Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses.

FIS-2005, ENSTA, Paris, 4-7 July 2005

Julio Michael Stern

BIOINFO and

Computer Science Dept.,

University of São Paulo, Brazil

jstern@ime.usp.br

www.ime.usp.br/ ~jstern

Autopoiesis (Maturana and Varela):

Autopoietic systems are non-equilibrium (dissipative) dynamic systems exhibiting (meta) stable structures, whose organization remains invariant over (long periods of) time, despite the frequent substitution of their components. Moreover, these components are produced by the same structures they regenerate.

For example, the macromolecular population of single cell can be renewed thousands of times during its lifetime.

The investigation of these regeneration processes in the autopoietic system production network leads to the definition of cognitive domain:

Maturana and Varela (1980):

"Our aim was to propose the characterization of living systems that explains the generation of all the phenomena proper to them. We have done this by pointing at Autopoiesis in the physical space as a necessary and sufficient condition for a system to be a living one."

"An autopoietic system is organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which:

(i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and

(ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as such a network."

Cognitive Domain (Maturana and Varela):

... Thus the circular organization implies the prediction that an interaction that took place once will take place again...

Every interaction is a particular interaction, but every prediction is a prediction of a class of interactions that is defined by those features of its elements that will allow the living system to retain its circular organization after the interaction, and thus, to interact again. This makes living systems inferential systems, and their domain of interactions a cognitive domain."

Objects as Tokens for Eigen-Solutions
(von Foerster):

"The meaning of recursion is to run through one's own path again. One of its results is that under certain conditions there exist indeed solutions which, when reentered into the formalism, produce again the same solution. These are called "eigen-values", "eigen- functions", "eigen-behaviors", etc., depending on which domain this formation is applied."

"Objects are tokens for eigen-behaviors.
Tokens stand for something else.
In exchange for money (a token itself for gold), ...tokens are used to gain admittance to the subway or to play pinball machines.
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."

Eigen-Solution Ontology
(von Foerster):

"Eigenvalues have been found ontologically to be discrete, 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."

a- Discrete (or sharp):

"There is an additional point I want to make, an important point. Out of an infinite continuum of possibilities, recursive operations carve out a precise set of discrete solutions. Eigenbehavior generates discrete, identifiable entities. Producing discreteness out of infinite variety has incredibly important consequences. It permits us to begin naming things. Language is the possibility of carving out of an infinite number of possible experiences those experiences which allow stable interactions of yourself with yourself."

It is important to realize that, in the sequel, the term "discrete", used by von Foerster to qualify eigen-solutions in general, should be replaced, depending on the specific context, by terms such as lower-dimensional, precise, sharp, singular, etc.

b- Equilibria (or stable):

A stable eigen-solution of the operator $Op(\ )$, defined by the fixed-point or invariance equation, $x_{inv} = Op(x_{inv})$, can be found (built or computed) as the limit, x_∞, of the sequence $\{x_n\}$, as $n \to \infty$, by the recursive application of the operator, $x_{n+1} = Op(x_n)$.

Under appropriate conditions (such as within a domain of attraction, for instance) the process convergence and its limit eigen-solution do not depend on the starting point, x_0.

We want to show, for statistical analysis in a scientific context, how the property of sharpness indicates that many, and perhaps some of the most relevant, scientific hypotheses are sharp, and how the property of stability, indicates that considering these hypotheses is natural and reasonable.

Coupling (Maturana and Varela):

"Whenever the conduct of two or more units is such that there is a domain in which the conduct of each one is a function of the conduct of the others, it is said that they are coupled in that domain."

"An autopoietic system whose autopoiesis entails the autopoiesis of the coupled autopoietic units which realize it, is an autopoietic system of higher order."

A typical example of a hierarchical system is a Beehive, a third order autopoietic system, formed by the coupling of individual Bees, the second order systems, which, in turn, are formed by the coupling of individual Cells, the first order systems.

