Logically-consistent hypothesis testing and the hexagon of oppositions

JULIO MICHAEL STERN*, *University of São Paulo, Rua do Matão, 1010, USP, São Paulo, SP, BR, 05508-090.*

RAFAEL IZBICKI**, *Departamento de Estatística, Federal University of São Carlos, UFSCar, Rodovia Washington Luis, km 235, São Carlos, SP, BR, 13565-905.*

LUIS GUSTAVO ESTEVES†, *University of São Paulo, Rua do Matão, 1010, USP, São Paulo, SP, BR, 05508-090.*

RAFAEL BASSI STERN‡, *Departamento de Estatística, Federal University of São Carlos, UFSCar, Rodovia Washington Luis, km 235, São Carlos, SP, BR, 13565-905.*

Abstract

Although logical consistency is desirable in scientific research, standard statistical hypothesis tests are typically logically inconsistent. To address this issue, previous work introduced agnostic hypothesis tests and proved that they can be logically consistent while retaining statistical optimality properties. This article characterizes the credal modalities in agnostic hypothesis tests and uses the hexagon of oppositions to explain the logical relations between these modalities. Geometric solids that are composed of hexagons of oppositions illustrate the conditions for these modalities to be logically consistent. Prisms composed of hexagons of oppositions show how the credal modalities obtained from two agnostic tests vary according to their threshold values. Nested hexagons of oppositions summarize logical relations between the credal modalities in these tests and prove new relations.

Keywords: Statistical hypothesis tests, hexagon of oppositions, logical consistency, agnostic hypothesis tests, probabilistic and alethic modalities.

1 Introduction

Logical consistency is desirable in scientific research. It can be hard to interpret inconsistent conclusions and to develop theories that explain them. As a result, they are detrimental to scientific communication and decision-making.

Despite the desirability of logical consistency, it is not satisfied by many methods of statistical hypothesis testing. For instance, consider that a researcher simultaneously tests two hypotheses, A and B, such that 'not B' implies 'not A'. In such a situation, one expects that the rejection of B would imply the rejection of A. However, this property does not apply to standard tests, such as the ones based on the p-value or the Bayes factor [17, 22, 24].

*E-mail: jmstern@hotmail.com
**E-mail: rizbicki@gmail.com
†E-mail: lesteves@ime.usp.br
‡E-mail: rbstern@gmail.com

doi:10.1093/jigpal/jzx024

TABLE 1. Modalities of agnostic hypothesis tests

Modality	Definition	Name	Equivalence	Interpretation
$\Box H$	$\mathcal{L}(H)=0$	Necessity (A)	$\Delta H \wedge \Diamond H$	H is accepted
$\neg\Diamond H$	$\mathcal{L}(H)=1$	Impossibility (E)	$\Delta H \wedge \neg\Box H$	H is rejected
∇H	$\mathcal{L}(H)=0.5$	Contingency (Y)	$\Diamond H \wedge \neg\Box H$	H not decided
$\Diamond H$	$\mathcal{L}(H)<1$	Possibility (I)	$\Box H \vee \nabla H$	H not rejected
$\neg\Box H$	$\mathcal{L}(H)>0$	Non-necessity (O)	$\neg\Diamond H \vee \nabla H$	H not accepted
ΔH	$\mathcal{L}(H)\neq 0.5$	Non-contingency (U)	$\Box H \vee \neg\Diamond H$	H is decided

Besides this flaw, one can find other logical inconsistencies in standard hypothesis testing procedures. da Silva et al. [9], Izbicki and Esteves [20] classify the types of logical inconsistencies in these procedures into failures to satisfy at least one of four of conditions of classical logics. If all conditions are satisfied by a test, then it is defined as logically consistent. This use of the term contrasts with the one in the literature of non-classical logics, in which (in)consistency often refers to forms of modifying classical notions of contradiction, negation, conflation, etc. so as to allow logical systems with greater flexibility than classical logic [6, 7, 26]. Izbicki and Esteves [20] shows that a standard test is logically consistent if and only if it is based on point estimation, which generally does not satisfy statistical optimality.

This grim result motivated Esteves et al. [16] to investigate *agnostic hypothesis tests*. Such a test is a function, $\mathcal{L}$, that assigns to each hypothesis, H, a value in $\{0, 0.5, 1\}$. H is accepted when $\mathcal{L}(H)=0$, H is rejected when $\mathcal{L}(H)=1$ and the test non-comitally neither accepts nor rejects H when $\mathcal{L}(H)=0.5$. In this article, we use agnostic tests to define *credal modalities*, as described in Table 1.

Using this framework, Esteves et al. [16] shows that the agnostic generalizations of some standard hypothesis tests can achieve logical consistency. Indeed, a hypothesis test is logically consistent if and only if it is based on a region estimator. As a result, there exist agnostic hypothesis tests that are both logically consistent and statistically optimal.

This positive result motivates the development of schematic diagrams for explaining agnostic tests and their logical properties. The relations between the credal modalities in Table 1 can be represented by the hexagon of oppositions [3, 4]. It follows from the definition in Table 1 that these credal modalities are defined by opposition relations in well-ordered structures, such as the ones in Béziau [1, 2]. Although this article considers solely total orders, extensions of the hexagon of oppositions to partial orders can be found in Demey and Smessaert [12].

Figure 1 represents the standard *hexagon of oppositions*. Each credal modality is associated with one vertex in the logical hexagon, namely: A (necessity), E (impossibility), I (possibility), O (non-necessity), U (non-contingency) and Y (contingency).[1] In Figure 1, logical relations between modalities are represented as edges, which are classified according to the following types:

[1]The four vowels used to label the vertices of the traditional square of opposition are chosen from the Latin word *AfIrmo*, for affirming universal (necessary) and particular (possible) modal propositions, and *nEgO* for negating them. [3] extended the square into the hexagon of opposition, using the two remaining French vowels to label the top and bottom of its six vertices. Some confusion comes from the historical use of four vowels (A, E, I and U) to describe which part(s) of a modal proposition (modus and dictum) are negated. This description follows the mnemonic rule: *nihil A, E dictum negat, Ique modum, U totum*, meaning: A- neither modus nor dictum is negated, E- dictum is negated I- modus is negated, U- modus and dictum are negated. Finally, each corner of the square of opposition is traditionally labelled by an anagram describing the (four) equivalent modal propositions obtained by using, respectively, the modality of — necessary, impossible, contingent and possible. Mnemonic anagrams are provided by the Latin words: (A) Purpurea — purple, (E) Iliace — colicky, (I) Amabimus — lovely, and (O) Edentuli — no-tooth; see [15, v.II, p.109–111].

