University of St Andrews

Detections of rare species lead citizen scientists to initiate data recording

Louis Backstrom, Rachel Drake, Hannah Worthington, Alison Johnston

Centre for Research into Ecological and Environmental Modelling,
School of Mathematics and Statistics
University of St Andrews, St Andrews, United Kingdom

BES Annual Meeting 2025 | 16 December 2025, Edinburgh

Anne Keymer Prize Candidate

Please judge my presentation

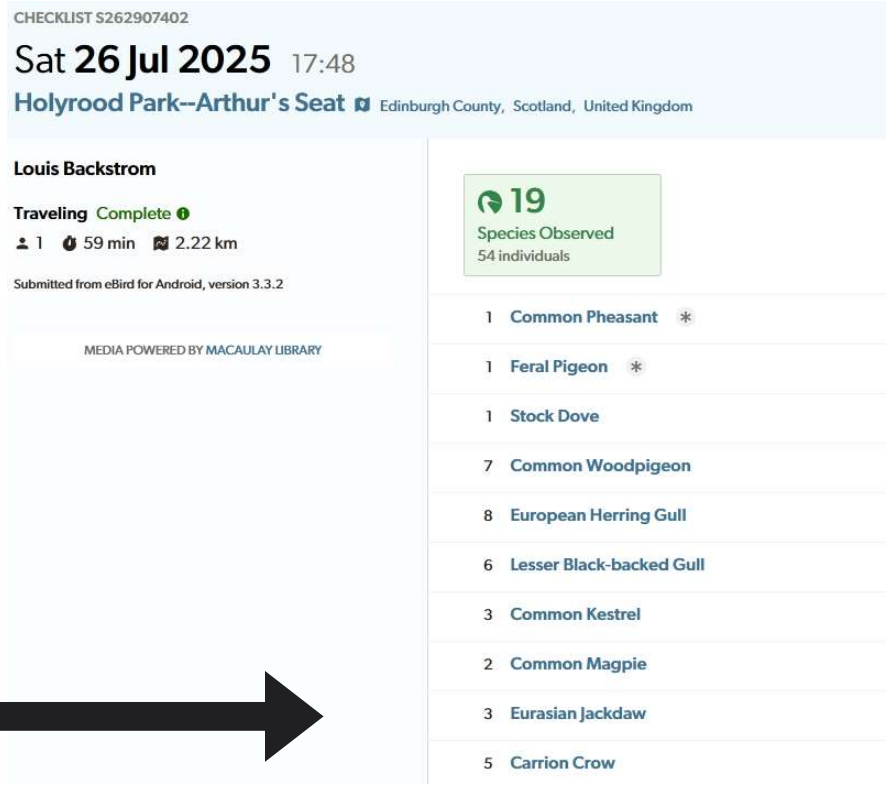
www.st-andrews.ac.uk

Motivation: Citizen Science

- Engaging the public in biological data collection has a long history
  - National Wildfowl Count in 1947 (now Wetland Bird Survey / WeBS)
- Explosion of popularity coincident with internet & smartphones
- Countless projects now active

BTO

eBird



Motivation: Observer Preferences

- Citizen scientists generate huge volumes of biodiversity data
  - eBird: > 2 billion observations
- But! They tend to survey in very non-random, unstructured ways
  - Sampling biases – spatial, temporal, **taxonomic**
- **Can we address these issues in analysis?**

Observers have preferences for some species over others

Boring 🥱



(common in Edinburgh)

