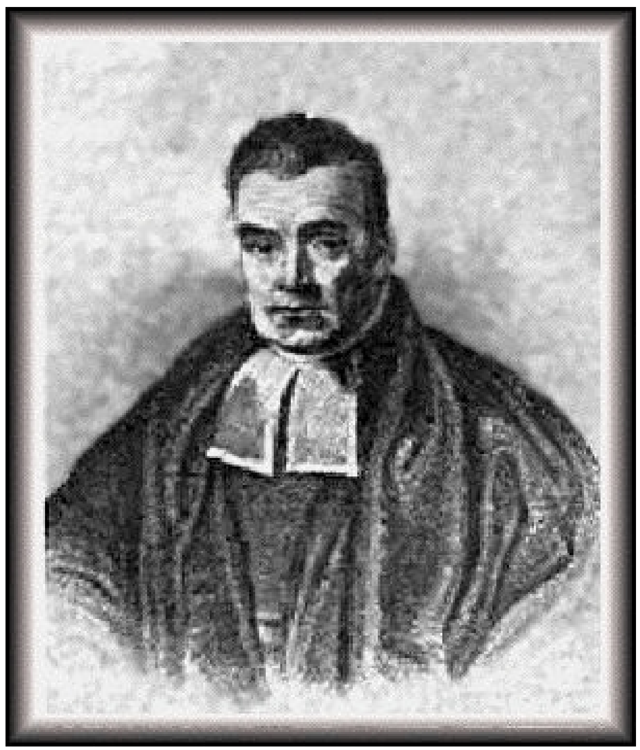
Why Health Economics Needs Bayesianism

Dale J. Poirier

University of California, Irvine



I am ***not*** a health economist, but ...

- "A Switching Simultaneous Equations Model of Physician Behaviour in Ontario," 1981, in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press), 392-421.

- "The Roles of Birth Inputs and Outputs in Predicting Health, Behavior, and Test Scores in Early Childhood" (with Kai Li), 2003, *Statistics in Medicine* 22, 3489-3514.

Comments

- Both health economics papers involve large simultaneous equation models involving 26 and 12 equations and hundreds of parameters.
    - The size reflects the large number of endogenous variables present in many health economics studies.
- Poirier (1981) relies on asymptotics in terms of ***sample size***.
    - In retrospect, would be easier to analyze from a Bayesian perspective.
- Li and Poirier (2003) provides exact finite sample posterior results based on asymptotics in the ***number of replications***.
    - Performing a compelling sensitivity analysis was challenging.

However, I am a Bayesian. Here are some of my introductory expositions:

"Exchangeability, Representation Theorems, and Subjectivity," 2009, in: Geweke, J., G. Koop, and H. Van Dijk, eds., *Handbook of Bayesian Econometrics* (Oxford), forthcoming.

"Bayesian Econometrics," in S. Derlauf and L. Blume, eds., *The New Palgrave Dictionary of Economics* (London: Palgrave Macmillan, second edition), forthcoming.

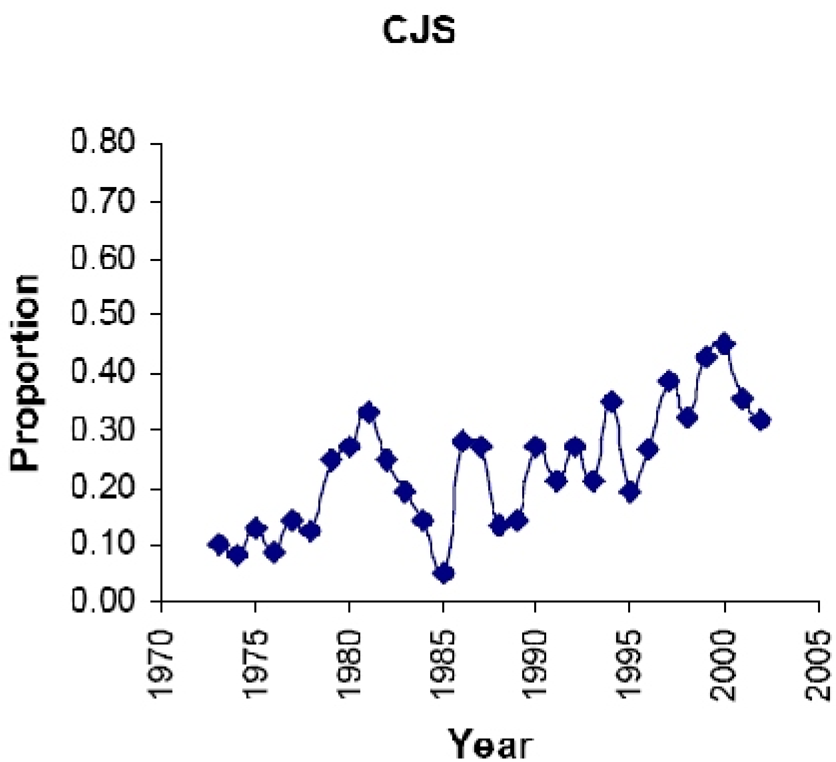
Bayesian Econometric Methods (Gary Koop and Justin Tobias), 2007, in K. Abadir, J. Magnus, and P. C. B. Phillips, eds., *Econometrics Exercises Series*, Vol. 7 (Cambridge).

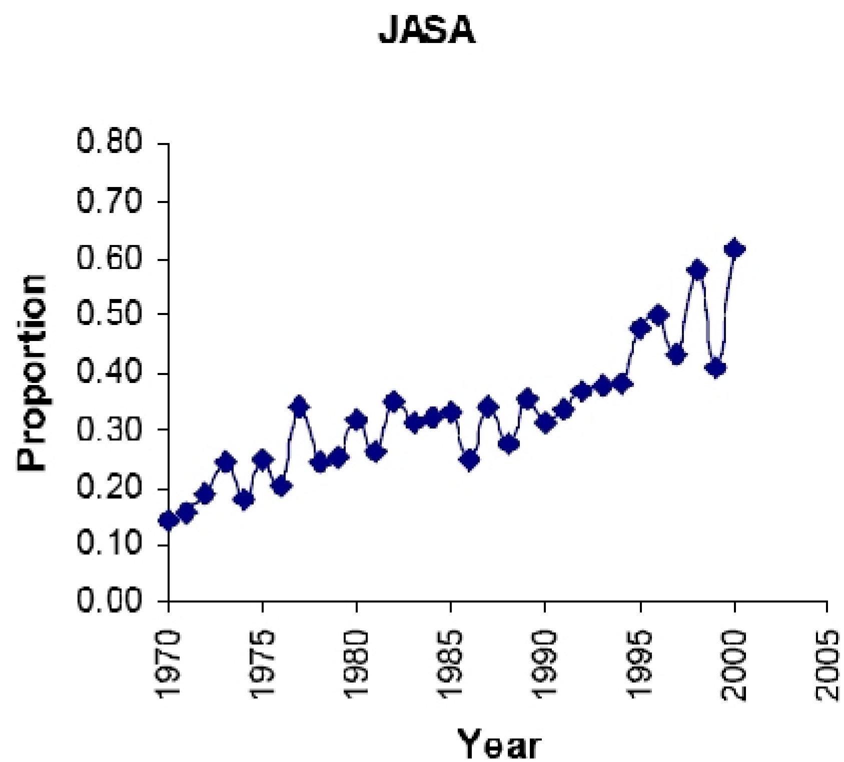
"Bayesian Econometrics," (Justin Tobias), 2006, in K. Patterson and T. C. Mills, eds., *Palgrave Handbooks of Econometrics*, Vol. 1, *Econometric Theory* (Palgrave Macmillan).

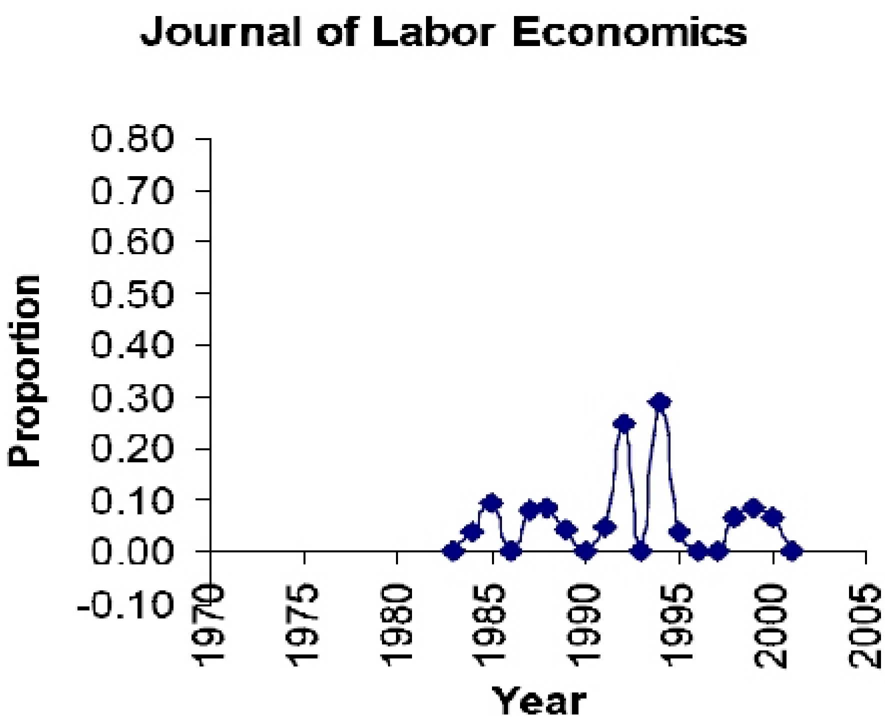
Intermediate Statistics and Econometrics: A Comparative Approach, 1995, (MIT Press).

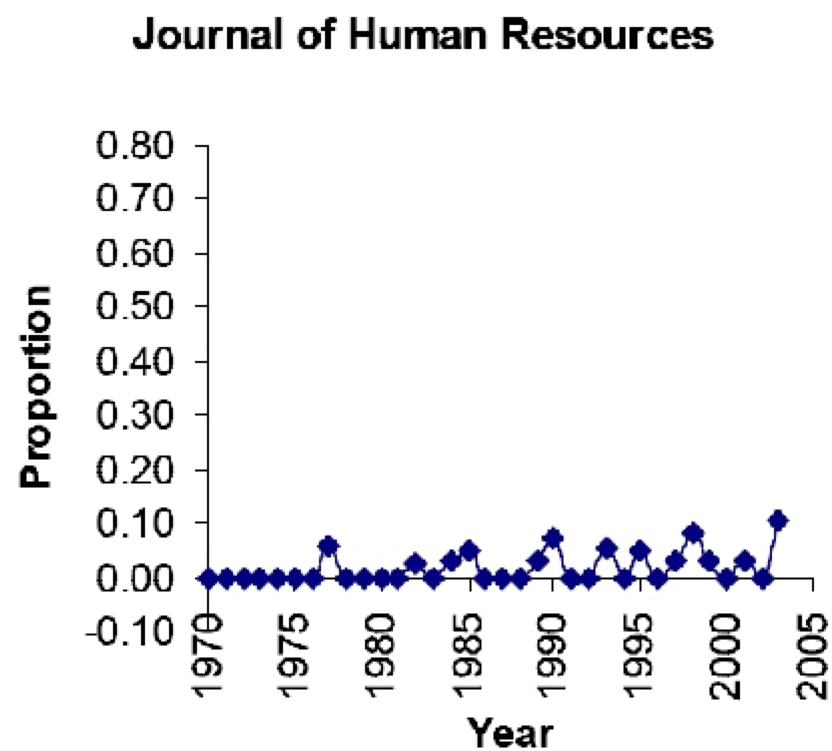
"Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economics" (with discussion), 1988, *Journal of Economic Perspectives* 2, 121-170.

