Holstonia Bigfoot Investigations
Perception Series No. 7

Observer Skill Stratification: Modeling Variability in Human Detection and Reporting

Daniel H. Kegley
holstonia-investigations.org
Version of Record: This document constitutes the authoritative version of this work. Please cite the version available at holstonia-investigations.org. Revised editions, if issued, will be explicitly identified.

Abstract

Human observers function as primary detection instruments in many field sciences, yet they differ in perceptual acuity, task familiarity, training, and interpretive discipline. These differences produce structured variability in detection outcomes. This paper examines observer heterogeneity as a measurable feature of the observation process rather than a contaminant to be dismissed. Drawing on research from wildlife biology, ecological monitoring, and citizen science, the analysis demonstrates that detection probability varies systematically across observers and that failure to model these differences can bias occupancy estimates, abundance calculations, and trend analyses. Rather than treating reports as uniformly reliable or unreliable, a stratified framework is proposed in which observers are understood as differently calibrated instruments embedded within

environmental and methodological constraints. Such an approach permits anomalous-report datasets to be evaluated within a mature measurement framework, shifting analytical focus from belief to detectability.

---

1. Introduction

Field sciences have long recognized that observation is not a passive act but a measurable process shaped by the capabilities of the observer. Variation among observers can alter population estimates, distort trend detection, and introduce systematic error into ecological inference (Nichols et al., 2000).

Despite this recognition in established disciplines, anomalous biological reporting has often treated witnesses as a homogeneous class — either trusted wholesale or dismissed categorically. Both responses obscure a more productive question:

What kind of instrument is a human observer?

Anomalous-report datasets cannot be responsibly interpreted without modeling heterogeneity in observer skill, perceptual training, environmental familiarity, and reporting discipline.

The goal is not to rank credibility but to describe measurement performance.

---

2. Observer Variation as a Scientific Problem

Interobserver variability has been documented across multiple ecological contexts. Research on detection of low-density populations demonstrates that experienced observers are more likely to detect organisms than less experienced participants, potentially biasing estimates when detection probability is not explicitly modeled (McClintock et al., 2010).

Similarly, ornithological studies employing double-observer methods show that single observers routinely miss animals that are present and that observer identity significantly influences detection probabilities (Nichols et al., 2000).

Detection estimates can vary dramatically even under controlled conditions. Alldredge et al. (2007) reported detection probabilities ranging from approximately 0.60 to 1.00 across observers for identical survey targets.

These findings establish a foundational principle:

Observation error is rarely random — it is structured.

Ignoring that structure risks attributing measurement artifacts to ecological reality.

---

3. Skill, Experience, and Perceptual Calibration

Experience often improves detection performance. In wildlife monitoring contexts, observers familiar with species-specific cues such as vocalizations or movement patterns demonstrate higher detection rates than novices (Garrard et al., 2008).

Yet skill effects are not strictly linear. Long-term survey research indicates that aging observers may exhibit declining detection ability, partially associated with sensory changes such as hearing loss (Farmer et al., 2014).

Even seemingly simple tasks such as estimating group size show substantial observer-dependent variability influenced by spatial arrangement, density, and visual complexity (Erwin, 1982).

Observers are therefore not interchangeable components within a monitoring system. They possess evolving perceptual profiles that shape what enters the dataset.

---

4. Citizen Science and the Myth of the Uniform Amateur

The expansion of citizen science has intensified interest in observer reliability. While concerns about data quality are common, empirical findings complicate binary distinctions between expert and amateur observers.

Comparative research demonstrates that trained volunteers can produce data closely aligned with professional scientists when protocols are clear and task complexity is controlled (Earp et al., 2018).

In some cases, non-professional observers perform comparably to trained research assistants on complex observational tasks, suggesting that structured methodology may matter more than formal credentials (SturtzSreetharan et al., 2021).

Unexpectedly, phenological studies have found that minimally trained participants occasionally record observations with accuracy equal to or exceeding that of trained observers, possibly due to interpretive drift introduced during training (Feldman et al., 2018).

These findings challenge the assumption of a uniformly unreliable amateur class and instead support a stratified model in which observer performance emerges from the interaction of training, protocol clarity, and environmental context.

---

5. Detectability as a Modeled Process

Modern ecological inference treats detection as an estimable parameter rather than an inconvenience. Sightability and double-observer models explicitly correct raw counts for undetected individuals, improving both abundance estimates and precision (Griffin et al., 2013).

Failure to model detection can bias species distribution analyses because organisms may go undetected despite being present — a problem directly addressed through occupancy modeling frameworks (MacKenzie et al., 2002).