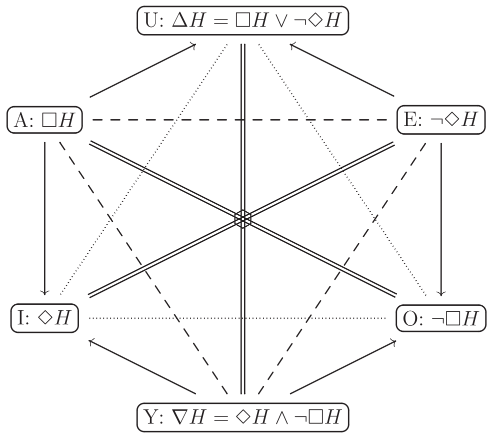

FIG. 1. The hexagon of oppositions for an agnostic hypothesis test.

- arrows: indicate *logical implications*, i.e., if modality J points to K and J is obtained, then K is also obtained.
- dashed-lines: indicate *contrariety*, i.e., modalities connected by these lines cannot both be true.
- dotted-lines: indicate *sub-contrariety*, i.e., modalities connected by these lines cannot both be false.
- double-lines: indicate *contradictions*, i.e., modalities connected by these lines cannot either be both true or be both false.

This article characterizes the credal modalities from the statistical theory developed in Esteves et al. [16] and uses the hexagon of oppositions to explain the logical properties that arise in these modalities. Section 2 uses polyhedra based on the hexagon of oppositions to present the four conditions for a test to be logically consistent. Next, hexagonal prisms illustrate the construction of the modalities that arise from standard hypothesis tests (Section 3) and that arise from a logically consistent agnostic test (Section 4). Section 5 uses a nested hexagon of oppositions to discuss the logical relations between the modalities obtained in Sections 3 and 4.

Note that Sections 3 and 4 examine two different types of credal modalities. To avoid ambiguity, the credal modality defined in each section is represented by a different symbol. Section 3 defines credal modalities based on threshold values on a *probability measure* over the space of hypotheses. Since probability is an additive measure, the corresponding *probabilistic modalities* are represented by superposing the plus sign to the standard modal symbols ($\boxplus$, $\Diamond\!\!\!\!+$ $\nabla\!\!\!\!+$ and $\Delta\!\!\!\!+$). In contrast, Section 4 defines credal modalities based on threshold values on a *possibilistic measure* over the space of hypotheses. The corresponding *alethic modalities* are denoted by the standard modal symbols ($\square$, $\Diamond$, ∇ and Δ). For a detailed explanation of probability and possibility measures see Borges and Stern [5], Dubois and Prade [13], Stern and Pereira [30].

2 Logical conditions on agnostic tests

To discuss the logical conditions on statistical tests, it is necessary to present some definitions. A *parameter* is an unknown or unobservable quantity with possible values in an arbitrary set, Θ, named the *parameter space*. A *statistical hypothesis*, H, is a statement about a parameter. It is of the form $H : \theta \in \Theta_H$, where Θ_H is an element of an arbitrary σ-field over Θ. For example, the parameter

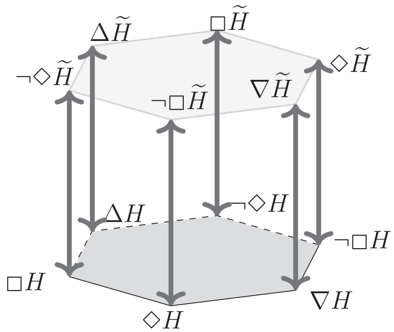

FIG. 2. A representation of invertibility using two hexagons of oppositions. $\widetilde{H}$ denotes the negation of a hypothesis H. Two modalities are connected by a bidirectional arrow if each one implies and is implied by the other.

space could be $\Theta = \mathbb{R}^d$, $d \in \mathbb{N}^*$, and a hypothesis could be stated as a linear equation, $H : A\theta = b$, corresponding to the hypothesis set $\Theta_H = \{\theta \in \mathbb{R}^d | A\theta = b\}$. Whenever there is no ambiguity, the symbol H is used as a shortcut for Θ_H, although the former is a statement and the latter is a set. A *statistical hypothesis test* is a function that attributes credal modalities to each statistical hypothesis.

Esteves et al. [16] introduces four logical consistency conditions for statistical hypothesis tests: invertibility, monotonicity, union consonance and intersection consonance. In the following, these properties are represented using geometric solids composed of hexagons of oppositions.

2.1 Invertibility

Invertibility restrains the conclusions that can be obtained when simultaneously testing a hypothesis, H, and its set complement, $\widetilde{H} = \Theta - H$. If invertibility holds, then either both hypotheses are undecided or one is accepted and the other is rejected. This restriction is represented symbolically in Definition 2.1:

DEFINITION 2.1 (Invertibility)

$$(\Box H \iff \neg\Diamond\widetilde{H}) \ \wedge \ (\nabla H \iff \nabla\widetilde{H}).$$

Figure 2 uses two hexagons of opposition to represent Definition 2.1. The hexagons for H (below) and $\widetilde{H}$ (above) are joined by double-arrowed edges to form a hexagonal prism. These edges illustrate that, if invertibility holds, then the connected modalities necessarily occur simultaneously. The modalities in the $\widetilde{H}$ hexagon are inverted in relation to the modalities in the H hexagon. As a result, if invertibility holds, then one obtains the following logical equivalences: $A \leftrightarrow \widetilde{\mathrm{E}}$, $E \leftrightarrow \widetilde{\mathrm{A}}$, $I \leftrightarrow \widetilde{\mathrm{O}}$, $O \leftrightarrow \widetilde{\mathrm{I}}$, $U \leftrightarrow \widetilde{\mathrm{U}}$, $Y \leftrightarrow \widetilde{\mathrm{Y}}$.

2.2 Monotonicity

Monotonicity restrains the conclusions that can be obtained when simultaneously testing nested hypotheses. Two hypotheses, H and H', are nested if H implies H' ($H \subseteq H'$). If monotonicity holds, then the conclusion obtained for H' is always at least as favourable as the conclusion obtained for

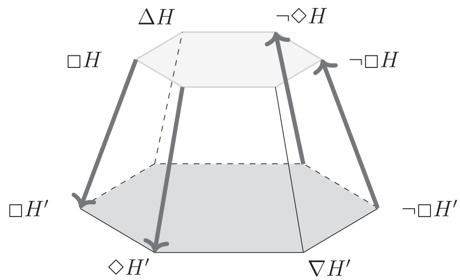

FIG. 3. Logical implications induced by monotonicity (blue arrows). The figure displays stacked hexagons of statistical modalities for a sequence of nested hypotheses. Each horizontal cut represents a different hypothesis; in the figure we display two of them: $H \subseteq H'$.

H. That is, if H is accepted, then H' is accepted and, if H is possible, then H' is possible. This restriction is represented symbolically in Definition 2.2.