Social Systems (Luhmann):

"Social systems use communication as their particular mode of autopoietic (re)production. Their elements are communications that are recursively produced and reproduced by a network of communications that are not living units, they are not conscious units, they are not actions. Their unity requires a synthesis of three selections, namely information, utterance and understanding (including misunderstanding)."

Differentiation (Luhmann):

Societys' strategy to deal with increasing complexity is the same one observed in most biological organisms, namely, differentiation.
Biological organisms differentiate in specialized systems, such as organs and tissues of a pluricellular life form (non-autopoietic or allopoietic systems), or specialized individuals in an insect colony (autopoietic system).
In fact, societies and organisms can be characterized by the way in which they differenciate into systems.

Modern societies are characterized by a vertical differenciation into autopoietic functional systems, where each system is characterized by its code, program and (generalized) media.

-The code gives a bipolar reference to the system, of what is positive, accepted, favored or valid, versus what is negative, rejected, disfavored or invalid.
-The program gives a specific context where the code is applied.
-The (generalized) media is the space in which the system operates.

Standard examples of social systems are:

Science: code: true/false; program: scientific theory; media: journals, proceedings.

Judicial: code: legal/illegal; program: laws and regulations; media: legal documents.

Religion: code: good/evil; program: sacred and hermeneutic texts; media: study, prayer and good deeds.

Economy: code: property / lack thereof; program: economic scenario, pricing method; media: money, assets.

Dedifferentiation (Luhmann):

Dedifferentiation (Entdifferenzierung) is the degradation of the system's internal coherence, through adulteration, disruption, or dissolution of its own autopoietic relations.
One form of dedifferentiation (in either biological or social systems) is the system's penetration by external agents who try to use system's resources in a way that is not compatible with the system's autonomy.

"Autopoieticists claim that the smooth functioning of modern societies depends critically on maintaining the operational autonomy of each and every one of its functional (sub) systems."

Each system may be aware of events in other systems (be cognitively open) but is required to maintain its differentiation (be operationally closed).

Scientific Production Diagram
(Krohn, Küppers and Nowotny):

Experiment				Theory
Experiment design	$\Leftarrow$	Operationalization	$\Leftarrow$	Hypotheses formulation
$\Downarrow$				$\Uparrow$
Effects observation		false/true eigensolution		Interpretation
$\Downarrow$				$\Uparrow$
Data acquisition	$\Rightarrow$	Explanation	$\Rightarrow$	Statistical analysis
Sample space				Parameter space

Scientific knowledge, structure and dynamics, as an autopoietc double feed-back system.

Predictive Probabilities:
(stated in the sample space)

- At the Experiment side of the diagram, the task of statistics is to make probabilistic statements about the occurrence of pertinent events, i.e. describe probabilistic distributions for what, where, when or which events can occur.

Statistical Support of Hypotheses:
(stated in the parameter space)

- At the Theory side of the diagram, the role of statistics is to measure the statistical support of (sharp) hypotheses, i.e. to measure, quantitatively, the hypothesis plausibility or possibility in the theoretical framework they were formulated, given the observed data.

Probabilistic Statements:

Frequentist probabilistic statements are made exclusively on the basis of the frequency of occurrence of an event in a (potentially) infinite sequence of observations generated by a random variable.

Epistemic probabilistic statements are made on the basis of the epistemic status (degree of belief, likelihood, truthfulness, validity) of an event from the possible (actual or potential) outcomes generated by a random variable.

Bayesian probabilistic statements are epistemic probabilistic statements generated by the (in practice, always finite) recursive use of Bayes formula:

$$p_n(\theta) \propto p_{n-1}(\theta)p(x_n|\theta) \ .$$

Frequentist (Classical) Statistics,
dogmatically demands that all probabilistic statements be frequentist. Therefore, any direct probabilistic statement on the parameter space is categorically forbidden.

