amazonian mixed-species bird flocks. *Biotropica*, **55**(2), 467–479. https://doi.org/10.1111/btp.13201

Flower, T. P., Gribble, M. & Ridley, A. R. (2014). Deception by flexible alarm mimicry in an African bird. *Science*, **344**(6183), 513–516. https://doi.org/10.1126/science.1249723

Goodale, E. & Kotagama, S. W. (2006a). Context-dependent vocal mimicry in a passerine bird. *Proceedings of the Royal Society B: Biological Sciences*, **273**(1588), 875–880. https://doi.org/10.1098/rspb.2005.3392

Goodale, E. & Kotagama, S. W. (2006b). Vocal mimicry by a passerine bird attracts other species involved in mixed-species flocks. *Animal Behaviour*, **72**(2), 471–477. https://doi.org/10.1016/j.anbehav.2006.02.004

Mokross, K., Potts, J. R., Rutt, C. L. & Stouffer, P. C. (2018). What can mixed-species flock movement tell us about the value of Amazonian secondary forests? Insights from spatial behavior. *Biotropica*, **50**(4), 664–673. https://doi.org/10.1111/btp.12557

Montaño-Centellas, F. A., Muñoz, J., Mangini, G. G., Ausprey, I. J., Newell, F. L., Jones, H. H., Fanjul, M. E., Tinoco, B. A., Colorado Z., G. J., Cahill, J. R. A., Arbeláez-Cortés, E., Marin-Gómez, O. H., Astudillo, P. X., Guevara, E. A., Ippi, S., McDermott, M. E., Rodewald, A. D., Matthysen, E. & Robinson, S. K. (2023). Network structure of avian mixed-species flocks decays with elevation and latitude across the Andes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **378**(1878), 20220099. https://doi.org/10.1098/rstb.2022.0099

Williams, S. M. & Lindell, C. A. (2019). The influence of a single species on the space use of mixed-species flocks in Amazonian Peru. *Movement Ecology*, **7**, 37. https://doi.org/10.1186/s40462-019-0181-5