Article

Pragmatic Hypotheses in the Evolution of Science

Luis Gustavo Esteves [1], Rafael Izbicki [2] , Julio Michael Stern [1],* and Rafael Bassi Stern [2]

[1] Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo 05508-090, Brazil
[2] Departamento de Estatística, Universidade Federal de São Carlos, São Paulo 13565-905, Brazil

* Correspondence: jstern@ime.usp.br

Abstract: This paper introduces pragmatic hypotheses and relates this concept to the spiral of scientific evolution. Previous works determined a characterization of logically consistent statistical hypothesis tests and showed that the modal operators obtained from this test can be represented in the hexagon of oppositions. However, despite the importance of precise hypothesis in science, they cannot be accepted by logically consistent tests. Here, we show that this dilemma can be overcome by the use of pragmatic versions of precise hypotheses. These pragmatic versions allow a level of imprecision in the hypothesis that is small relative to other experimental conditions. The introduction of pragmatic hypotheses allows the evolution of scientific theories based on statistical hypothesis testing to be interpreted using the narratological structure of hexagonal spirals, as defined by Pierre Gallais.

Keywords: hypothesis tests; precise hypotheses; pragmatic hypotheses

1. Introduction

Standard hypothesis tests can lead to immediate logical incoherence, which makes their conclusions hard to interpret. This incoherence occurs because such tests have only two possible outcomes. Indeed, Izbicki and Esteves [1] shows that there exists no two-valued test that satisfies desirable statistical properties and is also logically coherent.

In order to overcome such an impossibility result, Esteves et al. [2] propose agnostic hypothesis tests, which have three possible outputs: (A) accept the hypothesis, say H, (E) reject H, or (Y) remain agnostic about H. These tests can be made logically coherent while preserving desirable statistical properties. For instance, both conditions are satisfied by the Generalized Full Bayesian Significance Test (GFBST). Furthermore, Stern et al. [3] show that the GFBST's modal operators and their respective negations can be represented by vertices of the hexagon of oppositions [4–9], which is depicted in Figure 1.
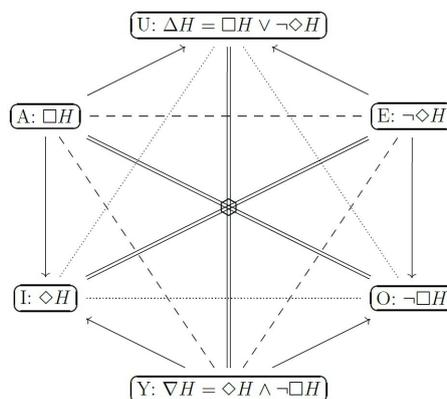

Figure 1. Hexagons of opposition for statistical modalities.

This paper complements the above static representation with an analysis of the GFBST in the dynamic evolution of scientific theories. The analysis is based on the metaphor of evolutive hexagonal spirals [10,11], in which the logical modalities associated to scientific theories change over time, as in Figure 2. Our key point in this paradigm is reconciling two apparently contradictory facts. On the one hand, precise or sharp hypotheses, that is, hypotheses that have a priori zero probability are central in scientific theories [12,13]. On the other hand, the GFBST never accepts (A) precise hypotheses. These observations lead to the apparent paradox that, if the GFBST were used to test scientific theories, then the acceptance step in the spiral of scientific theories would be forfeited.

We overcome this paradox by proposing the concept of a "pragmatic hypothesis" associated to a precise hypothesis. Although precise hypotheses are commonly obtained from mathematical theories used in areas of science and technology [12,13], the associated pragmatic hypothesis is an imprecise hypothesis that is sufficiently good from the practical purpose of an end-user of the theories. For instance, Newtonian theory assumes a gravitational force of magnitude given by the equation $F = G\, m_1\, m_2\, d^{-2}$, where the gravitational constant G has a precise value. However, the current Committee on Data for Science and Technology (CODATA) value for the gravitational constant is $G = 6.67408(31) \times 10^{-11} \mathrm{m}^3\, \mathrm{kg}^{-1}\, \mathrm{s}^{-2}$, which includes a standard deviation for the last significant digits, 408 ± 31. Thus, it may be reasonable for a given end-user to assume that the theoretical form of the last equation is exact, but that, pragmatically, the constant G can only be known up to a chosen precision. As a result, one might wish to test an imprecise hypothesis associated to the scientific hypothesis of interest [14,15].

This article advocates for the conceptual distinction between a precise scientific theory and an associated pragmatic hypotheses. The alternate use of precise and pragmatic versions of corresponding statistical hypotheses enables the GFBST to (pragmatically) accept scientific hypotheses. Moreover, this alternate use allows the GFBST to track the evolution of scientific theories, as interpreted in the context of Gallais' hexagonal spirals.

Our main goal in this paper is to formalize testing procedures for a theory taking into consideration the level of precision that is appropriate for a given end-user. To handle this problem, we consider the end-user's predictions about an experiment of his interest. The variation in these predictions can be explained by a combination of the level of imprecision in the theory and by properties of the end-user's experiment. For instance, the latter source of variation is influenced by properties of the equipment, including the precision, accuracy, and resolution of measuring devices [16,17], and also error bounds for fundamental constants and calibration factors [18–26]. We propose to choose a pragmatic hypothesis in such a way that the imprecision in the end-user's predictions is mostly due to his experimental conditions and not due to the level of imprecision in the theory that he uses.

In order to develop this argument, Section 2 first adapts Gallais's metaphor of hexagonal spirals to the evolution of science. Next, Section 3 proposes three methods of decomposing the variability in an end-user's predictions into the level of precision of the theory he uses and his experimental conditions. Sections 3.1 and 3.2 use these decompositions in order to build pragmatic hypothesis. They build pragmatic hypotheses for simple hypotheses and then prove that there exists a single way of extending this construction to composite hypotheses while preserving logical coherence in simultaneous hypothesis testing. This methodology is illustrated in Section 4. All proofs are found in Appendix A.

2. Gallais' Hexagonal Spirals and the Evolution of Science

Following a well-established tradition in structural semantics and narratology [27,28], Gallais and Pollina [10] propose that many classical medieval tales follow the same organizational pattern. More precisely, these narratives exhibit an underlying "intellectual structure" and are organized according to an underlying archetypal format or prototypical pattern. This pattern includes both static and a dynamical aspects. From a static perspective, the logical structure of the narrative is such that each arch is represented by a vertex of the "hexagon of oppositions" [4]. The static hexagon of oppositions

is depicted in Figure 1 and represents in each vertex a modal operator among necessity ($\Box$), possibility ($\Diamond$), contingency (Δ), and their negations ($\neg$). These modal operators are structured according to three axes of opposition, ($===$); a triangle of contrariety, ($---$); another triangle of sub-contrariety, ($\cdots$); and several edges of subalteration ($\longrightarrow$). From a dynamical perspective, the temporal evolution of the narrative follows a spiral (Figure 2) that unwinds (se déroule) around concentric and expanding hexagons of opposition [10,11].
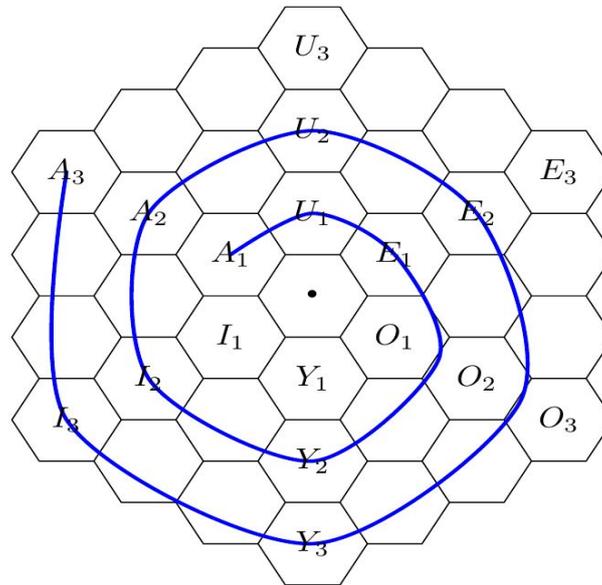

Figure 2. Gallais' evolutionary spiral.

Because the evolution of science can also be conceived as following a spiral pattern [29], its analysis can benefit from the structure in the works by the authors of [10,11]. From a static perspective, the logical modalities induced by agnostic hypothesis tests [3] can be represented in the hexagon of oppositions. From a dynamic perspective, scientific theories evolve as a spiral which unwinds around the following states:

- A_1 - Extant thesis: This vertex represents a standing paradigm, an accepted theory using well-known formalisms and familiar concepts, relying on accredited experimental means and methods, etc. In fact, the concepts of a current paradigm may become so familiar and look so natural that they become part of a reified ontology. That is, there is a perceived correspondence between concepts of the theory and "dinge-an-sich" (things-in-themselves) as seen in nature [29,30].
- U_1 - Analysis: This vertex represents the moment when some hypotheses of the standing theory are put in question. At this moment, possible alternatives to the standing hypotheses may still be only vaguely defined.
- E_1 - Antithesis: This vertex represents the moment when some laws of the standing theory have to be rejected. Such a rejection of old laws may put in question the entire world-view of the current paradigm, opening the way for revolutionary ideas, as described in the next vertex.
- O_2 - Apothesis/ Prosthesis: This vertex is the locus of revolutionary freedom. Alternative models are considered, and specific (precise) forms investigated. There is intellectual freedom to set aside and dispose of (apothesis) old preconceptions, prejudices and stereotypes, and also to explore and investigate new paths, to put together (prosthesis) and try out new concepts and ideas.
- Y_2 - Synthesis: It is at this vertex that new laws are formulated; this is the point of Eureka moment(s). A selection of old and and new concepts seem to click into place, fitting together in the form of new laws, laws that are able to explain new phenomena and incorporate objects of an expanded reality.
- I_2 - Enthesis: At this vertex new laws, concepts and methods must enter and be integrated into a consistent and coherent system. At this stage many tasks are performed in order to combine

novel and traditional pieces or to accommodate original and conventional components into an well-integrated framework. Finally, new experimental means and methods are developed and perfected, allowing the new laws to be corroborated.