Scientific hypotheses are epistemic statements about the parameters of a statistical model. Hence, frequentist statistics can not make any direct statement about the statistical significance (truthfulness) of hypotheses.

Strictly speaking it can only make statements at the Experiment side of the diagram.
The frequentist way of dealing with questions on Theory side of the diagram, is to embed them some how into the Experiment side.

A p-value (of the data bank, not of the hypothesis) is the probability of getting a sample that is more extreme (incompatible with H) than the one we got.

Bayesian statistics,
allows probabilistic statements on the parameter space, and also, of course, in the sample space. Thus it seems that Bayesian statistics is the right tool for the job, and so it is!

Nevertheless, we must first examine the role played by DecTh (decision theory) in the foundations of orthodox Bayesian statistics (stratified in two layers):

- First layer: DecTh provides a coherence system for the use of probability statements, see Finetti (1974, 1981). FBST use of probability theory is fully compatible with DecTh in this layer, Madruga (2001).

- Second layer: DecTh provides an epistemological framework for the interpretation of statistical procedures:

Betting on Theories (Savage):

"Gambling problems ... seem to embrace the whole of theoretical statistics according to the decision-theoretic view of the subject.

... the gambler in this problem is a person who must act in one of two ways (two guesses),... appropriate under (H_0) or its negation (H_1).

...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 ... seems to be important but obscure. ...I cannot give a very satisfactory analysis... nor say clearly how it is related to testing as defined in (DecTh) theoretical discussions."

In the DecTh framework we speak about the betting odds for "getting the hypothesis on a gamble taking place in the parameter space".

But sharp hypotheses are zero (Lebesgue) measure sets, so our betting odds must be null, i.e. sharp hypotheses are (almost) surely false.

If we accept the ConsTh view that an important class of hypotheses concern the identification of eigen-solutions, and that those are ontologically sharp, we have a paradox!

The Full Bayesian Significance Test, FBST, was specially designed to give an evidence value supporting a sharp hypothesis, H. This support function, Ev(H, p_n), is based on the posterior pobability measure of a set called the tangential set $T(H, p_n)$, which is a non zero measure set (so no null probability paradoxes). Furthermore Ev( ) has many necessary or desirable properties for a statistical support function, such as:

1- Give an intuitive and simple measure of significance for the (null) hypothesis, ideally, a *probability* defined directly in the original or *natural parameter space*.

2- Be able to provide a *consistent* test for a given sharp hypothesis.

3- Have an intrinsically geometric definition, independent of any non-geometric aspect, like the particular parameterization of the hypothesis (manifold) being tested, or the particular coordinate system chosen for the parameter space, i.e., be an *invariant* procedure.

4- Require *no ad hoc artifice* like assigning a positive prior probability to zero measure sets, or setting an arbitrary initial belief ratio between hypotheses.

5- Give a *possibilistic support structure* for hypotheses, and so comply with the Onus Probandi juridical principle (In Dubito Pro Reo rule), i.e. consider in the "most favorable way" the claim stated by the hypothesis.

6- Obey the *likelihood principle* : The information gathered from observations should be represented (only) by the likelihood function.

7- Allow the incorporation of previous experience or expert's opinion via *"subjective" prior distributions*.

8- Be an *exact* procedure, i.e., make no use of "large sample" asymptotic approximations.

9- Give a measure of significance that is smooth, i.e. *continuous and differentiable*, on the hypothesis parameters and sample statistics, under appropriate regularity conditions.

Semantic Degradation:

Hopefully it's now clear that several technical difficulties of testing (sharp) hypotheses in the traditional statistical paradigms are symptoms of problems with much deeper roots.
Regarding the abuse of (pseudo) economical analyses, see Luhmann (1989):

"In this sense, it is meaningless to speak of "non-economic" costs. This is only a metaphorical way of speaking that transfers the specificity of the economic mode of thinking indiscriminately to other social systems."

Once the forces pushing for systemic degradation are exposed, we hope one can understand our (aphoristic, double) plea for sanity:

Preserve systemic autopoiesis and semantic integrity, for de-differentiation is in-sanity itself.

