Cybernetics And Human Knowing. Vol. 14, no. 1, pp. 9-36

Cognitive Constructivism, Eigen-Solutions, and Sharp Statistical Hypotheses

Julio Michael Stern[1]

In this paper epistemological, ontological and sociological questions concerning the statistical significance of sharp hypotheses in scientific research are investigated within the framework provided by Cognitive Constructivism and the FBST (Full Bayesian Significance Test). The constructivist framework is contrasted with the traditional epistemological settings for orthodox Bayesian and frequentist statistics provided by Decision Theory and Falsificationism.
Keywords: Autopoiesis, Cognitive constructivism, Epistemology, Ontology, Scientific hypotheses, Statistical significance tests, Social systems, Systems' theory.

1 Introduction

In this paper, a few epistemological, ontological and sociological questions concerning the statistical significance of sharp hypotheses in the scientific context are investigated within the framework provided by cognitive constructivism, or the constructivist theory as presented in Maturana and Varela (1980), Foerster (2003) and Luhmann (1989, 1990a, 1990b, 1995). Several conclusions of the study, however, remain valid, mutatis mutandis, within various other organizations and systems, see for example Bakken and Hernes (2002), Christis (2001), Mingers (2000), and Rasch (1998).

The author's interest in this research topic emerged from his involvement in the development of the Full Bayesian Significance Test (FBST), a novel Bayesian solution to the statistical problem of measuring the support of sharp hypotheses, first presented in Pereira and Stern (1999). The problem of measuring the support of sharp hypotheses poses several conceptual and methodological difficulties for traditional statistical analysis under both the frequentist (classical) and the orthodox Bayesian approaches. The solution provided by the FBST has significant advantages over traditional alternatives, in terms of its statistical and logical properties. Since these properties have already been thoroughly analyzed in previous papers (see references), the focus herein is directed exclusively to epistemological and ontological questions.
Despite the fact that the FBST is fully compatible with decision theory, as shown in Madruga, Esteves, & Wechsler (2001), which, in turn, provides a strong and coherent epistemological framework to orthodox Bayesian Statistics, its logical properties open the possibility of using and benefiting from alternative epistemological settings. In this article, the epistemological framework of constructivist theory is counterpoised to that of decision theory. The contrast, however, is limited in scope by our interest in

1. Institute of Mathematics and Statistics, University of São Paulo, Brazil. Email: jstern@ime.usp.br

statistics and is carried out in a rather exploratory an non exhaustive form. The epistemological framework of constructivist theory is also counterpoised to that of falsificationism, the epistemological framework within which classical frequentist statistical test of hypotheses are often presented, as shown in Boyd (1991) and Popper (1959, 1963).

In section 2, the fundamental notions of autopoiesis and eigen-solutions in autopoietic systems are reviewed. In section 3, the same is done with the notions of social systems and functional differentiation and in section 4, a constructivist theory view of science is presented. In section 5, the material presented in sections 2, 3 and 4 is related to the statistical significance of sharp scientific hypotheses and the findings therein are counterpoised to traditional interpretations such as those of decision theory. In section 6, a few sociological analyses for differentiation phenomena are reviewed. In sections 7 and 8, the final conclusions are established.

In sections 2, 3, 4, and 6, well established concepts of the constructivist theory are presented. However, in order to overcome an unfortunately common scenario, an attempt is made to make them accessible to a scientist or statistician who is somewhat familiar with traditional frequentist, and decision-theoretic statistical interpretations, but unfamiliar with the constructivist approach to epistemology. Rephrasing these concepts (once again) is also avoided. Instead, quoting the primary sources is preferred whenever it can be clearly (in our context) and synthetically done. The contributions in sections 5, 7 and 8, relate mostly to the analysis of the role of quantitative methods specifically designed to measure the statistical support of sharp hypotheses. A short review of the FBST is presented in Appendix A.

2 Autopoiesis and Eigen-Solutions

The concept of autopoiesis tries to capture an essential characteristic of living organisms (auto=self, poiesis=production). Its purpose and definition are stated by Maturana and Varela:

> 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. (Maturana & Varela, 1980, p. 84).

> 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. (Maturana & Varela, 1980, pp. 78-79).

Autopoietic systems are non-equilibrium (dissipative) dynamical 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 a single cell can be renewed thousands of times during its lifetime, see Bertalanffy (1969). The investigation of these regeneration processes in the autopoietic system production network leads to the definition of cognitive domain, as stated in Maturana and Varela (1980, p. 10):

> The circularity of their organization continuously brings them back to the same internal state (same with respect to the cyclic process). Each internal state requires that certain conditions (interactions with the environment) be satisfied in order to proceed to the next state. Thus the circular organization implies the prediction that an interaction that took place once will take place again. If this does not happen the system maintains its integrity (identity with respect to the observer) and enters into a new prediction. In a continuously changing environment these predictions can only be successful if the environment does no change in that which is predicted. Accordingly, the predictions implied in the organization of the living system are not predictions of particular events, but of classes of inter-actions. 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.

The characteristics of this circular (cyclic or recursive) regenerative processes and their eigen (auto, equilibrium, fixed, homeostatic, invariant, recurrent, recursive) -states, both in concrete and abstract autopoietic systems, are further investigated in Foerster (2003) and Segal (2001):

> 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 – in the domain of numbers, in functions, in behaviors, etc. (Segal, 2001, p. 145).

The concept of eigen-solution for an autopoietic system is the key to distinguish specific objects in a cognitive domain. Von Foerster also establishes several essential properties of eigen-solutions that will support the analyses conducted in this paper and conclusions established herein:

> Objects are tokens for eigen-behaviors. Tokens stand for something else. In exchange for money (a token itself for gold held by one's government, but unfortunately no longer redeemable), 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. (Segal, 2001, p. 127).

> 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. (Foerster, 2003, p. 266).

The arguments used in this study rely heavily on two qualitative properties of eigen-solutions, referred by von Foerster by the terms *discrete* and *equilibria*. In what follows, the meaning of these qualifiers, as they are understood by von Foerster and used herein, are examined:

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. Eigen-behavior 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 your-self with yourself. (Segal, 2001, p. 128).

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. Even in the familiar case of linear algebra, if we define the eigen-vectors corresponding to a singular eigen-value c of a linear transformation $T(\ )$ only by its essential property of directional invariance, $T(x) = cx$, we obtain one dimensional sub-manifolds which, in this case, are subspaces or lines trough the origin. Only if we add the usual (but non essential) normalization condition, $\|x\| = 1$, do we get discrete eigen-vectors.

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 $\{x_n\}$, defined by recursive application of the operator, $x_{n+1} = Op(x_n)$. Under appropriate conditions, such as within a domain of attraction, the process convergence and its limit eigen-solution will not depend on the starting point, x_0. In the linear algebra example, using almost any staring point, the sequence generated by the recursive relation $x_{n+1} = T(x_n)/\|T(x_n)\|$ that is the application of T followed by normalization, converges to the unitary eigen-vector corresponding to the largest eigen-value.

In sections 4 and 5 it is shown, 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. The statistical consequences of these findings will be discussed in sections 7 and 8. Before that, however, a few other constructivist theory concepts must be introduced in sections 3 and 6.

Autopoiesis found its name in the work of Maturana and Varela (1980), together with a simple, powerful and elegant formulation using the modern language of system's theory. Nevertheless, some of the basic theoretical concepts, such as those of self-organization and autonomy of living organisms, have long historical grounds that some authors trace back to Kant. As seen in Kant (1790, sec. 65) for example, an *organism* is characterized as one in which,

> every part is thought as 'owing' its presence to the 'agency' of all the remaining parts, and also as existing 'for the sake of the others' and of the whole, that is as an instrument, or organ.

> Its parts must in their collective unity reciprocally produce one another alike as to form and combination, and thus by their own causality produce a whole, the conception of which, conversely,—in a being possessing the causality according to conceptions that is adequate for such a product—could in turn be the cause of the whole according to a principle, so that, consequently, the nexus of 'efficient causes' (progressive causation, nexus effectivus) might be no less estimated as an 'operation brought about by final causes' (regressive causation, nexus finalis).

For further historical comments we refer the reader to Zeleny (1980).

3 Functional Differentiation

In order to give appropriate answers to environmental complexities, autopoietic systems can be hierarchically organized as Higher Order Autopoietic Systems. As in Maturana and Varela (1980), this notion is defined via the concept of Coupling:

> 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. (p. 107)

> An autopoietic system whose autopoiesis entails the autopoiesis of the coupled autopoietic units which realize it, is an autopoietic system of higher order. (p. 109)

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.

The philosopher and sociologist Niklas Luhmann applied this notion to the study of modern human societies and its systems. Luhmann's basic abstraction is to look at social systems only at its higher hierarchical level, in which it is seen as an autopoietic communications network. In Luhmann's terminology, a communication event consists of: *utterance*, the form of transmission; *information*, the specific content; and *understanding*, the relation to future events in the network, such as the activation or suppression of future communications.

