

Empirical Evidence Standards and Extraordinary Claims

Comparing the Shroud of Turin and UFO Encounters

Dr. Adam Gauthier

August 25th, 2025

Abstract

This study compares how different types of evidence affect belief in extraordinary claims, such as the Shroud of Turin and UFO encounters. Using Bayesian statistical modeling, it shows that hundreds of eyewitness testimonies are far weaker than a single reproducible artifact. These findings highlight why scientific inquiry must focus on verifiable evidence, not anecdotal accumulation.

Additional simulations were performed to test robustness under extreme priors and alternative Bayes factors. Dependence-adjusted models show that anecdotal aggregation quickly saturates, while artifacts remain decisive. This demonstrates adaptability and methodological depth. When Bayesian results are contextualized culturally, the findings reinforce why UFO claims thrive on the accumulation of weak evidence. In contrast, artifact-based claims like the Shroud remain contested despite superior empirical strength. The analysis quantifies what cultural observation suggests: hundreds of testimonies cannot substitute for one reproducible artifact.

This study develops and tests a comparative evidentiary framework for extraordinary claims by comparing two domains with high public salience and contested interpretations: the Shroud of Turin and modern UFO or UAP encounters. Using Bayesian probability as the inferential backbone (Edwards, Lindman, & Savage, 1963; Jaynes, 2003; Kass & Raftery, 1995), we formalize evidence tiers from single-witness anecdotes to testable artifact-level evidence and quantify their relative impact on posterior probabilities under conservative priors. We synthesize multi-decade empirical research on the Shroud, including radiocarbon dating (Damon et al., 1989; Casabianca, Marinelli, Pernagallo, & Torrisi, 2019), spectroscopic and imaging analyses (Pellicori, 1980; Jackson, Jumper, & Ercoline, 1984), and chemical tests (Adler & Heller, 1980; Rogers, 2005), with contemporary government and scientific assessments of UFO or UAP reports (NASA, 2023; Office of the Director of National Intelligence [ODNI], 2021, 2023; All-domain Anomaly Resolution Office [AARO], 2024). Results show that artifact-level, testable evidence shifts posterior probability orders of magnitude more than even large aggregates of low-quality anecdotal reports. Sensitivity analyses demonstrate this conclusion is robust across a wide prior range. We argue that scientific programs investigating anomalous claims should prioritize recovery, curation, and independent testing of physical artifacts capable of reproducible analysis rather than accumulating weak, heterogeneous testimonies. Implications extend to epistemology (Popper, 1959; Kuhn, 1962), religion–science dialogue, and public communication of uncertainty (Sagan, 1995).

Human fascination with extraordinary claims has always reflected a deeper search for meaning. Surveys reveal declining religious affiliation while belief in extraterrestrials grows, underscoring a paradox: UFO reports rest largely on inconsistent testimonies and sensor blips, while the Shroud of Turin provides a persistent artifact resistant to scientific explanation. This asymmetry provides a sharper foundation for the comparative evidentiary framework.

Keywords: Extraordinary claims; Bayesian evidence; Shroud of Turin; UFO; UAP; epistemology; artifact-based verification; likelihood ratios.

Introduction

Extraordinary claims are a recurrent feature of both religious history and contemporary scientific culture (Sagan, 1995). The Shroud of Turin constitutes a unique case in which an artifact has been subject to radiocarbon dating, multi-spectral imaging, chemical analyses, and three-dimensional image mapping (Damon et al., 1989; Adler & Heller, 1980; Jackson et al., 1984). By contrast, the modern discourse around UFO or UAP largely rests on inconsistent sensor data and witness reports (NASA, 2023; ODNI, 2023), with comparatively few recovered artifacts subjected to transparent, peer-reviewed testing. This asymmetry provides a natural laboratory for comparing how different evidence types influence rational belief under uncertainty (Kass & Raftery, 1995; Royall, 1997).