DEFINITION 2.2 (Monotonicity)

$$H \subseteq H' \Rightarrow \begin{cases} \Box H \Rightarrow \Box H' \\ \Diamond H \Rightarrow \Diamond H'. \end{cases}$$

Figure 3 represents monotonicity through the hexagon of oppositions. The hexagon for H' is larger than the one for H as a means to illustrate that $H \subseteq H'$. These hexagons are connected by additional edges to form a pyramidal frustum. Each horizontal cut of this solid can be thought to represent a new hypothesis. These hypotheses imply the ones below them and are implied by the ones above them. The arrows that go from the top of the frustum to its bottom represent that, if a hypothesis is below another, then the conclusion obtained for the lower one should be at least as favourable as the one obtained for the upper one. The arrows that go in the opposite direction are obtained by taking the negation of the previous implications. As a result of monotonicity, one obtains the following logical implications: A $\rightarrow$ A', I $\rightarrow$ I', O' $\rightarrow$ O, E' $\rightarrow$ E.

2.3 Consonance

Consonance restricts the conclusions that one can obtain when simultaneously testing a set of hypotheses and their unions or intersections. Let $\{H_i\}_{i \in I}$ be an arbitrary collection of hypotheses. Under strong union consonance, if one concludes that $\cup_{i \in I} H_i$ is possible, then at least one H_i is possible. Similarly, under strong intersection consonance, if one concludes that $\cap_{i \in I} H_i$ is not necessary, then at least one H_i is not necessary. These restrictions are represented symbolically in Definition 2.3 and 2.4.

DEFINITION 2.3 (Strong union consonance)
For every I,

$$\Diamond(\cup_{i \in I} H_i) \rightarrow \exists i \in I \text{ s.t. } \Diamond H_i.$$

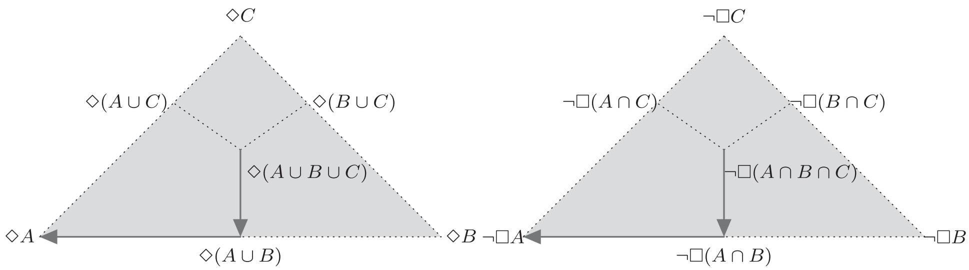

FIG. 4. Logical implications induced by strong union consonance (left) and strong intersection consonance (right). If one obtains the conclusion in the centre of the triangle, then one obtains at least one of the conclusions in the middle of the edges of the triangle and another in the vertices connected to this edge. That is, there exists at least one path such as the one illustrated by the blue arrows.

Strong union consonance is equivalent to

$$\forall i \in I, \neg\diamond H_i \Rightarrow \neg\diamond(\cup_{i\in I} H_i)$$

that is, the $\neg\diamond$-modality is closed under the union.

DEFINITION 2.4 (Strong intersection consonance)
For every I,

$$\neg\Box(\cap_{i\in I} H_i) \Rightarrow \exists i \in I \text{ s.t. } \neg\Box H_i.$$

Strong intersection consonance is equivalent to

$$\forall i \in I, \Box H_i \Rightarrow \Box(\cap_{i\in I} H_i),$$

that is, the $\Box$-modality is closed under the intersection.

Figure 4 illustrates Definition 2.3 and 2.4. For union consonance, consider a triangle in which each vertex represents the possibility of a single hypothesis. If one obtains the centre of the triangle and the union of all hypotheses is possible, then one also obtains at least one of the midpoints of the edge, representing the possibilities of pairwise unions, and at least one of the vertices connected to this edge. Figure 4 also illustrates the symmetry between strong union consonance and strong intersection consonance. Indeed, if invertibility holds, then these types of consonance are equivalent.

2.4 The logic of credal modalities in logically consistent hypothesis tests

A hypothesis test is *logically consistent* if it satisfies invertibility, monotonicity, union and intersection consonance, and $\Box\Theta$ is obtained. Logically consistent tests have properties that improve their computation and interpretation. For example, it follows from invertibility and Table 1 that, for every hypothesis H, $\Box H$ and $\Box\widetilde{H}$ cannot be simultaneously obtained. That is, a logically consistent hypothesis test admits no contradictions.

Logically consistent tests also allow the use of classical logic to obtain deductions over decided hypotheses. To show this property, we use the functionally complete 'nand' operator ($\uparrow$), either as defined over truth values or as defined over hypotheses (sets). Over hypotheses, $H_1 \uparrow H_2 :=$

$\Theta - (H_1 \cap H_2)$. It follows that, for example, $H_1 \wedge H_2 = H_1 \cap H_2$, $H_1 \vee H_2 = H_1 \cup H_2$ and $\neg H_1 = \Theta - H_1$. Over truth values, $\Box H_1 \uparrow \Box H_2$ is obtained if and only if either $\Box H_1$ is not obtained or $\Box H_2$ is not obtained. Lemma A.1 in the Appendix shows that, if a test is logically consistent and H_1 and H_2 are decided (ΔH_1 and ΔH_2 are obtained), then $\Box(H_1 \uparrow H_2)$ is obtained if and only if $\Box H_1 \uparrow \Box H_2$ is obtained. In other words, if a test is logically consistent and $H_1, \ldots, H_n$ are decided, then, for every logical proposition, P, $\Box P(H_1, \ldots, H_n)$ is obtained if and only if $P(\Box H_1, \ldots, \Box H_n)$ is obtained. Once one has decided the credal modality of $H_1, \ldots, H_n$, computing $P(\Box H_1, \ldots, \Box H_n)$ is usually cheaper than computing $\Box P(H_1, \ldots, H_n)$ by directly applying the statistical test. Also, this ability to treat decided hypothesis as if they were true or false makes the outcomes of the test easier to interpret.

Besides the application of classical logic to decided hypotheses, one can also use it between different credal modalities. For example, the implications in Figures 1 and 4 are transitive. This can be used to obtain new relations by combining the implications in each of the figures. For example, it follows from Figure 2 that $(\neg\Diamond H) \rightarrow (\Box \widetilde{H})$. Also, it follows from Figure 1 that $(\Box \widetilde{H}) \rightarrow (\Diamond \widetilde{H})$. Therefore, one can conclude that $(\neg\Diamond H) \rightarrow (\Box \widetilde{H})$. Similarly, it follows from Figure 2 that $(\Box H) \rightarrow \neg(\Diamond \widetilde{H})$. Also, it follows from the contrariety relation in Figure 1 that is obtained from Table 1 that $(\Diamond \widetilde{H}) \rightarrow (\neg\nabla \widetilde{H})$. By combining these results one can conclude that, if $\Box H$ is obtained, then $\nabla \widetilde{H}$ is not obtained.

