Significance Levels as Ambiguity Aversion in Optimal Testing

Simone Manganelli*

March 2026

Abstract

Wald's statistical decision framework is defined by three primitives: a statistical model, an action space, and a loss function. The Bayesian and frequentist paradigms agree on the loss function, but diverge on the object of judgment: parameters versus actions. By decomposing the level of significance into *ex ante* and *ex post* components, this paper provides a unified representation of these traditions: within the class of interpolating rules, any Bayesian decision is nested within a frequentist double conditioning representation. Bayesian decisions are the boundary case with zero structural inertia and *ex ante* significance level equal to 1. The *ex ante* significance level is uniquely identified from observed behavior as the *ex ante* rejection probability, providing a measurable index of uncertainty aversion across a continuum from uncertainty neutrality to complete inertia. This threshold structure arises independently from Neyman–Pearson optimality and Gilboa–Schmeidler minmax expected utility, and admits direct elicitation through controlled experiments.

Keywords: Statistical Decision Theory; Hypothesis Testing; Uncertainty Aversion.

JEL Codes: C12; C13; D81.

*European Central Bank (ECB), simone.manganelli@ecb.int.

1 Introduction

Wald (1950) provided a rigorous foundation for statistical decision theory, based on three primitives: a statistical model, an action space, and a loss function. Yet, his framework lacks a formal definition of *judgment*, leaving unaddressed how to reconcile the objective evidence contained in a data sample with the private information of the decision-maker. While both the Bayesian and frequentist traditions share the loss function as their evaluation criterion, they differ in the primitive to which judgment is applied: the Bayesian approach applies judgment to the parameters through a prior distribution; the frequentist approach applies it to the actions through an anchor and an associated significance level.

The central contribution of this paper is the decomposition of the significance level into *ex ante* and *ex post* components, from which three results follow. First, any Bayesian decision in the interpolating class is representable as a frequentist double conditioning mechanism. In this representation, the Bayesian update corresponds to the boundary case of zero structural inertia, where the *ex ante* significance level equals 1. Second, the *ex ante* significance level serves as a universal index of uncertainty aversion. It maps the Bayesian, frequentist, and robust Bayesian paradigms onto a single measurable continuum, rendering them formally comparable. Third, this *ex ante* significance level is theory-robust: the threshold inertia it governs emerges from two independent traditions. In the Neyman–Pearson framework, it arises as a property of optimal hypothesis testing under the monotone likelihood ratio condition. In the Gilboa–Schmeidler minmax expected utility framework, the same structure is generated by preference axioms over acts, a result established formally in a Gaussian-quadratic environment. Because these distinct justifications converge on a single behavioral primitive, the framework establishes the formal basis for direct elicitation of the *ex ante* significance level through controlled experiments.

The link between Bayesian updating, shrinkage-based decision rules, and data-dependent frequentist procedures is made precise through a double conditioning mechanism. First, the procedure for evaluating the optimality of the judgmental decision is itself conditional on the

sample realization. Second, the level of significance is revised upon observing the data. The double conditioning reveals that the Bayesian posterior decision, by setting the expectation of the gradient to zero at a point that necessarily differs from the prior-based decision, can be formally represented as a frequentist test that rejects the optimality of the anchor for every realization of the data. In the language of hypothesis testing, the Bayesian decision-maker corresponds to the boundary case in which the decision-maker has no control on Type I errors. This characterization identifies Bayesian updating as a limiting case within the broader class of significance-based decision rules, in which the judgmental anchor is never retained.

This representation is one-directional. Every interpolating Bayesian decision maps to a decision with judgment, but the converse fails: rules that retain the judgmental anchor with positive probability cannot be rationalized by a single proper prior. This asymmetry establishes the decision with judgment as a strictly more flexible framework, capable of nesting behaviors, such as structural inertia, that Bayesian models axiomatically exclude. The *ex ante* significance vector emerges as the fundamental primitive indexing a decision-maker's attitude toward uncertainty, from full inertia to uncertainty neutrality.

The paper's most general result is a representation theorem establishing that any interpolating decision rule satisfying a pre-committed statistical threshold (the frequentist definition of significance) is a decision with judgment. This theorem places subjective belief-revision (Savage, 1954), objective error-control (Neyman and Pearson, 1933), and axiomatic ambiguity aversion (Gilboa and Schmeidler, 1989) along a single continuum indexed by the *ex ante* significance level $\bar{\alpha}$. At one boundary ($\bar{\alpha} = 1$) lies the uncertainty-neutral Bayesian decision-maker, who imposes no structural inertia and abandons the anchor for any empirical evidence. At the other ($\bar{\alpha} = 0$) lies the full-inertia decision-maker, who retains the anchor regardless of the data. This framework shows that $\bar{\alpha}$ is uniquely identified from observed behavior as the *ex ante* rejection probability, and that it maps monotonically to the size of the inertia zone.