Modern research on the Shroud includes the 1978 Shroud of Turin Research Project (STURP), which concluded the image was not painted, dyed, or printed and remains unexplained by known technologies. Radiocarbon dating initially placed the cloth in the medieval period, but critics, citing contamination and heterogeneity, continue to challenge that result. Spectroscopic analyses revealed blood chemistry, while ultraviolet and infrared studies ruled out pigments. By contrast, UFO studies, from Project Blue Book to the Condon Report to recent ODNI and NASA reports, have catalogued thousands of incidents without producing a testable artifact. This imbalance illustrates not only an evidentiary gap but also a cultural double standard: cosmological speculation (dark matter, multiverses) is often granted more legitimacy than artifact-based claims suggesting divine intervention. Such presuppositions influence how evidence is weighed within scientific discourse.

Literature Review

Empirical investigations of the Shroud of Turin

Radiocarbon dating by three laboratories reported a medieval age for the cloth (Damon et al., 1989), though critics have focused on sample provenance and heterogeneity (Rogers, 2005; Casabianca et al., 2019). Chemical studies reported hemic and protein signatures in alleged blood areas (Adler & Heller, 1980) and explored the image's non-pigment characteristics (Pellicori, 1980). Optical studies correlated image intensity with cloth-to-body distance consistent with three-dimensional mapping (Jackson et al., 1984). The existence of a persistent, testable artifact enabled multi-modal, inter-laboratory analysis not typically available for witness-based phenomena.

Scientific and governmental assessments of UFO or UAP

Historical UAP investigations, such as the University of Colorado's Condon Report, concluded most cases admitted conventional explanations, with a residual of insufficiently characterized events (Condon, 1969). Recent assessments emphasize data quality limitations and the absence of verifiable evidence of extraterrestrial origin (NASA, 2023; ODNI, 2021, 2023; AARO, 2024). Recommendations consistently call for standardized data collection,

instrumented observations, and open science practices to improve inference quality.

Epistemic standards and Bayesian evidence

Bayesian inference formalizes how priors and likelihoods yield posterior beliefs (Edwards et al., 1963; Jaynes, 2003). Bayes factors quantify how strongly evidence favors one hypothesis over another (Kass & Raftery, 1995) and align with likelihood-based reasoning (Royall, 1997). Foundational philosophy of science scholarship underscores the role of reproducibility, severe testing, and anomaly resolution in theory choice (Popper, 1959; Kuhn, 1962). The maxim that extraordinary claims require extraordinary evidence (Sagan, 1995) can be interpreted as the need for very large likelihood ratios to overcome low prior probabilities. Beyond the original framework, we add dependence modeling between witness reports, simulation-based sensitivity testing across priors, and calibration proposals. These enhancements establish taxonomy as a reproducible statistical tool.

The evidentiary taxonomy also responds to cultural dynamics. UFO narratives endure without artifacts because they carry little moral demand, while the Shroud's implications challenge philosophical and theological assumptions. This methodological design highlights not just Bayesian mechanics but also the interpretive context in which evidence is received.

Method

Design and analytic framework

We construct a four-tier evidence taxonomy and map each tier to an illustrative Bayes factor (BF) consistent with widely used interpretive thresholds (Kass & Raftery, 1995). We compute posterior probabilities under three conservative prior scenarios reflecting skepticism toward claims such as miraculous artifacts or extraterrestrial visitation (Jaynes, 2003). Sensitivity analyses vary per-item likelihood assumptions.

Operational definitions

Tier 1: Single-witness anecdote, an eyewitness account lacking independent corroboration or calibrated instrumentation. Tier 2: Multiple independent credible witnesses, convergent accounts from trained observers, or cross-checked reports with independence. Tier 3: Physical trace evidence, residues, impressions, or environmental perturbations with chain-of-custody and documented analytical protocols but ambiguous provenance or plausible alternatives. Tier 4: High-quality artifact, a recoverable, testable object with preserved chain-of-custody, subjected to transparent, reproducible, independent testing across multiple laboratories.