What is the Current Impact of Bayesianism? [Poirier(2006, *BA*)]

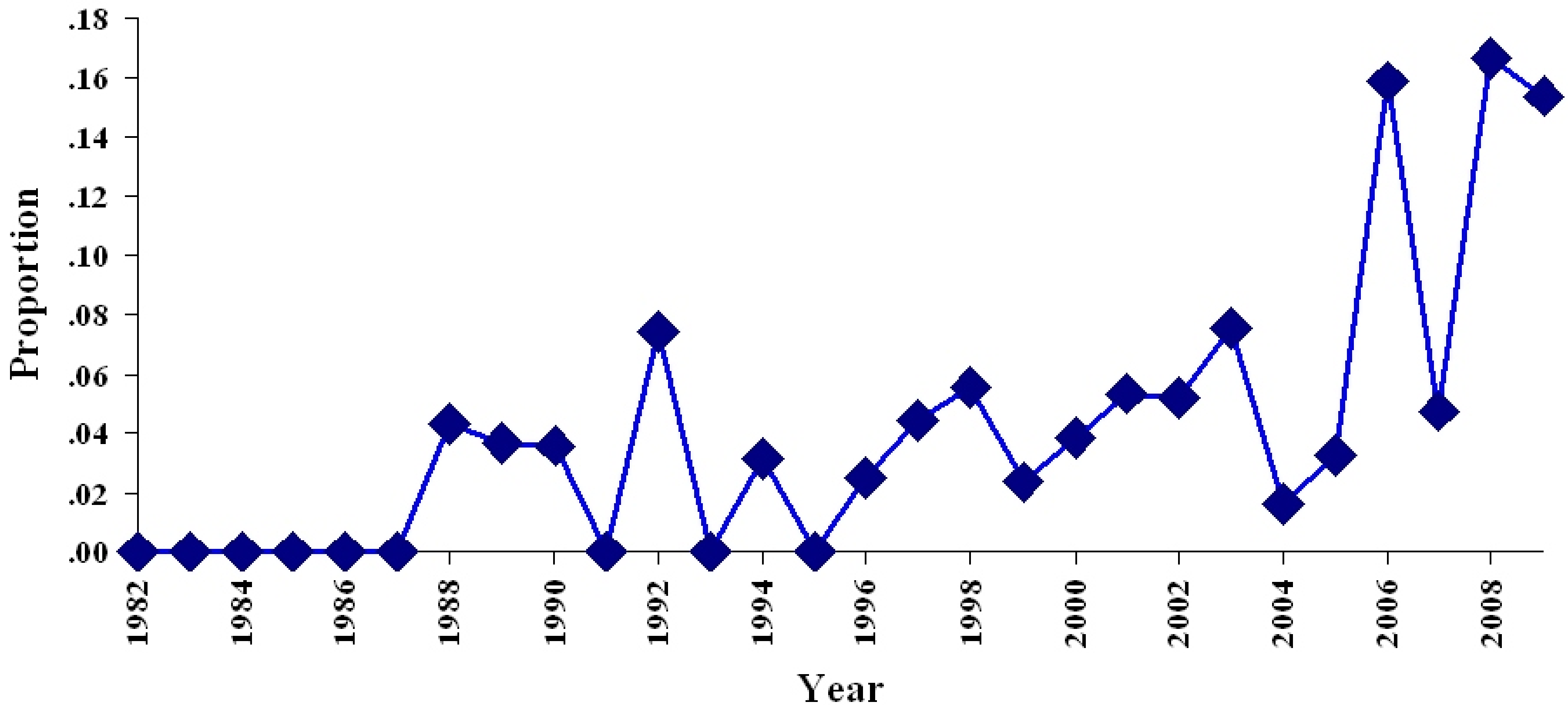





Journal of Labor Economics
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
Proportion
1970
1975
1980
1985
1990
1995
2000
2005
Year

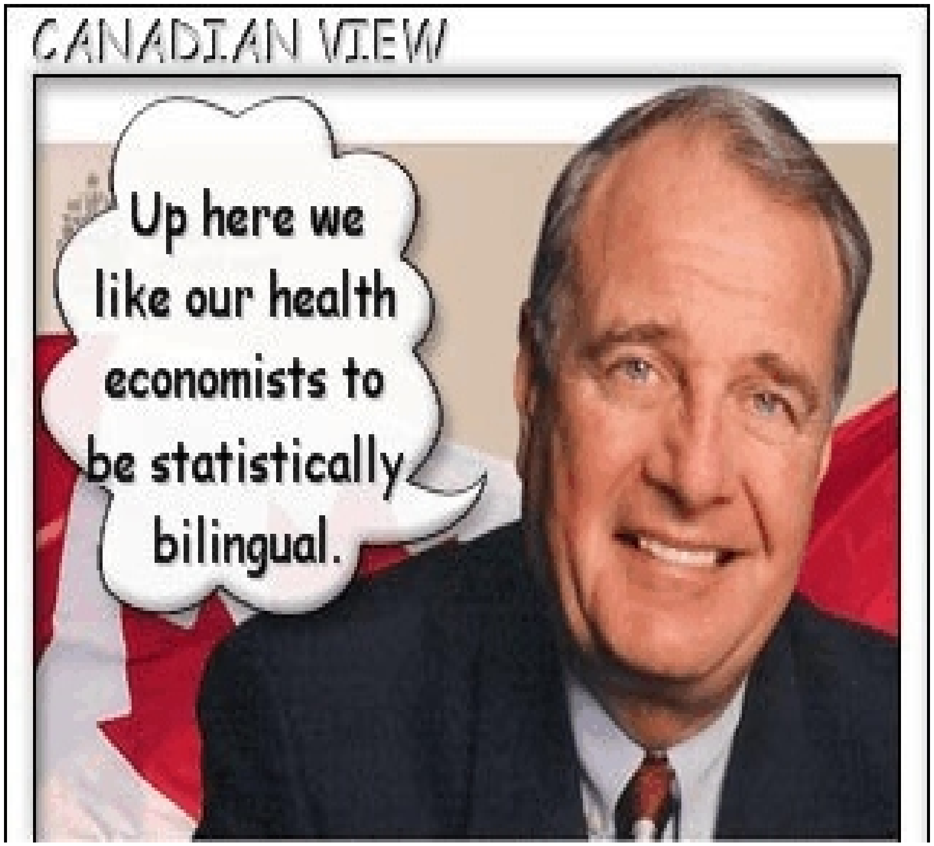

Journal of Human Resources
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
-0.10
Proportion
1970
1975
1980
1985
1990
1995
2000
2005
Year

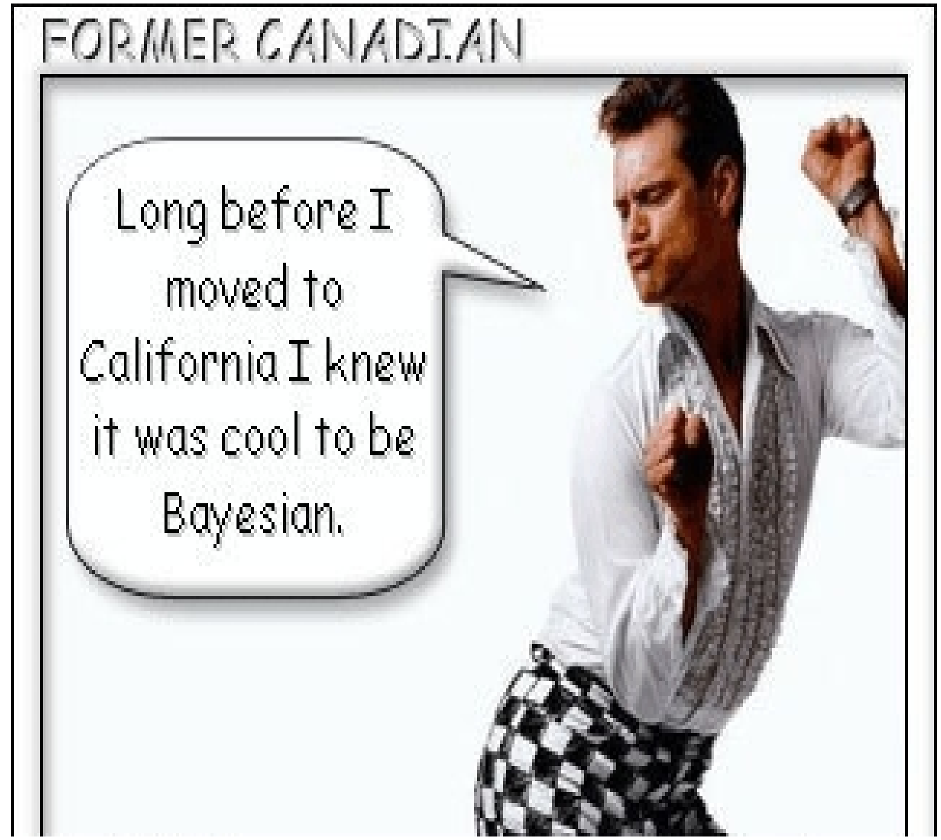
Bayesian Content in *JHE*



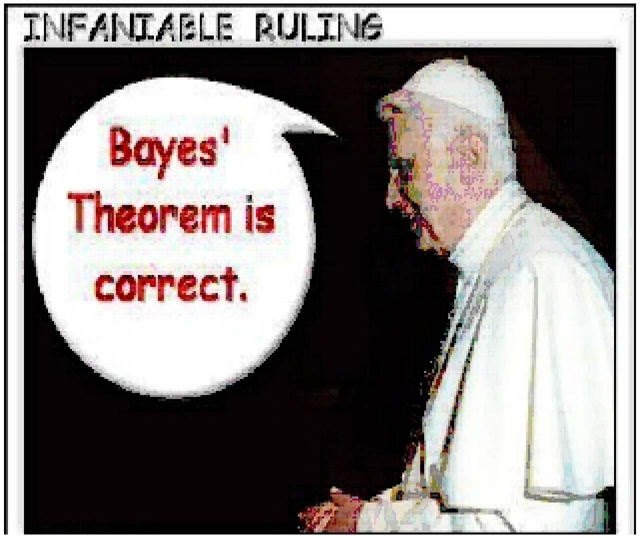
So, ***why*** does health economics need Bayesianism?

- Is it something special about health economics? Probably not.
  - Coherent
  - Evidential, i.e., conditions on what we observe
  - Exact finite sample inference
- Am I the only proponent? Hardly, see http://www.shef.ac.uk/chebs/.
- There have been some Canadians ...


CANADIAN VIEW
Up here we like our health economists to be statistically bilingual.


FORMER CANADIAN
Long before I moved to California I knew it was cool to be Bayesian.

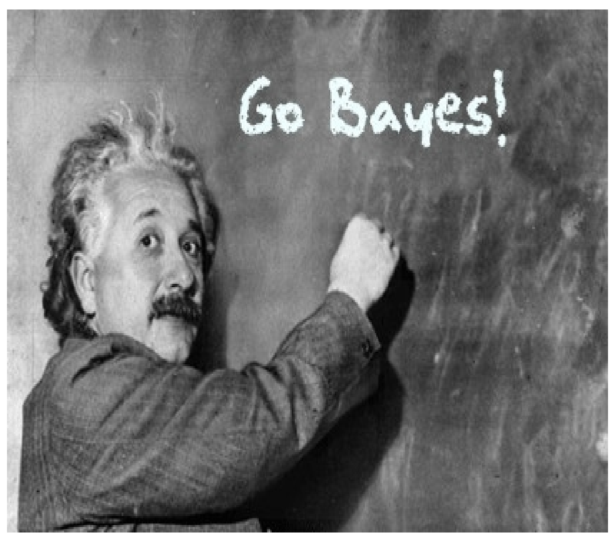
As well as others ...