- A_2- New Thesis: At this vertex, the new theory is accepted as the standard paradigm that succeeds the preceding one (A_1). Acceptance occurs after careful determination of fundamental constants and calibration factors (including their known precision), metrological and instrumentational error bounds, etc. At later stages of maturity, equivalent theoretical frameworks may be developed using alternative formalisms and ontologies. For example, analytical mechanics offers variational alternatives that are (almost) equivalent to the classical formulation of Newtonian mechanics [31]. Usually, these alternative worldviews reinforce the trust and confidence on the underlying laws. Nevertheless, the existence of such alternative perspectives may also foster exploratory efforts and investigative works in the next cycle in evolution.

Table 1 applies this spiral structure to the evolution of the theories of orbital astronomy and chemical affinity. The evolution of orbital astronomy has been widely studied [32]. The evolution of chemical affinity is presented in greater detail in Stern [29], Stern and Nakano [33].

Table 1. Evolution of orbital astronomy and chemical affinity.

Vertex	Orbital Astronomy	Chemical Affinity
I_1- Enthesis/ A_1- Thesis	Ptolemaic/ Copernican cycles and epicycles	Geoffroy affinity table and highest rank substitution
U_1- Analysis	Circular or oval orbits?	Ordinal or numeric affinity?
E_1- Antithesis	Non-circular orbits	Non-ordinal affinity
O_2- Apothesis /Prosthesis	Elliptic planetary orbits, focal centering of sun	Integer affinity values, for arithmetic recombination
Y_2- Synthesis	Kepler laws!	Morveau rules and tables!
I_2- Enthesis A_2- Thesis	Vortex physics theories, Keplerian astronomy	Affinity + stoichiometry substitution reactions
U_2- Analysis	Tangential or radial forces?	Total or partial reaction?
E_2- Antithesis	Non-tangential forces	Non-total substitutions
O_3- Apothesis /Prosthesis	Radial attraction forces, inverse square of distance	Reversible reactions, equilibrium conditions
Y_3- Synthesis	Newton laws!	Mass-Action kinetics!
I_3- Enthesis/ A_3- Thesis	Newtonian mechanics & variational equivalents	Thermodynamic theories for reaction networks

The above spiral structure highlights that a statistical methodology should be able to obtain each of the six modalities in the hexagon of oppositions. Before an acceptance vertex (A) in the hexagon is reached by the spiral of scientific evolution, theoretically precise or sharp hypotheses must be formulated. However, a logically coherent hypothesis test, such as the GFBST, can choose solely between rejecting or remaining agnostic (i.e., corroborating) such sharp hypotheses. Once the evolving theory becomes (part of) a well-established paradigm, the GFBST can be used with the goal of accepting non-sharp hypotheses in the context of the same paradigm, a context that includes fundamental constants and calibration factors (and their respective uncertainties), metrological error bounds, specified accuracies of scientific intrumentation, etc. The non-sharp versions of sharp hypotheses used in such tests are called pragmatic, and their formulation is developed in the following sections.

3. Pragmatic Hypotheses

In order to derive pragmatic hypotheses from precise ones, it is necessary to define an idealized future experiment. Let θ be an unknown parameter of interest, which is used to express scientific hypotheses and that takes values in the parameter space, Θ. A scientific hypothesis takes the form $H_0 : \theta \in \Theta_0$, where $\Theta_0 \subset \Theta$. Whenever there is no ambiguity, H_0 and Θ_0 are used interchangeably. Also, the determination of θ is useful for predicting an idealized future experiment, **Z**, which takes

values in $\mathcal{Z}$. The uncertainty about Z depends on θ by means of P_{θ^*}, the probability measure over $\mathcal{Z}$ when it is known that $\theta = \theta^*$, $\theta^* \in \Theta$.

Often, it is sufficient for an end-user to determine a pragmatic hypothesis, that is, that the parameter lies in a set of plausible values, which is larger than the null hypothesis. This set can be chosen in such a way that the variation over predictions about a future experiment is mostly due to experimental conditions rather than to the imprecision in the value of the parameter. This section formally develops a methodology for determining these pragmatic hypotheses.

In order to compare two parameter values, we use a "predictive dissimilarity", $d_{\mathbf{Z}}$, which is a function, $d_{\mathbf{Z}} : \Theta \times \Theta \to \mathbb{R}^+$, such that $d_{\mathbf{Z}}(\theta_0, \theta^*)$ measures how much the predictions made for $\mathbf{Z}$ based on θ^* diverge from the ones made based on θ_0. We define and compare three possible choices for such a dissimilarity.

Definition 1. *The Kullback–Leibler predictive dissimilarity, $KL_{\mathbf{Z}}$, is*

$$KL_{\mathbf{Z}}(\theta_0, \theta^*) = KL(\mathbb{P}_{\theta^*}, \mathbb{P}_{\theta_0}) = \int_{\mathcal{Z}} \log \left(\frac{d\mathbb{P}_{\theta^*}}{d\mathbb{P}_{\theta_0}} \right) d\mathbb{P}_{\theta^*},$$

that is, $KL_{\mathbf{Z}}(\theta_0, \theta^*)$ is the relative entropy between $\mathbb{P}_{\theta^*}$ and $\mathbb{P}_{\theta_0}$.

Example 1 (Gaussian with known variance.)**.** *Let $\mathbf{Z} = (Z_1 \ldots, Z_d) \sim N(\theta, \Sigma_0)$ be a random vector with a multivariate Gaussian distribution:*

$$\frac{d\mathbb{P}_\theta(\mathbf{z})}{d\mathbf{z}} = \|2\pi\Sigma_0\|^{-0.5} \exp\left(-0.5(\mathbf{z} - \theta)^t \Sigma_0^{-1} d(\mathbf{z} - \theta) \right)$$

$$KL_{\mathbf{Z}}(\theta_0, \theta^*) = \int_{\mathbb{R}^d} \log \left(\frac{d\mathbb{P}_{\theta^*}(\mathbf{z})}{d\mathbb{P}_{\theta_0}(\mathbf{z})} \right) d\mathbb{P}_{\theta^*}(\mathbf{z}) = 0.5(\theta_0 - \theta^*)^t \Sigma_0^{-1} (\theta_0 - \theta^*),$$

When $d = 1$ and $\Sigma_0 = \sigma_0^2$,

$$KL_{\mathbf{z}}(\theta_0, \theta^*) = \frac{(\theta_0 - \theta^*)^2}{2\sigma_0^2} \tag{1}$$

The KL dissimilarity evaluates the distance between the predictive probability distributions for the future experiment under two parameter values, θ_0 and θ^*. Although the KL dissimilarity is general, it can be challenging to interpret. In particular, it can be hard to establish the quality of the predictions for $\mathbf{Z}$ based on θ^* when $\mathbf{Z}$ is actually generated from θ_0 and $KL_{\mathbf{Z}}(\theta_0, \theta^*) \leq \epsilon$. A more interpretable dissimilarity is obtained by taking $d_{\mathbf{Z}}(\theta_0, \theta^*)$ to measure how far are the best predictions for $\mathbf{Z}$ based on θ^* and θ_0. In this case, if one makes a prediction for $\mathbf{Z}$ based on θ^*, $\mathbf{z}_*$, and $\mathbf{Z}$ was actually generated using θ_0, then $d_{\mathbf{Z}}(\theta_0, \theta^*) \leq \epsilon$ guarantees that $\mathbf{z}_*$ will be at most ϵ apart from the best possible prediction. Such a dissimilarity is discussed in the following definition.