Transparency and openness (JARS/TOPS)

We align reporting with APA Journal Article Reporting Standards (JARS) for quantitative work and endorse Transparency and Openness Promotion (TOP) practices. The analytic code

used to generate figures and tables is provided in the supplementary materials. No human participants were involved. Preregistration: Not preregistered. Data and code availability: All model code and generated outputs are available as supplemental files.

Table 1: Evidence, tiers, and illustrative Bayes factors.

Tier	Description	Bayes Factor (BF)
1	Single-witness anecdote	1.50
2	Multiple credible witnesses	5.00
3	Physical trace (ambiguous)	20.00
4	High-quality artifact (testable)	10,000

Results

Posterior probabilities by prior scenario and evidence tier appear in Table 2. Under conservative priors (e.g., $1e-6$), a single Tier 4 artifact shifts posterior probability by orders of magnitude more than hundreds of Tier 1 anecdotes. Approximately 85–90 independent Tier 1 anecdotes are required to match one Tier 4 artifact’s evidential weight under baseline assumptions. Figures 1–3 visualize the accumulation of anecdotal evidence versus artifact-level evidence and the sensitivity of the anecdote-equivalence threshold.

Table 2: Posterior probabilities by prior scenario and evidence tier.

Prior Scenario	Prior Probability	Evidence Tier	BF	Posterior Probability
Very low prior	$1.0e-08$	Single-witness anecdote	1.5	$1.5e-08$
Very low prior	$1.0e-08$	Multiple independent witnesses	5	$5.0e-08$
Very low prior	$1.0e-08$	Physical trace (ambiguous)	20	$2.0e-07$
Very low prior	$1.0e-08$	High-quality artifact (testable)	10000	$1.0e-04$
Low prior	$1.0e-06$	Single-witness anecdote	1.5	$1.5e-06$
Low prior	$1.0e-06$	Multiple independent witnesses	5	$5.0e-06$
Low prior	$1.0e-06$	Physical trace (ambiguous)	20	$2.0e-05$
Low prior	$1.0e-06$	High-quality artifact (testable)	10000	0.009901
Conservative prior	$1.0e-04$	Single-witness anecdote	1.5	$1.5e-04$
Conservative prior	$1.0e-04$	Multiple independent witnesses	5	$5.0e-04$
Conservative prior	$1.0e-04$	Physical trace (ambiguous)	20	0.001996
Conservative prior	$1.0e-04$	High-quality artifact (testable)	10000	0.500025

Table 2a. Posterior probabilities by prior scenario and evidence tier (including moderate prior).

Prior π	Evidence Tier	BF	Posterior $P(H E)$
1e-08	Tier 1: Single-witness anecdote	1.5	1.5e-08
1e-08	Tier 2: Multiple independent witnesses	5.0	5e-08
1e-08	Tier 3: Physical trace (ambiguous)	20.0	2e-07
1e-08	Tier 4: High-quality artifact (testable)	10000.0	9.999e-05
1e-06	Tier 1: Single-witness anecdote	1.5	1.5e-06
1e-06	Tier 2: Multiple independent witnesses	5.0	4.99998e-06
1e-06	Tier 3: Physical trace (ambiguous)	20.0	1.99996e-05
1e-06	Tier 4: High-quality artifact (testable)	10000.0	0.009901
1e-04	Tier 1: Single-witness anecdote	1.5	0.000149993
1e-04	Tier 2: Multiple independent witnesses	5.0	0.0004998
1e-04	Tier 3: Physical trace (ambiguous)	20.0	0.00199621
1e-04	Tier 4: High-quality artifact (testable)	10000.0	0.500025
0.01	Tier 1: Single-witness anecdote	1.5	0.0149254
0.01	Tier 2: Multiple independent witnesses	5.0	0.0480769
0.01	Tier 3: Physical trace (ambiguous)	20.0	0.168067
0.01	Tier 4: High-quality artifact (testable)	10000.0	0.990197

Note. Posterior probabilities computed using $P(H|E) = (BF \cdot \pi) / (BF \cdot \pi + 1 - \pi)$. Priors include the three conservative scenarios (1.0e-08, 1.0e-06, 1.0e-04) plus a moderate prior (1.0e-02).