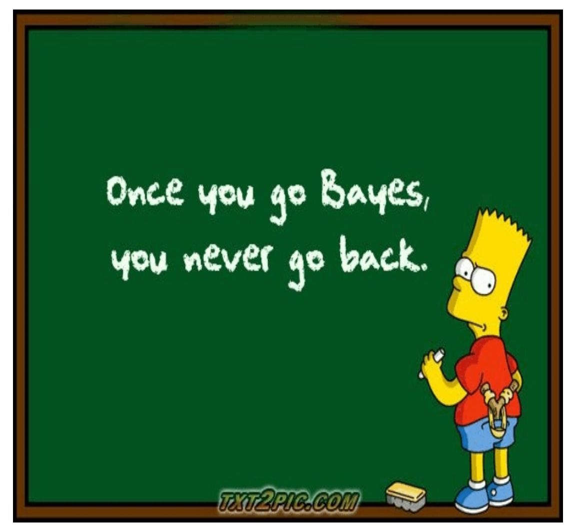

INFALIABLE RULING
Bayes' Theorem is correct.


Go Bayes!


MAO REFLECTS
Hmm. Maybe I should have updated my prior in light of the data?


I have gone Bayes. Have you?
Microsoft


Go Bayes!


Once you go Bayes, you never go back.
TXT2PIC.COM

I don’t want to make this a frequentist-bashing talk, but ...

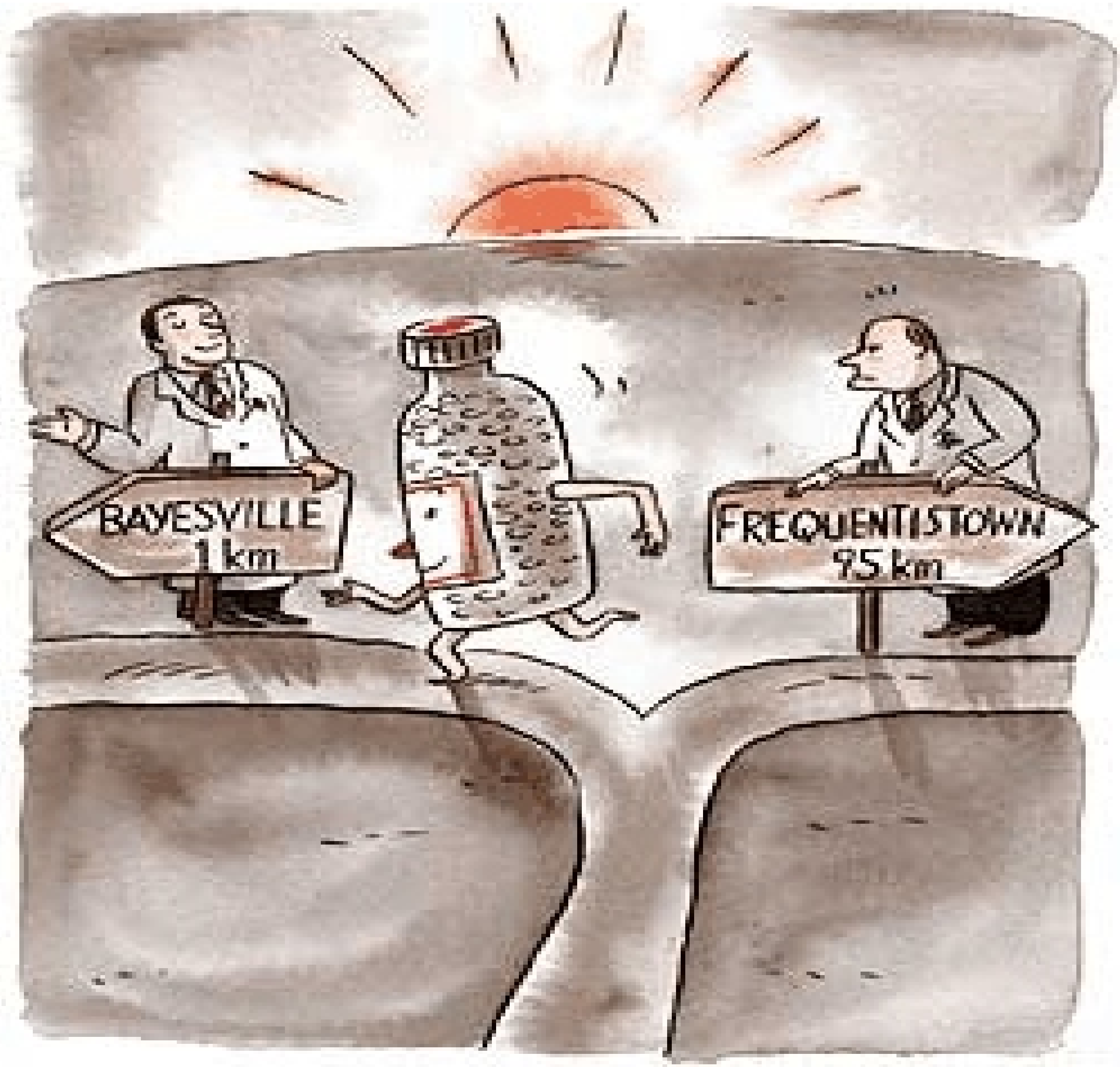

I wish I could go hunting with a frequentist.

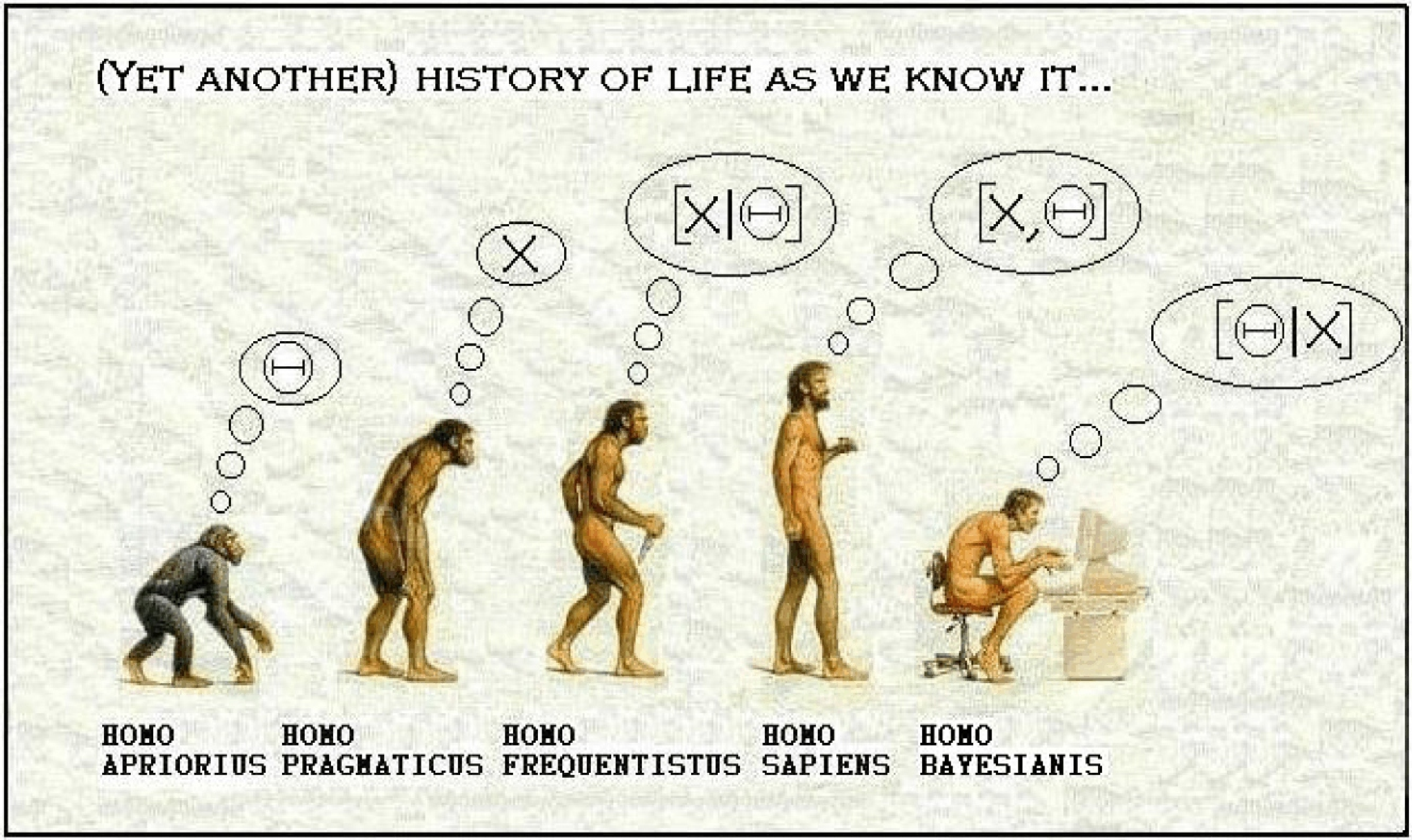
Computation

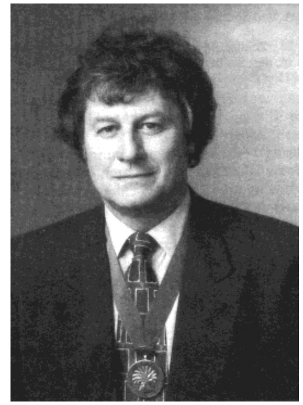
In the 1980s I often heard: "Dale, I agree with you in theory. I want to be a Bayesian. But the computations are too hard in practice."

- Well, the MCMC revolution is now twenty years old. We can simulate draws from most any posterior distribution. Then it is mostly just descriptive statistics.

- This is no longer a credible excuse.


BAYESVILLE
1 km
FREQUENTISTOWN
95 km


(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
[Θ]
[X]
[X|Θ]
[X,Θ]
[Θ|X]
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS



In their comprehensive treatise, Bernardo and Smith (1994, *Bayesian Theory*) offer the following summarization of ***Bayesian statistics***:

> "Bayesian Statistics offers a rationalist theory of personalistic beliefs in contexts of uncertainty, with the central aim of characterizing how an individual ***should*** act in order to avoid certain kinds of undesirable behavioral inconsistencies."

Bayesianism is the active promotion of Bayesian statistics.

- The framework establishes ***expected loss minimization*** as the basis for rational decision making and ***Bayes' Theorem*** as the key to the way beliefs should fit together in the light of changing evidence.

- The goal is to establish rules for ***disciplined uncertainty accounting***.

  - The theory is not ***descriptive***.

  - The theory is ***prescriptive***: if you wish to avoid the possibility of undesirable consequences you must act in the following way.

Components of Bayesian Analysis

- **Prior:** $f(\theta)$

- **Likelihood (window):** $\mathscr{L}(\theta; z)$

- **Bayes Theorem:** $f(\theta|z) \propto f(\theta)\,\mathscr{L}(\theta; z)$

- **Loss function:** $C(\hat{\theta}, \theta)$

- **Commandment:** Thou shalt ***minimize expected posterior loss.*** In point estimation:

$$\min_{\hat{\theta}} E_{\theta|z}[C(\hat{\theta}, \theta)]$$

- **Prediction of future z***:

$$f(z^*|z) = E_{\theta|z}[f(z^*|z, \theta)] = \int_{\Omega} f(z^*|z, \theta)\, f(\theta|z)\, d\theta$$

Predictive loss function: $C(\hat{z}^*, z^*)$

- **Sensitivity analysis with respect to prior, loss function, and likelihood.**

Bruno de Finetti 1906–1985

- Bruno de Finetti was famous for the aphorism:

"Probability does not exist."

- ◦ ***Subjectivist*** interpretation: Probability reflects an individual's beliefs about reality, rather than a property of reality itself.