> 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). (Luhmann, 1990b, p. 3).

For Luhmann, society's best strategy to deal with increasing complexity is the same as one observes 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 differentiate into systems. For Luhmann, modern societies are characterized by a vertical differentiation 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, and the media is the space in which the system operates.

Standard examples of social systems are:

- Science: with a true/false code, working in a program set by a scientific theory, and having articles in journals and proceedings as its media;
- Judicial: with a legal/illegal code, working in a program set by existing laws and regulations, and having certified legal documents as its media;
- Religion: with a good/evil code, working in a program set by sacred and hermeneutic texts, and having study, prayer and good deeds as its media;
- Economy: with a property/lack thereof code, working in a program set by economic planning scenarios and pricing methods, and having money and money-like financial assets as its media.

Before ending this section, a notion related to the break-down of autopoiesis is introduced: *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. In Luhmann's conception of modern society each system may be aware of events in other systems, that is, be cognitively open, but is required to maintain its differentiation, that is, be operationally closed.

4. Eigen-Solutions and Scientific Hypotheses

The interpretation of scientific knowledge as an eigensolution of a research process is part of a constructivistic approach to epistemology. Figure 1 presents an idealized structure and dynamics of knowledge production. This diagram represents, on the Experiment side (left column) the laboratory or field operations of an empirical science, where experiments are designed and built, observable effects are generated and measured, and the experimental data bank is assembled. On the Theory side (right column), the diagram represents the theoretical work of statistical analysis, interpretation and (hopefully) understanding according to accepted patterns. If necessary, new hypotheses (including whole new theories) are formulated, motivating the design of new experiments. Theory and experiment constitute a double feed-back cycle making it clear that the design of experiments is guided by the existing theory and its interpretation, which, in turn, must be constantly checked, adapted or modified in order to cope with the observed experiments. The whole system constitutes an autopoietic unit, as seen in Krohn and Küppers:

> The idea of knowledge as an eigensolution of an operationally closed combination between argumentative and experimental activities attempts to answer the initially posed question of how the construction of knowledge binds itself to its construction in a new way. The coherence of an eigensolution does not refer to an objectively given reality but follows from the operational closure of the construction. Still, different decisions on the selection of couplings may lead to different, equally valid eigen-solutions. Between such different solutions no reasonable choice is possible unless a new operation of knowledge is constructed exactly upon the differences of the given solutions. But again, this frame of reference for explicitly relating different solutions to each other introduces new choices with respect to the coupling of operations and explanations. It does not reduce but enhances the dependence of knowledge on decisions. On the other hand, the internal restrictions imposed by each of the chosen couplings do not allow for any arbitrary construction of results. Only few are suitable to mutually serve as inputs in a circular operation of knowledge. (1990, p. 214)

Experiment				Theory
Experiment design	$\Leftarrow$	Operation-alization	$\Leftarrow$	Hypotheses formulation
$\Downarrow$				$\Uparrow$
Effects observation		false / true eigen-solution		Inter-pretation
$\Downarrow$				$\Uparrow$
Data acquisition	$\Rightarrow$	Explanation	$\Rightarrow$	Statistical analysis
Sample space				Parameter space

Figure 1: Scientific Production Diagram.

5. Sharp Statistical Hypotheses

Statistical science is concerned with inference and application of probabilistic models. From what has been presented in the preceding sections, it becomes clear what the role of statistics in scientific research is, at least in the constructivist theory view of scientific research: Statistics has a dual task, to be performed both in the Theory and the Experiment sides of the diagram in Figure 1:

- At the Experiment side of the diagram, the task of statistics is to make probabilistic statements about the occurrence of pertinent events, that is describe probabilistic distributions for what, where, when or which events can occur. If the events are to occur in the future, these descriptions are called predictions, as is often the case in the natural sciences. It is also possible (more often in social sciences) to deal with observations related to past events, that may or may not be experimentally generated or repeated, imposing limitations to the quantity and/or quality of the available data. Even so, the habit of calling this type of statement "predictive probabilities" will be maintained.

- At the Theory side of the diagram, the role of statistics is to measure the statistical support of hypotheses, that is to measure, quantitatively, the hypotheses plausibility or possibility in the theoretical framework where they were formulated, given the observed data. From the material presented in the preceding sections, it is also clear that, in this role, statistics is primarily concerned with measuring the statistical support of sharp hypotheses, for hypotheses sharpness (precision or discreteness) is an essential characteristic of eigen-solutions.

Let us now examine how well the traditional statistical paradigms, and in contrast the FBST, are able to take care of this dual task. In order to examine this question, the first step is to distinguish what kind of probabilistic statements can be made. We make use of tree statement categories: frequentist, epistemic and Bayesian:

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 outcomes generated by a random variable. This generation may be actual or potential, that is, may have been realized or not, may be observable or not, may be repeated an infinite or finite number of times. Bayesian probabilistic statements are epistemic probabilistic statements generated by the (in practice, always finite) recursive use of Bayes formula:

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

In standard models, the parameter θ, a non observed random variable, and the sample x, an observed random variable, are related through their joint probability distribution, $p(x, \theta)$. The prior distribution, $p_0(\theta)$, is the starting point for the Bayesian recursion operation. It represents the initial available information about θ. In particular, the prior may represent no available information, like distributions obtained via the maximum entropy principle, see Dugdale (1996) and Kapur (1989). The posterior distribution, $p_n(\theta)$, represents the available information on the parameter after the n-th "learning step," in which Bayes formula is used to incorporate the information carried by observation x_n. Because of the recursive nature of the procedure, the posterior distribution in a given step is used as prior in the next step.

Frequentist 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. One way of doing this is by using a construction in which the whole data acquisition process is viewed as a single outcome of an imaginary infinite meta random process, and then make a frequentist statement, on the meta process, about the frequency of unsatisfactory outcomes of some incompatibility measure of the observed data bank with the hypothesis. This is the classic (and often forgotten) rationale used when stating a p-value. So we should always speak of the p-value of the data bank (not of the hypothesis). The resulting conceptual confusion and frustration (for most working scientists) with this kind of convoluted reasoning is captured by a wonderful parody of Galileo's dialogues in Rouanet et al. (1998).

A p-value is the probability of getting a sample that is more extreme than the one we got. We should therefore specify which criterion is used to define what we mean by more extreme, that is, how do we order the sample space, and usually there are several possible criteria to do that (e.g., Pereira & Wechsler, 1993).

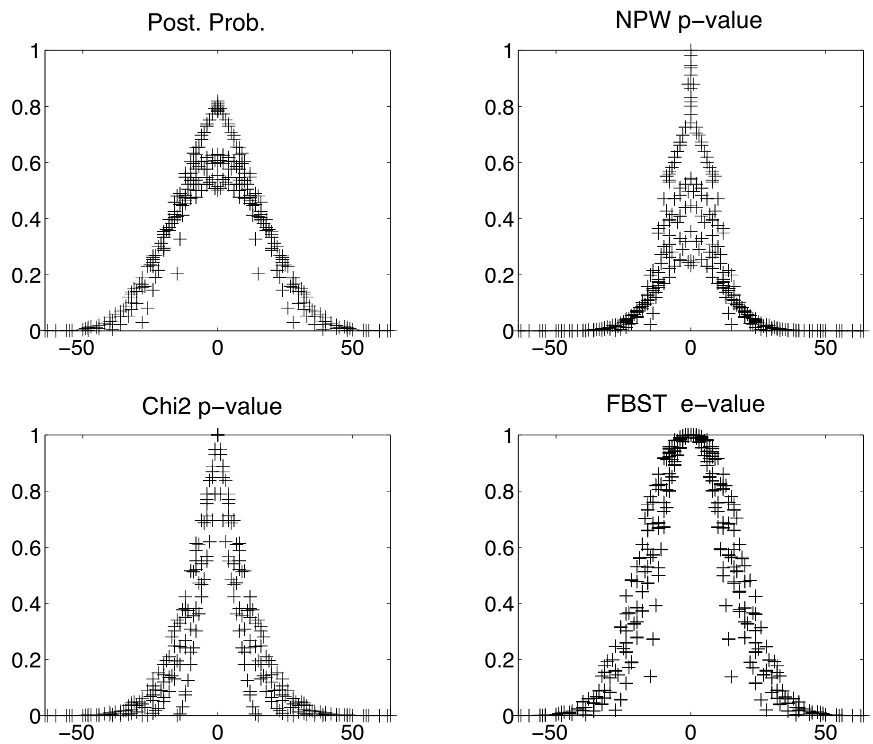


Figure 2: Independence Hypothesis, n=16

Figure 2 compares four statistics, namely, orthodox Bayesian posterior probabilities, Neyman-Pearson-Wald (NPW) p-values, Chi-square approximate p-values, and the FBST evidence value in favor of H. In this example H is the independence hypothesis in a *2 x 2* contingency table, for sample size n = 16, that is:

$$p_n(\theta \mid x) \propto \theta_1^{x_1+y_1}\theta_2^{x_2+y_2}\theta_3^{x_3+y_3}\theta_4^{x_4+y_4}, \; p_0(\theta) \propto \theta_1^{y_1}\theta_2^{y_2}\theta_3^{y_3}\theta_4^{y_4}, \; y = [0,0,0,0],$$

$$\Theta = \{\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 + \theta_4 = 1\} \;, \quad H : \theta_{1,1} = (\theta_{1,1} + \theta_{1,2})(\theta_{1,1} + \theta_{2,1}) \; .$$