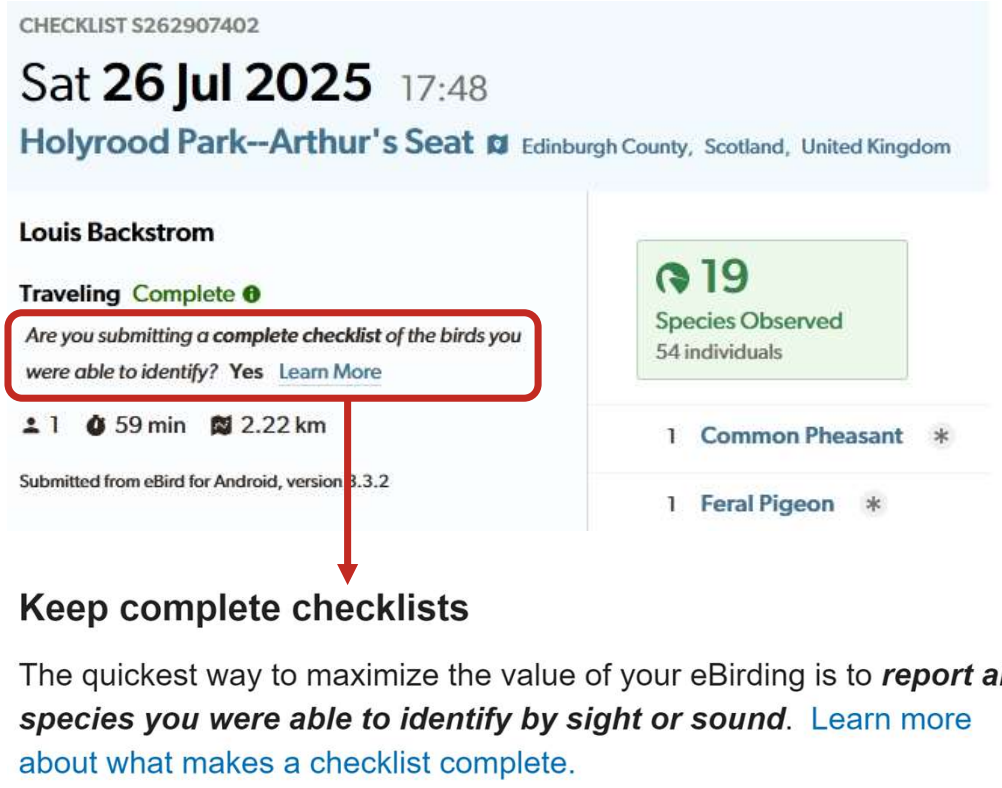
Interesting 😮



(rare in Edinburgh)

Motivation: Project Structure

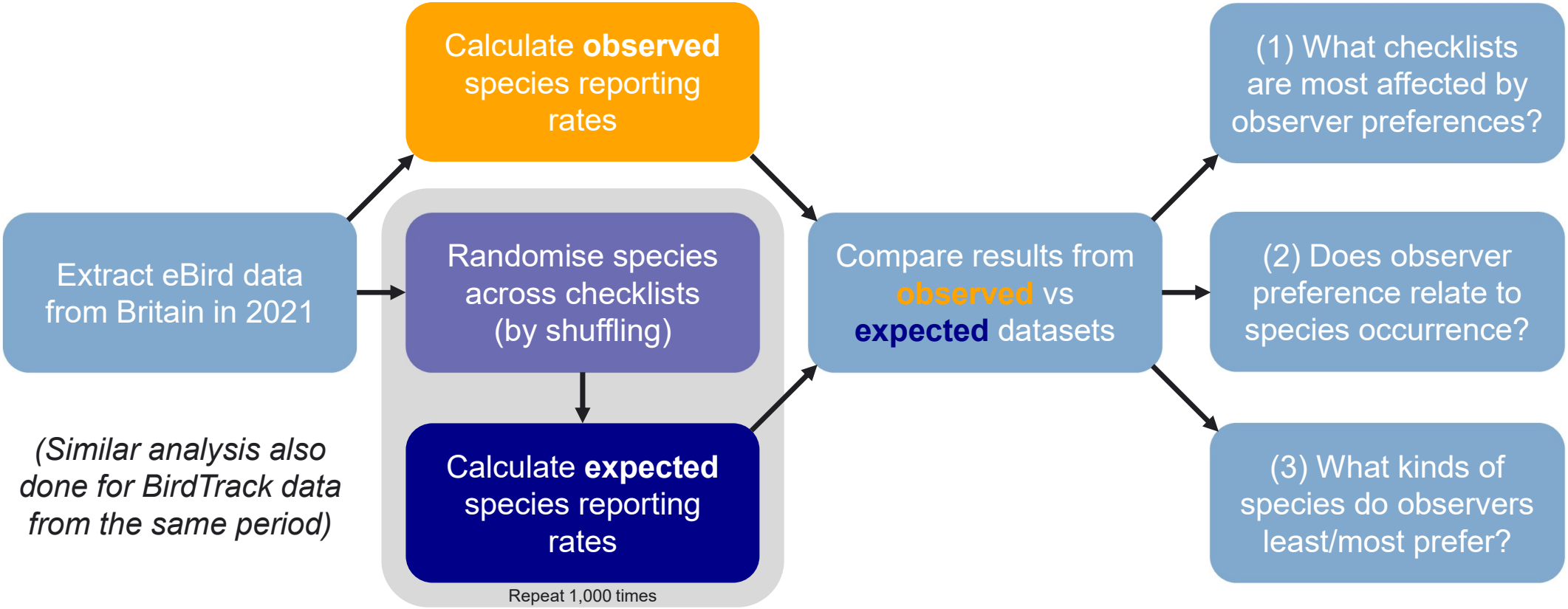
- Projects can impose **structure** or collect **auxiliary metadata**
  - Mitigates the effect of preferences
- Many projects (e.g. eBird) ask observers to report everything
  - “Complete” checklists
- Limits observers’ species preferences
  - **Or hides them?**