Chose the right tool for each job: "If you only have a hammer, everything looks like a nail".

Competing Sharp Hypotheses (Good):

Never test a single sharp hypothesis, an unfair faith of the poor sharp hypothesis standing all alone against everything else in the world. Instead, always confront a sharp hypothesis with a competing sharp hypotheses, a fair game.

“...If by the truth of Newtonian mechanics we mean... approximately true, we could obtain strong evidence that it is true; but if we mean... exactly true, then it has already been refuted.”

“...I think that the initial probability is positive for every self-consistent scientific theory... since the number of statable theories is at most countably infinite (enumerable).”

“...It is difficult to decide... numerical probabilities, but it is not so difficult to judge the ratio of initial probabilities of two theories by comparing their complexities. ...(that’s) why history of science is scientificaly important.”

The competing sharp hypotheses argument does not contradict the ConsTh epistemological framework, and it may be appropriate in certain circumstances. It may also mitigate or partially remediate the paradoxes of testing sharp hypotheses in the traditional frequentist or orthodox Bayesian settings (Bayes factors, Jeffreys' tests). However, we do not believe that having it is neither a necessary condition for good science practice, nor an accurate description of science history.

Quickly examine the very first major incident in the tumultuous debacle of Newtonian mechanics (just to stay with Good's example). This incident was Michelson's experiment on the effect of "etherial wind" over the speed of light. Michelson found no such effect, i.e. he found the speed of light to be constant, invariant with the relative speed of the observer (with no competing theory at that time).

Noether's Theorem:

For every continuous symmetry in a physical theory, there must exist an invariant quantity or conservation law:

Galileo's group $\Rightarrow$ Newtonian mechanics $\Rightarrow$ Newtonian objects (ex. momentum, enery).

Lorentz' goup $\Rightarrow$ Einstein's relativity $\Rightarrow$ Relativistic objects (ex. energy-momentum).

Conserv. laws are "ideal" sharp hypotheses.

Careful with competing theories historical analyses "ex post facto" or "post mortem".

Complex experiments require careful error and fluctuation analysis, and realistic (complex) statistical models. All statistically significant influences should be incorporated into the model.

De Finetti type Theorems,
given an invariance transformation group
(like permutability, spherical symmetry, etc.),
provide invariant distributions, that can in turn provide prototypical sharp hypotheses in many application areas.

Physics has its own heavy apparatus to deal with the all important issues of invariance and symmetry.

Statistics, via de Finetti theorems, can provide such an apparatus for other areas, even in situations that are not naturally embedded in a heavy mathematical formalism.

Compositionality

(von Foerster):
"...Eigenvalues have been found ontologically to be ...separable and composable...

(Luhmann, evolution of science):
"...something that idealization, mathematization, abstraction, etc. do not describe adequately. It concerns the increase in the capacity of decomposition and recombination, a new formulation of knowledge as the product of analysis and synthesis. ...(that) uncovers an enormous potential for recombination."

FBST - Full Bayesian Significance Test

Bayesian paradigm: the posterior density, $p_n(\theta)$, is proportional to the product of the likelihood and a prior density,

$$p_n(\theta) \propto L(\theta \mid x)\, p_0(\theta).$$

(Null) Hypothesis: $H: \ \theta \in \Theta_H$,

$$\Theta_H = \{\theta \in \Theta \mid g(\theta) \leq \mathbf{0} \wedge h(\theta) = \mathbf{0}\}$$

Precise hypothesis: $\dim(\Theta_H) < \dim(\Theta)$.