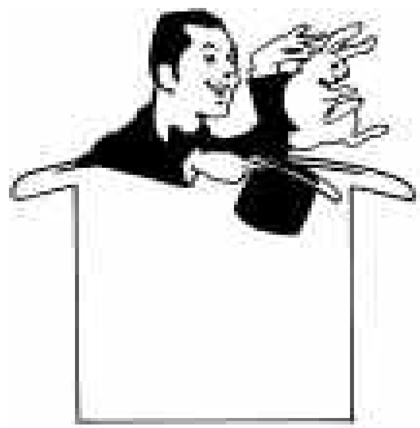
- ◦ This interpretation corresponds to the everyday use of "probability."

- The only universally required restriction that a probability cdf P(·) must satisfy is ***coherence***, i.e., use of P(·) avoids being made a sure loser (***Dutch Book***) regardless of the outcomes.

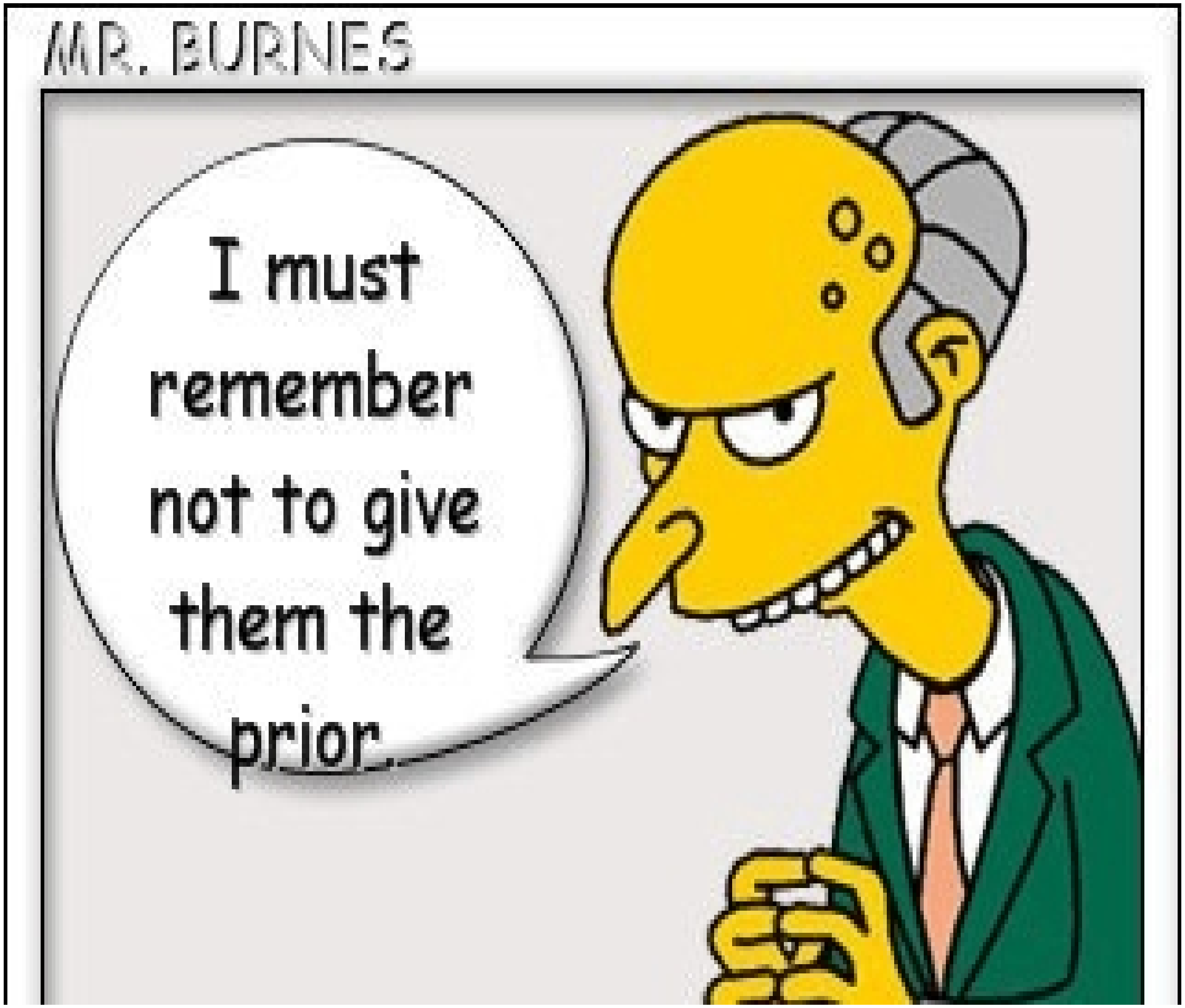
  - ◦ This implies that P(·) obeys the usual axioms of probability (at least up to finite additivity).

- For the subjectivist, individuals “know” ***their own*** beliefs.

    - Subjectivist theory takes such knowledge as a ***primitive assumption***, the same way the rational expectations framework assumes agents know the “true model.”



    - Whether these beliefs are ***well calibrated*** (i.e., in empirical agreement with reality) or ***easily articulated*** is a different issue.


MR. BURNES
I must remember not to give them the prior.

- de Finetti assigned a fundamental role to the concept of ***exchangeability***.
  - de Finetti was not the first to study exchangeability, but he popularized the subject.
  - Cifarelli and Regazzini (1996) report that it dates to a communication by Jules Haag at the International Mathematical Congress held in ***Toronto*** (August 1924) and published in 1928.

- Given a finite sequence $\{Z_n\}_{n=1}^{n=N}$ of observable random quantities taking values in a sample space $\mathcal{Z}$, suppose the ***subscripts are uninformative*** in the sense that

$$P(z_1, z_2, ..., z_N) = P(z_{\pi(1)}, z_{\pi(2)}, ..., z_{\pi(N)}),$$

where $\pi(n)$ ($n = 1, 2, ..., N$) is a permutation of $\{1, 2, ..., N\}$.

  - Such beliefs are defined to be ***exchangeable***.
  - **Example:** In the case of N = 3 exchangeable Bernoulli trials, the sequences 011, 101, and 110 are assigned the same probability.
  - Beliefs are ***infinitely exchangeable*** iff every finite subsequence is exchangeable.

- Like random (iid) sequences, the quantities in an exchangeable sequence are identically distributed.

  - Unlike iid sequences, such quantities need ***not*** be independent.

  - The dependency is what enables the researcher to learn from experience.

- Random sampling is the foundation of frequentist statistics; exchangeability is the foundation for Bayesian statistics.

- Exchangeability is a form of ***symmetry***.

  - It provides an operational meaning to the weakest possible notion of a sequence of "similar" random quantities.

  - It is "operational" in that it only requires probability assignments for ***observable*** quantities, albeit arbitrarily long sequences.

- Adding restrictions (e.g., exchangeability) on P(·) beyond coherence leads to ***de Finetti's Representation Theorem*** and its generalizations.

  - Such restrictions should ***not*** be thought of as "true" or "false."

  - They are ***not*** properties of reality; they are restrictions on beliefs about reality.

    - Others may or may not find such restrictions compelling.

    - One goal of empirical work is to articulate restrictions that other researchers are willing to entertain if not outright adopt, i.e., to obtain ***inter-subjective*** agreement among a bevy of Bayesians.

- Kreps (1988) opines that de Finetti's Theorem is the ***fundamental theorem of statistical inference***.

Theorem 1 (de Finetti's Representation Theorem): Let $\{Z_n\}_{n=1}^{n=\infty}$ be an infinitely exchangeable sequence of Bernoulli quantities with joint cdf P(·). Define $\bar{Z}_N = (Z_1 + Z_2 + \ldots + Z_N)/N$, and let $z = [z_1, z_2, \ldots, z_N]'$ denote realized values. Then there ***exists*** a cdf F(·) such that

$$p(z) = \int_{\Omega} \mathcal{L}(\theta; z)\, dF(\theta), \qquad \theta \in \Omega = [0, 1],$$

where

$$\mathcal{L}(\theta; z) = \prod_{n=1}^{N} \theta^{z_n}(1 - \theta)^{1 - z_n}, \qquad \theta \in \Omega,$$

$\Theta = \underset{N\to\infty}{\text{Limit}}\ \bar{Z}_N$ P-almost surely, and F(·) is the cdf of Θ under P(·), i.e.,

$$F(\theta) \equiv \underset{N\to\infty}{\text{Limit}}\ P(\bar{Z}_N \leq \theta). \ ■$$

- It is ***as if***, given $\Theta = \theta$, $\{Z_n\}_{n=1}^{n=\infty}$ are iid Bernoulli trials, where the probability Θ of a success is assigned a prior cdf $F(\theta)$ that can be interpreted as the researcher's beliefs about the long-run relative frequency of $\bar{Z}_N \leq \theta$ as $N \to \infty$.

- ***From de Finetti's standpoint,*** Θ, $\mathcal{L}(\theta; z)$, ***and independence are "mathematical fictions" implicit in the researcher's subjective assessment of arbitrarily long sequences of observable successes and failures.***

- This theorem and its generalizations:
    - provide connections between Bayesian and frequentist reasoning,
    - endogenize the choice of likelihood functions,
    - prove the ***existence*** of priors,
    - provide a different interpretation of parameters:
        - a convenient index for a distribution,
        - "lubricants" for fruitful thinking and communication
        - induce ***conditional independence*** for observables,
    - produce Bayes' Theorem as a corollary,
    - produce the LP and the SRP as corollaries, and
    - provide a solution to Hume's problem of induction.

- These are a large number of results rarely discussed in econometrics, ***eh***?

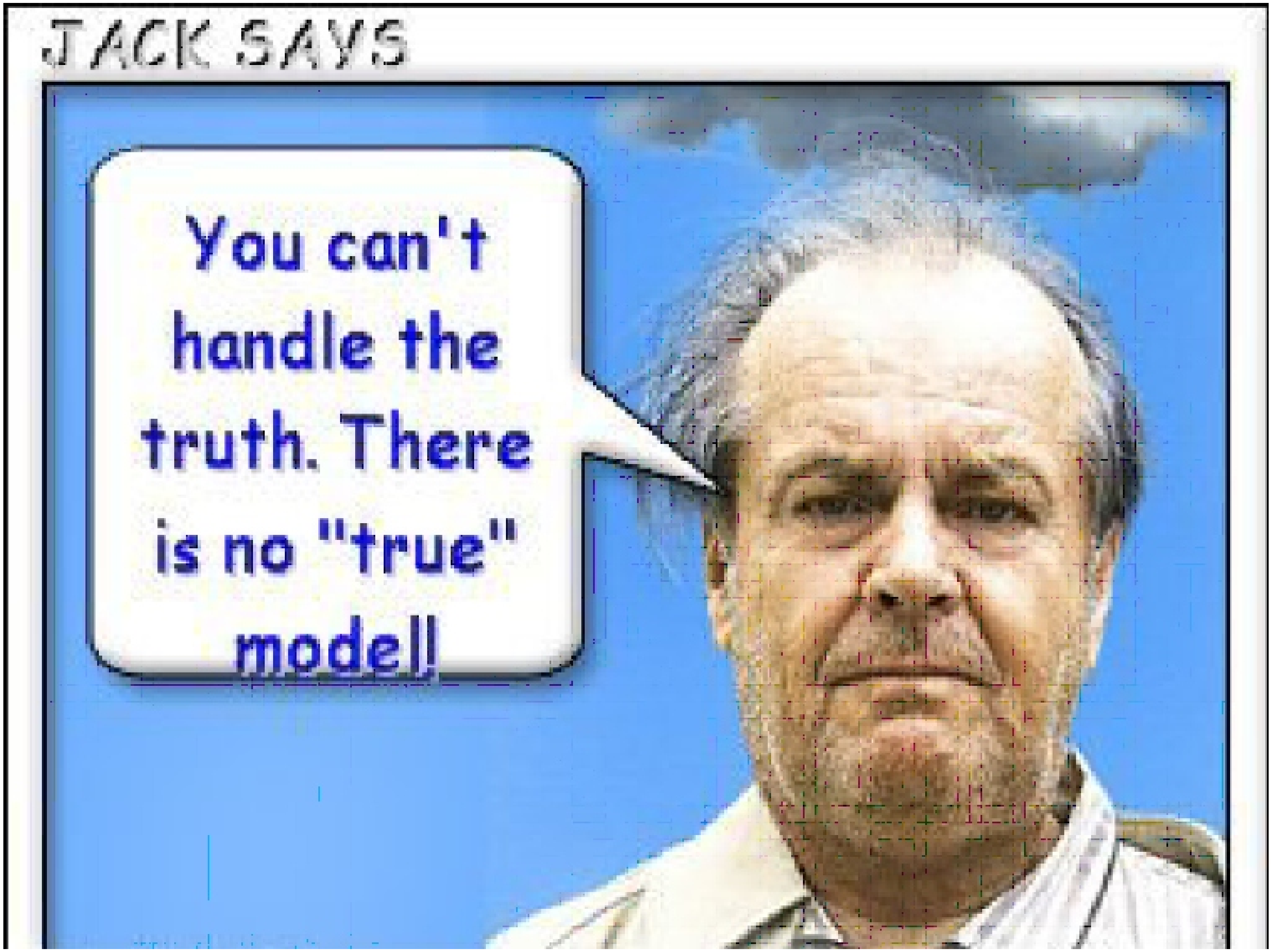
- In cases where $F(\theta)$ is absolutely continuous with pdf $f(\theta)$, then

$$p(z) = \int_{\Omega} \mathcal{L}(\theta; z) f(\theta)\, d\theta.$$

  - Most researchers think in terms of the right-hand side.