Definition 2 (Best prediction dissimilarity—BP.)**.** *Let $\hat{\mathbf{Z}} : \Theta \to \mathcal{Z}$ be such that $\hat{\mathbf{Z}}(\theta_0)$ is the best prediction for $\mathbf{Z}$ given that $\theta = \theta_0$. For example, one can take*

$$\hat{\mathbf{Z}}(\theta_0) = \arg \min_{\mathbf{z} \in \mathcal{Z}} \delta_{\mathbf{Z},\theta_0}(\mathbf{z}),$$

where $\delta_{\mathbf{Z},\theta_0} : \mathcal{Z} \to \mathbb{R}$ is such that $\delta_{\mathbf{Z},\theta_0}(\mathbf{z})$ measures how bad $\mathbf{z}$ predicts $\mathbf{Z}$ when $\theta = \theta_0$. The "best prediction dissimilarity", $BP_{\mathbf{Z}}(\theta_0, \theta^*)$, measures how badly $\hat{\mathbf{Z}}(\theta^*)$ predicts $\mathbf{Z}$ relatively to $\hat{\mathbf{Z}}(\theta_0)$ when $\theta = \theta_0$. Formally,

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(\frac{\delta_{\mathbf{Z},\theta_0}(\hat{\mathbf{Z}}(\theta^*)) - \delta_{\mathbf{Z},\theta_0}(\hat{\mathbf{Z}}(\theta_0))}{\delta_{\mathbf{Z},\theta_0}(\hat{\mathbf{Z}}(\theta_0))} \right),$$

where $g : \mathbb{R} \longrightarrow \mathbb{R}$ is a motononic function. The choice of g in a particular setting aims at improving the interpretation of the best prediction dissimilarity criterion.

Example 2 (BP under quadratic form.). *Let* $\mathcal{Z} = \mathbb{R}^d$, $\mu_{\mathbf{Z},\theta} = E[\mathbf{Z}|\theta]$, $\Sigma_{\mathbf{Z},\theta} = \mathbb{V}[\mathbf{Z}|\theta]$ *and* $\mathbf{S}$ *be a positive definite matrix. Define the quadratic form induced by* $\mathbf{S}$ *to be* $\|\mathbf{z}\|_{\mathbf{S}}^2 = \mathbf{z}^T \mathbf{S} \mathbf{z}$ *and*

$$\delta_{\mathbf{Z},\theta_0}(\mathbf{z}) = E\left[\|\mathbf{Z} - \mathbf{z}\|_{\mathbf{S}}^2 | \theta = \theta_0\right]$$

The optimal prediction under θ^* *is* $\hat{\mathbf{Z}}(\theta^*) = \mu_{\mathbf{Z},\theta^*}$. *It follows that*

$$\delta_{\mathbf{Z},\theta_0}(\hat{\mathbf{Z}}(\theta^*)) = E\left[\|\mathbf{Z} - \mu_{\mathbf{Z},\theta^*}\|_{\mathbf{S}}^2 | \theta = \theta_0\right]$$
$$= \|\mu_{\mathbf{Z},\theta_0} - \mu_{\mathbf{Z},\theta^*}\|_{\mathbf{S}}^2 + E\left[\|\mathbf{Z} - \mu_{\mathbf{Z},\theta_0}\|_{\mathbf{S}}^2 | \theta = \theta_0\right]$$

In particular, $\delta_{\mathbf{Z},\theta_0}(\hat{\mathbf{Z}}(\theta_0)) = E\left[\|\mathbf{Z} - \mu_{\mathbf{Z},\theta_0}\|_{\mathbf{S}}^2 | \theta = \theta_0\right]$. *Therefore,*

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(\frac{\|\mu_{\mathbf{Z},\theta_0} - \mu_{\mathbf{Z},\theta^*}\|_{\mathbf{S}}^2}{E\left[\|\mathbf{Z} - \mu_{\mathbf{Z},\theta_0}\|_{\mathbf{S}}^2 | \theta = \theta_0\right]}\right) \tag{2}$$

In this example, $BP_{\mathbf{Z}}$ *can be put in the same scale as* $\mathbf{Z}$ *by taking* $g(x) = \sqrt{x}$. *Also, two choices of* $\mathbf{S}$ *are of particular interest. When* $\mathbf{S} = \mathbb{V}[\mathbf{Z}|\theta = \theta_0]^{-1}$, *Equation (2) simplifies to*

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(d^{-1}\|\mu_{\mathbf{Z},\theta_0} - \mu_{\mathbf{Z},\theta^*}\|_{\Sigma_{\mathbf{Z},\theta_0}^{-1}}^2\right) \tag{3}$$

Similarly, when $\mathbf{S}$ *is the identity matrix, Equation (2) simplifies to*

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(\frac{\|E[\mathbf{Z}|\theta = \theta_0] - E[\mathbf{Z}|\theta = \theta^*]\|_2^2}{tr(\mathbb{V}[\mathbf{Z}|\theta = \theta_0])}\right) \tag{4}$$

Equation (4) admits an intuitive interpretation. The larger the value of $tr(\mathbb{V}[\mathbf{Z}|\theta = \theta_0])$, *the more* $\mathbf{Z}$ *is dispersed and the harder it is to predict its value. Also,* $\|E[\mathbf{Z}|\theta = \theta_0] - E[\mathbf{Z}|\theta = \theta^*]\|_2^2$ *measures how far apart are the best prediction for* $\mathbf{Z}$ *under* $\theta = \theta_0$ *and* $\theta = \theta^*$. *That is,* $BP_{\mathbf{Z}}(\theta_0, \theta^*)$ *captures that, if one predicts* $\mathbf{Z}$ *assuming that* $\theta = \theta^*$ *when it is actually* $\theta = \theta_0$, *then the error with respect to the best prediction is increased as a function of the distance between the predictions over the dispersion of* $\mathbf{Z}$.

Example 3 (Gaussian with known variance.). *Consider Example 1 and let* $\delta_{\mathbf{Z},\theta_0}(\mathbf{z})$ *be as in Example 2. It follows from Equation (4) that when* $\mathbf{S}$ *is the identity matrix,*

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(\frac{\|\theta_0 - \theta^*\|_2^2}{tr(\Sigma_0)}\right) \tag{5}$$

Similarly, it follows from Equation (3) that when $\mathbf{S} = \Sigma_0^{-1}$,

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(d^{-1}(\theta_0 - \theta^*)^t \Sigma_0^{-1}(\theta_0 - \theta^*)\right) \tag{6}$$

Conclude from Equation (6) that, if $\mathbf{S} = \Sigma_0^{-1}$ *and* $g(x) = x$, *then* $BP_{\mathbf{Z}}(\theta_0, \theta^*) = 2d^{-1}KL_{\mathbf{Z}}(\theta_0, \theta^*)$. *Also, when* $d = 1$, $\Sigma_0 = \sigma_0^2$ *and* $g(x) = \sqrt{x}$, *both Equation (5) and Equation (6) simplify to*

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = \sigma_0^{-1}|\theta_0 - \theta^*| \tag{7}$$

In some situations, $\mathbf{Z}$ is the average of m independent observations distributed as $N(\theta, \Sigma_0)$. In this case, $\mathbf{Z} \sim N(\theta, m^{-1}\Sigma_0)$. It follows from Equation (5) that $BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(\frac{m\|\theta_0 - \theta^*\|_2^2}{tr(\Sigma_0)} \right)$ when $\mathbf{S}$ is the identity, and $BP_{\mathbf{Z}}(\theta_0, \theta^*) = g\left(md^{-1}(\theta_0 - \theta^*)^t \Sigma_0^{-1}(\theta_0 - \theta^*) \right)$ when $\mathbf{S} = \Sigma_0^{-1}$.

Although $BP_{\mathbf{Z}}$ is more interpretable then $KL_{\mathbf{Z}}$, it also relies on more tuning variables, such as δ, $\hat{Z}$, and g. A balance between these features is obtained by a third predictive dissimilarity, which evaluates how easy it is to recover the value of θ between θ_0 or θ^* based on $\mathbf{Z}$.

Definition 3 (Classification distance—CD.). *Let $\hat{\theta}_{\theta_0, \theta^*} : \mathcal{Z} \to \Theta$ be such that*

$$\hat{\theta}_{\theta_0, \theta^*}(\mathbf{z}) = \arg\max_{\theta \in \{\theta_0, \theta^*\}} f_{\mathbf{Z}}(\mathbf{z}|\theta)$$

$\hat{\theta}_{\theta_0, \theta^*}$ assigns to each possible outcome of the future experiment $\mathbf{z}$, in which the values of θ, θ_0, or θ^* make the experimental result more likely. The classification distance between θ_0, θ^*, and $CD(\theta_0, \theta^*)$ is defined as

$$CD(\theta_0, \theta^*) = 0.5\mathbb{P}\left(\hat{\theta}_{\theta_0, \theta^*}(\mathbf{Z}) = \theta_0 | \theta_0 \right) + 0.5\mathbb{P}\left(\hat{\theta}_{\theta_0, \theta^*}(\mathbf{Z}) = \theta^* | \theta^* \right) - 0.5$$

$CD(\theta_0, \theta^*) + 0.5$ is the best Bayes utility in an hypothesis test of θ_0 against θ^* using a uniform prior for θ and the 0/1 utility [15]. By subtracting 0.5 from this quantity, $CD(\theta_0, \theta^*)$ varies between 0 and 0.5 and is a distance. Also,

$$CD(\theta_0, \theta^*) = 0.5TV(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta^*}) = 0.25\|\mathbb{P}_{\theta_0} - \mathbb{P}_{\theta^*}\|_1,$$

where $TV(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta^*}) = \sup_A |\mathbb{P}_{\theta_0}(A) - \mathbb{P}_{\theta^*}(A)|$ and $\|\mathbb{P}_{\theta_0} - \mathbb{P}_{\theta^*}\|_1 = \int_{\mathcal{Z}} |\mathbb{P}_{\theta_0}(z) - \mathbb{P}_{\theta^*}(z)|dz$ is the $\mathcal{L}_1$-distance between probability measures.