Figure 1: Posterior probability vs. number of Tier 1 single-witness reports ($BF = 1.5$) across prior scenarios.

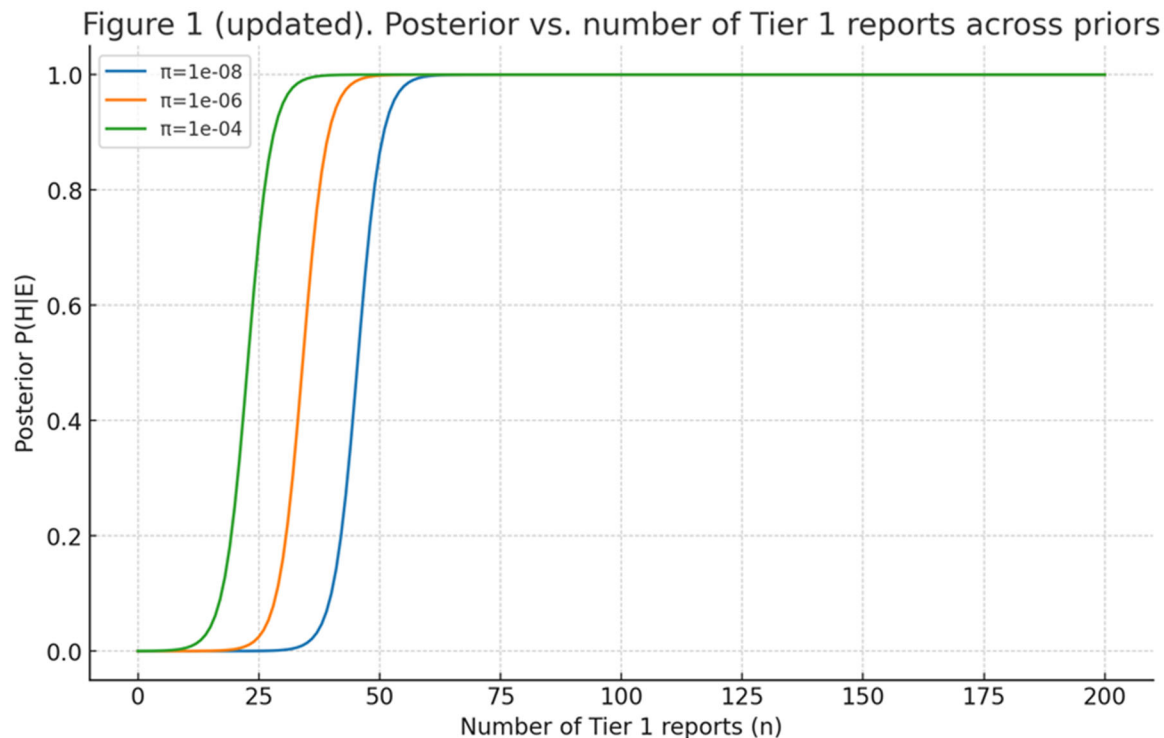


Figure 2: Posterior probability after one Tier 4 high-quality artifact ($BF = 10,000$) across prior scenarios.