Despite the advantages of working with logically consistent hypothesis tests, several standard tests available in the statistical literature are logically inconsistent. Indeed, standard hypothesis tests can fail basic logical properties. The following section investigates examples of logical inconsistency in frequentist and Bayesian statistics.

3 Logical inconsistency of standard tests

Hypothesis tests are often used to determine beliefs. Using Table 1, one can translate each possible result of an hypothesis test to a credal modality. In case a hypothesis test does not satisfy all the properties in Section 2, then we say the resulting logical system is incoherent. Even though it is challenging to communicate and interpret these systems, they are the ones that are generally obtained from standard hypothesis tests. The next subsection revisits some known examples of logical inconsistencies in classical tests.

3.1 Classical tests

- **Failure of invertibility**: Tests based on p-values treat the null and the alternative hypotheses differently. Indeed, the p-value is calculated using solely probabilities obtained under the null hypothesis; probabilities obtained under alternative hypotheses are left out. As a result, a test based on a p-value generally will not be invertible.
- **Failure of monotonicity**: Under regularity conditions, one can assume that if two individuals have the same genotype, then they will also have the same corresponding phenotypical characteristics. However, Izbicki et al. [21] shows that the likelihood ratio test can reject the hypothesis that two individuals have the same phenotype and not reject the hypothesis that they have the same genotype.
- **Failure of consonance**: Consider the standard framework of analysis of variance with three groups. Let α_1, α_2 and α_3 be the unknown means of each group. Izbicki and Esteves [20] shows that, using the generalized likelihood ratio test, one can conclude that '$\alpha_1 = \alpha_2 = \alpha_3 = 0$' is impossible and, at the same time, conclude that '$\alpha_i = 0$' is possible, for every $i \in \{1, 2, 3\}$.

TABLE 2. Example of a loss function for a Bayesian hypothesis test

Decision \ Truth	H	$\neg H$
$\boxplus H$	0	1
$\neg$ ⟐H	a	0

As an alternative to the above classical procedures, one might consider a Bayesian hypothesis test. One such hypothesis test obtains the credal modality for a hypothesis from its posterior probability. Specifically, the credal modality is chosen according to cutoff levels of the posterior probability. For simplicity, these tests are called 'posterior probability cutoff tests'. Since posterior probability tests are coherent from the perspective of Statistical Decision Theory, one might expect that they are also logically coherent. This is not the case, as shown in the following section.

3.2 Posterior probability cutoff tests

Posterior probabilities can be used to define *probabilistic modal operators* ($\boxplus$, ⟐, ⍒ and ⍋), as explained in the following paragraphs. We use these special symbols to distinguish the probabilistic operators from the standard (alethic) operators that were used so far: $\square$, $\diamond$, ∇ and Δ.

A Bayesian statistical model considers two types of random variables: a parameter θ in the parameter space Θ, and data, X, in the sample space, χ. While parameters correspond to unknown quantities, the data correspond to observable or observed quantities.

The uncertainty about the model's variables is described through probability statements. Before observing the data, the joint distribution for θ and X is denoted by $p(\theta, x)$. By integrating x out of the joint distribution, one obtains

$$p(\theta) = \int_{\chi} p(\theta, x) dx.$$

The distribution $p(\theta)$ is called the *prior* distribution for θ, since it represents the uncertainty about θ before observing X.

After the value x of X is observed, the uncertainty about θ is updated based on this observation. One's uncertainty about θ changes from $p(\theta)$ to $p(\theta|x)$, the probability of θ given x. The latter term is called the *posterior* distribution for θ and can be obtained using Bayes Theorem:

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{\int_{\Theta} p(\theta, x) d\theta}. \tag{1}$$

To attribute a credal modality to a hypothesis under scrutiny, one can use Bayesian decision theory [10]. In this context, one chooses a credal modality by minimizing the expected penalty that derives from a given *loss function*. Table 2 presents a loss function that is typically used in Bayesian hypothesis tests. If H is true and one decides that H is necessary, then the loss is 0. Similarly, if H is false and one decides that H is impossible, then the loss is 0. However, if H is true and one decides that H is impossible, then the loss is a, where $a > 0$. This type of decision is called a type-1 error. Similarly, if H is false and one decides that H is necessary, then the loss is 1. This type of decision is called a type-2 error. The constant a defines how much a type-1 error is worse than a type-2 error. When $a = 1$, both types of error are equally undesirable.

TABLE 3. Example of a loss function for an agnostic Bayesian hypothesis test.

Decision \ Truth	H	$\neg H$
$\boxplus H$	0	1
$\triangledown H$	b	b
$\neg \oplus H$	a	0

The optimal decision for the loss function in Table 2 is to choose $\boxplus H$, when $p(H|x) > (1+a)^{-1}$, and $\neg \oplus H$, when $p(H|x) < (1+a)^{-1}$. In other words, the credal modality for each hypothesis is decided based on a cutoff of the posterior probability. If the posterior probability of a hypothesis is sufficiently high, then one concludes that it is necessary. Otherwise, one concludes that the hypothesis is impossible.

The loss function in Table 2 can be extended to agnostic hypothesis tests, as presented in Table 3. In agnostic tests, one can choose to let the hypothesis be undecided, i.e., one can choose $\triangledown H$. Whenever this option is chosen, one commits a type-3 error and incurs in a loss of b, where $0 < b < \min(a, 1)$. This range of values for b implies that, by remaining agnostic, one loses less than by taking a wrong decision and more than by taking a correct decision.

The optimal decision for this loss function is similar to the one that was discussed for standard tests. By taking $c_1 = \max((1+a)^{-1}, 1-b)$ and $c_2 = \min((1+a)^{-1}, \frac{b}{a})$, one obtains that the optimal test is:

$$\text{Choose credal modality} \begin{cases} \boxplus H & \text{, if } p(H|x) > c_1 \\ \neg \oplus H & \text{, if } p(H|x) < c_2 \\ \triangledown H & \text{, otherwise.} \end{cases} \tag{2}$$

In words, the probabilistic modal operator is obtained by comparing the posterior probability of the hypothesis, H, to the cutoffs, c_1 and c_2. If $p(H|x)$ is larger than the upper cutoff, c_1, then the optimal decision is that H is necessary. If $p(H|x)$ is smaller than the lower cutoff, c_2, then the optimal decision is that H is impossible. Finally, for intermediate values of $p(H|x)$, H remains undecided. This test is illustrated in Figure 5.