This threshold structure is a property that emerges independently from two distinct

traditions. In the Neyman-Pearson framework, the threshold rule is the uniformly most powerful (UMP) accept-or-reject decision for any rule pre-committing to a Type I error level. In the Gilboa-Schmeidler minmax utility framework, a similar threshold behavior arises from preference axioms over acts. These two derivations are conceptually independent, yet both converge on the same behavioral primitive $\bar{\alpha}$: a pre-committed probability of departing from the judgmental anchor that is invariant to the underlying theoretical motivation.

Finally, $\bar{\alpha}$ admits direct elicitation through a controlled experiment in the spirit of Ellsberg (1961). The *ex ante* significance level is the maximum probability a decision-maker tolerates of performing worse than their anchor when that anchor is optimal. The elicitation experiment asks the decision-maker to choose $\bar{\alpha}$ to cap the number of "loss-inducing" balls in one urn (representing the null), which, by the power property of the optimal test, simultaneously determines the minimum number of "gain-inducing" balls in a second urn (representing the alternative). This choice problem identifies the degree of uncertainty aversion as a behavioral primitive, providing empirical discipline for the model without requiring the full apparatus of prior elicitation.

This framework relates to several classic strands of literature. The landmark complete class theorems of Wald (1950) and Brown (1981) establish that admissible procedures are Bayes or limits of Bayes, providing the foundational boundaries for optimal decision-making. Building on this tradition, Manganelli (2025) demonstrates that gradient-based tests form an essentially complete class, thereby establishing their admissibility. This paper provides a constructive mechanism that connects these two results: the double conditioning gradient-based representation allows for a formal frequentist operationalization of Bayesian and shrinkage decisions (James and Stein 1961 and Efron and Morris 1973). Our results are closely related to those of Bewley (2011), whose inertia zones generated by *status quo* bias are here mapped to identifiable significance thresholds. At the most basic level, this paper identifies the source of the Bayesian-frequentist tension within Wald's own framework: the two traditions impose judgment on different primitives of the same decision problem, and the

ex ante significance level is the parameter that measures how far a decision-maker sits from the Bayesian boundary. By providing a common unit of measurement for statistical evidence and subjective judgment, the paper contributes to the agenda of placing empirical evidence and decision theory on a unified foundation (Manski, 2021).

Importantly, this representation extends beyond Bayesian environments to the modern literature on ambiguity aversion. The multiple priors model of Gilboa and Schmeidler (1989) generates inertia zones whose behavioral implications are here characterized along a continuum indexed by the *ex ante* significance level $\bar{\alpha}$. In Gaussian-quadratic environments, the inertia zone of a minmax agent coincides exactly with the threshold level set of our significance test. This provides an empirically elicitable threshold for the robust control and business-cycle models (Hansen and Sargent 2024 and Ilut and Schneider 2023), which are hard to operationalize because of the difficulty of identifying the underlying sets of priors.

Throughout the paper, scalar variables and parameters are denoted in regular italic (x, θ, α); vectors and matrices are in boldface ($\mathbf{x}$, $\boldsymbol{\theta}$, $\mathbf{A}$). Column vectors are the default; $\boldsymbol{\lambda}^\top$ denotes the transpose. Estimated quantities are denoted with a hat ($\hat{\theta}$, $\hat{\mathbf{a}}$) and judgmental decisions with a tilde ($\tilde{a}$, $\tilde{\mathbf{a}}$). The symbol $\odot$ denotes the Hadamard (element-wise) product; $\mathbf{1}$ denotes a conformable vector of ones; and Φ denotes the cumulative distribution function of the standard normal distribution.

The paper is structured as follows. Section 2 introduces a stylized Gaussian-quadratic environment to illustrate the double conditioning mechanism and the observational equivalence between Bayesian and frequentist decisions. Section 3 develops the general multivariate theory, establishing the formal representation of Bayesian decisions as significance-based rules within an interpolating class. Section 4 provides the general representation theorem that places all such rules on a continuum indexed by the *ex ante* significance vector, establishes the theory-robustness of this primitive, and proposes a controlled experiment for its elicitation. Section 5 concludes.