The horizontal axes shows the "diagonal asymmetry" statistics (difference between the diagonal products). The statistic D is an estimator of an unormalized version of Pearson correlation coefficient, ρ. For detailed explanations see Irony et al. (1995, 2000), Stern and Zacks (2002), and Madruga, Pereira, and Stern (2003).

$$D = x_{1,1}x_{2,2} - x_{1,2}x_{2,1} \;, \quad \rho = \frac{\sigma_{1,2}}{\sigma_{1,1}\sigma_{2,2}} = \frac{\theta_{1,1}\theta_{2,2} - \theta_{1,2}\theta_{2,1}}{\sqrt{\theta_{1,1}\theta_{1,2}\theta_{2,1}\theta_{2,2}}} \; .$$

Samples that are "perfectly compatible with the hypothesis," that is, having no asymmetry, are near the center of the plot, with increasingly incompatible samples to the sides. The envelope curve for the resulting FBST e-values, to be commented later in this section, is smooth (differentiable) and therefore level at its maximum, where it reaches the value 1. In contrast the envelope curves for the p-values take the form of a cusp, that is a pointed curve, that is broken (non differentiable) at its maximum, where it also reaches the value one. The acuteness of the cusp also increases with increasing sample size. In the case of NPW p-values we see, at the top of the cusp, a *ladder* or spike, with several samples with no asymmetry, but having different outcome probabilities, "competing" for the higher p-value.

This is a typical collateral effect of the artifice that converts a question about the significance of H, asking for a probability in the parameter space as an answer, into a question, conditional on H being truth, about the outcome probability of the observed sample, offering a probability in the sample space as an answer. This qualitative analysis of the p-value methodology gives us an insight on the frequent abuses of expressions like "increase sample size to reject." In the words of I. J. Good (1983, p. 135):

> Very often the statistician doesn't bother to make it quite clear whether his null hypothesis is intended to be sharp or only approximately sharp. ... It is hardly surprising then that many Fisherians (and Popperians) say that you can't get (much) evidence in favor of the null hypothesis but can only refute it.

In Bayesian statistics we are allowed to make 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 decision theory in orthodox Bayesian statistics. Since the pioneering work of de Finetti, Savage and many others, orthodox Bayesian Statistics has developed strong and coherent foundations grounded on decision theory, where many basic questions could be successfully analyzed and solved.

This foundations can be stratified in two layers:

- In the first layer, decision theory provides a coherence system for the use of probability statements, in the sense of Finetti (1974, 1981, 1991). In this context, the FBST use of probability theory is fully compatible with decision theory, as shown in Madruga et al. (2001).
- In the second layer, decision theory provides an epistemological framework for the interpretation of statistical procedures. The FBST logical properties open the possibility of using and benefiting from alternative epistemological settings such as constructivist theory. Hence, decision theory does not have to be "the tool for all trades."

We claim that, in the specific case of statistical procedures for measuring the support (significance tests) for sharp scientific hypotheses, constructivist theory provides a more adequate epistemological framework than decision theory. This point is as important as it is subtle. In order to understand it let us first remember the orthodox paradigm, as it is concisely stated in Dubins and Savage (1965, 12.8). In a second quote, from Savage (1954, 16.3) we find that sharp hypotheses, even if important, make little sense in this paradigm, a position that is accepted throughout decision theoretic Bayesian statistics, as can also be seen in Levi (1974) and Maher et al. (1993).

> Gambling problems in which the distributions of various quantities are prominent in the description of the gambler's fortune seem to embrace the whole of theoretical statistics according to one view (which might be called the decision-theoretic Bayesian view) of the subject. … From the point of view of decision-theoretic statistics, the gambler in this problem is a person who must ultimately act in one of two ways (the two guesses), one of which would be appropriate under one hypothesis (H_0) and the other under its negation (H_1). … Many problems, of which this one is an instance, are roughly of the following type. A person's opinion about unknown parameters is described by a probability distribution; he is allowed successively to purchase bits of information about the parameters, at prices that may depend (perhaps randomly) upon the unknown parameters themselves, until he finally chooses a terminal action for which he receives an award that depends upon the action and parameters. (Durbins & Savage, 1965, sec.12.8, p.229-230)

> I turn now to a different and, at least for me, delicate topic in connection with applications of the theory of testing. Much attention is given in the literature of statistics to what purport to be tests of hypotheses, in which the null hypothesis is such that it would not really be accepted by anyone. ... extreme (sharp) hypotheses, as I shall call them... 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. (Savage, 1954, sec.16.3, p.254).

As it is clearly seen, in the decision theory framework we speak about the betting odds for "the hypothesis wining on a gamble taking place in the parameter space." But since sharp hypotheses are zero (Lebesgue) measure sets, our betting odds must be null, that is sharp hypotheses must be (almost surely) false. If we accept the constructivist theory view that an important class of hypotheses concern the identification of eigen-solutions, and that those are ontologically sharp, we have a paradox!

From these considerations it is not surprising that frequentist and decision theory orthodoxy consider sharp hypotheses, at best as anomalous crude approximations used when the scientist is incapable of correctly specifying error bounds, cost, loss or utility functions, etc., or then just consider them to be just plain silly. In the words of D. Williams (2002, p. 234):

> Bayesian significance of sharp hypothesis: a plea for sanity: … It astonishes me therefore that some Bayesian now assign non-zero prior probability that a sharp hypothesis is exactly true to obtain results which seem to support strongly null hypotheses which frequentists would very definitely reject. (Of course, it is blindingly obvious that such results must follow).

But no matter how many times statisticians reprehend scientist for their sloppiness and incompetence, they keep formulating sharp hypotheses, as if they where magnetically attracted to them. From the constructivist theory plus FBST perspective they are, of course, just doing the right thing!

Decision theoretic statistics has also developed methods to deal with sharp hypotheses, posting sometimes a scary caveat emptor for those willing to use them. The best known of such methods are Jeffreys' tests, based on Bayes Factors, assigning a positive prior probability mass on the sharp hypothesis. This positive prior mass is supposed to work like a handicap system designed to balance the starting odds and make the game "fair." Out of that we only get new paradoxes, like the well documented Lindley's paradox. In opposition to its frequentist counterpart, this is an "increase sample size to accept" effect, see Shafer (1982).

The FBST e-value or evidence value supporting the hypothesis, $\mathrm{ev}(H)$, was specially designed to effectively evaluate the support for a sharp hypothesis, H. This support function is based on the posterior probability measure of a set called the tangential set, $\overline{T}(H)$, which is a non zero measure set (so no null probability paradoxes), see Pereira and Stern (1999), Madruga et al. (2003) and subsection A1 of the appendix.

Although $\mathrm{ev}(H)$ is a probability in the parameter space, it is also a possibilistic support function. The word *possibilistic* carries a heavy load, implying that $\mathrm{ev}(H)$ complies with a very specific logic (or algebraic) structure, as seen in Darwishe and

Ginsberg (1992), Stern (2003, 2004), and subsection A3 of the appendix. Furthermore the e-value has many necessary or desirable properties for a statistical support function, such as:

1. Give an intuitive and simple measure of significance for the hypothesis in test, ideally, a *probability* defined directly in the original or *natural parameter space*.
2. Have an intrinsically geometric definition, independent of any non-geometric aspect, like the particular parameterization of the (manifold representing the) null hypothesis being tested, or the particular coordinate system chosen for the parameter space, that is, be an *invariant* procedure.
3. Give a measure of significance that is smooth, that is *continuous and differentiable*, on the hypothesis parameters and sample statistics, under appropriate regularity conditions of the model.
4. Obey the *likelihood principle*, that is, the information gathered from observations should be represented by, and only by, the likelihood function.
5. Require *no ad hoc artifice* like assigning a positive prior probability to zero measure sets, or setting an arbitrary initial belief ratio between hypotheses.
6. Be a *possibilistic* support function.
7. Be able to provide a *consistent* test for a given sharp hypothesis.
8. Be able to provide *compositionality* operations in complex models.
9. Be an *exact* procedure, not requiring the use of "large sample" asymptotic approximations.
10. Allow the incorporation of previous experience or expert's opinion via (subjective) *prior distributions*.

For a careful and detailed explanation of the FBST definition, its computational implementation, statistical and logical properties, and several already developed applications, the reader is invited to consult some of the articles in the reference list. Appendix A provides a short review of the FBST, including its definition and main properties.