Example 4 (Gaussian with known variance.). *Consider Examples 1 and 3, when $d = 1$, $\Sigma_0 = \sigma_0^2$, obtain*

$$CD_{\mathbf{Z}}(\theta_0, \theta^*) = \Phi\left(\frac{|\theta_0 - \theta^*|}{2\sigma_0} \right) - \frac{1}{2} \tag{8}$$

Note that, in this case, CD would be the same as BP if, instead of taking $g(x) = \sqrt{x}$, one chose $g(x) = \Phi(0.5\sqrt{x}) - 0.5$.

Although analytical expressions for CD are generally not available, it is possible to approximate it via numerical integration methods.

The choice between predictive dissimilarity functions depends on the type of guarantee the end-user wishes to obtain. In particular, *BP*, *KL*, and *CD* is not an exhaustive list of dissimilarities. However, some of their properties can be useful in obtaining a choice. For instance, although *BP* yields a metric on the parameter space, *KL* and *CD* are dissimilarities between probability functions. That is, although *BP* will generate pragmatic hypothesis that have parameter values numerically close to a given θ_0, *KL* and *CD* will yield pragmatic hypothesis that have parameter values lead to similar prediction about $\mathbf{Z}$. Also, although *KL* evaluates similarity between predictions from an information theoretic perspective, *CD* evaluates them from a perspective of hypothesis tests.

3.1. Singleton Hypotheses

We start by defining the pragmatic hypothesis associated to a singleton hypothesis. A singleton hypothesis is one in which the parameter assumes a single value, such as $H_0 : \theta = \theta_0$. In this case, the pragmatic hypothesis associated to H_0 is the set of points whose dissimilarity to θ_0 is at most ϵ, as formalized below.

Definition 4 (Pragmatic hypothesis for a singleton.)**.** *Let $H_0 : \theta = \theta_0$, $d_{\mathbf{Z}}$ be a predictive dissimilarity function and $\epsilon > 0$. The pragmatic hypothesis for H_0, $Pg(\{\theta_0\}, d_{\mathbf{Z}}, \epsilon)$, is*

$$Pg(\{\theta_0\}, d_{\mathbf{Z}}, \epsilon) = \{\theta^* \in \Theta : d_{\mathbf{Z}}(\theta_0, \theta^*) \leq \epsilon\}$$

Note that for $\epsilon_1 < \epsilon_2$, $Pg(\{\theta_0\}, d_{\mathbf{Z}}, \epsilon_1) \subseteq Pg(\{\theta_0\}, d_{\mathbf{Z}}, \epsilon_2)$.

Example 5 (Gaussian with known variance)**.** *Consider Examples 1 and 3 when $d = 1$, $\Sigma_0 = \sigma_0^2$ and $g(x) = \sqrt{x}$. It follows from Equations (1), (7) and (8) that*

$$Pg(\{\theta_0\}, BP_{\mathbf{Z}}, \epsilon) = [\theta_0 - \epsilon\sigma_0, \theta_0 + \epsilon\sigma_0]$$
$$Pg(\{\theta_0\}, KL_{\mathbf{Z}}, \epsilon) = \left[\theta_0 - \sqrt{2\epsilon}\sigma_0, \theta_0 + \sqrt{2\epsilon}\sigma_0\right]$$
$$Pg(\{\theta_0\}, CD_{\mathbf{Z}}, \epsilon) = \left[\theta_0 - 2\Phi^{-1}(0.5 + \epsilon)\sigma_0, \theta_0 + 2\Phi^{-1}(0.5 + \epsilon)\sigma_0\right]$$

Note that the size of each of the pragmatic hypothesis is proportional to σ_0. This occurs because each predictive dissimilarity functions makes the prediction error due to the unknown parameter value small with respect to that due to the data variability, σ_0^2.

3.2. Composite Hypotheses

Next, we consider pragmatic hypotheses for general hypotheses $H_0 : \theta \in \Theta_0$, where $\Theta_0 \subset \Theta$.

Definition 5. *For each hypothesis $\Theta_0 \subseteq \Theta$, predictive dissimilarity $d_{\mathbf{Z}}$ and $\epsilon > 0$, $Pg(\Theta_0, d_{\mathbf{Z}}, \epsilon)$ is the pragmatic hypothesis associated to Θ_0 induced by $d_{\mathbf{Z}}$ and ϵ. Whenever $d_{\mathbf{Z}}$ and ϵ are clear or not relevant to the result, we write $Pg(\Theta_0)$ instead of $Pg(\Theta_0, d_{\mathbf{Z}}, \epsilon)$.*

In order to construct these pragmatic hypotheses, we use logically coherent agnostic hypothesis tests. For each hypothesis, an agnostic hypothesis test can either reject it (1), accept it (0), or remain agnostic (1/2) [34]. Esteves et al. [2] shows that an agnostic hypothesis test is logically coherent if and only if it is based on a region estimator. Such tests are presented in Definition 7 and illustrated in Figure 3.
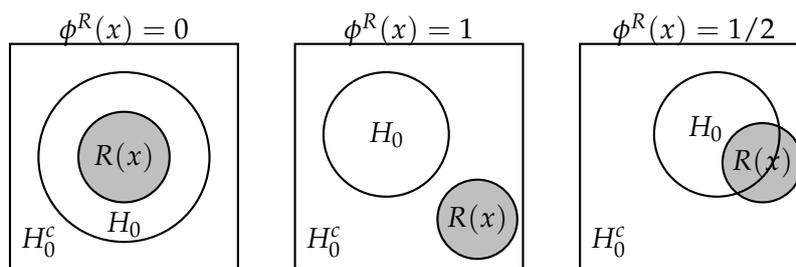
Figure 3. $\phi(x)$ is an agnostic test based on the region estimator $R(x)$ for testing H_0.

Definition 6. *Let $\mathcal{X}$ denote the sample space of the data used to test a hypothesis. A region estimator is a function, $R : \mathcal{X} \longrightarrow \mathcal{P}(\Theta)$, where $\mathcal{P}(\Theta)$ is the power set of Θ.*

Definition 7 (Agnostic test based on a region estimator.)**.** *The agnostic test based on the region estimator R for testing H_0, $\phi_{H_0}^R$, such that*

$$\phi_{H_0}^R(x) = \begin{cases} 0 & , \text{if } R(x) \subseteq H_0 \\ 1 & , \text{if } R(x) \subseteq H_0^c \\ \frac{1}{2} & , \text{otherwise.} \end{cases}$$

Besides the logical conditions on the hypothesis test, one might also impose logical restraints on how pragmatic hypotheses are constructed. For instance, let A and B be two hypothesis, such that B logically entails A, that is, $B \subseteq A$. If a logically coherent test accepts B, then it also accepts A. This property is called monotonocity [1,35–37]. One might also impose that Pg, such that if a logically coherent hypothesis test accepts $Pg(B)$, then it should also accept $Pg(A)$. Similarly, let $(A_i)_{i \in I}$ be a collection of hypothesis which cover A, that is, $A \subseteq \cup_{i \in I} A_i$. If a logically coherent hypothesis test rejects every A_i, then it rejects A. This property is called union consonance. One might also impose that Pg is such that, if a logically coherent hypothesis test rejects $Pg(A_i)$ for every i, then it should also reject $Pg(A)$. The above conditions define the logical coherence of a procedure for constructing pragmatic hypotheses.

Definition 8. *A procedure for constructing pragmatic hypothesis, Pg, is logically coherent if, for every logically coherent hypothesis test ϕ and sample point x:*

1. *If $\phi_{Pg(B)}(x) = 0$ for some $B \subseteq A$, then $\phi_{Pg(A)}(x) = 0$.*
2. *If $\phi_{Pg(A_i)}(x) = 1$ for every $i \in I$ and $A \subseteq \cup_{i \in I} A_i$, then $\phi_{Pg(A)}(x) = 1$.*

In order to motivate the above definition, consider that the frequencies of AA, AB, and BB in a given population are θ_1, θ_2, and θ_3, respectively. Note that $B := \{0.25, 0.5, 0.25\}$ is a subset of $A = \{(p^2, 2p(1-p), (1-p)^2) : p \in [0,1]\}$, which denotes the Hardy–Weinberg equilibrium. That is, if the frequencies AA, AB, and BB are, respectively, 0.25, 0.5, and 0.25, then the population follows the Hardy–Weinberg equilibrium. As a result, if one pragmatically accepts that the population satisfies the specified proportions, then one might also wish to pragmatically accept that the population follows the Hardy–Weinberg. Similarly, if one pragmatically rejects for every $p \in [0,1]$ that the frequencies of AA, AB, and BB are, respectively, p^2, $2p(1-p)$, and $(1-p)^2$, then one might also wish to pragmatically reject that the population follows the Hardy–Weinberg equilibrium. These conditions are assured in Definition 8.

In a logically coherent procedure for constructing pragmatic hypotheses, the pragmatic hypothesis associated to a composite hypothesis is completely determined by the pragmatic hypotheses associated to simple hypotheses. This result is presented in Theorem 1.

