# Constructive Verification, Empirical Induction, and Falibilist Deduction: A Threefold Contrast

## Interpretation of Bayesian *e*-values

**Julio Michael Stern***,
* Institute of Mathematics and Statistics
of the University of Sao Paulo
*jstern@ime.usp.br*

EBL - 2011
Encontro Brasileiro de Lógica

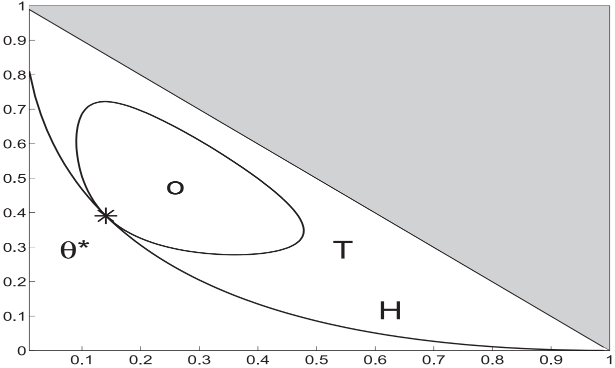
# Previous Work of IME-USP Bayesian Group

Statistical significance, in empirical science, is the measure of belief or credibility or the truth value of an hypothesis.

1. Pereira and Stern (1999), Pereira et al. (2008):
   Statistical Theory of *e*-values - ev($H$) or ev($H \mid X$)
   epistemic value of hypothesis $H$ given de data $X$
   or evidence given by $X$ in support of $H$.
2. Stern (2003, 2004), Borges and Stern (2007):
   "Logical" theory for composite *e*-valyes
   Compound Statistical Hypotheses in HDNF -
   Homogeneous Disjunctive Normal Form.
   (no such thing for *p*-values or Bayes factors)
3. Stern (2007a, 2007b, 2008a, 2008b):
   Epistemological Framework given by
   Cognitive Constructivism.

# Statistical Sharp Hypothesis

States that the true value of the parameter, $\theta$, of the sampling distribution, $p(x \mid \theta)$, lies in a low dimension set: The Hypothesis set, $\Theta_H = \{\theta \in \Theta \mid g(\theta) \leq 0 \wedge h(\theta) = 0\}$, has Zero volume (Lebesgue measure) in the parameter space.




Hardy-Weinberg Hypothesis

Bayesian setup:

- $p(x \mid \theta)$: *Sampling distribution* of an observed (vector) random variable, $x \in \mathcal{X}$, indexed by the (vector) *parameter* $\theta \in \Theta$, regarded as a latent (unobserved) random variable.
- The model's joint distribution can be factorized either as the *likelihood function* of the parameter given the observation times the *prior* distribution on $\theta$, or as the *posterior* density of the parameter times the observation's marginal density,

$$p(x, \theta) = p(x \mid \theta)p(\theta) = p(\theta \mid x)p(x) \ .$$

- $p_0(\theta)$: The *prior* represents our initial information.
- The *posterior* represents the available information about the parameter after 1 observation (unormalized potential),

$$p_1(\theta) \propto p(x \mid \theta)p_0(\theta) \ .$$

  Normalization constant $c_1 = \int_\theta p(x \mid \theta)p_0(\theta)d\theta$
- Bayesian learning is a recursive and comutative process.

- Hardy-Weinberg genetic equilibrium, see Pereira and Stern (1999).
  $n$ , sample size, $x_1, x_3$ , homozygote,
  $x_2 = n - x_1 - x_3$ , heterozygote count.