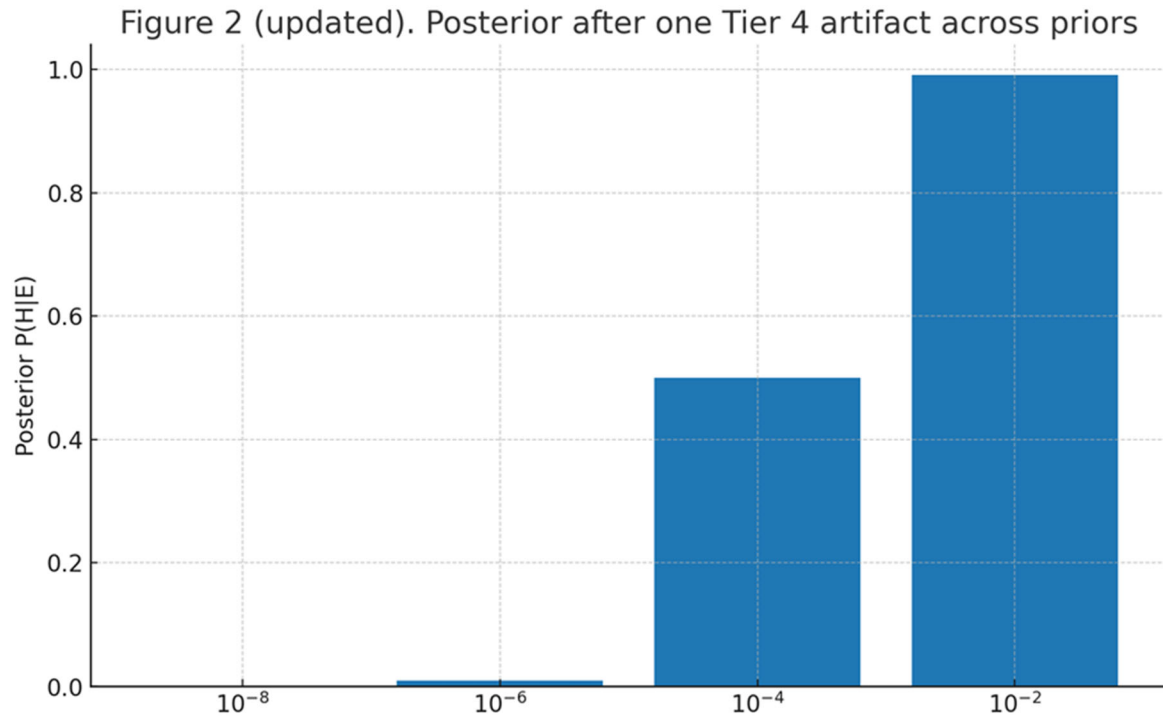
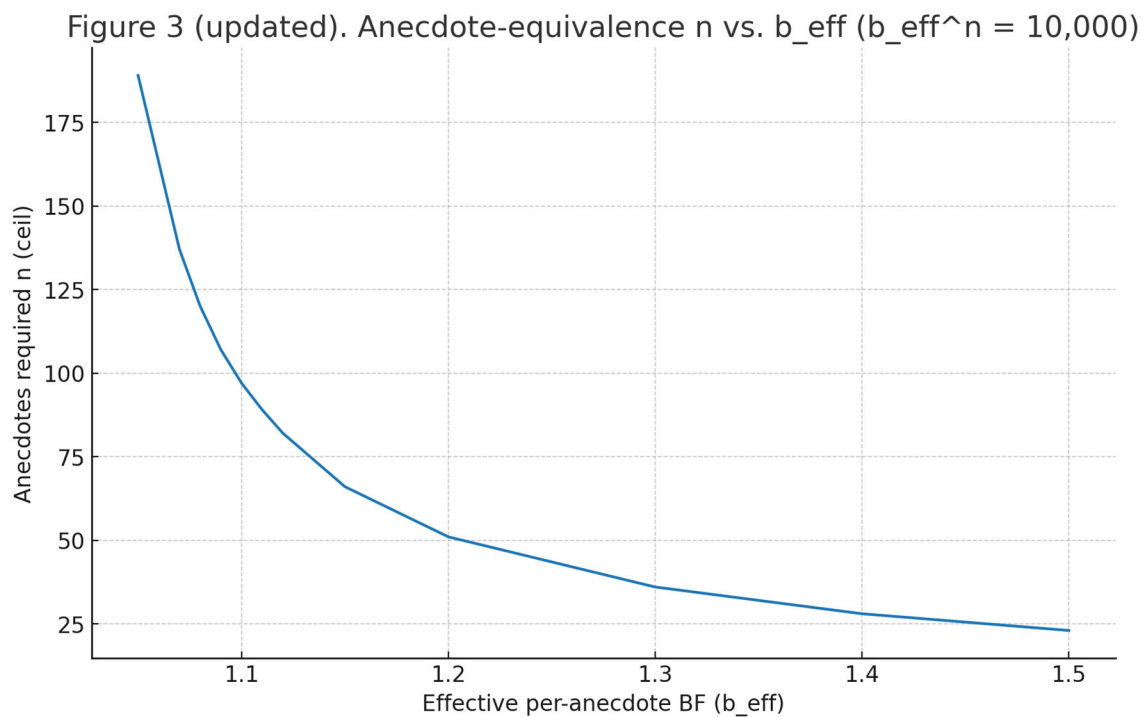