Theorem 1. *A procedure for constructing pragmatic hypothesis, Pg, is logically coherent if and only if, for every hypothesis, Θ_0, $Pg(\Theta_0) = \cup_{\theta \in \Theta_0} Pg(\{\theta\})$.*

Using Theorem 1, it is possible to determine a logically coherent procedure for constructing pragmatic hypotheses by determining only the pragmatic hypothesis associated to simple hypothesis, such as in Section 3.1. Theorem 1 is illustrated in Section 4. One can also obtain the following general relation between predictive dissimilarities.

Lemma 1. $Pg(\Theta_0, KL_{\mathbf{Z}}, 8\epsilon^2) \subseteq Pg(\Theta_0, CD_{\mathbf{Z}}, \epsilon)$.

Besides being logically coherent, it is often desirable in statistics [38,39] and in science [12,13] for a procedure to be invariant to reparametrization, so as to ensure that the procedure reaches the same conclusions whatever the coordinate system is used to specify both the sample and the parameter spaces. For instance, the pragmatic hypothesis that is obtained using the International metric system should be compatible to the one that is obtained using the English metric system. Invariance to reparametrization is formally presented in Definition 10.

Definition 9. $\left(\mathbb{P}^*_{\theta^*}\right)_{\theta^* \in \Theta^*}$ *is a reparameterization of $(\mathbb{P}_\theta)_{\theta \in \Theta}$ if there exists a bijective function, $f : \Theta \to \Theta^*$, such that for every $\theta \in \Theta$, $\mathbb{P}_\theta = \mathbb{P}^*_{f(\theta)}$.*

Definition 10. *Let $\left(\mathbb{P}_{\theta^*}^*\right)_{\theta^* \in \Theta^*}$ be a reparametrization of $(\mathbb{P}_\theta)_{\theta \in \Theta}$ by a bijective function, $f : \Theta \to \Theta^*$. Also, let $d_\mathbf{Z}$ and $d_\mathbf{Z}^*$ be predictive dissimilarity functions. The functions $d_\mathbf{Z}$ and $d_\mathbf{Z}^*$ are invariant to the reparametrization if for every logically coherent procedure for constructing pragmatic hypotheses, Pg,*

$$f[Pg(\Theta_0, d_\mathbf{Z}, \epsilon)] = Pg(f[\Theta_0], d_\mathbf{Z}^*, \epsilon),$$

Definition 10 states that, if Θ_0 is an hypothesis and invariance to reparametrization holds, then the pragmatic hypothesis obtained in a reparametrization of Θ_0, say $Pg(f[\Theta_0])$, is the same as the transformed pragmatic hypothesis associated to Θ_0, $f[Pg(\Theta_0)]$. Theorem 2 presents a sufficient condition for obtaining invariance to reparametrization.

Theorem 2. *Let $\left(\mathbb{P}_{\theta^*}^*\right)_{\theta^* \in \Theta^*}$ be a reparameterization of $(\mathbb{P}_\theta)_{\theta \in \Theta}$ given by a bijective function, f. If $d_\mathbf{Z}$ and $d_\mathbf{Z}^*$ satisfy $d_\mathbf{Z}(\theta_0, \theta) = d_\mathbf{Z}^*(f(\theta_0), f(\theta))$, then $d_\mathbf{Z}$ and $d_\mathbf{Z}^*$ are invariant to this reparametrization.*

Corollary 1. *If $d_\mathbf{Z}$ and $d_\mathbf{Z}^*$ are the same choice between KL, BP, or CD, then $d_\mathbf{Z}$ and $d_\mathbf{Z}^*$ are invariant to every reparametrization.*

The procedures for constructing pragmatic hypotheses induced by *KL* and *CD* also satisfy an additional property given by Theorem 3.

Theorem 3. *Let $\mathbf{Z}_m = (Z_1, \ldots, Z_m)$, where Z_i's are i.i.d. F_θ and $(F_\theta)_{\theta \in \Theta}$ is identifiable [40,41]. Also, let KL_m and CD_m be the dissimilarities calculated using $\mathbf{Z}_m$. If Pg is logically coherent, then, for every $\Theta_0 \subseteq \Theta$ and $\epsilon > 0$,*

(i) $(Pg(\Theta_0, KL_m, \epsilon))_{m \geq 1}$ and $(Pg(\Theta_0, CD_m, \epsilon))_{m \geq 1}$ are non-increasing sequences of sets
(ii) $Pg(\Theta_0, KL_m, \epsilon) \xrightarrow{m \to \infty} \Theta_0$ and $Pg(\Theta_0, CD_m, \epsilon) \xrightarrow{m \to \infty} \Theta_0$.

Theorem 3 states that the sequence of pragmatic hypotheses for Θ_0 induced by $d_{\mathbf{Z}_m}$ is non-increasing if the dissimilarity is evaluated by either KL or CD. The greater the number of observable quantities $\mathbf{Z}_m$, the easier it is to distinguish two parameter values θ_0 and θ^*, and therefore the smaller the amount of parameters that are taken as close to θ_0. Also, as the sample size goes to infinity, the pragmatic hypothesis associated to Θ_0 converges to to Θ_0. In other words, for each $\theta_0 \in \Theta_0$, no other parameter value can predict infinitely many observable quantities with a precision sufficiently close to that of θ_0.

4. Applications

In the following, pragmatic hypotheses for standard statistical problems are derived. Coscrato et al. [42] provide additional examples and methods for obtaining pragmatic hypotheses.

Example 6 (Gaussian with unknown variance.)**.** *Consider the setting from Example 5, but with σ^2 unknown and $0 < \sigma^2 \leq M^2$. In this case, the parameter is $\theta = (\mu, \sigma^2)$. Consider the composite hypothesis $H_0 : \{\mu_0\} \times (0, M^2]$, which is often written as $H_0 : \mu = \mu_0$. In this case, let $\theta_0 = (\mu_0, \sigma_0^2)$ and $\Theta_0 = \{\mu_0\} \times (0, M^2]$. Proceeding as in Example 5, it follows that*

$$Pg(\{\theta_0\}, BP_\mathbf{Z}, \epsilon) = [\mu_0 - \epsilon\sigma_0, \mu_0 + \epsilon\sigma_0] \times (0, M^2]$$
$$Pg(\Theta_0, BP_\mathbf{Z}, \epsilon) = [\mu_0 - \epsilon M, \mu_0 + \epsilon M] \times (0, M^2] \qquad \text{Theorem 1}$$

The rectangular shape of these pragmatic hypotheses seems to be unreasonable, as, for instance, whether a point (μ, σ^2) is close to (μ_0, σ_0^2) does not depend on σ_0^2. This is a consequence of the choice of δ in Example 5.

Figure 4 presents the pragmatic hypotheses for $H_0 : \mu = 0, \sigma^2 = 1$ and $H_0 : \mu = 0$ when $\epsilon = 0.1$ and $M^2 = 2$, and using the KL and CD dissimilarities. Contrary to BP, the hypotheses obtained from these dissimilarities do not have a rectangular shape. In particular, the triangular shape of the pragmatic hypotheses

for $H_0 : \mu = 0$ is such that the closer σ^2 is to 0, the smaller the range of values for μ that are included in the pragmatic hypothesis. This behavior might be desirable, as when σ^2 is small, there is little uncertainty about the value of $\mathbf{Z}$, and consequently a narrow interval of values of μ can predict Z with precision ϵ.
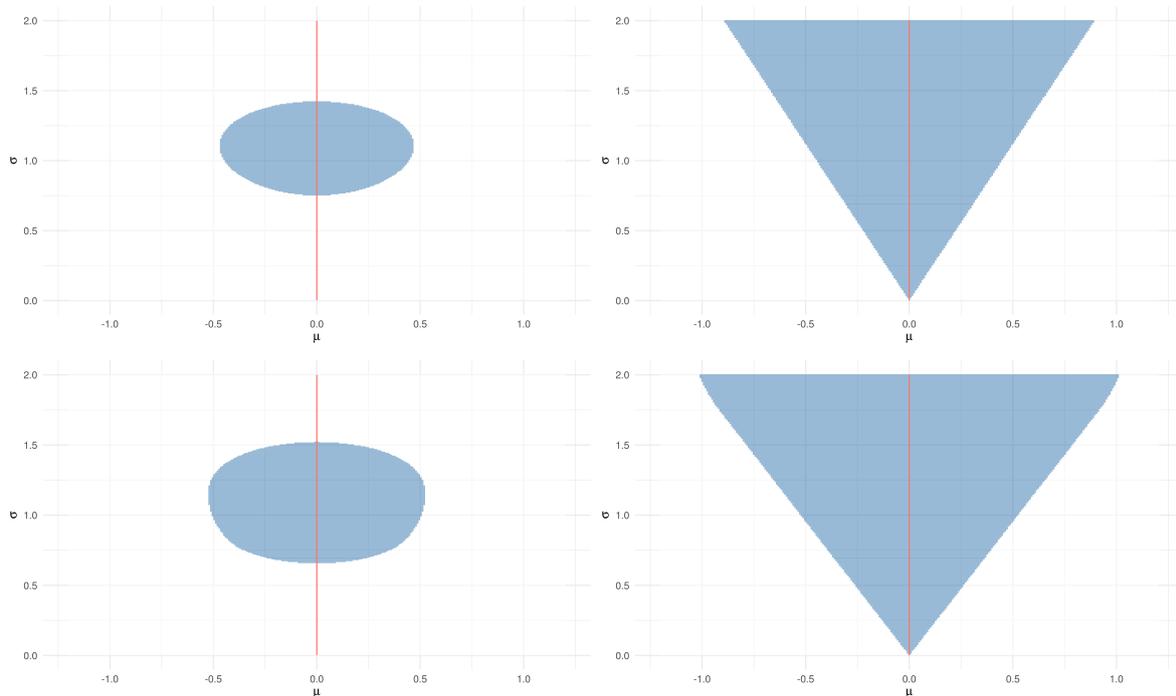
Figure 4. Pragmatic hypotheses in Example 6 for $H_0 : \mu = 0$ with KL (upper), CD (lower), $\epsilon = 0.1$, and $M^2 = 2$. H_0 is represented by a red line in all figures.