    - Non-Bayesians implicitly do so with a degenerate distribution for Θ that treats Θ equal to a constant θ_o with probability one, i.e., a degenerate "prior" distribution for Θ at the "true value" θ_o.

  - ***I advocate an attitude that emerges from the left-hand side of but which can help researchers work on the right-hand side***.


JACK SAYS
You can't handle the truth. There is no "true" model!

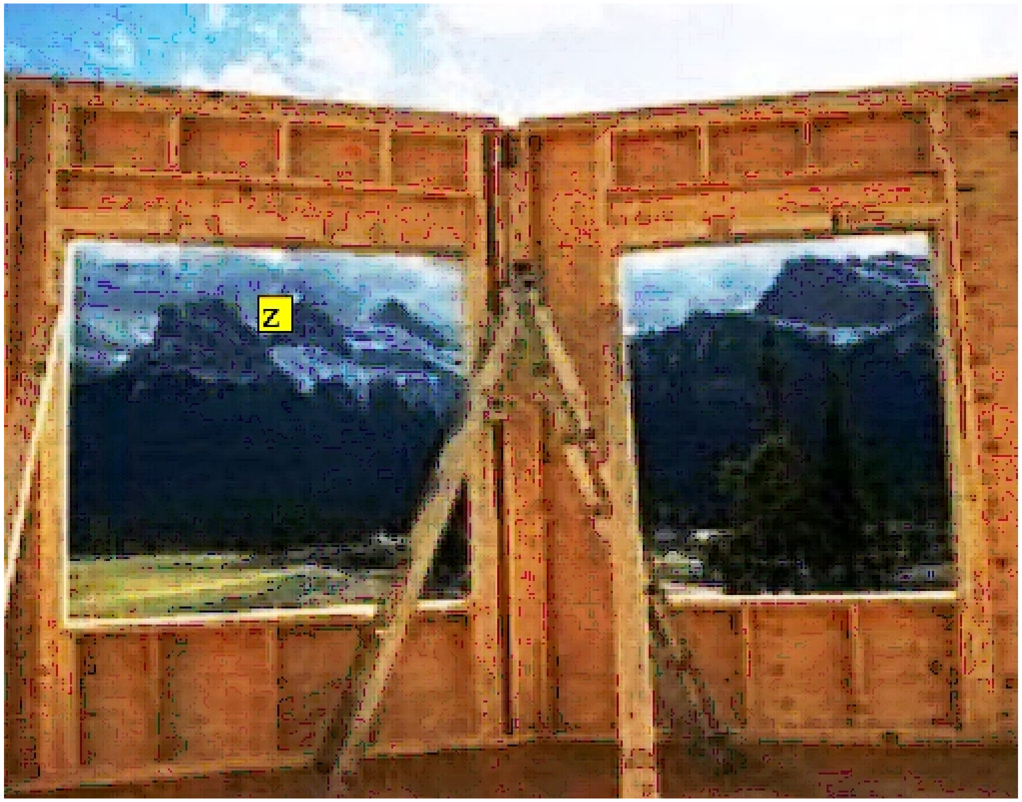
- Putting further restrictions on the observable Bernoulli quantities $\{Z_n\}_{n=1}^{n=\infty}$ beyond infinite exchangeability can pin down the prior $F(\theta)$.

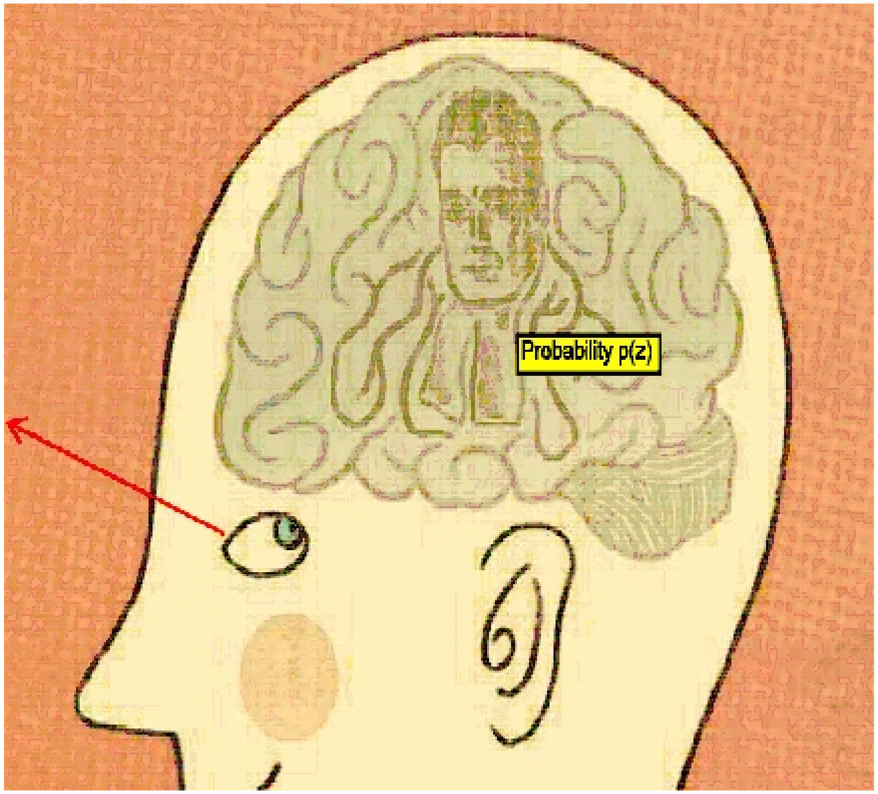
    - The assumption that $\{Z_n\}_{n=1}^{n=\infty}$ correspond to draws from a ***Polya urn process*** implies the prior $F(\cdot)$ belongs to the conjugate beta family [Freedman (1965)].

- Rather than elaborate on other representation theorems, let me summarize my views with a diagram.

model ≡ likelihood & prior (both subjective)

$\mathcal{L}(\theta_1; z, \mu = 1)$ $\mathcal{L}(\theta_2; z, \mu = 2)$





$f(\theta_1|\mu = 1)$ $f(\theta_2|\mu = 2)$

$$p(z) = p(\mu = 1)\int_0^1 \mathcal{L}(\theta_1; z) f(\theta_1|\mu = 1)\, d\theta_1 + p(\mu = 2)\int_0^1 \mathcal{L}(\theta_2; z) f(\theta_2|\mu = 2)\, d\theta_2$$

Pragmatic Principles of Model Building

Poirier (1988, *JEP*)

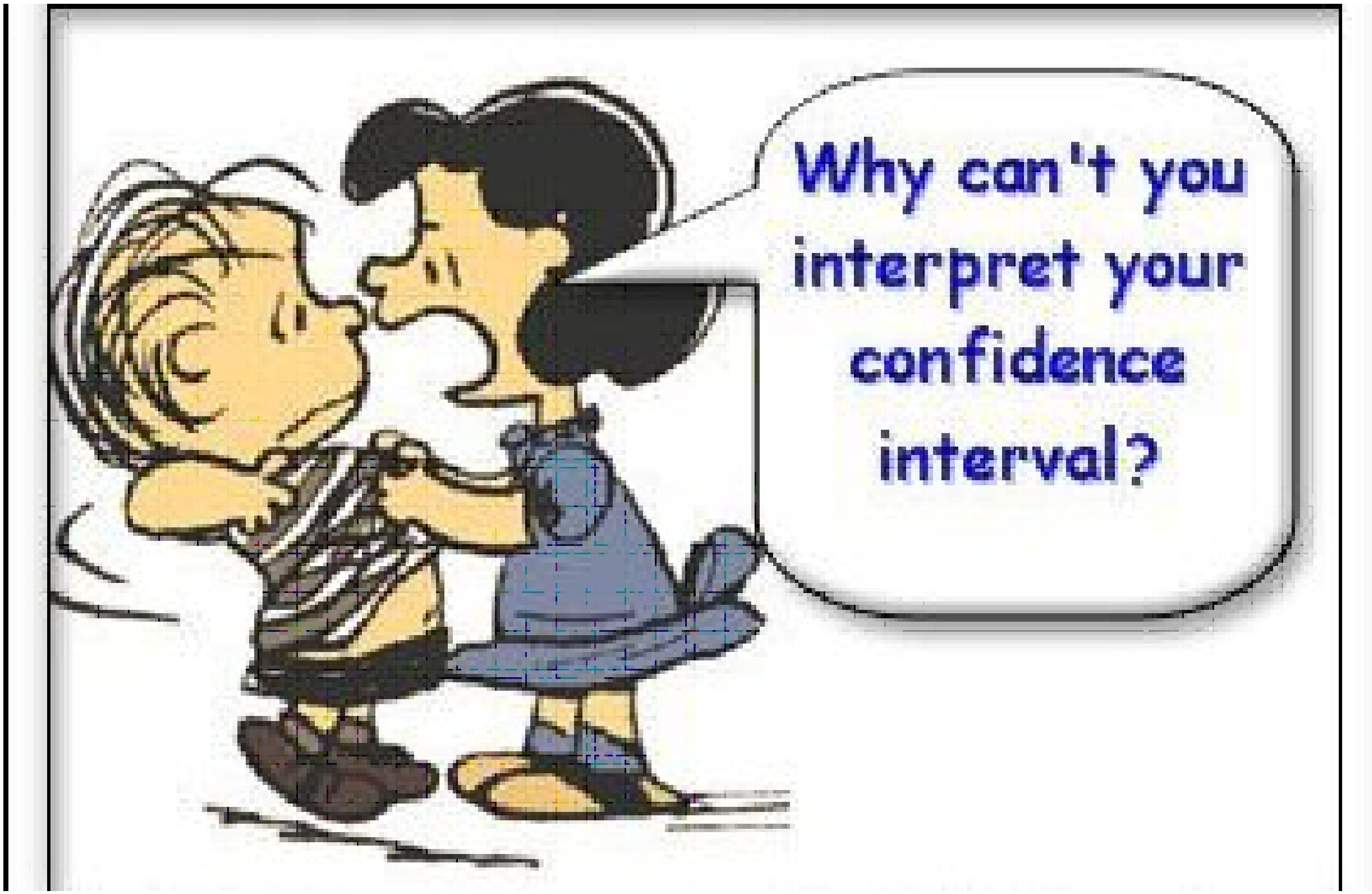
PPMB1 (LP): Given the likelihood, inferences regarding its unknown parameters should be made ***conditional on the observed data*** rather than averaged over all possible data sets.