Motivation: The Problem

What if an observer only starts a checklist because they've seen an interesting (e.g. rare) bird?

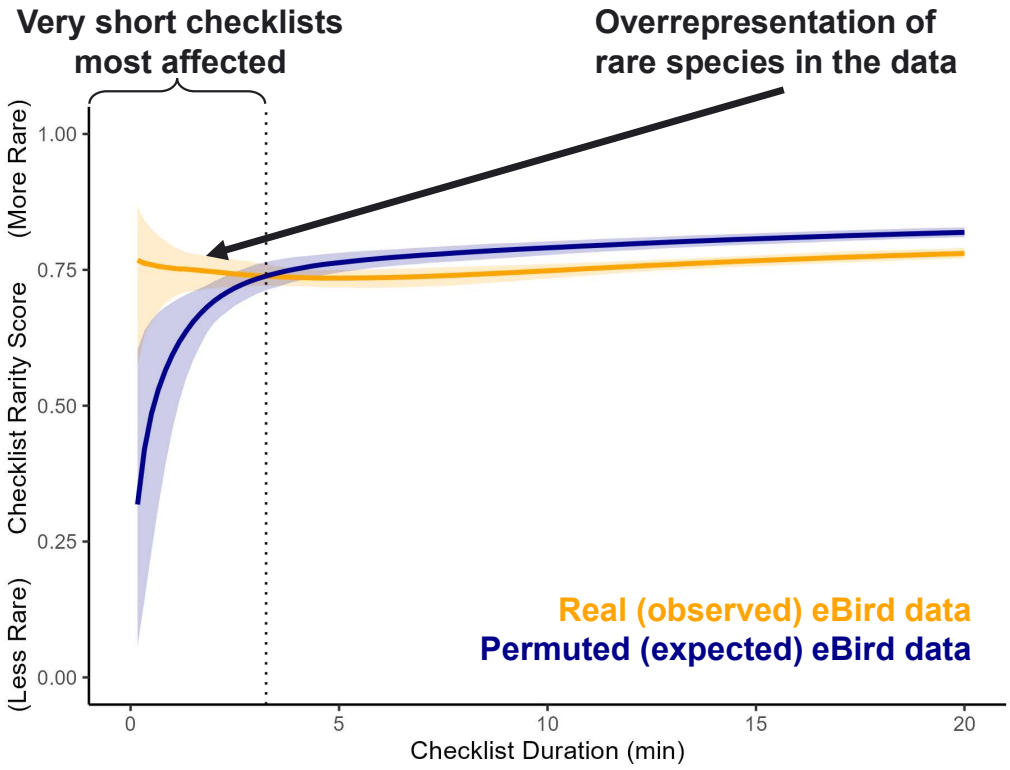
Methods



(Similar analysis also done for BirdTrack data from the same period)