Reference density, $r(\theta)$, interpreted as a representation of no information in the parameter space, or the limit prior for no observations, or the neutral ground state for the Bayesian operation. Standard (possibly improper) uninformative references include the uniform and maximum entropy densities, see Dugdale (1996) and Kapur (1989)

FBST evidence value supporting and against the hypothesis H, $\mathrm{Ev}(H)$ and $\overline{\mathrm{Ev}}(H)$,

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

$$\widehat{s} = s(\widehat{\theta}) = \mathrm{sup}_{\theta \in \Theta} \, s(\theta) \ ,$$

$$s^* = s(\theta^*) = \mathrm{sup}_{\theta \in H} \, s(\theta) \ ,$$

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

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

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

$s(\theta)$ is the posterior surprise relative to $r(\theta)$.
The tangential set $\overline{T}$ is a HRSS.
(Highest Relative Surprise Set)
$W(v)$ is the cumulative surprise distribution.
$\mathrm{Ev}(H)$ is invariant under reparameterizations.
If $r \propto 1$ then $s(\theta) = p_n(\theta)$ and $\overline{T}$ is a HPDS.

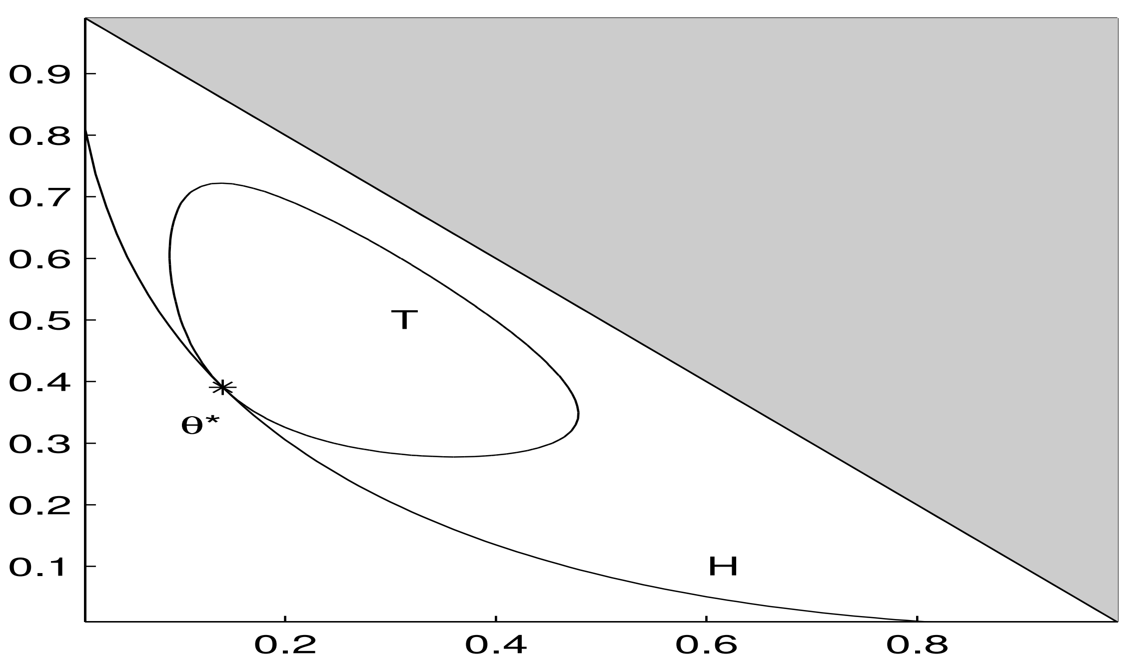
Hardy-Weinberg genetic equilibrium,
see (Pereira and Stern 1999).
n , sample size;
x_1, x_3 , homozygote sample counts;
$x_2 = n - x_1 - x_3$, heterozygote count.

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

$$r(\theta) \propto \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3} \ , \quad y = [-1, -1, -1] \ ,$$

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

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



Abstract Belief Calculus, ABC,
see Darwiche, Ginsberg (1992),
and Stern (2003).