Example 7 (Hardy–Weinberg equilibrium). *Let $\mathbf{Z} \sim Multinomial(m, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, $\theta_i \geq 0$, and $\sum_{i=1}^{3} \theta_i = 1$. The Hardy–Weinberg (HW) hypothesis [43], H_0, which is depicted in the red curve in Figure 5 satisfies*

$$H_0 : \theta \in \Theta_0, \qquad \Theta_0 = \left\{ \left(p^2, 2p(1-p), (1-p)^2 \right) : 0 \leq p \leq 1 \right\}$$

If $\boldsymbol{\theta}_0^p = (p^2, 2p(1-p), (1-p)^2)$, $\delta_{\mathbf{Z}}(\mathbf{z}) = E[\|\mathbf{Z} - \mathbf{z}\|_2^2 | \boldsymbol{\theta} = \boldsymbol{\theta}_0^p]$ and $g(x) = \sqrt{x}$, then it follows from Example 2 that

$$BP_{\mathbf{Z}}(\boldsymbol{\theta}_0^p, \boldsymbol{\theta}^*) = \left(m \times \frac{(\theta_1 - p^2)^2 + (\theta_2 - 2p(1-p))^2 + (\theta_3 - (1-p)^2)^2}{p^2(1-p^2) + 2p(1-p)(1-2p(1-p)) + (1-p)^2(1-(1-p)^2)} \right)^{0.5}$$

The pragmatic hypotheses that are obtained using KL, BP, and CD for the HW hypothesis are depicted in Figure 5. The choice between BP or KL and CD has a large impact over the shape of the pragmatic hypotheses. Although, for BP, the width of the pragmatic hypothesis is approximately uniform along the HW curve, the width of the pragmatic hypotheses obtained using KL and CD is smaller towards the edges of the HW curve. This behavior could be expected, as towards the edges of the HW curve, $\mathbf{Z}$ has the smallest variability. The figure also depicts the challenge in calibrating KL. Although the pragmatic hypotheses for BP and CD have similar sizes when using $\epsilon = 0.1$, this result was obtained for KL while using $\epsilon = 0.01$.
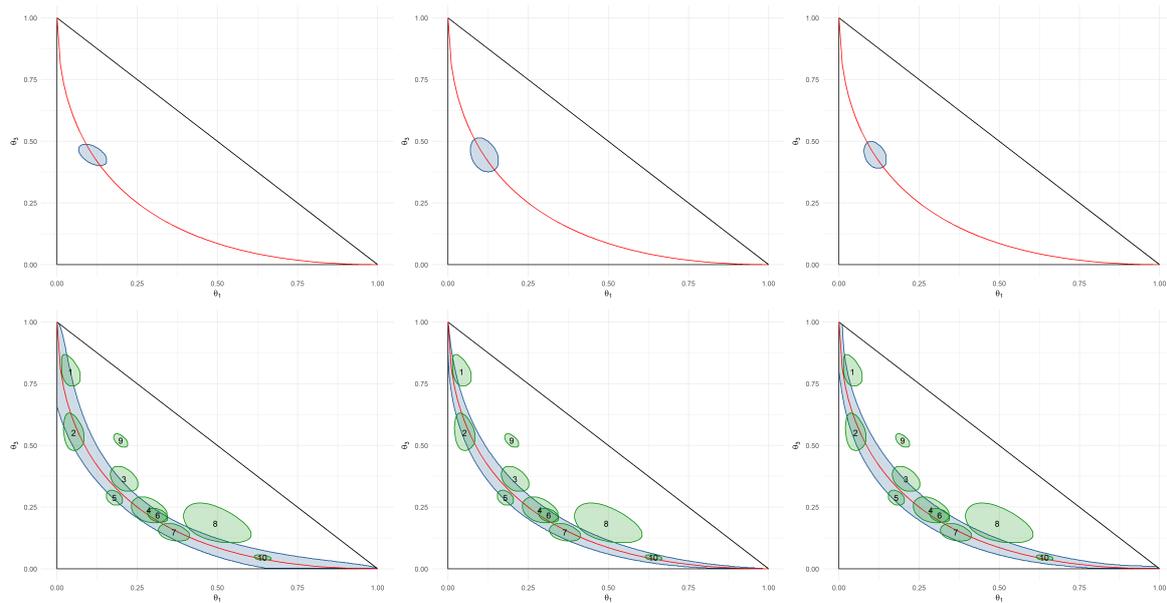
Figure 5. Pragmatic hypotheses obtained for the HW equilibrium, depicted in red, using $m = 20$, $\epsilon = 0.1$ for BP and CD and $\epsilon = 0.01$ for KL. The blue regions indicate the pragmatic hypothesis for HW and $p = \frac{1}{3}$ (top) and for HW (bottom). The lower, middle, and right panels were obtained, respectively, with BP, KL, and CD. The green regions in the right panels represents 80% HPD regions for the genotype distribution of each of the eight groups collected by Brentani et al. [44] and two simulated datasets.

The pragmatic hypotheses in Figure 5 are further tested using data from Brentani et al. [44], which is presented in Table 2. This study had the goal of verifying association between the APOE-ε4 gene and Alzheimer disease. The lower panels of Figure 5 present the 80% HPD regions for the distribution of this gene in each of the eight groups observed in the study. Additionally, they present two simulated datasets, 9 and 10. Groups 9 and 10 were generated by populations that were, respectively, not under and under the HW equilibrium. Group 9 and 10 fall, respectively, outside and inside of the pragmatic hypothesis.

Table 2. Genotype counts for the eight groups in Brentani et al. [44]. Also, the decision of the GFBST agnostic hypothesis test [2] for testing in each group the pragmatic Hardy–Weinberg equilibrium hypothesis with $m = 20$. The decisions are the same for *KL*, *BP*, and *CD*.

	AA	**AD**	**DD**	**Decision**
1	4	18	94	Agnostic
2	6	53	74	Accept
3	57	118	100	Agnostic
4	58	97	48	Agnostic
5	120	361	194	Agnostic
6	206	309	142	Accept
7	110	148	44	Accept
8	34	22	12	Agnostic
9	198	282	520	Reject
10	641	314	45	Accept

Example 8 (Bioequivalence). *Assume that $\mathbf{Z} = (X, Y) \sim N((\mu_1, \mu_2), \sigma^2 \mathbb{I}_2)$, with σ known. We derive the pragmatic hypothesis for $H_0 : \mu_1 = \mu_2$, that is, for $\{(\mu_1, \mu_2) \in \mathbb{R}^2 : \mu_1 = \mu_2\}$. Such a test might be used in a bioequivalence study, where X and Y are the concentrations of an active ingredient in a generic (test) drug medication and in the brand name (reference) medication [45], respectively. As H_0 is composite, it helps to derive the pragmatic hypothesis of its constituents.*

In order to do so, let $\theta_0 = (\mu_0, \mu_0)$, $\mu_0 \in \mathbb{R}$, $\theta^* = (\mu_1^*, \mu_2^*)$, and $H^{\theta_0} : \theta = \theta_0$. If $\delta_{\mathbf{Z},\theta^*}(\mathbf{z}) = E\left[(X - \mathbf{z}_1)^2 + (Y - \mathbf{z}_2)^2 | \theta = \theta^*\right]$ and $g(x) = \sqrt{x}$, then

$$BP_{\mathbf{Z}}(\theta_0, \theta^*) = \sqrt{\frac{(\mu_1^* - \mu_0)^2 + (\mu_2^* - \mu_0)^2}{2\sigma^2}}$$

Hence, $Pg(\{\theta_0\}, BP_{\mathbf{Z}}, \epsilon) = \left\{(\mu_1^*, \mu_2^*) : (\mu_1^* - \mu_0)^2 + (\mu_2^* - \mu_0)^2 \leq 2\epsilon^2\sigma^2\right\}$, which is a circle with center (μ_0, μ_0) and radius $\sqrt{2}\epsilon\sigma$, as depicted on the left panel of Figure 6. In this case, the pragmatic hypothesis is the Tier 1 Equivalence Test hypothesis suggested by the US Food and Drug Administration [45]. The pragmatic hypothesis for $H_0 : \mu_1 = \mu_2$ is obtained by taking the union of the pragmatic hypotheses associated to its constituents, as illustrated in the right panel of Figure 6. Specifically,

$$Pg(H_0, BP_{\mathbf{Z}}, \epsilon) = \left\{(\mu_1^*, \mu_2^*) : |\mu_2^* - \mu_1^*| \leq \epsilon\sigma\right\}$$