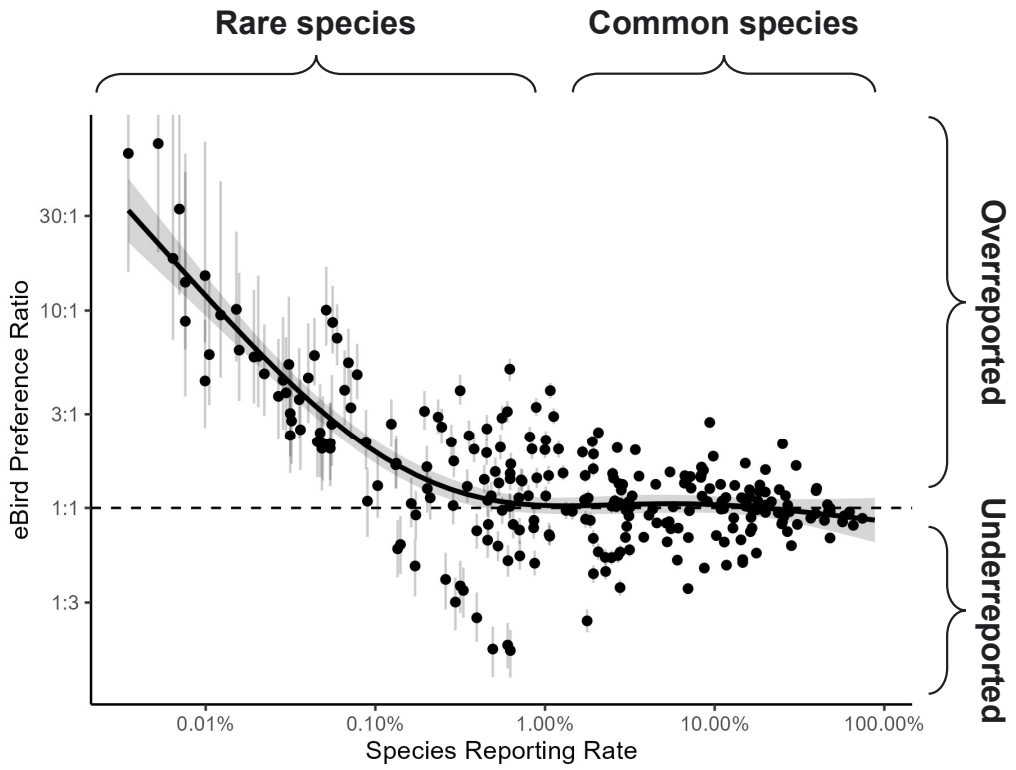
Results: (1) Checklist Rarity

- Model rarity in **observed** & **expected** datasets
- **Checklist Rarity Score** is a normalised metric of how rare the rarest species observed is
- **Observers are more likely to start an eBird checklist when they find a rare bird!**
  - Same behaviour also seen in BirdTrack data



Results: (2) Species Preference

- Compare **observed** & **expected** species occurrence rates on most affected checklists
- **eBird Preference Ratio** is the ratio of these occurrence rates
- Rare species overpreferred
- Common species vary greatly in level of observer preference



Results: (3) Species Rankings

- Among common birds, the most-preferred species tend to be **distinctive** and **detectable**
  - Raptors, large waterfowl



- Least-preferred species more obscure, often cryptic
  - Seabirds, ducks, drab passerines



Implications

Citizen scientists exhibit **species-specific preferences**

These **preferences affect how observers survey**
(even in semi-structured projects like eBird and BirdTrack)

Limited evidence that these preferences led to biased inference
from occupancy models fit using eBird data (more work needed!)

Recommendations

For analysts:

Remove very short duration checklists when analysing citizen science data

10 minutes a safe threshold for most datasets (e.g. eBird, BirdTrack)

For project managers:

Design citizen science projects to **track the order in which observers record species**

Removal of first-recorded species would limit most of the bias explored here

Acknowledgements

- Thanks to Ali Johnston, Hannah Worthington, and Rachel Drake
- PhD and travel funding from CREEM



Paper in *Diversity and Distributions*

(Open Access)

Download these slides

Email me at ljb38@st-andrews.ac.uk

Anne Keymer Prize Candidate

Please judge my presentation