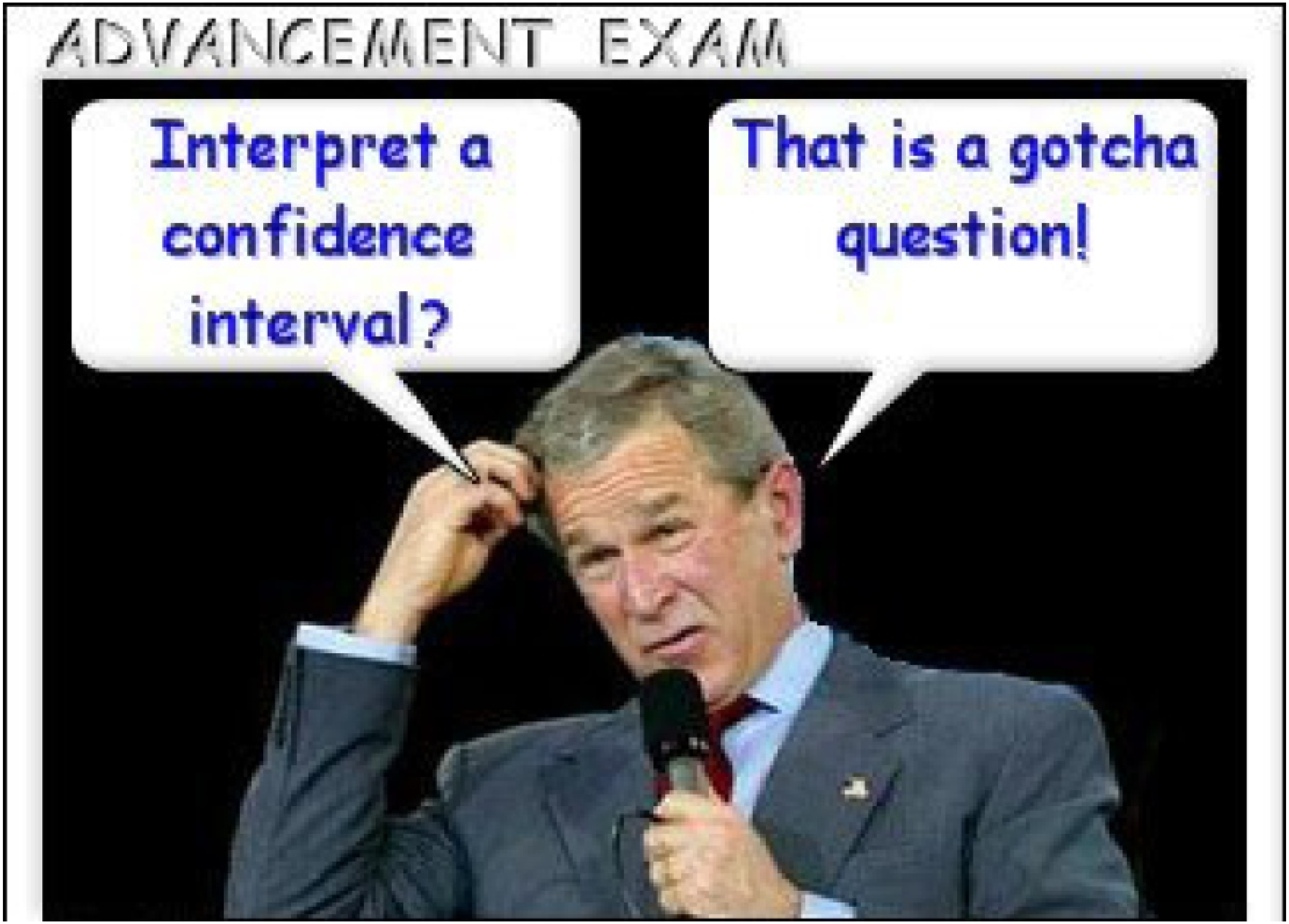
- ***LP***: Two experiments involving the same θ which yield proportional realized likelihoods contain the same evidence regarding θ.

- Classical statistics is ***ex ant***e $(z|\theta)$; Bayesian statistics is ***ex post*** $(\theta|z)$.


FREQUENTIST MEDICAL ADVICE
Sorry. You have been assigned to the control group. But upon repeated sampling ...

- The ***ex ant***e vs. ***ex post*** distinction underlies many exam questions.




ADVANCEMENT EXAM
Interpret a confidence interval?
That is a gotcha question!

- The ***LP*** implies the ***SRP***.

Poirier (1995, Example 6.2.3): In 1991 the *NY Times* reported on a surgical technique designed to clear clogged neck arteries leading to the brain which had been found to be effective in preventing strokes in patients suffering from a severe case of blockage (Robarts Research Institute in London, Ont.).

- Patients with blockage, were randomly split into two groups: 331 patients were treated with aspirin and the anticoagulant warfarin, and 328 patients underwent the surgical technique known as ***carotoid endarterectomy***.
- In the first group 26% of the patients had a subsequent stroke compared to only 9% in the second group.
- The length of the initial experiment was terminated early because preliminary results indicated that "the patients receiving the surgery were doing so well that it would be ***unethical*** to continue to endorse conventional medical therapy."

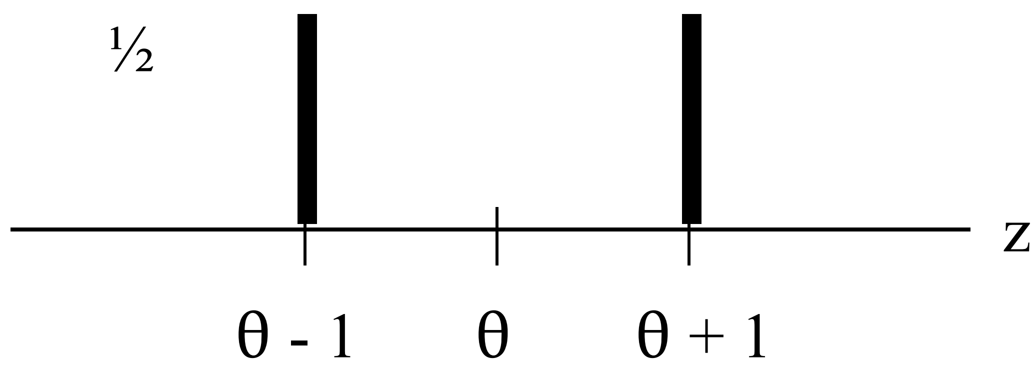
- To many the ***SRP*** (loosely speaking, it is OK to look at the data and deciding when to stop sampling) at first sounds like scientific heresy since it seems to condone sampling to a foregone conclusion.

  ◦ Such possibilities are indeed possible for some statistical hypothesis testing techniques, however, that fact should draw into doubt the techniques involved (classical techniques) rather than the ***SRP*** itself.

- If statisticians could agree on conditioning, there would be far fewer controversies in statistics since it is the basis for most debates (e.g., unit roots).

Poirier (1995, Example 6.10.1) [Berger and Wolpert (1988, pp. 5-6)]:
Given θ ($-\infty < \theta < \infty$), suppose Z_t (t = 1, 2) are iid with pmf

$$P(Z_t = \theta - 1|\theta) = P(Z_t = \theta + 1|\theta) = ½.$$



- A 75% confidence set of smallest size is

$$C(z) = \begin{cases} ½(z_1 + z_2), & \text{if } z_1 \neq z_2 \\ z_1 - 1, & \text{if } z_1 = z_2 \end{cases}.$$

In repeated sampling, $C(Z) = \theta$ with probability of .75.

- Note, however, when observing $y_1 \neq y_2$, it is ***absolutely certain*** that $\theta = \frac{1}{2}(z_1 + z_2)$, and when observing $y_1 = y_2$, it is ***equally uncertain*** whether $\theta = z_1 - 1$ or $\theta = z_2 + 1$ (assuming no prior knowledge about θ).

  - ***Ex post*** the "confidence" in using C(z) is either 100% or 50%, and depending on whether $z_1 \neq z_2$ or $z_1 = z_2$, we know which level it is.

  - From the ***ex ante*** perspective, however, C(z) is 75% confidence set: an average of the two ***ex post*** levels.

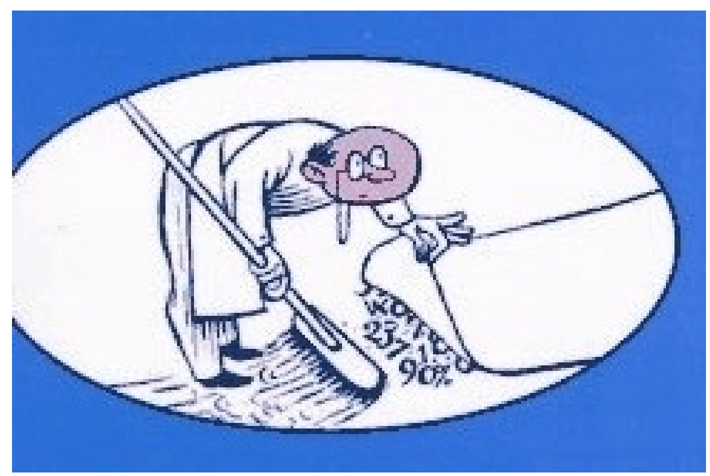
  - The embarrassing question for the pure frequentist is: why use the realized value of C(z) and report the ***ex ante*** confidence level 75% instead of the appropriate ***ex post*** measure of uncertainty?

PPMB2: Subjective prior beliefs have a role to play in scientific research. Bayesian analysis involves ***formal*** consideration of both prior information and loss, and such concepts play a ***central*** role.

- The existence of priors are guaranteed by representation theorems.

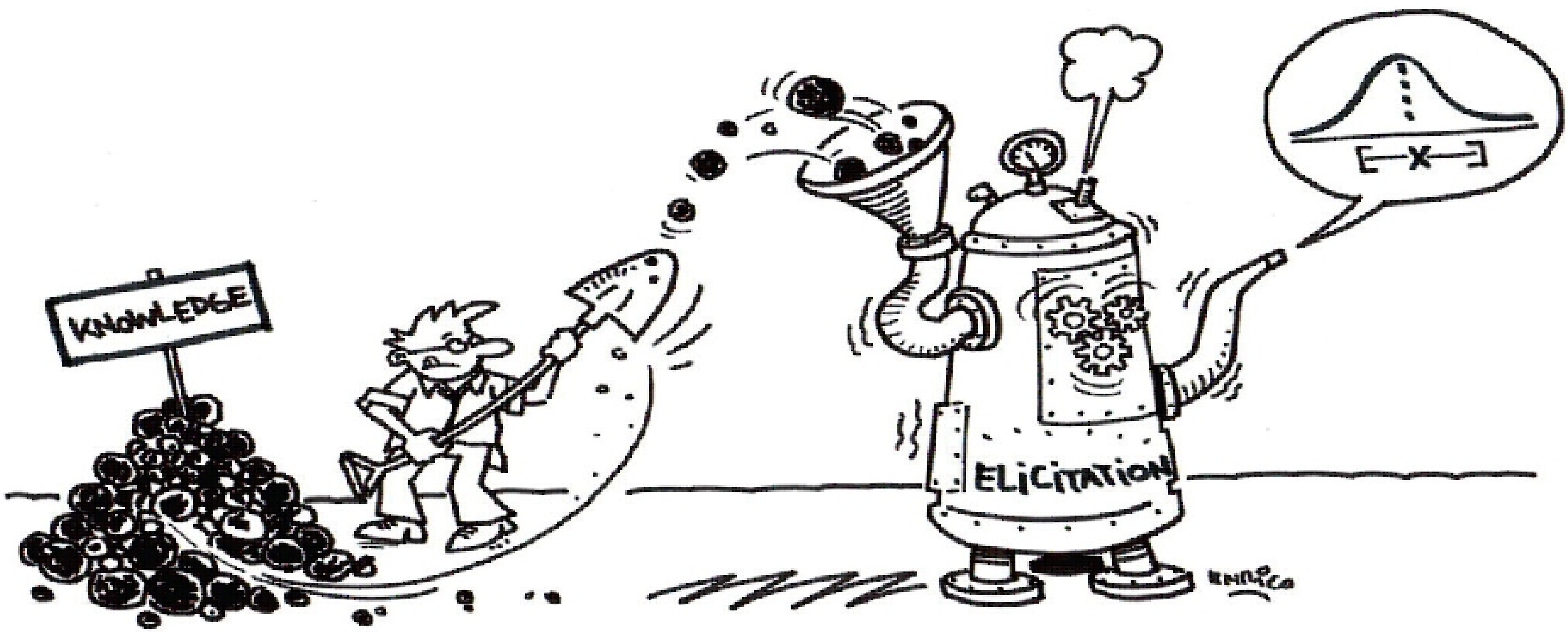
- While many researchers are reluctant to admit they entertain "subjective" non-data information, they also argue that they bring valuable "insight" and "wisdom" to data analysis in their field of expertise.