$$p_0(\theta) \propto \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3} \ ,$$

| | |
|---|---|
| $y = [0, 0, 0]$, | Flat or uniform prior, |
| $y = [-1/2, -1/2, -1/2]$, | Invariant Jeffreys' prior, |
| $y = [-1, -1, -1]$, | Maximum Entropy prior. |

$$p_n(\theta \mid x) \propto \theta_1^{x_1+y_1} \theta_2^{x_2+y_2} \theta_3^{x_3+y_3} \ ,$$

$$\Theta = \{\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 = 1\} \ ,$$

$$H = \{\theta \in \Theta \mid \theta_3 = (1 - \sqrt{\theta_1}\ )^2\} \ .$$

# 1- Full Bayesian Significance Test

- $r(\theta)$, the reference density, is a representation of no, minimal or vague information about the parameter $\theta$. If $r \propto 1$ then $s(\theta) = p_n(\theta)$ and $\overline{T}$ is a HPDS.
- $r(\theta)$ defines the information metric in $\Theta$, $dl^2 = d\theta' J(\theta) d\theta$, directly from the Fisher Information Matrix,

$$J(\theta) \equiv -\mathrm{E}_{\mathcal{X}} \frac{\partial^2 \log p(x \mid \theta)}{\partial \theta^2} = \mathrm{E}_{\mathcal{X}} \left( \frac{\partial \log p(x \mid \theta)}{\partial \theta} \frac{\partial \log p(x \mid \theta)}{\partial \theta} \right) .$$

- $s(\theta) = p_n(\theta)/r(\theta)$, posterior surprise relative to $r(\theta)$.
- $\overline{T}(v)$, the tangential set, is the HRSS, Highest Relative Surprise Set, above level $v$.
- $W(v)$, the truth (Wahrheit) function, is the cumulative surprise distribution.

- FBST evidence value supporting and against the hypothesis $H$, $\text{ev}(H)$ and $\overline{\text{ev}}(H)$,

$$s(\theta) = p_n(\theta) / r(\theta) \ ,$$

$$\widehat{s} = s(\widehat{\theta}) = \sup_{\theta \in \Theta} s(\theta) \ ,$$

$$s^* = s(\theta^*) = \sup_{\theta \in H} s(\theta) \ ,$$

$$T(v) = \{\theta \in \Theta \,|\, s(\theta) \leq v\} \ , \ \ \overline{T}(v) = \Theta - T(v) \ ,$$

$$W(v) = \int_{T(v)} p_n(\theta) \, d\theta \ , \ \ \overline{W}(v) = 1 - W(v) \ ,$$

$$\text{ev}(H) = W(s^*) \ , \ \ \overline{\text{ev}}(H) = \overline{W}(s^*) = 1 - \text{ev}(H) \ .$$

- $H$ in Homogeneous Disjunctive Normal Form;
  Independent statistical Models $j = 1, 2, \ldots$
  with stated Hypotheses $H^{(i,j)}$, $i = 1, 2 \ldots$
  Structures: $M^{(i,j)} = \{\Theta^j, H^{(i,j)}, p_0^j, p_n^j, r^j\}$ .

$$\text{ev}(H) = \text{ev}\left(\bigvee_{i=1}^{q} \bigwedge_{j=1}^{k} H^{(i,j)}\right) =$$

$$\max_{i=1}^{q} \text{ev}\left(\bigwedge_{j=1}^{k} H^{(i,j)}\right) =$$

$$W\left(\max_{i=1}^{q} \prod_{j=1}^{k} s^{*(i,j)}\right) ,$$

$$W = \bigotimes_{1 \le j \le k} W^j .$$

- Composition operators: max and $\bigotimes$ (Mellin convolution).
- If all $s^* = 0 \vee \widehat{s}$, ev $= 0 \vee 1$, classical logic.

# Wittgenstein's concept of Logic

- We analyze the relationship between the credibility, or truth value, of a complex hypothesis, $H$, and those of its elementary constituents, $H^j$, $j = 1 \ldots k$. This is the *Compositionality* question (ex. in analytical philosophy).
- According to Wittgenstein, (*Tractatus*, 2.0201, 5.0, 5.32):
  - Every complex statement can be analyzed from its elementary constituents.
  - Truth values of elementary statements are the results of those statements' truth-functions.
  - All truth-function are results of successive applications to elementary constituents of a finite number of truth-operations.
- Wahrheitsfunktionen, $W^j(s)$;
  Wahrheitsoperationen, $\bigotimes$, max.