The posterior probability cutoff tests are logically incoherent. Although the values of a and b can be chosen so that the tests are invertible and monotonic, these tests are not consonant. The main argument for the lack of consonance can be presented in two steps. First, the hypothesis $H_0 : \theta \in \Theta$ has probability 1. Second, H_0 can generally be partitioned into a large collection of hypotheses, $H_1, \ldots, H_n$, such that each $P(H_i|x)$ is arbitrarily small. As a result, the cutoff tests obtain that H_0 is necessary and $H_1, \ldots, H_n$ are all impossible, i.e., they lack consonance with respect to the union operation. A similar example shows that these tests also are not consonant with respect to intersection.

The results of this section show that several standard hypothesis tests cannot be made logically consistent. In the following section, we discuss how to obtain logically consistent hypothesis tests.

4 Logically consistent agnostic hypothesis tests

A logically consistent agnostic hypothesis test satisfies all the properties described in Section 2: invertibility, monotonicity, union and intersection consonance, and $\Box\Theta$ is obtained. Esteves et al. [16] shows that every logically consistent agnostic testing scheme can be obtained from a *region*

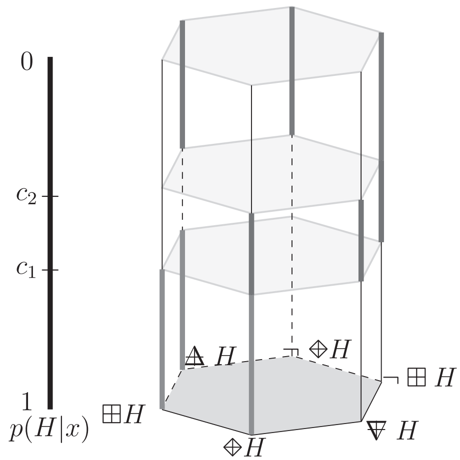

FIG. 5. Typical hexagon of statistical modalities for the posterior probability cutoff test, given by Table 3. The coloured edges indicate the chosen modalities as a function of the posterior probability of H, $p(H|x)$.

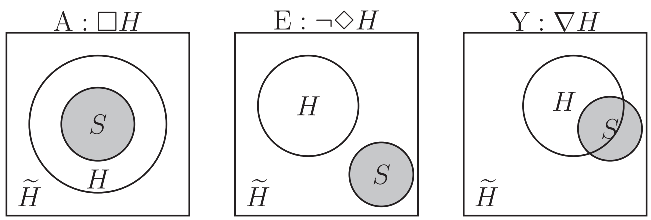

FIG. 6. A test for hypothesis H based on the region estimator for θ, S.

estimator (a set of plausible values for θ). More precisely, for every logically consistent test, there exists a region estimator, $S \subseteq \Theta$, such that the test is of the form

$$\text{Choose credal modality} \begin{cases} \Box H & \text{if } S \subseteq H \\ \neg\Diamond H & \text{if } S \subseteq \widetilde{H} \\ \nabla H & \text{if } S \cap H \neq \varnothing \text{ and } S \cap \widetilde{H} \neq \varnothing. \end{cases}$$

The tests that are of the above type are called *region-based tests*. Figure 6 illustrates such tests. No matter what region estimator, S, is used, the test based on S is logically consistent [16]. For the test to have additional statistical properties, S might be built using the Bayesian or the frequentist frameworks. In the following, we describe a particular type of region-based test: the *Generalized Full Bayesian Significance Test* (GFBST). This test is an extension of the Full Bayesian Significance Test [23] (FBST) to agnostic hypothesis tests.

4.1 FBST and GFBST

Credal modalities that are obtained from a possibilistic measure [13, 14] are defined as *possibilistic* (or alethic) modal operators ($\Box$, $\Diamond$, ∇ and Δ). Note that these symbols are different from the ones used for the probability modal operators. Hereafter, we study the possibilistic modal operators that are obtained from the FBST and the GFBST.

The FBST is based on the *epistemic value* of the hypothesis of interest given the observed data, $\text{ev}(H|x)$. The epistemic value is a transformation of a probability into a possibility measure [5, 30]. Specifically, $\text{ev}(H|x)$ is obtained through the steps:

(1) Define a surprise function, $s(\theta|x) = \frac{p(\theta|x)}{r(\theta)}$, where $r(\theta)$ is a *reference* density over Θ [27]. The *reference density* can be interpreted as an invariant measure under a relevant transformation group. For example, the reference density may be obtained from the information geometry according to a metric on the parameter space [28].
(2) Define the tangent set to H, $T(H)$, as $T(H)$, where,

$$T(H) = \{\theta_1 \in \Theta \,|\, \forall \theta_0 \in H, s(\theta_1|x) > s(\theta_0|x)\}.$$

Note that $T(H) \cap H = \emptyset$, that is, the tangent set to an hypothesis and the hypothesis are disjoint. This aspect of the e-value's definition is motivated by the legal principle known as *in dubio pro reo* (most favorable interpretation, benefit of the doubt, or presumption of innocence) [25]. Intuitively, only parameter points that are more probable than every point of a given hypothesis can be admitted as witnesses against this hypothesis.
(3) Obtain $\text{ev}(H|x) = 1 - p(T(H)|x)$.
Alernatively, Lemma A.2 in the Appendix shows how $\text{ev}(H|x)$ can be obtained as a function of $\text{ev}(\{\theta_0\}|x)$, for each $\theta_0 \in H$.

The FBST chooses $\Box H$ when $\text{ev}(H|x) > c$, and $\neg\Diamond H$ when $\text{ev}(H|x) \leq c$, where $c \in (0,1)$ is a cutoff defined by the practitioner. This procedure is similar to region-based tests. Indeed, if $S = \{\theta_0 \in \Theta : \text{ev}(\{\theta_0\}|x) > c\}$, then Theorem A.3 shows that the FBST has the form

$$\text{Choose credal modality} \begin{cases} \neg\Diamond H & \text{, if } S \subseteq \widetilde{H} \\ \Box H & \text{, otherwise.} \end{cases}$$

The GFBST extends the FBST into a region-based agnostic hypothesis test. Let S be such as in the previous paragraph. The GFBST has the form:

$$\text{Choose credal modality} \begin{cases} \neg\Diamond H & \text{, if } S \subseteq \widetilde{H} \ (S \cap H = \emptyset) \\ \Box H & \text{, if } S \subseteq H \ (S \cap \widetilde{H} = \emptyset) \\ \nabla H & \text{, if } S \cap H \neq \varnothing \text{ and } S \cap \widetilde{H} \neq \varnothing. \end{cases}$$