    - Formal introduction of subjective information is more ***intellectually honest*** than traditional ways of hiding it from the reader.



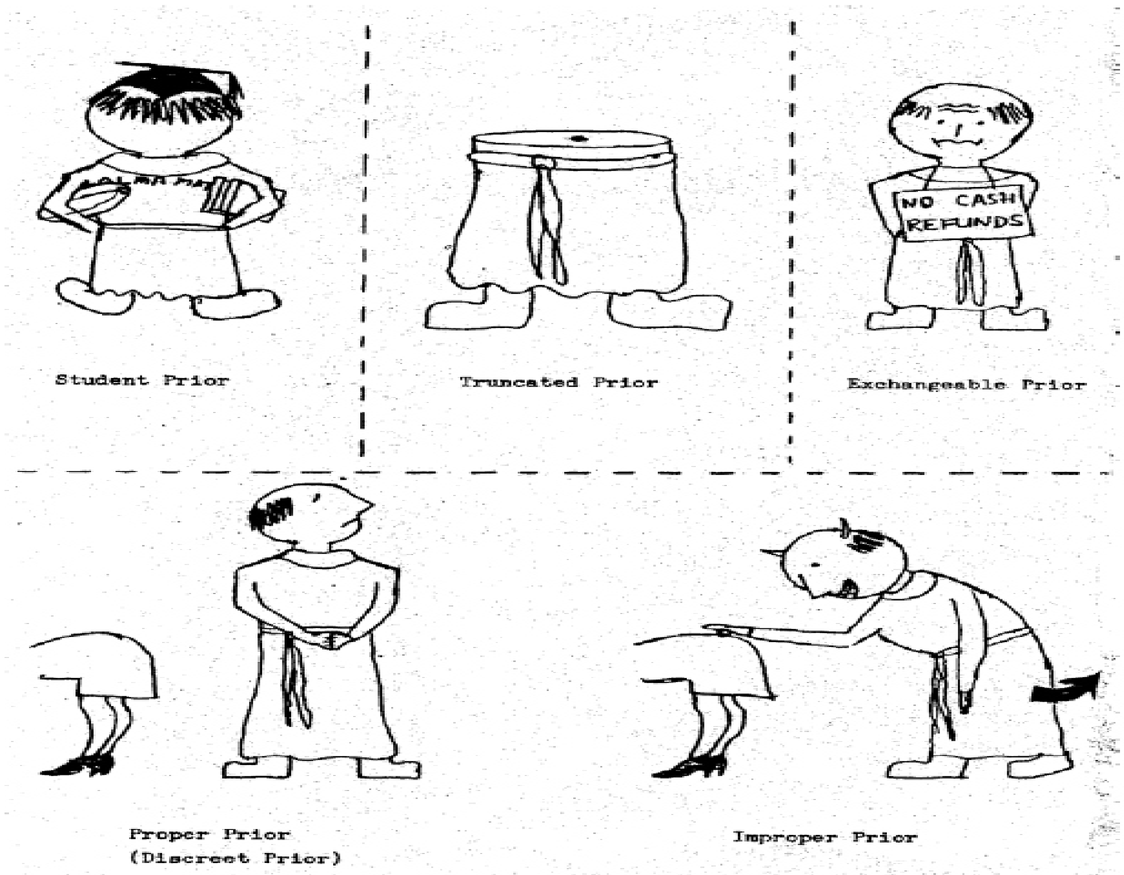
- But how do we articulate these priors?

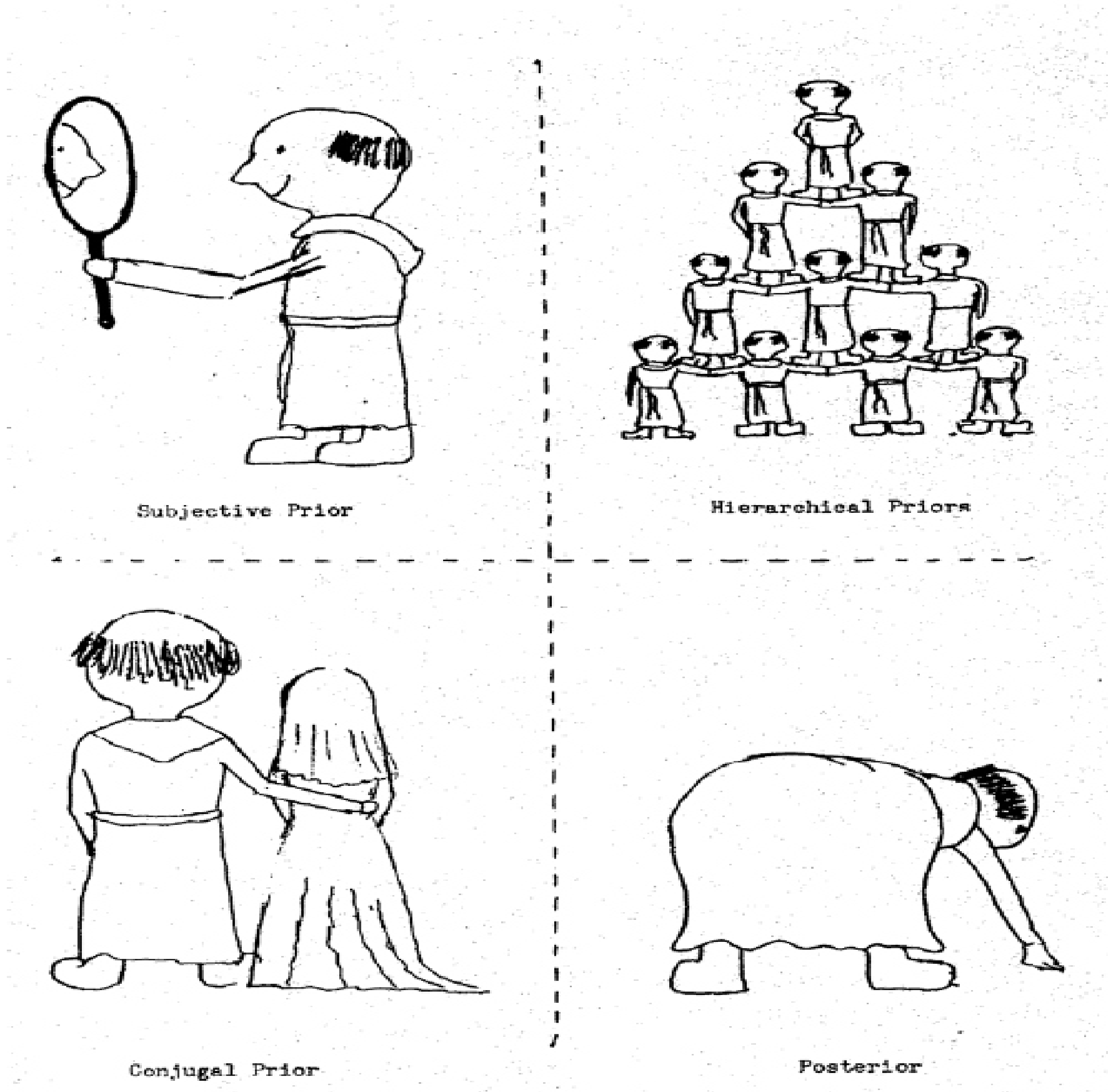
- We need a prior elicitation machine.



O'Hagan, A., 2005, "Elicitation," *Significance*, 84-86.

- There are many types of priors. Here are some metaphors.




Subjective Prior
Hierarchical Priors
Conjugal Prior
Posterior

- An elicitation method forms a bridge between an expert's opinions and an expression of these opinions in a statistically useful form.

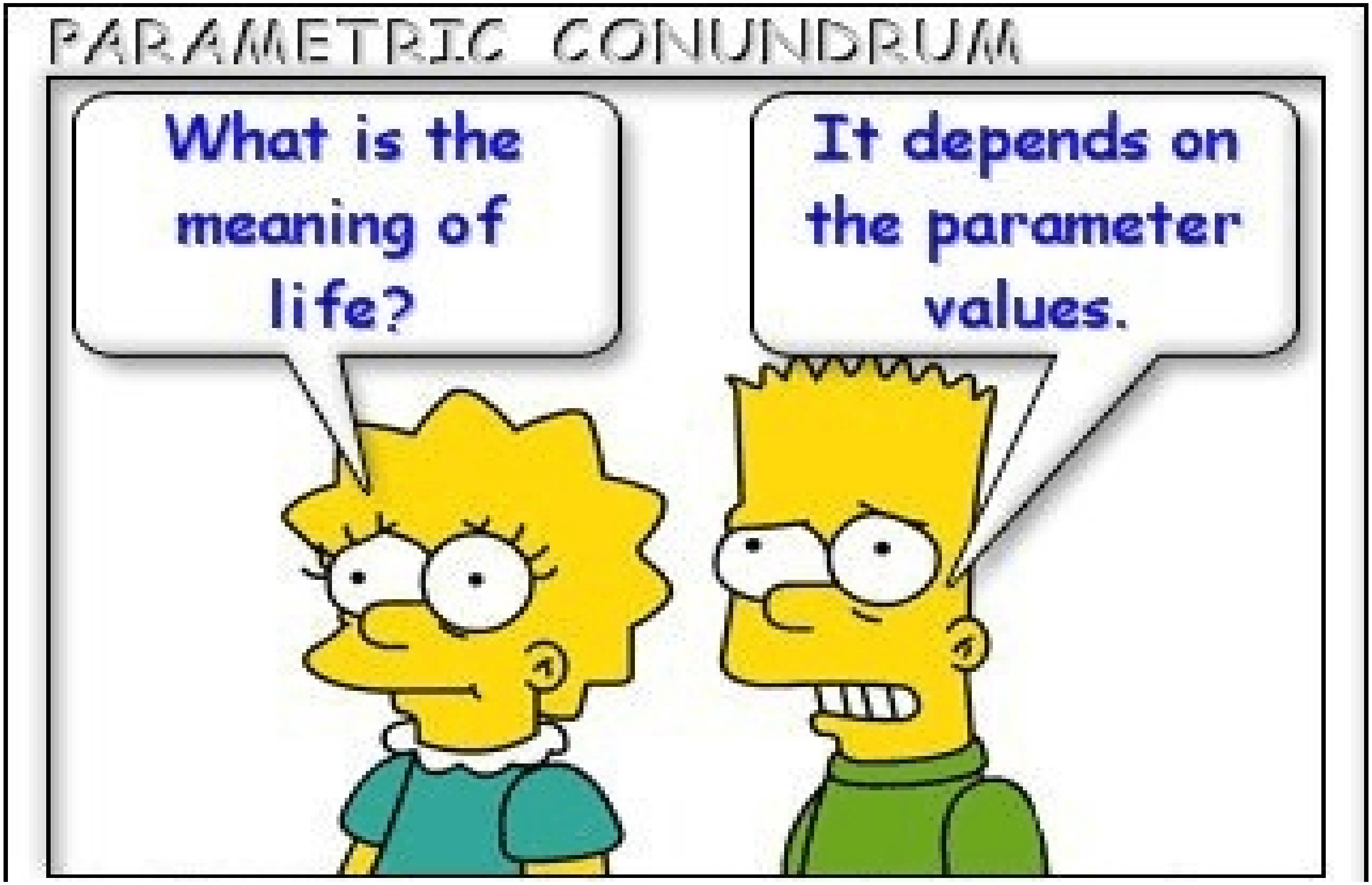
    - ***Structural elicitation***: analyst is able to state a prior directly based on subject matter considerations.