# Birnbaum's Logic for Reliability Eng.

- In reliability engineering, (Birnbaum, 1.4):
  *"One of the main purposes of a mathematical theory of reliability is to develop means by which one can evaluate the reliability of a structure when the reliability of its components are known. The present study will be concerned with this kind of mathematical development. It will be necessary for this purpose to rephrase our intuitive concepts of structure, component, reliability, etc. in more formal language, to restate carefully our assumptions, and to introduce an appropriate mathematical apparatus."*
- Composition operations:
  - Series and parallel connections;
- Belief values and functions:
  - Survival probabilities and functions.

# Abstract Belief Calculus - ABC

- Darwiche, Ginsberg (1992).
- $\langle \Phi, \oplus, \oslash \rangle$ , Support Structure;
  $\Phi$ , Support Function, for statements on $\mathcal{U}$;
  $\mathcal{U}$, Universe of valid statements;
  **0** and **1**, Null and Full support values;
  $\oplus$ , Support Summation operator;
  $\oslash$ , Support Scaling or Conditionalization.

  $\otimes$, Support Unscaling, inverse of $\oslash$.
  $\langle \Phi, \oplus \rangle$ , Partial Support Structure.

- $\oplus$, gives the support value of the disjunction of any two logically disjoint statements from their individual support values,

$$\neg(A \wedge B) \Rightarrow \Phi(A \vee B) = \Phi(A) \oplus \Phi(B) \; .$$

- $\oslash$, gives the conditional support value of $B$ given $A$ from the unconditional support values of $A$ and the conjunction $C = A \wedge B$,

$$\Phi_A(B) = \Phi(A \wedge B) \oslash \Phi(A) \; .$$

- $\otimes$, unscaling: If $\Phi$ does not reject $A$,

$$\Phi(A \wedge B) = \Phi_A(B) \otimes \Phi(A) \; .$$

- Support structures for some belief calculi, Probability, Possibility, Classical Logic, Disbelief. $a = \Phi(A)$, $b = \Phi(B)$, $c = \Phi(C = A \wedge B)$.

| ABC | $\Phi(\mathcal{U})$ | $a \oplus b$ | $\mathbf{0}$ | $\mathbf{1}$ | $a \preceq b$ | $c \oslash a$ | $a \otimes b$ |
|---|---|---|---|---|---|---|---|
| Pr | $[0,1]$ | $a+b$ | $0$ | $1$ | $a \leq b$ | $c/a$ | $a \times b$ |
| Ps | $[0,1]$ | $\max(a,b)$ | $0$ | $1$ | $a \leq b$ | $c/a$ | $a \times b$ |
| CL | $\{0,1\}$ | $\max(a,b)$ | $0$ | $1$ | $a \leq b$ | $\min(c,a)$ | $\min(a,b)$ |
| DB | $\{0..\infty\}$ | $\min(a,b)$ | $\infty$ | $0$ | $b \leq a$ | $c-a$ | $a+b$ |

- FBST setup: two belief calculi are in simultaneous use: ev constitutes a possibilistic (partial) support structure in the hypothesis space coexisting in harmony with the probabilistic support struct. given by the posterior probability measure in the parameter space; see Zadeh (1987) and Klir (1988) for nesting prop.of $T(v)$.

# 3- Epistemological Frameworks

Statistical significance, in empirical science, is the measure of belief or credibility or the truth value of an hypothesis.
There are (at least) three competing statistical theories on how to compute a significance measure.
Each of these theories has co-evolved with a specific epistemological framework, and a basic metaphor of truth.

- Decision Theory and The Scientific Casino:
  Bayesian posterior probability of hypothesis $H$ given the observed data-base $X$, $p$, or the corresponding Bayes factor, the Betting Odds $p/(1-p)$.
- Falsificationism and The Scientific Tribunal:
  Frequentist statistics' $p$-value of the observed data-base, $X$, given the hypothesis $H$.
- Cognitive Constructivism and Objects as Eigen-Solutions:
  Bayesian epistemic value of hypothesis $H$ given data $X$.