6. Semantic Degradation

In this section some constructivist analyses of dedifferentiation phenomena in social systems are reviewed. If the conclusions in the last section are correct, it is surprising how many times decision theory, sometimes with a very narrow pseudo-economic interpretation, was misused in scientific statistical analysis. The difficulties of testing sharp hypotheses in the traditional statistical paradigms are well documented, and extensively discussed in the literature, see for example the articles in Harlow, Muliak, and Steiger (1997). We hope the material in this section can help us understand these difficulties as symptoms of problems with much deeper roots. By no means the author is the first to point out the danger of analyses carried out by blind transplantation of categories between heterogeneous systems. In particular, regarding the abuse of economical analyses, Luhmann (1989) states:

> 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. (Luhmann, 1989, p. 164).

For a sociological analysis of this phenomenon in the context of science, see for example Fuchs (1996) and DiMaggio and Powell (1991):

> higher-status sciences may, more or less aggressively, colonize lower-status fields in an attempt at reducing them to their own First Principles. For particle physics, all is quarks and the four forces. For neurophysiology, consciousness is the aggregate outcome of the behavior of neural networks. For sociobiology, philosophy is done by ants and rats with unusual large brains that utter metaphysical nonsense according to acquired reflexes. In short, successful and credible chains or reductionism usually move from the top to the bottom of disciplinary prestige hierarchies. (Fuchs, 1996, p. 310).

> This may explain the popularity of giving an "economical understanding" to processes in functionally distinct areas even if (or perhaps because) this semantics is often hidden by statistical theory and methods based on decision theoretic analysis. This also may explain why some areas, like ecology, sociology or psychology, are (or where) far more prone to suffer this kind of dedifferentiation by semantic degradation than others, like physics. (DiMaggio & Powell, 1991, p. 63).

Once the forces pushing towards systemic degradation are clearly exposed, we hope one can understand the following corollary of von Foerster famous ethical and aesthetical imperatives:

- Theoretical imperative: Preserve systemic autopoiesis and semantic integrity, for de-differentiation is in-sanity itself.
- Operational imperative: Choose the right tool for each job: "If you only have a hammer, everything looks like a nail."

7. Competing Sharp Hypotheses

In this section we examine the concept of competing sharp hypotheses. This concept has several variants, but the basic idea is that a good scientist should never test a single sharp hypothesis, for it would be an unfair faith of the poor sharp hypothesis standing all alone against everything else in the world. Instead, a good scientist should always confront a sharp hypothesis with a competing sharp hypotheses, making the test a fair game. As seen in Good (1983):

> Since I regard refutation and corroboration as both valid criteria for this demarcation it is convenient to use another term, Checkability, to embrace both processes. I regard checkability as a measure to which a theory is scientific, where checking is to be taken in both its positive and negative senses, confirming and disconfirming. (Good, 1983, p. 167)

> If by the truth of Newtonian mechanics we mean that it is approximately true in some appropriate well defined sense we could obtain strong evidence that it is true; but if we mean by its truth that it is exactly true then it has already been refuted. (Good, 1983, p. 135)

> I think that the initial probability is positive for every self-consistent scientific theory with consequences verifiable in a probabilistic sense. No contradiction can be inferred from this assumption since the number of statable theories is at most countably infinite (enumerable). (Good, 1983, p. 126)

> It is very difficult to decide on numerical values for the probabilities, but it is not quite so difficult to judge the ratio of the subjective initial probabilities of two theories by comparing their complexities. This is one reason why the history of science is scientifically important. (Good, 1983, p.126).

The competing sharp hypotheses argument does not directly contradict the epistemological framework presented in this article, and it may be appropriate under certain circumstances. It may also mitigate or partially remediate the paradoxes pointed out in the previous sections when testing sharp hypotheses in the traditional frequentist or orthodox Bayesian settings. However, the author does not believe that having competing sharp hypotheses is neither a necessary condition for good science practice, nor an accurate description of science history.

Just to stay with Good's example, let us quickly examine the very first major incident in the tumultuous debacle of Newtonian mechanics. This incident was Michelson's experiment on the effect of "aethereal wind" over the speed of light, see Michelson and Morley (1887) and Lorentz et al. (1952). A clear and lively historical account to this experiment can be found in Jaffe (1960). Actually Michelson found no such effect, that is he found the speed of light to be constant, invariant with the relative speed of the observer. This result, a contradiction in Newtonian mechanics, is easily explained by Einstein's special theory of relativity. The fundamental difference between the two theories is their symmetry or invariance groups: Galileo's group for Newtonian mechanics, Lorentz' group for special relativity. A fundamental result of physics, Noether's Theorem, states that for every continuous symmetry in a physical theory, there must exist an invariant quantity or conservation law. For detail the reader is referred to Doncel, Hermann, Michel, and Pais (1987), Fleming (1979), Gruber and Millman (1980), Gruber and Lenczewski (1986), Gruber and Iachello (1989), Gruber and Yopp (1990), Houtappel, Dam, and Wigner (1965), French (1968), Landau and Lifchitz (1966), Noether (1918), Wigner (1970), Weyl (1952). Conservation laws are sharp hypotheses ideally suited for experimental checking. Hence, it seems that we are exactly in the situation of competing sharp hypotheses, and so we are today, from a far away historical perspective. But this is a post-mortem analysis of Newtonian mechanics. At the time of the experiment there was no competing theory. Instead of confirming an effect, specified only within an order of magnitude, Michelson found, for his and everybody else's astonishment, an, up to the experiment's precision, null effect.

Complex experiments like Michelson's require a careful analysis of experimental errors, identifying all significant source of measurement noise and fluctuation. This kind of analysis is usual in experimental physics, and motivates a brief comment on a secondary source of criticism on the use of sharp hypotheses. In the past, one often had to work with over simplified statistical models. This situation was usually imposed by limitations such as the lack of better or more realistic models, or the unavailability of the necessary numerical algorithms or the computer power to use them. Under these limitations, one often had to use minimalist statistical models or approximation techniques, even when these models or techniques were not recommended. These models or techniques were instrumental to provide feasible tools for statistical analysis, but made it very difficult to work (or proved very ineffective) with complex systems, scarce observations, very large data sets, and so forth. The need to work with complex models, and other difficult situations requiring the use of sophisticated statistical methods and techniques, is very common (and many times inescapable) in research areas dealing with complex systems like biology, medicine, social sciences, psychology, and many other fields, some of them distinguished with the mysterious appellation of "soft" science. A colleague once put it to me like this: "It seems that physics got all the easy problems ... ."

If there is one area where the computational techniques of Bayesian statistics have made dramatic contributions in the last decades, that is the analysis of complex models. The development of advanced statistical computational techniques like Markov Chain Monte Carlo (MCMC) methods, Bayesian and neural networks, random fields models, and many others, make us hope that most of the problems related to the use of over simplified models can now be overcome. Today good statistical practice requires all statistically significant influences to be incorporated into the model, and one seldom finds an acceptable excuse not to do so; see also Pereira and Stern (2001).

8 Final Remarks

It should once more be stressed that most of the material presented in sections 2, 3, 4, and 6 is not new in constructivist theory. Unfortunately constructivist theory has had a minor impact in statistics, and sometimes provoked a hostile reaction from the ill-informed. One possible explanation of this state of affairs may be found in the historical development of constructivist theory. The constructivist reaction to a dogmatic (metaphysical) realism prevalent in hard sciences, specially in the XIX and the beginning of the XX century, raised a very outspoken rhetoric intended to make explicitly clear how naive and fragile the foundations of this over simplistic realism were. This rhetoric was extremely successful, quickly awakening and forever changing the minds of those directly interested in the fields of history and philosophy of science, and spread rapidly into many other areas. Unfortunately the same rhetoric could, in a superficial reading, make constructivist theory be perceived as either hostile or intrinsically incompatible with the use of quantitative and statistical methods, or leading to an extreme forms of subjectivism.

In constructivist theory, or (Objective) Idealism as presented in this article, neither does one claim to have access to a "thing in itself" or "Ding an sich" in the external environment, as do dogmatic forms of realism, nor does one surrender to solipsism, as do skeptic forms of subjectivism, including some representatives of the subjectivist school of probability and statistics. For the use of some of these terms in Kantian philosophy see Bonaccini (2000) and Caygill (1995), for subjectivism in statistics see Finetti (1974, 1.11, 7.5.7). In fact, it is the role of the external constraints imposed by the environment, together with the internal autopoietic relations of the system, to guide the convergence of the learning process to precise eigen-solutions, these being at the end, the ultimate or real objects of scientific knowledge. As stated by Luhmann (1990b, 1995):

> constructivism maintains nothing more than the unapproachability of the external world "in itself" and the closure of knowing - without yielding, at any rate, to the old skeptical or "solipsistic" doubt that an external world exists at all. (Luhmann, 1990a, p. 65).