$\langle \Phi, \oplus, \oslash \rangle$, Support Structure,
Φ , Support Function, for statements on $\mathcal{U}$.
Null and full support values are **0** and **1**.
$\oplus$, Support Summation operator,
$\oslash$, Support Scaling or Conditionalization,
$\langle \Phi, \oplus \rangle$, 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 $C = A \wedge B$,

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

Support structures for some belief calculi, $c = \Phi(C = A \wedge B)$.

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

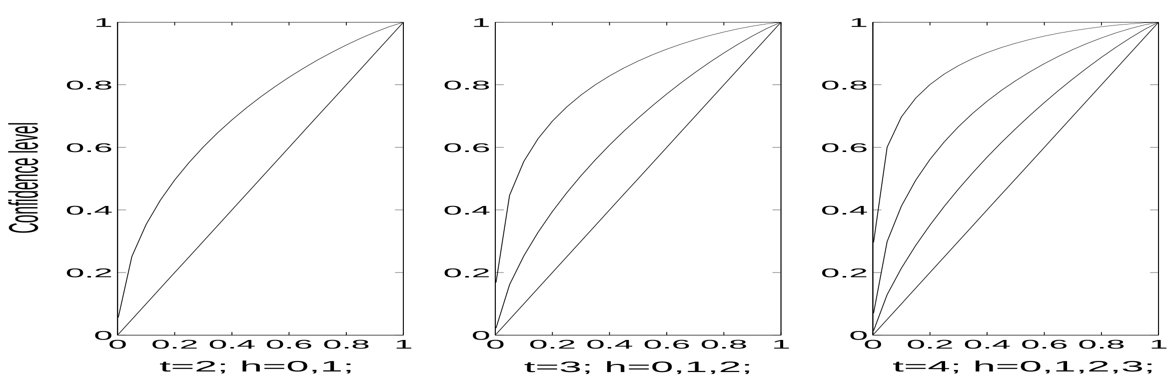
CL= Classical Logic, Pr= Probability,
Ps= Possibility, DB= Disbelief.

In the FBST setup, two belief calculi are in simultaneous use: Ev constitutes a possibilistic partial support structure coexisting in harmony with the probabilistic support structure given by the posterior probability measure in the parameter space.

Consistency:
$\overline{V}(c) = \Pr(\overline{\mathsf{Ev}} \leq c)$, the cumulative distribution of $\overline{\mathsf{Ev}}(H)$, given θ^0, the true parameter value.
Let $t = \dim(\Theta)$ and $h = \dim(H)$.
Under appropriate regularity conditions,
for increasing sample size, $n \to \infty$,

- If H is false, $\theta^0 \notin H$, then $\overline{\mathsf{Ev}}(H)$ converges (in probability) to one, $\overline{V}(c) \to \delta(1)$.
- If H is true, $\theta^0 \in H$, then $\overline{V}(c)$, the confidence level, is approximated by the function

$$\mathsf{Q}\left(t-h, \mathsf{Q}^{-1}(t,c)\right) \ .$$



Test τ_c critical level vs. confidence level

FBST Hypotheses Compositionality,
(or Composability of Eigen-Solutions):

Disjunction of (homogeneous) hypotheses $\Rightarrow$
Possibilistic support structure:

$$\text{Ev}\left(\bigvee_{i=1}^{q} H^i\right) = F\left(\max_{i=1}^{q} s^{*i}\right) = \max_{i=1}^{q}\left(\text{Ev}(H^i)\right) \ ,$$

Conjunction of (homogeneous) hypotheses $\Rightarrow$
Mellin convolution (truth operation):

$$\text{Ev}\left(\bigwedge_{j=1}^{k} H^j\right) = W(s^*) = \bigotimes_{1\leq j\leq k} W^j \left(\prod_{j=1}^{k} s^{*j}\right) \ ,$$

Ev, W, $\otimes$: Truth value, function, operation.
If all $s^* = 0 \vee \widehat{s}$, $\text{Ev} = 0 \vee 1$, classical logic.
Stern (2003) and Borges and Stern (2005).