While in the FBST, $S \cap H \neq \emptyset$ implies $\Box H$, in the GFBST this case can lead to either $\Box H$ or ∇H. Since this difference makes the GFBST a region-based test, it is logically consistent. By applying Theorem A.3, the GFBST can also be written as a function of $\text{ev}(H|x)$ and $\text{ev}(\widetilde{H}|x)$:

$$\text{Choose credal modality} \begin{cases} \neg\Diamond H & \text{, if } \text{ev}(H|x) \leq c \\ \Box H & \text{, if } \text{ev}(\widetilde{H}|x) \leq c \\ \nabla H & \text{, otherwise.} \end{cases} \tag{3}$$

Remark: In previous references, the tangent set to H was defined as $T^*(H) = \{\theta_1 \in \Theta : s(\theta_1|x) > \sup_{\theta_0 \in H} s(\theta_0|x)\}$. Under regularity conditions that are usually found in statistical models [20, ex.3.23], $p(T^*(H)|x) = p(T(H)|x)$. Hence, in these cases, both definitions lead to the same results. However, only $T(H)$ allows the assumption-free characterization in Equation (3).

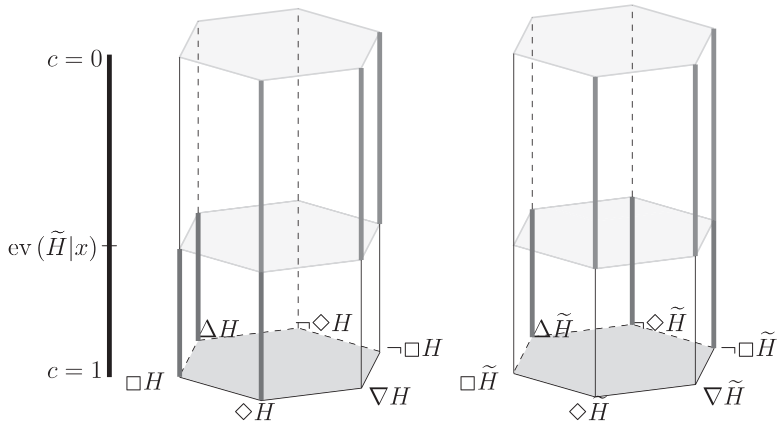

FIG. 7. Example of the behaviour of the GFBST for a fixed sample x as a function of the cutoff, c, when the supremum of the surprise function is obtained only in H. The prisms on the left and right side represent the results obtained by the GFBST when applied, respectively, to H and $\widetilde{H}$. The hexagon obtained from each horizontal cut represents the GFBST test that uses the corresponding cutoff. The coloured edges represent the credal modalities that are obtained from the GFBST as a function of the cutoff value.

The GFBST can be made more interpretable using some properties of ev. Note that, for every H, either $T(H)=\emptyset$ or $T(\widetilde{H})=\emptyset$. Therefore, either $\text{ev}(H|x)=1$ or $\text{ev}(\widetilde{H}|x)=1$. When $\text{ev}(H|x)=1$ and $\text{ev}(\widetilde{H}|x)=1$, no matter what the value of c is, ∇H and $\nabla \widetilde{H}$ are obtained. When $\text{ev}(H|x)=1$ and $\text{ev}(\widetilde{H}|x)<1$, either $\square H$ and $\neg\diamond \widetilde{H}$ are obtained, or ∇H and $\nabla \widetilde{H}$ are obtained. Similarly, if $\text{ev}(\widetilde{H}|x)=1$ and $\text{ev}(H|x)<1$, then either $\square \widetilde{H}$ and $\neg\diamond H$ are obtained or ∇H and $\nabla \widetilde{H}$ are obtained. Section 7 illustrates this behaviour when $\text{ev}(H|x)=1$ and $\text{ev}(\widetilde{H}|x)<1$. For every $c\in(0,1)$, $\neg\diamond H$ is not obtained. This behaviour is different from the one in Figure 5.

5 Hybrid hexagons: relations between probabilistic and possibilistic modal operators

Although the posterior probability test and the GFBST have different logical properties, It follows from the characterization of the GFBST in Section 4, that $\square H$ can be made a more stringent modality than $\diamondplus H$. Indeed, if $c<0.5$ in Equation (3), then Theorem A.4 in the Appendix proves that

$$\square H \Rightarrow p(H|x)\geq 1-c \Rightarrow \diamond H. \tag{4}$$

Therefore, in Equation (2), the cutoffs $c_1=1-c$ and $c_2=c$ yield

$$\square H \Rightarrow \boxplus H \Rightarrow \diamondplus H \Rightarrow \diamond H.$$

These relationships are summarized in the nested hexagon of Figure 8. The above implications show that it is possible to combine the two types of credal modality into a single hexagon of oppositions. Since the hexagon of oppositions has two degrees of freedom and the implication $\square H \Rightarrow \diamondplus H$ holds, it is possible to use $\square H$ as the necessity modality and $\diamondplus H$ as the possibility modality in a new hexagon of oppositions. Similarly, since $\boxplus H \Rightarrow \diamond H$ holds, one can use $\boxplus H$ as the necessity modality and

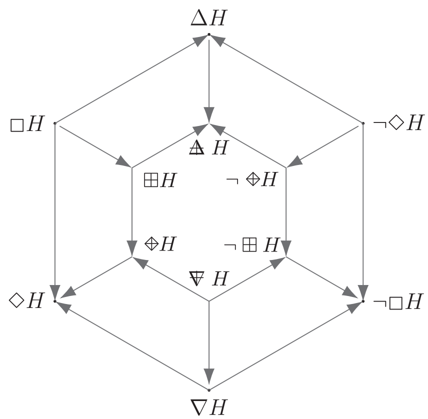

FIG. 8. Nested hexagon of opposition that shows the implication relations between the probabilistic and possibilistic modal operators.

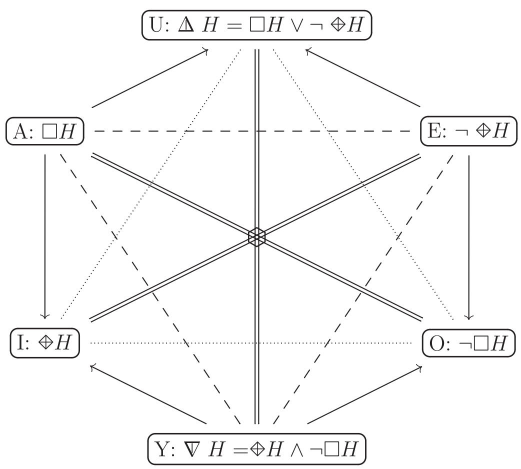

FIG. 9. Hybrid hexagon of oppositions that is obtained from combining probabilistic and possibilistic modalities.