> at least in systems theory, they (statements) refer to the real world. Thus the concept of system refers to something that in reality is a system and thereby incurs the responsibility of testing its statements against reality. (Luhmann, 1995, p.12).

> both subjectivist and objectivist theories of knowledge have to be replaced by the system/environment distinction, which then makes the distinction subject/object irrelevant. (Luhmann, 1990a, p. 66).

The author hopes to have shown that constructivist theory not only gives a balanced and effective view of the theoretical/experimental aspects of scientific research but also that it is well suited (or even better suited) to give the necessary epistemological foundations for the use of quantitative methods of statistical analysis needed in the practice of science. It should also be stressed, according to author's interpretation of constructivist theory, the importance of measuring the statistical support for sharp hypotheses. In this setting, the author believes that, due to its statistical and logical characteristics, the FBST is the right tool for the job, and hopes to have motivated the reader to find more about the FBST definition, theoretical properties, efficient computational implementation, and several of the already developed applications, in some of the articles in the reference list. This perspective opens interesting areas for further research. Among them, we mention the following two.

8.1 Noether and de Finetti Theorems

The first area for further research has to do with some similarities between Noether theorems in physics, and de Finetti type theorems in statistics. Noether theorems provide invariant physical quantities or conservation laws from symmetry transformation groups of the physical theory, and conservation laws are sharp hypotheses by excellence. In a similar way, de Finetti type theorems provide invariant distributions from symmetry transformation groups of the statistical model. Those invariant distributions 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, see Feller (1971, ch.7) and also Diaconis (1987, 1988), Eaton (1989), Nachbin (1965) and Ressel (1987).

8.2 Compositionality

The second area for further research has to do with one of the properties of eigen-solutions mentioned by von Foerster that has not been directly explored in this article, namely that eigen-solutions are "composable," see Borges and Stern (2005) and section A4. Compositionality properties concern the relationship between the credibility, or truth value, of a complex hypothesis, H, and those of its elementary constituents, H^j, $j = 1 \ldots k$. Compositionality questions play a central role in analytical philosophy (e.g., see Conde, 1988)

According to Wittgenstein (1961):

> Every complex statement can be analyzed from its elementary constituents…(sec. 2.0201)
>
> Truth values of elementary statement are the results of those statements' truth-functions (Wahrheitsfunktionen). (sec. 5.0)
>
> All truth-function are results of successive applications to elementary constituents of a finite number of truth-operations (Wahrheitsoperationen). (sec. 5.32)

Compositionality questions also play a central role in far more concrete contexts, like that of reliability engineering. Birnbaum, Esary, and Saunders state:

> 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. (1961, sec. 1.4)

In Luhmann we find the following remark on the evolution of science that directly hints the importance of this property:

> After the (science) system worked for several centuries under these conditions it became clear where it was leading. This is 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. In this case analysis is what is most important because the further decomposition of the visible world into still further decomposable molecules and atoms, into genetic structures of life or even into the sequence human/

role/action/ action-components as elementary units of systems uncovers an enormous potential for recombination. (Luhmann, 1989, p.79)

In the author's view, the composition (or re-combination) of scientific knowledge and its use, so relevant in technology development and engineering, can give us a different perspective (perhaps a, bottom-up, as opposed to the top-down perspective in this article) on the importance of sharp hypotheses in science and technology practice. It can also provide some insight on the valid forms of iteration of science with other social systems or, in Luhmann's terminology, how science does (or should) "resonate" in human society.

Acknowledgements

The author has benefited from the support of FAPESP, CNPq, BIOINFO, the Computer Science Department at São Paulo University, Brazil, and the Mathematical Sciences Department at SUNY-Binghamton, USA. The author is grateful to many people for helpful discussions, most specially, Wagner Borges, Carlos Humes, Joseph Kadane, Luis Gustavo Esteves, Marcelo Lauretto, Fabio Nakano, Carlos Alberto de Braganç a Pereira, Rafael Bassi Stern, Sergio Wechsler, and Shelemyahu Zacks. The author also received several interesting comments and suggestions from the participants of FIS-2005, the Third Conference on the Foundations of Information Science.

References

Bakken, T., & Hernes, T. (2002). *Autopoietic organization theory: Drawing on Niklas Luhmann's social systems perspective*. Copenhagen: Copenhagen Business School.

Bertalanffy, L.von (1969). *General system theory.* New York: George Braziller.

Birnbaum, Z.W., Esary, J. D., & Saunders, S. C. (1961). Multicomponent systems and structures, and their reliability. *Technometrics, 3*, 55-77.

Bonaccini, J. A. (2000). *Kant e o Problema da Coisa Em Si no Idealismo Alemão*. São Paulo: Relume Dumarà.

Borges, W., & Stern, J. M. (2005). On the truth value of complex hypothesis. *CIMCA-2005 International Conference on Computational Intelligence for Modelling Control and Automation*. Los Alamitos, CA: IEEE.

Boyd, R., Gasper, P., & Trout, J. D. (1991). *The philosophy of science*. Cambridge, MA: The MIT Press.

Caygill, H. (1995). *A Kant dictionary.* Oxford: Blackwell.

Christis, J. (2001). Luhmann's theory of knowledge: Beyond realism and constructivism? *Soziale Systeme, 7*, 328–349.

Conde, M. L. (1998). *Wittgenstein: Linguagem e Mundo*. São Paulo: Annablume.

Costa, N. C., Abe, J. M., Murolo, A. C., & Silva, J. I. da, Casemiro, C.F.S. (1999). *Lógica Paraconsistente Aplicada*. São Paulo: Atlas.

Darwiche, A. Y. (1993). *A symbolic generalization of probability theory.* Unpublished dissertation, Stanford University, Stanford, CA.

Darwiche, A. Y., & Ginsberg, M. L. (1992). A Symbolic Generalization of Probability Theory. In *Proceedings of the 10th Conference of the American Association for Artificial Intelligence* (pp. 622-627).San Jose, CA: AAAI.

Diaconis, P., & Freeman, D. (1987). A Dozen de Finetti Style Results in Search of a Theory. *Ann. Inst. Poincaré Probab. Stat*., 23, 397–423.

Diaconis, P. (1988). *Group representation in probability and statistics*. Hayward, CA: IMA.

DiMaggio, P., & Powell, W. W. (1991). *New institutionalism in organizational analysis*. Chicago: Chicago Univ. Press.

Doncel, M. G., Hermann, A., Michel, L., & Pais, A. (1987). *Symmetries in physics (1600-1980).* Seminari d'Història des les Ciénces. Universitat Autò noma de Barcelona.

Dubins, L. E., Savage, L. J. (1965). *How to gamble if you must: Inequalities for stochastic processes*. New York: McGraw-Hill.

Dugdale, J. S. (1996). *Entropy and Its physical meaning*. London: Taylor & Francis.

Eaton, M.L. (1989). *Group invariance applications in statistics*. Hayward, CA: IMA.

Evans, M. (1997). Bayesian inference procedures derived via the concept of relative surprise. *Communications in Statistics*, *26*, 1125–1143.

Feller, W. (1971). *An introduction to probability theory and its applications,* Vol. 2 (2nd ed.) New York: Wiley.

Finetti, B. de (1974). *Theory of probability,* V1 and V2. London: Wiley.

Finetti, B. de (1981& 1991). *Scritti*. V1 and V2. Padova, Italy: CEDAM.

Fleming, H. (1979). As Simetrias como Instrumento de Obtenç ão de Conhecimento. *Ciência e Filosofia, 1*, 99–110.

Foerster, H. von (2003). *Understanding understanding: Essays on cybernetics and cognition*. New York: Springer. (The following articles (chapters) are of special interest: (1) On Self-Organizing Systems and their Environments, 1–19; (8) On Constructing a Reality, 211–227; (11) Objects: Tokens for Eigen-Behaviors, 261–271; (15) For Niklas Luhmann: How Recursive is Communication? 305–323.)

French, A. P. (1968). *Special relativity.* New York: Chapman and Hall.

Fuchs, S. (1996). The new wars of truth: Conflicts over science studies as differential modes of observation. *Social Science Information, 35*, 307–326.

Gaskins, R. H. (1992). *Burdens of proof in modern discourse*. Princeton: Yale Univ. Press.

Ghosh, J. K. (1988). Statistical information and likelihood. *Lecture Notes in Statistics, 45: A collection of critical essays by Dr. D. Basu*. Berlin: Springer.

Good, I. J. (1981). *Good thinking*. Minneapolis:University of Minnesota Press.

Gruber, B., & Millman, R. S. (Eds.) (1980). *Symmetries in science*. New York: Plenum Press.

Gruber, B., Lenczewski ,R. (Eds.) (1986). *Symmetries in science II*. New York: Plenum Press.

Gruber, B., Iachello, F. (Eds.) (1989). *Symmetries in science III*. New York: Plenum Press.