The pragmatic hypothesis for H_0 using KL is obtained similarly. Note that

$$KL_{\mathbf{Z}}(\theta_0, \theta^*) = 0.5 BP_{\mathbf{Z}}^2(\theta_0, \theta^*)$$

Therefore, $Pg(\{\theta_0\}, KL_{\mathbf{Z}}, \epsilon) = \left\{(\mu_1^*, \mu_2^*) : (\mu_1^* - \mu_0)^2 + (\mu_2^* - \mu_0)^2 \leq 2\epsilon\sigma^2\right\}$ and

$$Pg(H_0, KL_{\mathbf{Z}}, \epsilon) = Pg(H_0, KL_{\mathbf{Z}}, 0.5\epsilon^2)$$

The pragmatic hypothesis for H_0 that is obtained using CD has no analytic expression. However, by observing that $N(\mu, \sigma^2) = \mu + \sigma N(0,1)$, it is possible to show that there exists a monotonically increasing function, $h : \mathbb{R} \longrightarrow \mathbb{R}$, such that

$$Pg(H_0, CD_{\mathbf{Z}}, 0.5\epsilon^2) = \left\{(\mu_1^*, \mu_2^*) : |\mu_2^* - \mu_1^*| \leq h(\epsilon)\sigma\right\}$$

(a) $H_0 : \mu_1 = \mu_2 = \mu_0$.   **(b)** $H_0 : \mu_1 = \mu_2$.
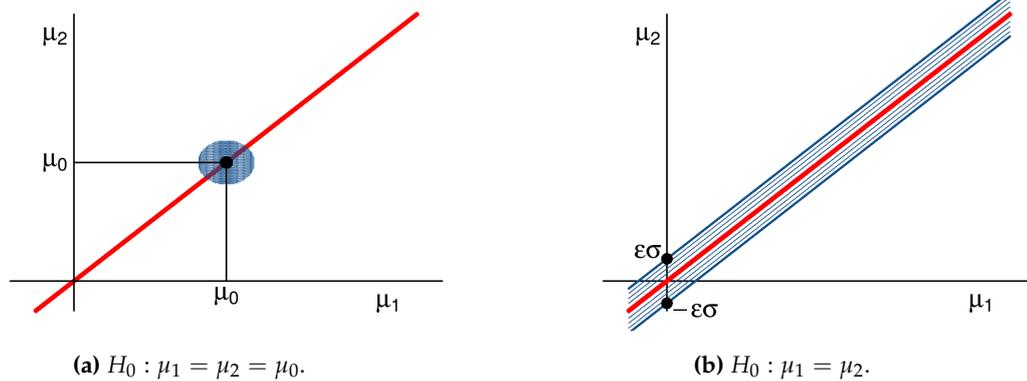
Figure 6. Pragmatic hypotheses using BP in Example 8 when σ is known.

That is, the pragmatic hypothesis associated to H_0 have the same shape as in the right panel of Figure 6. They differ solely on how many standard deviations correspond to the width of the pragmatic hypothesis.

5. Final Remarks

The spiral structure studied in the work by the authors of [10] can be used to describe scientific evolution. However, in order for the analogy to be complete, it is necessary to indicate what types of scientific theories or hypotheses are effectively tested in the acceptance vertex of the hexagon of oppositions. We defend that these are pragmatic hypotheses, which are sufficiently precise for the end-user of the theory.

In order to make this statement formal, we introduce three methods for constructing a pragmatic hypothesis associated to a precise hypothesis. These methods are based on three predictive dissimilarity functions: KL, BP and CD. Each of these methods have different advantages. For instance, the scale of BP and CD is more interpretable than KL, making it easier to determine whether the former are large or small. On the other hand, BP relies on the definition of more functions than KL and CD, such as $\delta_{\mathbf{Z},\theta_0}(\mathbf{z})$ in Definition 2. If these function are chosen inadequately, then the shape of the resultant pragmatic hypothesis might be counterintuitive or meaningless. Finally, CD often does not have an analytic expression. It relies on numerical integration over the sample space, which can be taxing in high dimensions.

The applications at Section 4 present adequate choices of metrics and bounds (defining pragmatic hypotheses) for some given theoretical and experimental setups. Nevertheless, the authors did not propose a general or automated recipe for making these choices, nor do they think this to be a feasible goal. In future research, the authors intend to explore a variety of application cases, some using historical data of important experiments, and discuss possible choices of metrics and bounds for each case. The authors hope that, in time, the accumulation of such examples will provide useful guidelines for the good use of methods developed in this paper, in the same way that the statistical literature provides useful guidelines for choosing good statistical models for practical applications.

Appendix A. Proofs

Proof of Lemma 1. Let $\theta \in Pg(\Theta_0, KL_{\mathbf{Z}}, 8\epsilon^2)$. It follows from Theorem 1 that there exist $\theta_0 \in \Theta_0$, such that $KL_{\mathbf{Z}}(\theta_0, \theta) \leq 0.5\epsilon^2$. Conclude from Pinsker's inequality that $CD_{\mathbf{Z}}(\theta_0, \theta) \leq \sqrt{8^{-1}KL_{\mathbf{Z}}(\theta_0, \theta)} = \epsilon$, that is, $\theta \in Pg(\Theta_0, CD_{\mathbf{Z}}, \epsilon)$. □

Proof of Theorem 1. Let Pg be logically coherent. Pick an arbitrary $\theta_0 \in \Theta_0$ and note that, if $R(x) \equiv Pg(\{\theta_0\})$, then $\phi^R_{Pg(\{\theta_0\})}(x) = 0$. Since Pg is logically coherent, conclude that $\phi^R_{Pg(\Theta_0)}(x) \equiv 0$, that is, $Pg(\{\theta_0\}) \subseteq Pg(\Theta_0)$. Since $\theta_0 \in \Theta_0$ was arbitrary, conclude that

$$\bigcup_{\theta_0 \in \Theta_0} Pg(\{\theta_0\}) \subseteq Pg(\Theta_0) \tag{A1}$$

Next, let $R(x) \equiv \bigcap_{\theta_0 \in \Theta_0} Pg(\{\theta_0\})^c$. For every $\theta_0 \in \Theta_0$, $\phi^R_{Pg(\{\theta_0\})}(x) = 1$. Since Pg is logically coherent, $\phi^R_{Pg(\Theta_0)} \equiv 1$, that is, $Pg(\Theta_0) \subseteq R^c \equiv \bigcup_{\theta_0 \in \Theta_0} Pg(\{\theta_0\})$. Conclude that

$$Pg(\Theta_0) \subseteq \bigcup_{\theta_0 \in \Theta_0} Pg(\{\theta_0\}) \tag{A2}$$

It follows from Equations (A1) and (A2) that $Pg(\Theta_0) = \bigcup_{\theta_0 \in \Theta_0} Pg(\{\theta_0\})$. It also follows from direct calculation that, if $Pg(\Theta_0) = \bigcup_{\theta_0 \in \Theta_0} Pg(\{\theta_0\})$, then Pg is logically coherent. □

Proof of Theorem 2. Let $\Theta_0 \subseteq \Theta$

$$
\begin{aligned}
Pg(f[\Theta_0], d_{\mathbf{Z}}^*, \epsilon) &= \{\theta^* \in \Theta^* : \exists \theta_0^* \in f[\Theta_0] \text{ s.t. } d_{\mathbf{Z}}^*(\theta^*, \theta_0^*) \leq \epsilon\} \\
&= \{\theta^* \in \Theta^* : \exists \theta_0^* \in f[\Theta_0] \text{ s.t. } d_{\mathbf{Z}}(f^{-1}(\theta^*), f^{-1}(\theta_0^*)) \leq \epsilon\} \\
&= f\left[\{\theta \in \Theta : \exists \theta_0 \in \Theta_0 \text{ s.t. } d_{\mathbf{Z}}(\theta, \theta_0) \leq \epsilon\}\right] \\
&= f\left[Pg(\Theta_0, d_{\mathbf{Z}}, \epsilon)\right]
\end{aligned}
$$

$\square$

Proof of Theorem 3. Since the Z_i's are i.i.d., $\mathrm{KL}_m(\theta_0, \theta^*) = m\mathrm{KL}_{Z_1}(\theta_0, \theta^*)$. It follows that

$$
\begin{aligned}
Pg(\Theta_0, KL_m, \epsilon) &= \bigcup_{\theta_0 \in \Theta_0} Pg(\{\theta_0\}, \mathrm{KL}_m, \epsilon) \\
&= \bigcup_{\theta_0 \in \Theta_0} Pg(\{\theta_0\}, m\mathrm{KL}_{Z_1}, \epsilon) = \bigcup_{\theta_0 \in \Theta_0} \left\{\theta^* \in \Theta : KL_{Z_1}(\theta_0, \theta^*) \leq m^{-1}\epsilon\right\}
\end{aligned}
$$