$\Diamond H$ as the possibility modality in another hexagon of oppositions. Similar hybrid hexagons have appeared in Carnielli and Pizzi [8], Demey and Smessaert [11].

Figure 9 illustrates the hybrid hexagon obtained using $\Box H$ and $\oplus H$. This hexagon summarizes several relations between $\Box H$ and $\oplus H$ that can be proved using the fact that $\Box H \Rightarrow \oplus H$. In particular, note that, if $p(H)=0$, then it follows from Equation (1) that $p(H|x)=0$. Therefore, it follows from Equation (2) that $\neg \oplus H$ holds. Using the contrariety relation in Figure 9, one can conclude that $\Box H$ does not hold. In summary, if a hypothesis, H, is such that $p(H)=0$, then H is not accepted by the GFBST.

The previous conclusion is compatible with a common idea in Statistics: if $p(H)=0$, then one cannot accept H; one can only fail to reject H [18, 19]. Indeed, if $p(H)=0$, then $\neg\Box H$ is obtained, i.e., the GFBST cannot accept H. However, even though $p(H)=0$, it is possible to obtain $\Diamond H$, i.e., the GFBST can fail to reject H. However, when $p(H)=0$, then $p(H|x)=0$ and $\neg \diamondplus H$ is obtained, that is, the posterior probability test always rejects H.

The above result might be surprising, given the importance of null probability hypotheses in the theoretical and empirical sciences [29]. For example, if $\Theta = \mathbb{R}^d$, then it is common to find hypotheses in the form of linear equations, $H: A\theta = b$, where matrix A is $h \times d$. This case includes, for example, the simple null hypothesis $H: \theta = c$. In case the distribution of θ is continuous, then $p(H)=0$ and, for every x, $p(H|x)=0$. Therefore, while the probability cutoff test always obtains $\neg \diamondplus H$ for precise hypotheses, the GFBST can either obtain ∇H or $\neg\Diamond H$. That is, while the credal modality obtained from the probability cutoff test for precise hypotheses is known beforehand, the GFBST allows one to use data to revisit their beliefs regarding these hypotheses.

6 Conclusions and final remarks

We show how the hexagon of oppositions can be a useful tool to explain consistency properties of statistical hypothesis tests. Indeed, geometric solids composed of hexagons of oppositions illustrate the conditions for a statistical hypothesis test to be logically consistent. Also, prisms composed of hexagons of oppositions explain the definitions of hypothesis tests such as the probability cutoff test and the GFBST. A hybrid hexagon of oppositions summarizes the logical relations between these tests. In summary, the hexagon of opposition can be used as a powerful form of diagrammatic representation that is helpful in organizing, displaying and explaining complex logical properties of multiple statistical tests of hypothesis and their intricate inter-relationships.

Acknowledgements

The authors are grateful for the support of IME-USP, the Institute of Mathematics and Statistics of the University of São Paulo, and the Department of Statistics of UFSCar — The Federal University of São Carlos. J. M.S. is grateful to FAPESP — the State of São Paulo Research Foundation (grants CEPID 2013/07375-0 and CEPID 2014/50279-4); and CNPq — the Brazilian National Counsel of Technological and Scientific Development (grants PQ 301206/2011-2 and PQ 301892/2015-6). R. I. is grateful for the support of FAPESP (grants 2014/25302-2 and 2017/03363-8). Finally, the authors are grateful for advice and comments received from anonymous referees, and from participants of the 5th World Congress on the Square of Opposition, held on November 11–15, 2016, at Hanga Roa, Rapa Nui, Chile, having as main organizer Jean-Yves Béziau; and the XVIII Brazilian Logic Conference, held on May 8–12, 2017, at Pirenópolis, Goiás, Brazil, co-chaired by Marcelo Coniglio, Itala D'Ottaviano and Wagner Sanz.

References

[1] J.-Y. Béziau. The power of the hexagon. *Logica Universalis*, **6**, 1–43, 2012.

[2] J.-Y. Béziau. Opposition and order. In *New Dimensions of the Square of Opposition*, J. Y. Bziau, J. and Gan-Krzywoszynska, K., eds, pp. 1–11. Philosophia Verlag, 2015.

[3] R. Blanché. Sur l'opposition des concepts. *Theoria*, 19:89–130, 1953.

[4] R. Blanché and G. Davy. *Structures intellectuelles: essai sur l'organisation systématique des concepts*. A la recherche de la vérité. J. Vrin, 1981.

[5] W. Borges and J. M. Stern. The rules of logic composition for the bayesian epistemic e-values. *Logic Journal of IGPL*, **15**, 401–420, 2007.

[6] W. Carnielli and M. E. Coniglio. *Paraconsistent Logic: Consistency, Contradiction and Negation*, Vol. 40 of *Logic, Epistemology, and the Unity of Science*. Springer, 2016.

[7] W. Carnielli, M. E. Coniglio and J. Marcos. Logics of formal inconsistency. In *Handbook of Philosophical Logic*, pp. 1–93. Springer, 2007.

[8] W. Carnielli and C. Pizzi. *Modalities and Multimodalities*, Vol. 12 of *Logic, Epistemology, and the Unity of Science*. Springer, 2008.

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

[10] M. H. DeGroot. *Optimal Statistical Decisions*, Vol. 82. John Wiley & Sons, 2005.

[11] L. Demey and H. Smessaert. The relationship between aristotelian and hasse diagrams. In *International Conference on Theory and Application of Diagrams*, pp. 213–227. Springer, 2014.

[12] L. Demey and H. Smessaert. Metalogical decorations of logical diagrams. *Logica Universalis*, **10**, 233–292, 2016.

[13] D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In *Fuzzy Information and Decision Processes*, M. Gupta and E. Sanchez, eds, pp. 167–181. Elsevier/North-Holland, 1982.

[14] D. Dubois and H. Prade. From Blanché's hexagonal organization of concepts to formal concept analysis and possibility theory. *Logica Universalis*, **6**, 149–169, 2012.

[15] A. Dumitriu. *History of Logic*. Abacus Press, 1977.

[16] L. G. Esteves, R. Izbicki, J. M. Stern and R. B. Stern. The logical consistency of simultaneous agnostic hypothesis tests. *Entropy*, **18**, 256, 2016.

[17] V. Fossaluza, R. Izbicki, G. M. da Silva and L. G. Esteves. Coherent hypothesis testing. *The American Statistician*, (just-accepted), 2017.

[18] G. Gigerenzer. Mindless statistics. *The Journal of Socio-Economics*, **33**, 587–606, 2004.

[19] H. F. Inman. Karl pearson and r.a. fisher on statistical tests: a 1935 exchange from nature. *The American Statistician*, **48**, 2–11, 1994.

[20] R. Izbicki and L. G. Esteves. Logical consistency in simultaneous statistical test procedures. *Logic Journal of IGPL*, **23**, 732–758, 2015.