Gruber, B., Yopp, J. H. (Eds.) (1990). *Symmetries in science IV.* New York: Plenum Press.

Harlow, L. L., Mulaik, S. A., Steiger, J. H. (1997). *What if there were no significance tests?* London: Lawrence Erlbaum Associates.

Houtappel, R., Dam, H. van, & Wigner, E. P. (1965). The conceptual basis and use of the geometric invariance principles. *Reviews of Modern Physics, 37*, 595–632.

Irony, T. Z., & Pereira, C. A. B. (1995), Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses, *Resenhas, 2*, 27-46.

Irony, T. Z., Pereira, C. A. B., & Tiwari, R. C. (2000). Analysis of opinion swing: Comparison of two correlated proportions, *The American Statistician, 54*, 57–62.

Jaffe, B. (1960). *Michelson and the speed of light*. New York: Anchor.

Kadane, J. B., &Winkler, R. L. (1987). De Finetti's methods of elicitation. In R. Viertl (Ed.), *Probability and Bayesian statistics* (pp. 279-284). New York: Plenum.

Kant, I. (1790). Critique of teleological judgement. In *The critique of pure reason: The critique of practical reason; the critique of judgment*. Encyclopaedia Britannica Great Books of Western World, 42, 1952.

Kapur, J. N. (1989). *Maximum entropy models in science and engineering*. New York: Wiley.

Krohn, W., & Küppers, G. (1990). The selforganization of science - Outline of a theoretical model. In W. Krohn, G. Küppers, & H. Nowotny (Eds.), *Selforganization: Portrait of a scientific revolution* (pp. 208–222). Dordrecht: Kluwer.

Landau, L. D., & Lifchitz, E. M. (1966). *Cours de Physique Théorique*. Moscow: MIR.

Lauretto, M., Pereira, C. A. B., Stern, J. M., & Zacks, S. (2003). Comparing parameters of two bivariate normal distributions using the invariant full Bayesian significance test. *Brazilian Journal of Probability and Statistics, 17*, 147-168.

Levi, I. (1974). *Gambling with truth: An essay on induction and the aims of science*. Cambridge, MA: The MIT Press.

Lorentz, H. A., Einstein, A., Minkowski, H., & Weyl, H. (1952). *The principle of relativity: A collection of original memoirs on the special and general theory of relativity.* New York: Dover.

Luhmann, N. (1989). *Ecological communication*. Chicago: Chicago Univ. Press.

Luhmann, N. (1990a). The cognitive program of constructivism and a reality that remains unknown. In W. Krohn, G. Küppers, & H. Nowotny (Eds.), *Selforganization: Portrait of a scientific revolution* (pp. 64–86). Dordrecht: Kluwer.

Luhmann, N. (1990b). *Essays on self-reference*. New York: Columbia Univ. Press.

Luhmann, N. (1995). *Social systems*. Stanford, CA: Stanford Univ. Press.

Madruga, M. R., Esteves, L. G., & Wechsler, S. (2001). On the Bayesianity of Pereira-Stern tests. *Test*, *10*, 291–299.

Madruga, M. R., Pereira, C. A. B., & Stern, J. M. (2003). Bayesian evidence test for precise hypotheses. *Journal of Statistical Planning and Inference, 117*, 185–198.

Maher, P., Skyrms, B., Adams, E. W., Binmore, K., Butterfield, J., Diaconis, P., & Harper, W. L. (1993). *Betting on theories*. New York: Cambridge Univ. Press.

Maturana, H. R.,& Varela, F. J. (1980). *Autopoiesis and cognition. The realization of the living*. Dordrecht: D. Reidel.

Michelson, A. A., & Morley, E. W. (1887). On the relative motion of the earth and the luminiferous ether. *American Journal of Physics, 34*, 333–345.

Mingers, J. (1995) *Self-producing systems: Implications and applications of autopoiesis*. New York: Plenum.

Noether, E. (1918). Invariante Varlationsprobleme. *Nachrichten der Könighche Gesellschaft der Wissenschaften zu Göttingen,* (pp. 235–257). (English version: *Transport Theory and Statistical Physics,* 1971, 1, 183–207).

Nachbin, L. (1965). *The Haar integral*. New York: Van Nostrand.

Pereira, C.A.B., Wechsler, S. (1993). On the concept of p-value. *Brazilian Journal of Probability and Statistics*, *7*, 159–177.

Pereira, C. A. B., & Stern, J. M. (1999). Evidence and credibility: Full Bayesian significance test for precise hypotheses. *Entropy Journal*, *1*, 69–80.
Pereira, C. A. B., & Stern, J. M. (2001). Model selection: Full Bayesian approach. *Environmetrics*, *12*, 559–568.
Popper, K. R. (1959). *The logic of scientific discovery.* New York: Routledge.
Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York: Routledge.
Rasch, W. (1998). Luhmann's Widerlegung des Idealismus: Constructivism as a two-front war. *Soziale Systeme, 4*, 151–161.
Ressel, P. (1987). A very general De Finetti type theorem. In: R. Viertl (Ed.), *Probability and Bayesian statistics* (pp. 403-414). New York: Plenum.
Rouanet, H., Bernard, J. M., Bert, M. C., Lecoutre, B., Lecoutre, M. P., & Roux, B. Le (1998). *New ways in statistical methodology: From significance tests to Bayesian inference*. Berne: Peter Lang.
Royall, R. (1997). *Statistical evidence: A likelihood paradigm*. London: Chapman & Hall.
Rubin, H. (1987). A weak system of axioms for "rational" behaviour and the non-separability of utility from prior. *Statistics and Decisions*, *5*, 47–58.
Savage, L. J. (1972): *The foundations of statistics*. New York: Dover. (Reprint of The foundations of statistics, by L. J. Savage, 1954, New York: Wiley.)
Segal, L. (2001). *The dream of reality: Heinz von Foerster's constructivism.* New York: Springer.
Shafer, G. (1982), Lindley's paradox. *J. American Statistical Assoc*., *77*, 325–51.
Springer, M. D. (1979). *The algebra of random variables*. New York: Wiley.
Stern, J. M. (2003). Significance tests, belief calculi, and burden of proof in legal and scientific discourse. Laptec'03, *Frontiers in Artificial Intelligence and its Applications*, *101*, 139–147.
Stern, J. M. (2004). Paraconsistent sensitivity analysis for Bayesian significance tests. SBIA'04, *Lecture Notes Artificial Intelligence, 3171*, 134–143.
Stern, J. M., & Zacks, S. (2002). Testing independence of poisson variates under the Holgate bivariate distribution: The power of a new evidence test. *Statistical and Probability Letters*, *60*, 313–320.
Weyl, H. (1952). *Symmetry.* Princeton: Princeton Univ. Press.
Wigner, E. P. (1970). *Symmetries and reflections*. Cambridge, MA: The MIT Press.
Williams, D. (2001) *Weighing the odds*. New York: Cambridge Univ. Press.
Williamson, R. C. (1989). *Probabilistic arithmetic*. Unpublished dissertation, University of Queensland, St. Lucia, Australia.
Wittgenstein, L. (1961). *Tractatus Logico Philosophicus* (Logisch-Philosophische Abhandlung). London: Routledge. (Text includes original German and English translation.)
Zeleny, M. (1980). *Autopoiesis, dissipative structures, and spontaneous social orders*. Washington, DC: American Association for the Advancement of Science

Appendix A: FBST Review

> (A) man's logical method should be loved and reverenced as his bride, whom he has chosen from all the world. He need not condemn the others; on the contrary, he may honor them deeply, and in doing so he honors her more. But she is the one that he has chosen, and he knows that he was right in making that choice.
> (C.S. Peirce, The Fixation of Belief, 1877).

The objective of this appendix is to provide a very short review of the Full Bayesian Significance Test (FBST), show a simple concrete example, and summarize the most important logical properties of the FBST support function. Several applications of the FBST, details of its efficient numerical and computational implementation, demonstrations of theoretical properties, comparison with other statistical tests for sharp hypotheses, and an extensive list of references can be found in the author's previous papers.

A.1 The Epistemic e-values

Let $\theta \in \Theta \subseteq R^p$ be a vector parameter of interest, and $L(\theta | x)$ be the likelihood associated to the observed data x, a standard statistical model. Under the Bayesian paradigm the posterior density, $p_n(\theta)$, is proportional to the product of the likelihood and a prior density,

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

The (null) hypothesis H states that the parameter lies in the null set, defined by inequality and equality constraints given by vector functions g and h in the parameter space.

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

From now on, we use a relaxed notation, writing H instead of Θ_H . We are particularly interested in sharp (precise) hypotheses, those in which $\dim(H) < \dim(\Theta)$, that is there is at least one equality constraint.