More recent statistical approaches incorporate participant ability into hierarchical models, estimating observer sensitivity and specificity to better characterize latent ecological processes (Chambert et al., 2015).

The implication is profound:

Observation is itself a measurable ecological process.

---

6. Toward a Stratified Observer Framework

Observer skill stratification does not imply that some reports should be dismissed outright. Instead, it proposes that datasets contain layered detection probabilities shaped by observer characteristics.

Relevant dimensions include:

- perceptual acuity
- domain familiarity
- training quality
- task complexity
- environmental conditions
- cognitive load

When modeled explicitly, these variables transform anecdotal reports into analyzable measurement events.

This shift mirrors developments in astronomy, wildlife biology, and epidemiology — disciplines that matured only after acknowledging instrument variability.

Here, the human observer is the instrument.

---

7. Implications for Anomalous Biological Reports

Anomalous-report research has often been constrained by debates over belief. Observer stratification reframes the discussion:

The central question becomes not whether a report is true, but **what detection conditions produced it.**

Such reframing yields several analytical advantages:

- Detection failure becomes expected rather than disconfirming.
- High-skill observers gain interpretive weight without invoking authority.
- Training protocols become scientifically meaningful.
- Citizen science transitions from informal participation to structured data collection.

Most importantly, uncertainty becomes quantifiable rather than rhetorical.

---

8. Limitations and Cautions

Stratification must avoid collapsing into witness hierarchies. The objective is not to rank human worth but to model observational performance.

Observer skill also interacts with ecological variables such as vegetation density, species behavior, and abundance — all of which influence detectability (Johnston et al., 2018).

No observer operates independently of the environment.

Measurement is always relational.

---

9. Conclusion

Observer heterogeneity is not a peripheral concern but a foundational feature of field-based knowledge production. Treating observers as differently calibrated instruments enables anomalous-report datasets to be evaluated within a disciplined inferential framework.

The transition from undifferentiated testimony to modeled detection marks an epistemic shift:

From asking whether observers should be believed
to understanding how observation itself functions.

Disciplines mature when they begin to measure their instruments.

For fields reliant on human perception, the observer is the first instrument that must be understood.

Recognizing observer variability is only the first step in constructing a disciplined observation science. Once heterogeneity in perceptual skill is acknowledged, the methodological question becomes unavoidable: how should observation systems be designed to incorporate these differences without collapsing into credibility hierarchies?

Addressing that question requires a shift from diagnosis to instrumentation. The calibrated observer framework, developed in *Holstonia Methods 6 - The Calibrated Observer*, advances calibration as a scientific alternative to witness ranking and outlines how distributed observation networks can model human detection as a measurable component of inference.

References

Alldredge, M. W., Pollock, K. H., Simons, T. R., Collazo, J. A., & Shriner, S. A. (2007). Time-of-detection method for estimating abundance from point-count surveys. *The Auk*, 124(2), 653–664.

Chambert, T., Miller, D. A. W., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology*, 96(2), 332–339.

Earp, H. S., Liconti, A., Yulita, K. S., & Andradi-Brown, D. A. (2018). Do citizen science data match the quality of professional data? *Citizen Science: Theory and Practice*, 3(1).

Erwin, R. M. (1982). Observer variability in estimating numbers: An experiment. *Journal of Field Ornithology*, 53(2), 159–167.

Farmer, R. G., Leonard, M. L., & Mills Flemming, J. E. (2014). Observer aging and long-term avian survey data quality. *Ecology and Evolution*, 4(13), 2563–2576.

Feldman, R. E., Zemaite, I., & Miller-Rushing, A. J. (2018). How training citizen scientists affects the accuracy and precision of phenological data. *International Journal of Biometeorology*, 62, 1421–1435.

Garrard, G. E., Bekessy, S. A., McCarthy, M. A., & Wintle, B. A. (2008). When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology*, 33(8), 986–998.

Griffin, P. C., Lubow, B. C., Jenkins, K. J., Vales, D. J., & Moeller, B. J. (2013). A hybrid double-observer sightability model for aerial surveys. *Journal of Wildlife Management*, 77(6), 1226–1234.

Johnston, A., Fink, D., Hochachka, W. M., & Kelling, S. (2018). Estimates of observer expertise improve species distributions from citizen science data. *Methods in Ecology and Evolution*, 9(1), 88–97.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), 2248–2255.

McClintock, B. T., Bailey, L. L., Pollock, K. H., & Simons, T. R. (2010). Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, 91(8), 2446–2454.

Nichols, J. D., Hines, J. E., Sauer, J. R., Fallon, F. W., Fallon, J. E., & Heglund, P. J. (2000). A double-observer approach for estimating detection probability and abundance from point counts. *The Auk*, 117(2), 393–408.