# Example of Inference by Ch.S.Peirce (1868)

Induction of letter frequencies and abduction of cipher codes.
- Given the English books $B_1, B_2, \ldots B_k$, compile letter frequency vectors $\lambda^1, \lambda^2, \ldots \lambda^k$. Realize that they all (approximately) agree with the average frequency vector, $\lambda^a$.
- Given a new English book, $B_{k+1}$, we may state, by Induction, that its not yet compiled letter frequency vector, $\lambda^{k+1}$, will also be (approximately) equal to $\lambda^a$.
- Given a coded book $C$, encrypted by a simple substitution cipher, compile its letter frequency vector, $\lambda^c$.

We realize that there is one and only one permutation vector, $\pi$, that can be used to (approximately) match vectors $\lambda^a$ and $\lambda^c$, that is, there is a unique bijection $\pi = [\pi(1), \pi(2), \ldots \pi(m)]$, where $m$ is the number of letters in the alphabet, such that $\lambda^a(j) \approx \lambda^c(\pi(j))$, for $1 \leq j \leq m$. We may state, by Abduction, the hypothesis that vector $\pi$ is the correct key for the cipher.

- A standard formulation for the induction part of this example includes parameter estimation (posterior distribution, likelihood or, at least, a point estimate and confidence interval) in an n-dimensional Dirichlet-Multinomial model, where $m$ is the number of letters in the English alphabet.
The parameter space of this model is the $(m-1)$-simplex, $\Lambda = \{\lambda \in [0,1]^m \,|\, \lambda \mathbf{1} = 1\}$.
- A possible formulation for the abduction part involves expanding the parameter space of the basic model to $\Theta = \Lambda \times \Pi$, where $\Pi$, the discrete space of $m$-permutations.
- Peirce's (abductive) hypothesis about the cipher proclaims the 'correct' or 'true' permutation vector, $\pi^0$. This hypothesis has an interesting peculiarity: The parameter space, $\Theta = \Lambda \times \Pi$, has a continuous sub-space, $\Lambda$, and a and a discrete (actually, finite) sub-space, $\Pi$. However, the hypothesis only (directly) involves the finite part. This peculiarity makes this hypothesis very simple, and amenable to the treatment given by Peirce.

- Assuming that the cypher key, $\pi^0$, was chosen with uniform prior probability $p_0(\pi) = 1/m!$,
- We can compute the posterior probability $p_n(\pi \mid \lambda^a, \lambda^c)$, where $n$ is the number of letters in the coded book, for each possible key $\pi$.
- Bayes rule operates the update from $p_0(\pi)$ to $p_n(\pi)$. “Inverse” probabilities - defined in the parameter space.
- Let $\pi^*$ be the key with highest posterior probability. As $n \to \infty$, $p_n(\pi^*) \to 1$ and $p_n(\pi^0) \to 1$. That is, we can be certain to select the correct key.
- $L(\pi, \pi^0)$: Gain-Loss function pricing correct-incorrect key selections: Morgenstern von Neumann Decision Theory teaches “How gamble if you must” (in science). $p_n(\pi)/(1 - p_n(\pi))$ are the hypotheses’ betting odds.

Lakatos (1978b,p.152): *Neoclassical empiricism had a central dogma: the dogma of the identity of (1) probabilities, (2) degree of evidential support (or confirmation), (3) degree of rational belief, and (4) rational betting quotients. This 'neoclassical chain of identities' is not implausible. For a true empiricist the only source of rational belief is evidential support: thus he will equate the degree of rationality of a belief with the degree of its evidential support. But rational belief is plausibility measured by rational betting quotients. It was, after all, to determine rational betting quotients that the probability calculus was invented.*

Dubins and Savage (1965,p.229): *Gambling problems... ...seem to embrace the whole of theoretical statistics.*

# Zero Probability Paradox

Shap Hypotheses, ex: Hardy-Weinberg genetic equilibrium.
Zero prior $+$ Multiplicative unscaling (Bayes rule)
$\Rightarrow$ Zero posterior (whatever the observed data $X$).

There are two neoclassical ways out of the ZPP conundrum:
(A) Fixing the mathematics to avoid the ZPP.

- The idea of subjective prior justifies any abuse.
- Jeffreys' tests: Singular measures on $H$ establishing apriori betting odds (handicap system for weak players).
  - It gives you nightmares, like Lindley's paradox.
- Amend the setup with artificial priors, like fractional (post) posteriors or other complicated oxymora, to no avail.
- Display a Caveat Emptor exempting responsibility.