The FBST defines $\text{ev}(H)$, actually $\text{ev}(H; p_n, r)$, the e-value or evidence value supporting (in favor of) the hypothesis H, and its complement, $\overline{\text{ev}}(H)$, the evidence value against H, as

$$s(\theta) = \frac{p_n(\theta)}{r(\theta)} \;, \quad s^* = s(\theta^*) = \sup_{\theta \in H} s(\theta) \;, \quad \widehat{s} = s(\widehat{\theta}) = \sup_{\theta \in \Theta} s(\theta) \;,$$

$$T(v) = \{\theta \in \Theta \mid s(\theta) \leq v\} \;, \quad W(v) = \int_{T(v)} p_n(\theta)\, d\theta \;, \quad \text{ev}(H) = W(s^*) \;,$$

$$\overline{T}(v) = \Theta - T(v) \;, \quad \overline{W}(v) = 1 - W(v) \;, \quad \overline{\text{ev}}(H) = \overline{W}(s^*) = 1 - \text{ev}(H) \;.$$

The function $s(\theta)$ is known as the posterior surprise relative to a given reference density, $r(\theta)$. $W(v)$ is the cumulative surprise distribution. The surprise function was used, among other statisticians, by Good (1983), Evans (1997) and Royall (1997). Its role in the FBST is to make $\text{ev}(H)$ implicitly invariant under suitable transformations on the coordinate system of the parameter space, see next subsection.

The tangential (to the hypothesis) set $\overline{T} = \overline{T}(s*)$, is a Highest Relative Surprise Set (HRSS). It contains the points of the parameter space with higher surprise, relative to the reference density, than any point in the null set of H. When $r(\theta) \propto 1$, the possibly improper uniform density, $\overline{T}$ is the Posterior's Highest Density Probability Set (HDPS) tangential to the null set of H. Small values of $\text{ev}(H)$ indicate that the hypothesis traverses high density regions, favoring the hypothesis.

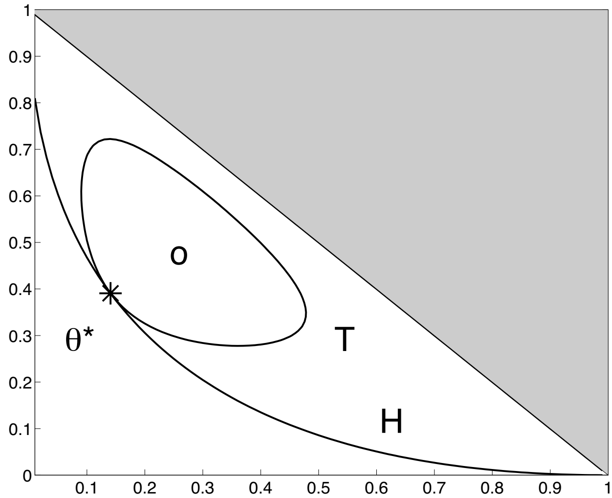


Figure A1: H-W Hypothesis and Tangential Set

The evidence value, defined above, has a simple and intuitive geometric characterization. Figure A1 shows the null set of H, the tangential HRSS $\overline{T}$, and the points of constrained and unconstrained maxima, θ^* and $\widehat{\theta}$, for testing Hardy-Weinberg equilibrium law in a population genetics problem, as discussed in (Pereira and Stern 1999). In this biological application n is the sample size, x_1 and x_3 are the two homozygote sample counts and $x_2 = n - x_1 - x_3$ is heterozygote sample count. $\theta = [\theta_1 , \theta_2, \theta_3]$ is the parameter vector. The posterior and maximum entropy reference densities for this trinomial model, the parameter space and the null set are:

$$p_n(\theta \mid x) \propto \theta_1^{x_1+y_1} \theta_2^{x_2+y_2} \theta_3^{x_3+y_3} \ , \quad r(\theta) \propto \theta_1^{y_1} \theta_2^{y_2} \theta_3^{y_3} \ , \quad y = [-1, -1, -1] \ ,$$

$$\Theta = \{\theta \geq 0 \mid \theta_1 + \theta_2 + \theta_3 = 1\} \ , \quad \Theta_H = \{\theta \in \Theta \mid \theta_3 = (1 - \sqrt{\theta_1}\ )^2\} \ .$$

In orthodox decision theoretic Bayesian statistics, a significance test is legitimate if and only if it can be characterized as an Acceptance (A) or Rejection (R) decision procedure defined by the minimization of the posterior expectation of a loss function, Λ. Madruga et al. (2001) gives the following family of loss functions characterizing the FBST. This loss function is based on indicator functions of θ being or not in the tangential set $\overline{T}$:

$$\Lambda(R, \theta) = a\, I(\theta \notin \overline{T}) \ , \quad \Lambda(A, \theta) = b + d\, I(\theta \in \overline{T}) \ .$$

Note that this loss function is dependent on the observed sample (via the likelihood function), on the prior, and on the reference density, stressing the important point of non-separability of utility and probability, see Kadane and Winkler (1987) and Rubin (1987).

Finally, consider the situation where the hypothesis constraint, $H: h(\theta) = h(\delta) = 0$, $\theta = [\delta, \lambda]$ is not a function of some of the parameters, λ. This situation is described by Basu and Ghosh:

> *If the inference problem at hand relates only to δ, and if information gained on λ is of no direct relevance to the problem, then we classify λ as the Nuisance Parameter. The big question in statistics is: How can we eliminate the nuisance parameter from the argument?* (1988, p. 115)

Basu and Ghosh list at least 10 categories of procedures to achieve this goal, like using max_λ or $\int d\lambda$, the maximization or integration operators, in order to obtain a projected profile or marginal posterior function, $f(\delta | x)$. The FBST does not follow the nuisance parameters elimination paradigm. In fact, staying in the original parameter space, in its full dimension, explains the *intrinsic regularization* property of the FBST, when it is used for model selection, see Pereira and Stern (2001).

A.2 Reference and Consistency

In the FBST the role of the reference density, $r(\theta)$ is to make $\overline{\text{ev}}(H)$ implicitly invariant under suitable transformations of the coordinate system. Invariance, as used in statistics, is a metric concept. The reference density can be interpreted as a compact and interpretable representation for the reference metric in the original parameter space. This metric is given by the geodesic distance on the density surface. The natural choice of reference density is an uninformative prior, 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 priors include the uniform and maximum entropy densities, for a detailed discussion the reader is referred to Dugdale (1996) and Kapur (1989). In the H-W example, using the notation above, the uniform density can be represented by $y = [0,0,0]$ observation counts, and the standard maximum entropy density can be represented by $y = [-1, -1, -1]$ observation counts.

Let us consider the cumulative distribution of the evidence value against the hypothesis, the confidence level function, $\overline{V}(c) = \Pr(\overline{\text{ev}} \le c)$, given θ^0, the true value of the parameter. Under appropriate regularity conditions, for increasing sample size, $n \to \infty$, we can say the following:

- If H is false, $\theta^0 \notin H$, then $\overline{\text{ev}}$ converges (in probability) to 1, that is, $\overline{V}(0 \le c < 1) \to 0$.
- If H is true, $\theta^0 \notin H$, then $\overline{V}(c)$, converges (in distribution) to

$$QQ(t,h,c) = \mathrm{Q}\left(t-h, \mathrm{Q}^{-1}(t,c)\right) \ , \quad \text{where}$$

$$\mathrm{Q}(k,x) = \frac{\Gamma(k/2, x/2)}{\Gamma(k/2,\infty)} \ , \quad \Gamma(k,x) = \int_0^x y^{k-1} e^{-y} dy \ ,$$

$t = \dim(\Theta)$, $h = \dim(H)$ and $Q(k, x)$ is the cumulative chi-square distribution with k degrees of freedom. Figure A2 portrays the function $QQ(t, h, c)$ for $t = 2 \ldots 4$ and $h = 0 \ldots t-1$.

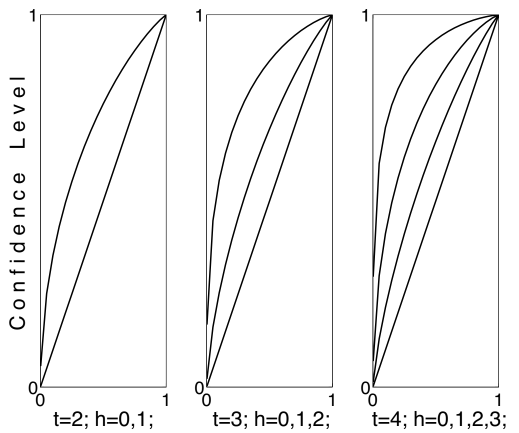


Figure A2: Test τ_c critical level vs. confidence level

Under the same regularity conditions, an appropriate choice of threshold or critical level, $c(n)$, provides a consistent test, τ_c, that rejects the hypothesis if $\overline{\mathrm{ev}}(H) > c$. The empirical power analysis developed in Stern and Zacks (2002) and Lauretto, Pereira, Stern, and Zacks (2003), provides critical levels that are consistent and also effective for small samples.