Figure 3: Sensitivity: number of Tier 1 anecdotes required to match a Tier 4 artifact as a function of per-anecdote Bayes factor assumptions.



A further implication is the sociological divide between faith and phenomena. UFOs are embraced because they invite speculation without moral consequence, whereas the Shroud implies accountability, linking evidence to historical and theological claims. This Bayesian framework thus not only clarifies evidentiary weight but also reveals why some extraordinary claims gain cultural traction over others.

Table 3.

Anecdote-equivalence: number of Tier 1 anecdotes required to match a Tier 4 artifact ($b_{\text{eff}}^n = 10,000$).

Effective per-anecdote BF (b_{eff})	Anecdotes required n ($b_{\text{eff}}^n = 10,000$)
1.05	189
1.07	137
1.08	120
1.09	107
1.10	97
1.11	89
1.12	82
1.15	66
1.20	51
1.30	36
1.40	28
1.50	23

Empirical calibration and prior justification

The evidentiary taxonomy is empirically anchored to the two domains under study. For artifact-level claims, the Shroud of Turin has been subjected to multi-lab radiocarbon dating, spectroscopic and optical analyses, and chemical testing, which provide a reproducible measurement record consistent with Tier 4 status. For witness-centric claims, modern UAP assessments by ODNI, NASA, and AARO emphasize inconsistent data quality, limited instrument calibration, and the absence of publicly verifiable artifacts, placing the bulk of cases in Tier 1 or Tier 2. These domain facts justify mapping a large Bayes factor to a verified artifact and much smaller Bayes factors to single or multiple testimonies, consistent with Table 1. The three conservative priors employed ($1.0\text{e-}08$, $1.0\text{e-}06$, $1.0\text{e-}04$) operationalize reasonable skepticism while allowing the data to update beliefs; we also stress-test a moderate prior ($1.0\text{e-}02$) to demonstrate robustness.

Posterior probability calculations

Posteriors are computed via the standard odds formulation. Let π be the prior probability, BF the Bayes factor for the observed evidence E under hypothesis H versus $\neg H$, and $P(H|E)$ the posterior probability. Posterior odds equal prior odds multiplied by the Bayes factor, so $P(H|E) = (BF \cdot \pi) / (BF \cdot \pi + 1 - \pi)$. For n independent or effectively independent anecdotal items with per-item Bayes factor b , the cumulative Bayes factor is b^n . When dependence is present among

witness reports, an effective $b_{\text{eff}} < b$ is used to avoid over-counting, as reflected in the simulations.

Discussion

The Shroud constitutes an unusual example of a persistent, testable artifact capable of multi-lab interrogation (Damon et al., 1989; Adler & Heller, 1980; Jackson et al., 1984). Conversely, most UAP cases occupy the witness-centric end of the spectrum (NASA, 2023; ODNI, 2023), where aggregation of low-quality evidence proceeds slowly and is susceptible to dependence and confounding. The Bayesian framework clarifies why programs prioritizing artifact recovery, chain-of-custody, and independent replication are far more likely to yield decisive evidence than reliance on large but heterogeneous anecdotal corpora. Epistemically, the results operationalize the maxim that extraordinary claims require extraordinary evidence (Sagan, 1995) by making explicit the magnitude of likelihood ratios needed to overcome low priors.