There are two neoclassical ways out of the ZPP conundrum:
(B) Forbidding the use of sharp hypotheses.

Savage (1954, 16.3, p.254):
*The unacceptability of extreme (sharp) null hypotheses is perfectly well known; it is closely related to the often heard maxim that science disproves, but never proves, hypotheses. The role of extreme (sharp) hypotheses in science and other statistical activities seems to be important but obscure. In particular, though I, like everyone who practice statistics, have often "tested" extreme (sharp) hypotheses, I cannot give a very satisfactory analysis of the process, nor say clearly how it is related to testing as defined in this chapter and other theoretical discussions.*

Lakatos (1978b,p.154):
*But then degrees of evidential support cannot be the same as degrees of probability [of a theory] in the sense of the probability calculus.*
*All this would be trivial if not for the powerful time-honored dogma of what I called the 'neoclassical chain' identifying, among other things, rational betting quotients with degrees of evidential support.*
*This dogma confused generations of mathematicians and of philosophers.*

# Frequentist p-values

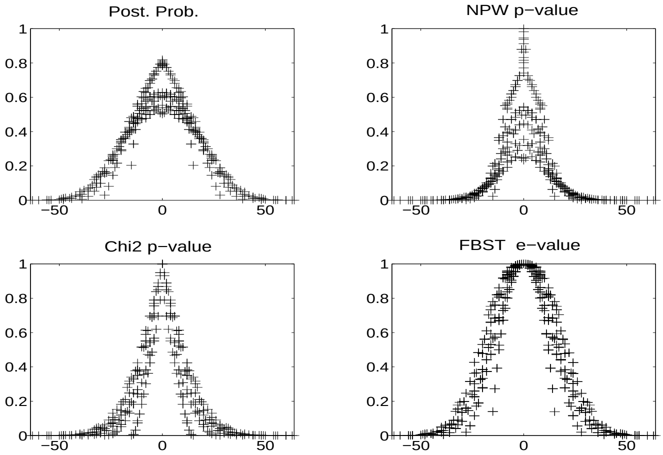
Peirce (1883): *[Kepler] traced out the miscellaneous consequences of the supposition that Mars moved in an ellipse, with the sun at the focus, and showed that both the longitudes and the latitudes resulting from this theory were such as agreed with observation. ...The term Hypothesis [means] a proposition believed in because its consequences agree with experience.*

$p$-value is the probability of getting a sample (data set), $X$, that is more (at least as) extreme (improbable) than the one we got, assuming that the (null) hypothesis is true.

*p*-value does not get in trouble with sharp hypotheses.
However, it may not do what one think it does...

- The *p*-value "shifts" a question about the parameter into a question about possible observations (assuming $H$).
$\Rightarrow$ Leads to difficult or false interpretations.
- For a singular hypothesis, $H = \{\theta^0\}$, the *p*-value is well defined. However, a composite hypothesis defines **no** probability order in the sample space. One may:
- Reduce $H$ to the constrained MAP (max.-a-posteriori) singular hypothesis, $H^* = \{\theta^* = \arg\max_H p(X \mid \theta)\}$,
- or compute the posterior average $\int_H p(X \mid \theta) p_n(\theta) d\theta$,
- or many other possible variations.
- Technical problems: May require a "stopping rule";
May use non invariant procedures in its calculations,
May not conform to the Likelihood Principle; etc., etc.

# Qualitative Comparison of Performance




Hardy-Weinberg Hypothesis, all samples for n=16.
Frequency Asymmetry, $FA = f_3 - (1 - \sqrt{f_1})^2$.

# Falibilist Deduction

- Scientific Tribunal proves guilt, never innocence.
  "Increase sample size to reject" (the theory).
- Probability calculus restricted to the sample space.
  Belief calculus in the parameter space: None!
- "Statistics is Prediction", a Weltanschauung shared by Decision Theoretic Bayesian statistics, together with
- Positivist disdain for parameters, theoretical concepts, and any other metaphysical entity.
  Parameters are intermediate (integration) variables used to compute predictive probabilities, risk, expected values, etc.