Stern (2004) presents an alternative approach, based on sensitivity analysis in the context of paraconsistent logic and bilattice structures, see also Costa, Abe, Murolo, Silva, and Casemiro (1999). This analysis in based on the inconsistency induced by a set of alternative reference densities, $r, r', r'' \ldots$, or a set of alternative priors, $p_0, p_0', p_0'' \ldots$ or a set of alternative likelihood power or "sample size" perturbation parameters, L^{γ}, $1 = \gamma > \gamma' > \gamma'' \ldots > 0$.

A.3 Belief Calculi and Support Structures

Many standard Belief Calculi can be formalized in the context of Abstract Belief Calculus, ABC, see Darwiche and Ginsberg (1992), Darwiche (1993) and Stern

(2003). In a Support Structure, $\langle \Phi, \oplus, \oslash \rangle$, the first element is a Support Function, Φ, on a universe of statements, $\mathcal{U}$. Null and full support values are represented by 0 and 1. The second element is a support Summation operator, $\oplus$, and the third is a support Scaling or Conditionalization operator, $\oslash$. A Partial Support Structure, $\langle \Phi, \oplus \rangle$, lacks the scaling operation.

The Support Summation operator, $\oplus$, gives the support value of the disjunction of any two logically disjoint statements from their individual support values, that is,

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

The Support Scaling operator, $\oslash$, gives the conditional support value of *B* given from the unconditional support values of *A* and the conjunction *C* = *A* Λ *B*, that is,

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

Support structures for some standard belief calculi are given in Table A1, where the support value of two statements their conjunction are given by *a* = Φ *(A)*, *b* = Φ *(B)*, *c* = Φ *(C = A Λ B)*.

Table A1: Support structures for some belief calculi, $c = \Phi(C = A \wedge B)$.

$\Phi(\mathcal{U})$	$a \oplus b$	**0**	**1**	$a \preceq b$	$c \oslash a$	Calculus
$\{0, 1\}$	$\max(a, b)$	0	1	$a \leq b$	$\min(c, a)$	Classical Logic
$[0, 1]$	$a + b$	0	1	$a \leq b$	c/a	Probability
$[0, 1]$	$\max(a, b)$	0	1	$a \leq b$	c/a	Possibility
$\{0 \ldots \infty\}$	$\min(a, b)$	∞	0	$b \leq a$	$c - a$	Disbelief

In Table A1, the relation *a* ≤ *b* indicates that the value *a* represents a stronger support than the value *b*. Darwiche and Ginsberg (1992) also gives a set of axioms defining the essential functional properties of a (partial) support function. Stern (2003) shows that the support Φ (*H*) = ev(*H*) complies with all Darwiche and Ginsberg axioms.

In the FBST, the support values, Φ (*H*) = ev(*H*), are computed using standard probability calculus on θ which has an intrinsic conditionalization operator. The computed e-values, on the other hand, have a possibilistic summation, that is, the evidence value in favor of a composite hypothesis (*H*) = A ∨ B, is the most favorable evidence value in favor of each of its terms, that is, ev(*H*) = max {ev(*A*), ev(*B*)}. It is impossible however to define a simple scaling operator for this possibilistic support function that is compatible with the FBST's e-value, ev( ), as it is defined.

Hence, two belief calculi are in simultaneous use in the Full Bayesian Significance Test setup: 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.

Stern (2003) comments the interpretation of this results in the juridical or legal context. In this context, the possibilistic structure corresponds to the *Onus Probandi* juridical principle, or the *In Dubito pro Reo* rule. These are "benefit of the doubt" type norms, requiring the statement presented by the defendant to be considered in most favorable manner, as seen in Gaskins (1992).

A4. Complex Models and Compositionality

The relationship between the credibility of a complex hypothesis, H, and those of its constituent elementary hypothesis, $H^{(i,j)}$, in the independent setup, can be analyzed under the FBST, see Borges and Stern (2005) for precise definitions, and detailed interpretation.

Let us consider elementary hypotheses, $H^{(i,j)}$, in k independent constituent models, M^j, and the complex or composite hypothesis H, equivalent to a (homogeneous) logical composition (disjunction of conjunctions) of elementary hypotheses, in the composite product model, M. The following result can be established, see Borges and Stern (2005, proposition 5.1):

If H is expressed in HDNF or Homogeneous Disjunctive Normal Form,

$$H = \bigvee_{i=1}^{q} \bigwedge_{j=1}^{k} H^{(i,j)} \ , \quad M^{(i,j)} = \{\Theta^j, H^{(i,j)}, p_0^j, p_n^j, r^j\} \ ,$$

$$M = \{\Theta, H, p_0, p_n, r\} \ , \quad \Theta = \prod_{j=1}^{k} \Theta^j \ , \quad p_n = \prod_{j=1}^{k} p_n^j \ , \quad r = \prod_{j=1}^{k} r^j \ ;$$

then the e-value supporting H is

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

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

where the cumulative surprise distribution of the composite model, $W(v)$, is given by the Mellin convolution operation, see Springer (1979) and Williamson (1989), defined as

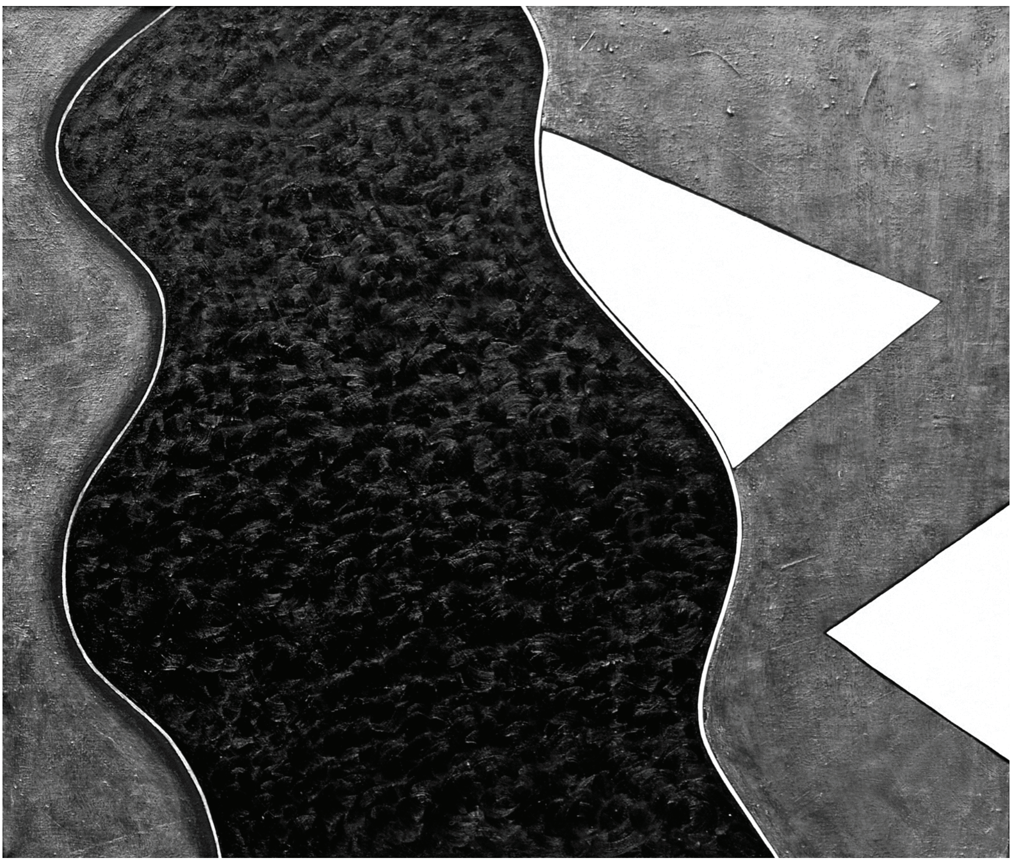
$$W = \bigotimes_{1 \leq j \leq k} W^j \quad , \quad W^1 \otimes W^2(v) \; = \; \int_0^\infty W^1(v/y) W^2(dy) \; .$$

The probability distribution of the product of two independent positive random variables is the Mellin convolution of each of their distributions. From this interpretation, we immediately see that $\otimes$ is a commutative and associative operator.

Mirroring Wittgenstein, in the FBST context, we can call the e-value, $\text{ev}(H)$, the cumulative surprise distribution, $W(v)$, and the Mellin convolution operation, $\otimes$, respectively, truth value, truth function, and truth operation.

Finally, we observe that, in the extreme case of null-or-full support, that is, when, for $1 \leq i \leq q$ and $1 \leq j \leq k$, $s^{*(i,j)} = 0$ or $s^{*(i,j)} = \widehat{s}^j$, the evidence values (or, in this context, truth values) of the constituent elementary hypotheses are either 0 or 1, and the conjunction and disjunction composition rules of classical logic hold.

.



Barrett, K. (n.d.). *Beldam*. 71 x 61 cm, acrylic on canvas.