

# Prior Knowledge and Base Rates: Necessary Constraint or Hidden Bias

Juan Segurola

3Dresyns by Resyner Technologies S.L.

© 2026 Juan Segurola. All Rights Reserved.

Correspondence: info@3Dresyns.com · www.3Dresyns.com

INF-006 · Scientific Inference and Experimental Reasoning · Version v0.2

## Abstract

Scientific inference never begins from a completely empty background. Existing theory, previous experiments, prevalence, mechanism, calibration data and known operating limits influence which hypotheses are considered plausible and how new evidence is interpreted. Prior knowledge can therefore improve inference by constraining implausible explanations and stabilising estimates, but it can also hide bias when prior assumptions are implicit, excessively concentrated, selected after seeing the data or protected from model checking. This article distinguishes base rates, prior distributions, background knowledge and theory constraints. It treats prior information as an auditable component of inference rather than as an embarrassment to be denied or a licence to encode preference. The article also separates base-rate information from likelihood-based evidential weight, preserving the INF-005 boundary. Its central requirement is transparency: prior constraints should be stated, justified, varied where defensible and exposed to conflict with the data.

Keywords: prior knowledge; base rate; prior distribution; Bayesian inference; background theory; model checking; bias

## 1. No inference is context-free

The choice of measurement, model, hypothesis set and plausible parameter range already incorporates background knowledge. Bayesian methods make one part of this background explicit through prior distributions, but frequentist models also contain structural assumptions and constraints. Greenland argues that prior assumptions can be made more transparent by expressing their information content rather than pretending standard analyses are assumption-free [1].

## 2. Base rates and prior distributions are related but not identical

A base rate is a frequency or prevalence relevant to the target population or process. A prior distribution is a mathematical representation of information or constraint about parameters or hypotheses before the current data are incorporated. Base rates can inform priors, but a prior can also encode physical bounds, previous experiments, hierarchical structure or regularisation. INF therefore avoids treating “prior” as a synonym for subjective opinion.

## 3. Why base rates matter

Evidence about a diagnostic signal or rare event can be interpreted very differently when the event is common or rare. Neglect of base rates has long been documented as a source of human judgment error [2]. The scientific lesson is not that every analysis must be Bayesian; it is that prevalence and prior plausibility can be relevant to the meaning of a result and should not be silently replaced by an implicit 50:50 starting point.

## 4. Prior information can be useful

- Physical constraints can exclude impossible parameter regions.
- Historical calibration or previous experiments can stabilise inference when new data are sparse.
- Hierarchical priors can pool information while allowing unit-to-unit variation.
- Plausibility constraints can prevent a model from placing substantial probability on scientifically incoherent states.

## 5. Prior information can also mislead

- A prior chosen after inspecting the data can disguise double use of evidence.
- An excessively concentrated prior can dominate weak data and create artificial certainty.
- A prior imported from a different population or measurement process can encode invalid transport.
- A theory-driven prior can protect a preferred model if model checking is not allowed to challenge the structure.

Gelman and Shalizi emphasise that model checking and model revision remain essential even in Bayesian practice [3]. A coherent posterior within a misspecified model is still conditional on that model.

## 6. Base rates versus evidential strength

INF-005 owns the strength of the present data in discriminating among hypotheses. INF-006 owns the background information that determines the starting plausibility or constraint. Bayes' theorem combines these components, but keeping them conceptually distinct makes it possible to ask whether a conclusion is driven mainly by the new data or by the prior structure. Kass and Raftery review Bayes factors as one way to separate model evidence from prior model odds [4,5].

## 7. Prior-data conflict

A prior should not be immune to contradiction. Large disagreement between prior expectations and observed data can indicate an unusual but real event, model misspecification, measurement failure or an inappropriate prior. The response should therefore be diagnostic rather than automatic: inspect the data-generating process, check model adequacy and perform sensitivity analysis across defensible prior choices.

## 8. Minimum reporting discipline

- Identify every material prior or background constraint.
- State its empirical, theoretical or physical basis.
- Distinguish population base rates from parameter priors.
- Report whether the conclusion changes under defensible alternative priors.
- Do not use a prior to encode information already counted in the likelihood unless the duplication is intentional and justified.
- Treat prior-data conflict as information requiring investigation, not as a reason to suppress one side.

## References

- [1] Greenland, S. Bayesian Perspectives for Epidemiological Research: I. Foundations and Basic Methods. *International Journal of Epidemiology* 35(3), 765-775 (2006). DOI: 10.1093/ije/dyi312.
- [2] Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* 185(4157), 1124-1131 (1974). DOI: 10.1126/science.185.4157.1124.
- [3] Gelman, A.; Shalizi, C.R. Philosophy and the Practice of Bayesian Statistics. *British Journal of Mathematical and Statistical Psychology* 66(1), 8-38 (2013). DOI: 10.1111/j.2044-8317.2011.02037.x.
- [4] Kass, R.E.; Raftery, A.E. Bayes Factors. *Journal of the American Statistical Association* 90(430), 773-795 (1995). DOI: 10.1080/01621459.1995.10476572.
- [5] Goodman, S.N. Toward Evidence-Based Medical Statistics. 2: The Bayes Factor. *Annals of Internal Medicine* 130(12), 1005-1013 (1999). DOI: 10.7326/0003-4819-130-12-199906150-00019.

## Data and materials statement

No new experimental dataset was generated for this article. The article is a methodological analysis based on the cited literature and the frozen ownership architecture of the INF series.