Metaphysical: Strict sense - not directly measurable;
Gnosiological (Aristotelic) sense - a basis for explanation.

# Aufhebung to Rational Metaphysics, a plea

Lakatos (1977b, p.31-32): *Neyman and Popper found a revolutionary way to finesse the issue by replacing inductive reasoning with a deductive process of hypothesis testing. They then proceeded to develop this shared central idea in different directions, with Popper pursuing it philosophically while Neyman (in his joint work with Pearson) showed how to implement it in scientific practice.*

Imre Lakatos; A Plea to Popper for a Whiff of Inductivism, in Schilpp (1974,Ch.5,p.258).
*With a positive solution to the problem of induction, however thin, methodological theories of demarcation can be turned from arbitrary conventions into rational metaphysics.*

# Cognitive Constructivist Ontology

Heinz von Foerster (2003): *Objects are tokens for eigen-behaviors.* (eigen-... = system's recurrent solution)

*Tokens stand for something else. In the cognitive realm, objects are the token names we give to our eigen-behavior. This is the constructivist's insight into what takes place when we talk about our experience with objects.* (ex: money, itself a token for gold).

*Eigenvalues have been found ontologically to be discrete (sharp), stable, separable and composable, while ontogenetically to arise as equilibria that determine themselves through circular processes.*
*Ontologically, Eigenvalues and objects, and likewise, ontogenetically, stable behavior and the manifestation of a subject's 'grasp' of an object cannot be distinguished."*

Hermann Weyl (1989, p.132): *Objectivity means invariance with respect to the group of automorphisms.*

| | | | | |
|---|---|---|---|---|
| Experiment | | | | Theory |
| Experiment design | $\Leftarrow$ | Operatio-nalization | $\Leftarrow$ | Hypotheses formulation |
| $\Downarrow$ | | | | $\Uparrow$ |
| Effects observation | | false/true eigensolution | | Inter-pretation |
| $\Downarrow$ | | | | $\Uparrow$ |
| Data acquisition | $\Rightarrow$ | Explanation | $\Rightarrow$ | Statistical analysis |
| Sample space | | | | Parameter space |

Scientific Production Diagram, after Krohn and Küppers (1990)
Dynamical structure as autopoietic double feed-back system.

# Cog-Con Aufhebung to Rational Metaphysics

The FBST solution to the problem of verification is indeed very thin, in the sense that the proposed epistemic support function, the *e*-value, although based on a Bayesian posterior probability measure, provides only a possibilistic (not a probabilistic) support measure for the hypothesis under scrutiny.

However, this apparent weakness is in fact the key to overcome the deadlocks of induction related to the ZPP .

Nevertheless, the simultaneous Cog-Con characterization of the supported objects (sharp or precise stable, separable and composable eigen-solutions) and their associated hypotheses, implies such a strong and rich set of essential properties, that the Cog-Con solution becomes also very positive.

# Sharp hypotheses - ZPP absolution

Sharp hypotheses are freed from the zero-support syndrome, and admitted as full citizens in the hypothesis space.

However, that does not warrant that there will ever be a sharp hypothesis in an empirical science with good support.
In fact, considering the original ZPP, finding such an outstanding (sharp) hypothesis should be really surprising, the scientific equivalent of a miracle!
What else should we call showing possible, what is almost surely (in the probability measure) infeasible?

Nevertheless, we know that miracles do exist.
(Non-believers must take Experimental Physics 101)

# Sharp hypotheses - Metaphysical redemption

The Cognitive Constructivism epistemological framework, equipped with the FBST / ev($H$) apparatus, not only redeems sharp or precise hypotheses from statistical damnation, but places them at the center stage of scientific activity. (The star role of any exact science will always be played by eigen-solutions represented by *somebody's equation*).

Hence, these equations, parameters, and metaphysical concepts they represent, receive a high ontological status.

Therefore, we believe that the Cog-Con framework provides important insights about the nature of empirical sciences, insights that, in important issues, penetrate deeper than some of the standard alternative epistemological frameworks.