        - Economic theorists, can you hear me now?

"CAN YOU HEAR ME NOW?"

    - ***Predictive elicitation***: Analysts "backs-out" $f(\theta)$ from

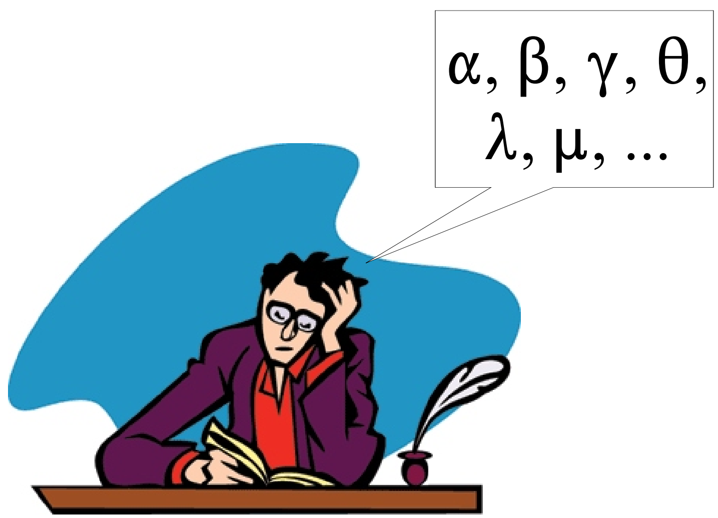
$$p(z) = \int_0^1 \mathcal{L}(\theta; z) f(\theta)\, d\theta.$$


PARAMETRIC CONUNDRUM
What is the meaning of life?
It depends on the parameter values.

PPMB3: Likelihoods and parameters are artifacts of the mind rather than intrinsic characteristics of the world. They provide useful windows through which like-minded researchers view the observable world and engage in prediction of future observables based on what has been observed in the past.

- Parametric models are not intended to be "literal" descriptions of reality; rather they are potentially useful windows for viewing the observable world and making inferences regarding future observables.

- Economists have a ***Hellenic fetish*** for the Greek alphabet.



- This principle is in the spirit of a representation theorem.

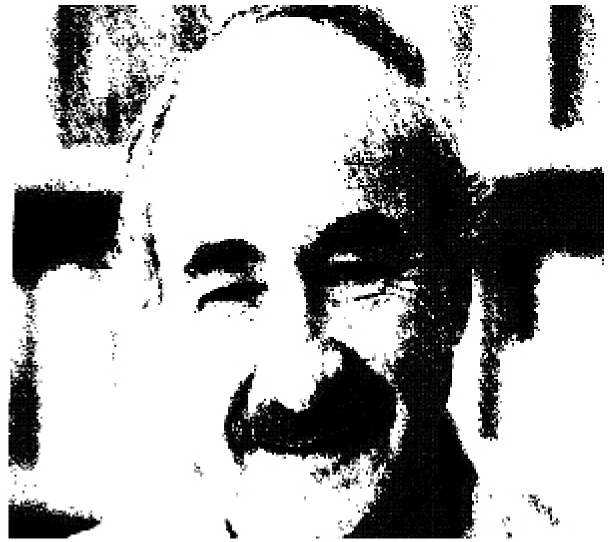
  - Representation theorems encourage use of ***proper*** priors.

- I believe most researchers admit when pressed that parameters are meant to be ***used, not worshiped***.

- The emphasis on parameters is sometimes a crude approximation for summarizing predictive effects.

PPMB4: Given an acceptable window through which to view the world, the major task facing the researcher is to conduct a ***sensitivity analysis*** of the DOI with respect to as wide a range of professionally interesting priors as possible.

- In other words: ***convince the reader that small changes in the analysis will not drastically change the primary results of interest***.
  - Analytical results are few.
  - This is part artistic in implementation.
- Some priors and hypotheses may be ***data-instigated***.



DeGroot (1980): "We open the newspaper in the morning and read some data on a topic we had not previously thought about. In order to process the data, we try to think about what our prior distribution would have been before we saw the data so we can calculate our posterior distribution. But we are too late. Who can say what our prior distribution would have been before we saw the data. We lost our virginity when we read the newspaper."

- ***How does one restore virginity lost?***

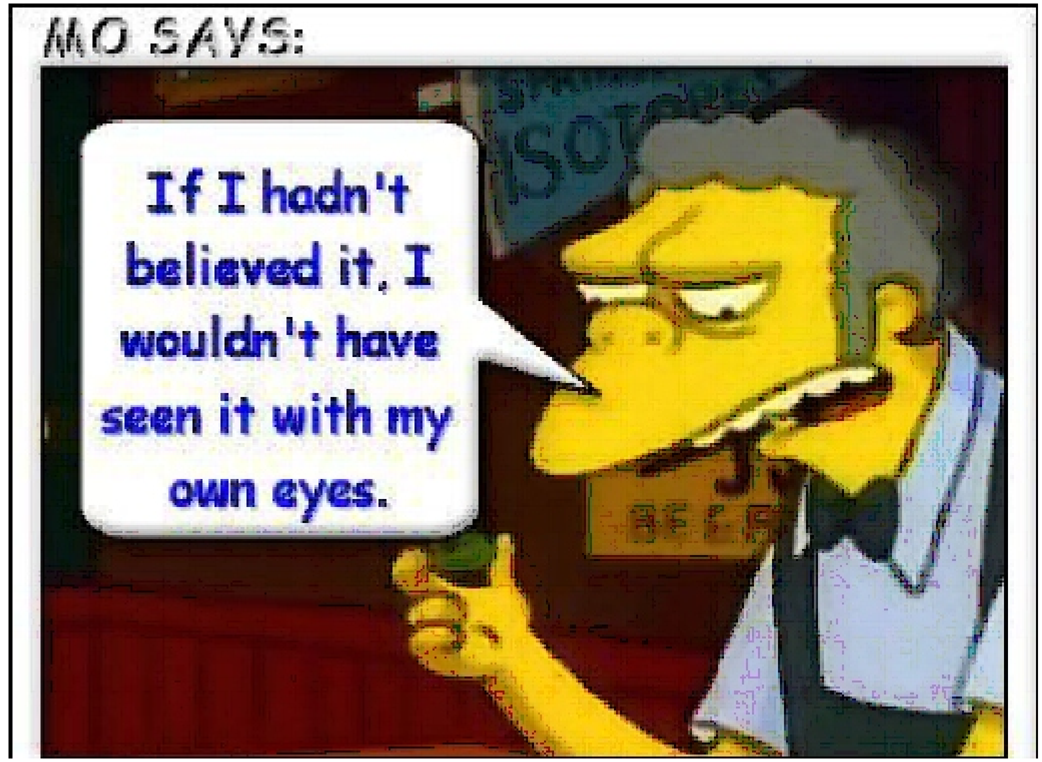
http://www.revirginizer.com/revirgCANADIAN.html

- Once the researcher gives up the ideal state of the "single-prior Bayesian" and admits the need for sensitivity analysis in public research, the researcher is left with the usual task of presenting a variety of mappings from ***interesting priors*** to posteriors.

    - Priors that have been contaminated by data can be presented as such - ***as always it remains for the reader to assess their plausibility***.

- Those who prefer ***virgin priors*** are likely ***virgin data analysts***.

PPMB5 (Cromwell's Rule):
Never assign a literal probability of unity to the window through which you choose to view the world.



- This appeals to the humble side of most seasoned empirical researchers.

- The essential message is: think critically about the models under consideration.

- **Corollary:** Admit the possibility that some other model, not yet introduced, may be deemed "better" in the future.

    - Consider models $m = 1, 2, ..., M$. Assign $1-\varepsilon$ prior probability to them, and reserve ε probability for model $M+1$ "something else."

    - Given priors for unknown parameters in each of the M models, and the relative prior probabilities $\text{Prob}(\mu = m | \mu = 1, 2, ..., M)$, standard posterior odds analysis permits comparison of the ***relative*** posterior probabilities of these M models ***without specifying*** ε.

    - If a new insight leads to specification of "something else," and a prior for the parameters of model $M+1$, analysis proceeds straightforwardly.

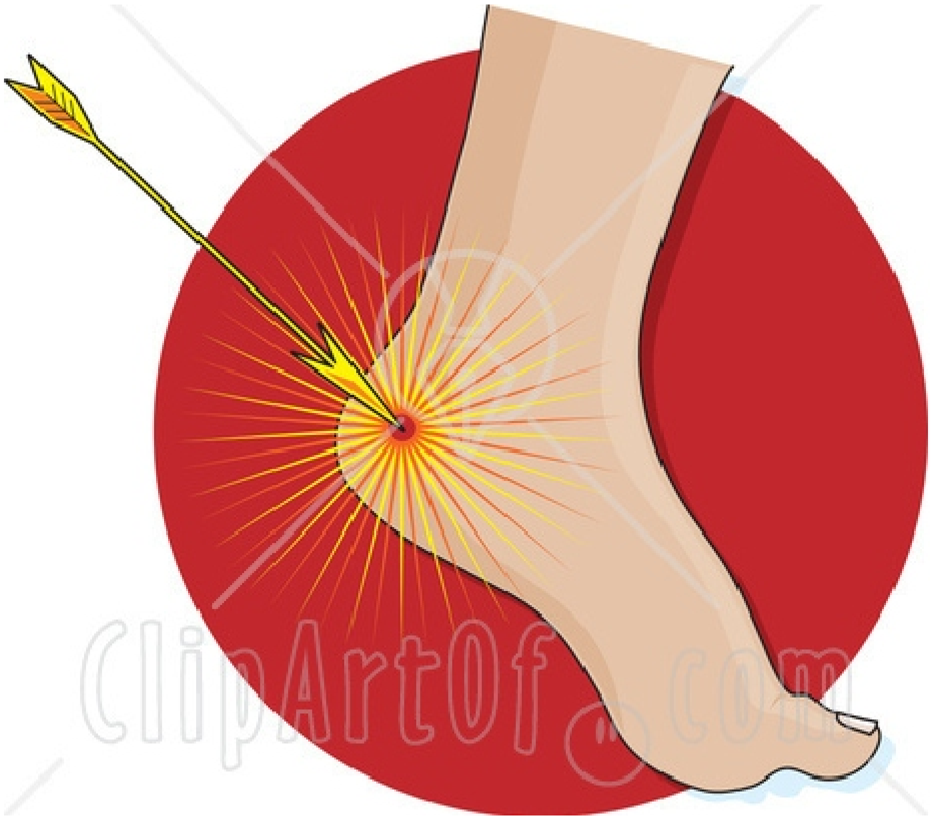
Conclusions

Charge: Frequentists are (among other things)

- Incoherent (e.g., linear probability models)
- Non-evidential, i.e., do not condition on observed data (e.g., confidence intervals).
- Bi-conditionalists (e.g., use p-values).
- Pretesters (e.g., dropping "statistically insignificant" parameters.
- Fixed-size (independent of sample size) hypothesis tests.
- Residents of the planet ***Asymptopia*** (you know when you get there because your parameters will be "statistically significant").


THE VERDICT IS IN!
I find the frequentists guilty as charged!

Care to get on board?



Bayesian: prior $f(\theta)$

Frequentist: sample space $\mathcal{Z}$