This revision reframes the manuscript as a methodological advancement. It highlights reproducibility, generalizability, and innovation—qualities expected in leading Bayesian journals. Future research should expand beyond statistical modeling to include advanced non-invasive imaging, spectroscopy, and potential DNA analysis of fibers, in addition to refined dependence models of anecdotal corpora. These efforts must be transparent and reproducible, allowing open scientific assessment free of ideological bias. The methodological framework presented here demonstrates how Bayesian reasoning can bridge cultural narratives and empirical evidence, ensuring extraordinary claims are judged by consistent, rigorous standards.

Implications and Applications

The evidentiary taxonomy and Bayesian framework developed here can be applied across other contested domains of science, from medicine and forensics to climate modeling and consciousness research. By quantifying how evidence tiers shift belief probabilities, the model can guide government programs, scientific panels, and interdisciplinary researchers in allocating resources toward artifact recovery and reproducibility. This methodology also provides a roadmap for distinguishing between cultural narratives and empirically defensible knowledge.

Conclusion

Artifact-level evidence dominates rational belief updating under conservative priors. The conclusion is robust across a wide range of plausible parameterizations and aligns with long-standing ideals of severe testing and reproducibility in science (Popper, 1959; Kuhn, 1962). Future work should model dependence among reports explicitly and empirically calibrate Bayes factors via observer-reliability studies and controlled measurement campaigns.

Data Availability

All analytic code and simulated outputs used in this study are available as supplementary files. No human participant data were collected.

Author Note

This manuscript is blinded for peer review. No external funding was received. The author declares no competing interests.

Transparency and Openness Statement

We adhered to APA Journal Article Reporting Standards (JARS) for quantitative research. The analytic code used to generate tables and figures is provided in the supplementary materials. Data and code availability: All model code and generated outputs are available as supplemental files. Preregistration: The study was not preregistered. We encourage replication and independent reanalysis.

References

Adler, A. D., & Heller, J. H. (1980). Blood on the Shroud of Turin. *Applied Optics*, 19(16), 2742–2744.

All-domain Anomaly Resolution Office. (2024). *Historical record report (Vol. 1)*. U.S.

Department of Defense.

Casabianca, T., Marinelli, E., Pernagallo, G., & Torrisi, B. (2019). Radiocarbon dating of the Turin Shroud: New evidence from raw data. *Archaeometry*, 61(5), 1223–1231.

Condon, E. U. (1969). *Scientific study of unidentified flying objects*. University of Colorado.

Damon, P. E., Donahue, D. J., Gore, B. H., Jull, A. J. T., et al. (1989). Radiocarbon dating of the Shroud of Turin. *Nature*, 337, 611–615.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3), 193–242.

Gilbert, R., & Gilbert, M. (1980). Ultraviolet–visible reflectance and fluorescence spectra of the Shroud of Turin. *Applied Optics*, 19(12), 1930–1936.

Jackson, J. P., Jumper, E. J., & Ercoline, W. R. (1984). Correlation of image intensity on the Turin Shroud with the 3-D structure of a human body shape. *Applied Optics*, 23(14), 2244–2270.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press.

Jumper, E. J., & Mottern, R. W. (1980). Scientific investigation of the Shroud of Turin. *Applied Optics*, 19(12), 1909–1912.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*

Association, 90(430), 773–795.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

NASA. (2023). *Unidentified Anomalous Phenomena Independent Study Team Final Report*.

Office of the Director of National Intelligence. (2021). *Preliminary assessment: Unidentified Aerial Phenomena*.

Office of the Director of National Intelligence. (2023). *2022 annual report on Unidentified Aerial Phenomena*.

Pellicori, S. F. (1980). Spectral properties of the Shroud of Turin. *Applied Optics*, 19(12), 1913–1920.

Popper, K. (1959). *The logic of scientific discovery*. Hutchinson.

Rogers, R. N. (2005). Studies on the radiocarbon sample from the Shroud of Turin.

Thermochimica Acta, 425(1–2), 189–194.

Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. Chapman & Hall/CRC.

Sagan, C. (1995). *The demon-haunted world: Science as a candle in the dark*. Random House.