# Aliis exterendum - a plea for humility

Others tresh (think, criticize), we (empiricists) harvest.
Others indulge in metaphysics, we access truth directly.

In the Cog-Con framework, the certification of a sharp hypothesis by *e*-values close to unity is a strong form of verification, akin to empirical confirmation or pragmatic authentication. (Popperian corroboration is only fail to refute).
Nevertheless, the *e*-value does **not** provide the inductive engine or truth-pump dreamed by the empiricist school.
There is a lot more to the understanding of science as an evolutionary process than the passive waiting for truthful theories to mushroom-up from well harvested data.
Actually, such an engine could become a real nightmare, draining all soul and conscience from research activity and extinguishing the creative spirit of scientific life.

# Mathematics as Quasi-Empirical

Lakatos (1978,V.2,p.40):
*Whether a deductive system is Euclidean or quasi-empirical is decided by the pattern of truth value flow in the system. The system is Euclidean if the characteristic flow is the transmission of truth from the set of axioms 'downwards' to the rest of the system - logic here is the organon of proof; it is quasi-empirical if the characteristic flow is 'upwards' towards the 'hypothesis' - logic here is an organon of criticism. We may speak (even more generally) of Euclidean versus quasi-empirical theories independently of what flows in the logical channels: certain or fallible truth or falsehoods, probability or improbability, moral desirability or undesirability, etc. It is the how of the flow that is decisive.*

If all all hypothesis have null or full support, rules of composition for *e*-values and classical logic coincide.

This property constitutes a bridge from physics to mathematics, from empirical to quasi-empirical science. From this perspective, mathematics can be seen as an idealized world of absolutely verified theories populated by hypotheses with either full or null support.

I will not venture into the discussion of whether or not good mathematics comes from heaven or "straight from The Book". I will only celebrate the revelation of this mystery. It represents the ultimate transmutation of the ZPP, from bad omen of confusion, to good augury of universal knowledge.

# Bibliography


- W.Borges, J.M.Stern (2007). The Rules of Logic Composition for the Bayesian Epistemic e-Values. *Logic J. IGPL*, 15, 401-420.
- R.Inhasz, J.M.Stern (2010). Emergent Semiotics in Genetic Programming and the Self-Adaptive Semantic Crossover. *Studies in Computational Intelligence*, 314, 381-392.
- M.R.Madruga, L.G.Esteves, S.Wechsler (2001). On the Bayesianity of Pereira-Stern Tests. *Test*, 10, 291–299.
- C.A.B.Pereira, J.M.Stern (1999). Evidence and Credibility: Full Bayesian Significance Test Precise Hypotheses. *Entropy*, 1, 69-80.
- C.A.B.Pereira, S.Wechsler, J.M.Stern (2008). Can a Significance Test be Genuinely Bayesian? *Bayesian Analysis,* 3, 79-100.
- J.M.Stern (2003). Significance Tests, Belief Calculi, and Burden of Proof in Legal and Scientific Discourse. UAI'03 and Laptec'03, *Frontiers in Artificial Intell. and its Applications*, 101, 139–147.
- J.M.Stern (2004). Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. SBIA'04, *LNAI*, 3171, 134–143.

- J.M.Stern (2007a). Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses. *Cybernetics and Human Knowing*, 14, 9-36.
- J.M.Stern (2007b). Language and the Self-Reference Paradox. *Cybernetics and Human Knowing*, 14, 71-92.
- J.M.Stern (2008a). Decoupling, Sparsity, Randomization, and Objective Bayesian Inference. *Cybernetics and Human Knowing*, 15, 49-68.
- J.M.Stern (2008b). *Cognitive Constructivism and the Epistemic Significance of Sharp Statistical Hypotheses.* Tutorial book for MaxEnt 2008, The 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. July 6-11 of 2008, Boracéia, São Paulo, Brazil.
- J.M.Stern (2011). Spencer-Brown vs. Probability and Statistics: Entropy's Testimony on Subjective and Objective Randomness. *Information,* 2, 277-301.
- J.M.Stern (2011). *Symmetry, Invariance and Ontology in Physics and Bayesian Statistics*. Submitted.