Thus, $(Pg(\Theta_0, KL_m, \epsilon))_{m \geq 1}$ is a non-increasing sequence of sets. It follows that

$$
\begin{aligned}
\lim_{m \to \infty} Pg(\Theta_0, KL_m, \epsilon) &= \bigcap_{m \geq 1} \bigcup_{\theta_0 \in \Theta_0} \left\{\theta^* \in \Theta : \mathrm{KL}_{Z_1}(\theta_0, \theta^*) \leq m^{-1}\epsilon\right\} \\
&= \bigcup_{\theta_0 \in \Theta_0} \bigcap_{m \geq 1} \left\{\theta^* \in \Theta : \mathrm{KL}_{Z_1}(\theta_0, \theta^*) \leq m^{-1}\epsilon\right\} \\
&= \bigcup_{\theta_0 \in \Theta_0} \left\{\theta^* \in \Theta : \mathrm{KL}_{Z_1}(\theta_0, \theta^*) = 0\right\} \\
&= \bigcup_{\theta_0 \in \Theta_0} \{\theta_0\} = \Theta_0
\end{aligned}
$$

where the next-to-last equality follows from the assumption that $(F_\theta)_{\theta \in \Theta}$ is identifiable. The proofs for the *CD* divergence follows from the fact that $\mathrm{TV}(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta^*}) \leq \sqrt{\mathrm{KL}(\mathbb{P}_{\theta_0}, \mathbb{P}_{\theta^*})}$. $\square$

References

1. Izbicki, R.; Esteves, L.G. Logical consistency in simultaneous statistical test procedures. *Log. J. IGPL.* **2015**, *23*, 732–758. [CrossRef]
2. Esteves, L.G.; Izbicki, R.; Stern, J.M.; Stern, R.B. The logical consistency of simultaneous agnostic hypothesis tests. *Entropy* **2016**, *18*, 256. [CrossRef]
3. Stern, J.; Izbicki, R.; Esteves, L.; Stern, R. Logically-Consistent Hypothesis Testing and the Hexagon of Oppositions. *Log. J. IGPL* **2017**, *25*, 741–757. [CrossRef]
4. Blanché, R. *Structures Intellectuelles: Essai sur l'Organisation Systématique des Concepts*; Vrin: Paris, France, 1966. (In French)
5. Béziau, J.Y. The power of the hexagon. *Log. Univers.* **2012**, *6*, 1–43. [CrossRef]
6. Béziau, J.Y. Opposition and order. In *New Dimensions of the Square of Opposition*; Béziau, J.Y., Gan-Krzywoszynska, K., Eds.; Philosophia Verlag: Munich, Germany, 2015; pp. 1–11.
7. Carnielli, W.; Pizzi, C. *Modalities and Multimodalities (Logic, Epistemology, and the Unity of Science)*, Springer: Berlin, Germany, 2008.
8. Dubois, D.; Prade, H. On several representations of an uncertain body of evidence. In *Fuzzy Information and Decision Processes*; Gupta, M., Sanchez, E., Eds.; Elsevier: Amsterdam, Netherlands, 1982; pp. 167–181.
9. Dubois, D.; Prade, H. From Blanché's Hexagonal Organization of Concepts to Formal Concept Analysis and Possibility Theory. *Log. Univers.* **2012**, *6*, 149–169. [CrossRef]
10. Gallais, P.; Pollina, V. Hegaxonal and Spiral Structure in Medieval Narrative. *Yale Fr. Stud.* **1974**, *51*, 115–132. [CrossRef]

11. Gallais, P. *Dialectique Du Récit Mediéval: Chrétien de Troyes et l'Hexagone Logique*; Rodopi: Amsterdam, Netherlands, 1982. (In French)

12. Stern, J.M. Symmetry, Invariance and Ontology in Physics and Statistics. *Symmetry* **2011**, *3*, 611–635. [CrossRef]

13. Stern, J.M. Continuous versions of Haack's Puzzles: Equilibria, Eigen-States and Ontologies. *Log. J. IGPL* **2017**, *25*, 604–631. [CrossRef]

14. DeGroot, M.H.; Schervish, M.J. *Probability and Statistics*; Pearson Education: London, UK, 2012.

15. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer Science & Business Media: Berlin, Germany, 2013.

16. Bucher, J.L. *The Metrology Handbook*, 2nd ed.; ASQ Quality Press: Milwaukee, WI, USA, 2012.

17. Czichos, H.; Saito, T.; Smith, L. *Springer Handbook of Metrology and Testing*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2011.

18. Cohen, R.; Crowe, K.; DuMond, J. *The Fundamental Constants of Physics*; CODATA/Interscience Publishers: Geneva, Switzerland, 1957.

19. Cohen, R. Mathematical Analysis of the Universal Physical Constants. *Il Nuovo Cimento* **1957**, *6*, 187–214. [CrossRef]

20. Lévy-Leblond, J. On the conceptual nature of the physical constants. *Il Nuovo Cimento* **1977**, *7*, 187–214.

21. Pakkan, M.; Akman, V. Hypersolver: A graphical tool for commonsense set theory. *Inform. Sci.* **1995**, *85*, 43–61. [CrossRef]

22. Akman, V.; Pakkan, M. Nonstandard set theories and information management. *J. Intell. Inf. Syst.* **1996**, *6*, 5–31. [CrossRef]

23. Wainwright, M.J. Stochastic Processes on Graphs with Cycles: Geometric and Variational Approaches. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, April 2002.

24. Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*; Springer: Heidelberg, Germany, 2006.

25. Iordanov, B. HyperGraphDB: A generalized graph database. In *International Conference on Web-Age Information Management*; Springer: Berlin, Germany, 2010, pp. 25–36.

26. Gelman, A.; Vehtari, A.; Jylänki, P.; Sivula, T.; Tran, D.; Sahai, S.; Blomstedt, P.; Cunningham, J.P.; Schiminovich, D.; Robert, C. Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data. *arXiv* **2014**, arXiv:1412.4869.

27. Greimas, A. *Structural Semantics: An Attempt at a Method*; University of Nebraska Press: Lincoln, NE, USA, 1983.

28. Propp, V. *Morphology of the Folktale*; University of Texas Press: Austin, TX, USA, 2000.

29. Stern, J.M. Jacob's Ladder and Scientific Ontologies. *Cybern. Human Knowing* **2014**, *21*, 9–43.

30. Stern, J.M. Constructive Verification, Empirical Induction, and Falibilist Deduction: A Threefold Contrast. *Information* **2011**, *2*, 635–650. [CrossRef]

31. Abraham, R.; Marsden, J.E. *Foundations of Mechanics*; Addison-Wesley: Boston, MA, USA, 2013.

32. Hawking, S. *The Illustrated On the Shoulders of Giants: The Great Works of Physics and Astronomy*; Running Press: Philadelphia, PA, USA, 2004.

33. Stern, J.M.; Nakano, F. Optimization Models for Reaction Networks: Information Divergence, Quadratic Programming and Kirchhoff's Laws. *Axioms* **2014**, *3*, 109–118. [CrossRef]

34. Coscrato, V.; Izbicki, R.; Stern, R.B. Agnostic tests can control the type I and type II errors simultaneously. *Braz. J. Probab. Stat.* **2019**. Available online: https://www.imstat.org/wp-content/uploads/2019/01/BJPS431.pdf (accessed on 9 September 2019).

35. Izbicki, R.; Fossaluza, V.; Hounie, A.G.; Nakano, E.Y.; de Braganca Pereira, C.A. Testing allele homogeneity: The problem of nested hypotheses. *BMC Genet.* **2012**, *13*, 103. [CrossRef] [PubMed]

36. da Silva, G.M.; Esteves, L.G.; Fossaluza, V.; Izbicki, R.; Wechsler, S. A bayesian decision-theoretic approach to logically-consistent hypothesis testing. *Entropy* **2015**, *17*, 6534–6559. [CrossRef]

37. Fossaluza, V.; Izbicki, R.; da Silva, G.M.; Esteves, L.G. Coherent hypothesis testing. *Am. Statist.* **2017**, *71*, 242–248. [CrossRef]

38. Pereira, C.A.d.B.; Stern, J.M.; Wechsler, S. Can a Signicance Test be Genuinely Bayesian? *Bayesian Anal.* **2008**, *3*, 79–100. [CrossRef]

39. Stern, J.M.; Pereira, C.A.D.B. Bayesian Epistemic Values: Focus on Surprise, Measure Probability! *Log. J. IGPL.* **2014**, *22*, 236–254. [CrossRef]

40. Casella, G.; Berger, R.L. *Statistical Inference*; Duxbury Press: Pacific Grove, CA, USA, 2002.

41. Wechsler, S.; Izbicki, R.; Esteves, L.G. A Bayesian look at nonidentifiability: A simple example. *Am. Statist* **2013**, *67*, 90–93. [CrossRef]

42. Coscrato, V.; Esteves, L.G.; Izbicki, R.; Stern, R.B. Interpretable hypothesis tests. *arXiv* **2019**, arXiv:1904.06605.

43. Hardy, G. Mendelian proportions in a mixed population. 1908. *Yale J. Biol. Med.* **2003**, *76*, 79. [PubMed]

44. Brentani, H.; Nakano, E.Y.; Martins, C.B.; Izbicki, R.; Pereira, C.A.d.B. Disequilibrium coefficient: A Bayesian perspective. *Stat. Appl. Genet. Mol.* **2011**, *10*. [CrossRef]

45. Chow, S.C.; Song, F.; Bai, H. Analytical similarity assessment in biosimilar studies. *AAPS J.* **2016**, *18*, 670–677. [CrossRef]