[21] R. Izbicki, V. Fossaluza, A. G. Hounie, E. Y. Nakano and C. A. B. Pereira. Testing allele homogeneity: the problem of nested hypotheses. *BMC genetics*, **13**, 103, 2012.

[22] M. Lavine and M. Schervish. Bayes factors: what they are and what they are not. *The American Statistician*, **53**, 119–122, 1999.

[23] C. A. d. B. Pereira and J. M. Stern. Evidence and credibility: full bayesian significance test for precise hypotheses. *Entropy*, **1**, 99–110, 1999.

[24] M. J. Schervish. P values: what they are and what they are not. *The American Statistician*, **50**, 203–206, 1996.

[25] J. M. Stern. Signicance tests, belief calculi, and burden of proof in legal and scientic discourse. In *Laptec-2003*, pp. 139–147. Frontiers in Artificial Intelligence and its Applications, 101, IOS Press, 2003.

[26] J. M. Stern. Paraconsistent sensitivity analysis for bayesian significance tests. In *Brazilian Symposium on Artificial Intelligence*, pp. 134–143. Springer, 2004.

[27] J. M. Stern. Constructive verification, empirical induction, and falibilist deduction: a threefold contrast. *Information*, **2**, 635–650, 2011a.

[28] J. M. Stern. Symmetry, invariance and ontology in physics and statistics. *Symmetry*, **3**, 611–635, 2011b.

[29] J. M. Stern. Continuous versions of haack's puzzles: equilibria, eigen-states and ontologies. *CLE e-Prints*, **15**, 1–25, 2015.
[30] J. M. Stern and C. A. D. B. Pereira. Bayesian epistemic values: focus on surprise, measure probability! *Logic Journal of IGPL*, **22**, 236–254, 2014.

Received 28 January 2017

Appendix

Lemma A.1
If ΔH_1 and ΔH_2 are obtained in a logically consistent hypothesis test, then $\Box(H_1 \uparrow H_2)$ is obtained if and only if $\Box H_1 \uparrow \Box H_2$ is obtained.

Proof. If $\Box(H_1 \uparrow H_2)$ is obtained, then $\Box(\Theta - (H_1 \cap H_2))$ is obtained. It follows from invertibility that $\neg\Diamond(H_1 \cap H_2)$ is obtained. Therefore, it follows from an implication relation in Figure 1 that $\neg\Box(H_1 \cap H_2)$ is obtained. Hence, using intersection consonance, either $\neg\Box H_1$ or $\neg\Box H_2$ is obtained. Conclude from a contradiction relation in Figure 1 that either $\Box H_1$ is not obtained or $\Box H_2$ is not obtained, i.e., $\Box H_1 \uparrow \Box H_2$ is obtained.

If $\Box H_1 \uparrow \Box H_2$ is obtained, then either $\Box H_1$ is not obtained or $\Box H_2$ is not obtained. Hence, using a contradiction relation in Figure 1, either $\neg\Box H_1$ or $\neg\Box H_2$ is obtained. Since by hypothesis ΔH_1 and ΔH_2 are obtained, it follows from Table 1 that either $\neg\Diamond H_1$ or $\neg\Diamond H_2$ is obtained. Therefore, using monotonicity, $\neg\Diamond(H_1 \cap H_2)$ is obtained. It follows from invertibility that $\Box(\Theta - (H_1 \cap H_2))$ is obtained, that is, $\Box(H_1 \uparrow H_2)$ is obtained. ∎

Lemma A.2
$\text{ev}(H|x) = \sup_{\theta_0 \in H} \text{ev}(\{\theta_0\}|x)$

Proof.

$$\begin{aligned}
\text{ev}(H|x) &= 1 - p(\{\theta_1 \in \Theta | \forall \theta_0 \in H, s(\theta_1|x) > s(\theta_0|x)\}\ |x) \\
&= 1 - p(\cap_{\theta_0 \in H}\{\theta_1 \in \Theta | s(\theta_1|x) > s(\theta_0|x)\}\ |x) \\
&= 1 - \inf_{\theta_0 \in H} p(\{\theta_1 \in \Theta | s(\theta_1|x) > s(\theta_0|x)\}\ |x) \\
&= 1 - \inf_{\theta_0 \in H} p(T(\{\theta_0\})|x) = \sup_{\theta_0 \in H} ev(\{\theta_0\}|x)
\end{aligned}$$

∎

Theorem A.3
Let $S = \{\theta_0 \in \Theta : \text{ev}(\{\theta_0\}|x) > c\}$. For every $H \subseteq \Theta$, $\text{ev}(H|x) \leq c$ if and only if $H \cap S = \emptyset$.

Proof. Note that $\text{ev}(H|x) \leq c$ if and only if $\sup_{\theta_0 \in H} \text{ev}(\{\theta_0\}|x) \leq c$ (Lemma A.2). That is, $\text{ev}(H|x) \leq c$ if and only if $S \cap H = \emptyset$. ∎

Theorem A.4
If $c < 0.5$ in Equation (3), then

$$\Box H \Rightarrow p(H|x) \geq 1 - c \Rightarrow p(H|x) > c \Rightarrow \Diamond H.$$

PROOF. For every $\Theta_H \subset \Theta$,

$$\begin{aligned} 1-\text{ev}(\widetilde{H}\,|x) &= p(\theta \in T(\widetilde{H})|x) \\ &= p(\theta \in \{\theta_1 \in \Theta\,|\,\forall \theta_0 \in \widetilde{H}, s(\theta_1|x) > s(\theta_0|x)\}\ |x) \\ &\leq p(\theta \in \Theta_H|x) = p(H|x), \end{aligned} \tag{5}$$

where the last inequality follows from the fact that

$$\{\theta_1 \in \Theta\,|\,\forall \theta_0 \in \widetilde{H}, s(\theta_1|x) > s(\theta_0|x)\} \subseteq \Theta - \Theta_{\widetilde{H}} = \Theta_H.$$

Similarly, one can obtain

$$1-\text{ev}(H|x) \leq p(\widetilde{H}\,|x) \tag{6}$$

Assume that $\Box H$ is obtained. It follows from Equation (3) that $\text{ev}(\widetilde{H}|x) \leq c$, that is, $1-\text{ev}(\widetilde{H}|x) \geq 1-c$. Conclude from Equation (5) that $p(H|x) \geq 1-c$. Then, since $c < 0.5$, it holds that $p(H|x) > c$. Therefore, $p(\widetilde{H}|x) < 1-c$. Conclude from Equation (6) that $1-\text{ev}(H|x) < 1-c$, that is, $\text{ev}(H|x) > c$. It follows from Equation (3) that $\Diamond H$ is obtained. ∎