2 Representation in a Gaussian-Quadratic Environment

This section sets up a simple statistical decision problem and solves it using Bayesian, frequentist and robust Bayesian decisions.

Decision-makers operate in the following simplified environment:

Assumption 2.1 (**Simplified Decision Environment**). *(i) The random variable* $X \sim \mathcal{N}(\theta, 1)$*, for an unknown* $\theta \in \mathbb{R}$*.*

(ii) A single sample realization x *from* X *is observed.*

(iii) The decision-maker chooses the action $a \in \mathbb{R}$*, incurring the loss* $L(\theta, a) = \frac{1}{2}(a - \theta)^2$*.*

2.1 The Bayesian Decision

The Bayesian approach assumes that the decision-maker uses subjective information in the form of a prior distribution over the unknown parameter θ.

Proposition 2.1 (**Bayesian Decision in the Simplified Environment**). *Suppose Assumption 2.1 holds. A Bayesian decision-maker with prior distribution* $\pi(\theta) = \mathcal{N}(0, 1)$ *over* θ *chooses the action:*

$$\delta^{\pi(\theta|x)}(x) = x/2. \tag{1}$$

Proof. See Appendix. □

2.2 The Decision with Judgment

The decision rule incorporating judgment of Manganelli (2025) assumes the decision-maker utilizes subjective information in the form of a judgmental decision $\tilde{a} \in \mathbb{R}$ and a level of significance $\alpha \in [0, 1]$, collectively denoted as the judgment $A \equiv \{\tilde{a}, \alpha\}$. Under this framework, the decision-maker adopts the benchmark action $\tilde{a}$ as a *status quo* and departs from it only if statistical evidence against its optimality is sufficiently strong.

In the event of rejection, the decision-maker moves from the judgmental decision toward the maximum likelihood estimate, which in this Gaussian-quadratic setting is $\hat{a}(x) = x$. This movement is parameterized using the following class of shrinkage:

$$a(\lambda; x) = \lambda x + (1 - \lambda)\tilde{a}, \quad \lambda \in [0, 1]. \tag{2}$$

Testing the optimality of the judgmental decision is equivalent to testing the null hypothesis $H_0 : \theta = \tilde{a}$, which corresponds to $\lambda = 0$. This procedure selects the shrinkage weight closest to the anchor consistent with the sample evidence at significance level α, formalized as follows:

Proposition 2.2 (**Decision with Judgment in the Simplified Environment**). *Suppose Assumption 2.1 holds. A decision-maker with judgment $A = \{\tilde{a}, \alpha\}$, for $\tilde{a} = 0$ and $\alpha \in [0, 1]$, chooses the action:*

$$\delta^A(x) = \hat{\lambda} x, \tag{3}$$

where $\hat{\lambda} = \max\{0, \lambda^*\}$ with $\lambda^* = 1 + c_{\alpha/2}/|x|$ if $x \neq 0$, and $c_{\alpha/2} = \Phi^{-1}(\alpha/2)$ is the $\alpha/2$-quantile of the standard normal distribution. When $x = 0$, $\delta^A(x) = 0$ for any $\hat{\lambda} \in [0, 1]$.

Proof. See Appendix. □

2.3 Representation of Bayesian Decisions as Decisions with Judgment

Establishing a formal connection between the two frameworks requires extending the judgment $A \equiv \{\tilde{a}, \alpha\}$ to allow for a data-dependent level of significance.

Proposition 2.3 (**p-value**). *Suppose Assumption 2.1 holds, the judgmental decision is $\tilde{a} = 0$ and $x \neq 0$. The* p-value *associated with the observed sample realization x is*

$$\tilde{\alpha}(x) = 2\Phi(-|x|), \tag{4}$$

where Φ *is the cdf of the standard normal distribution.*

Proof. See Appendix. □

The test statistic (24) in the proof of Proposition 2.3 illustrates the first dimension of double conditioning. In the decision with judgment, the construction of the test statistic is inherently tied to the sample realization x because the shrinkage path $a(\lambda; x)$ in equation (2) is itself a function of the data (Manganelli, 2025). To test the optimality of the judgment $\tilde{a}$ without logical circularity, one must distinguish between the x that defines the direction of the action and the randomness of the estimator used to evaluate that action. By introducing the auxiliary random variable $Y \sim \mathcal{N}(\theta, 1)$, we characterize the distribution of the gradient at the specific point x identified by the data, effectively conditioning the test on the observed direction of adjustment.

The formal reconciliation of Bayesian and frequentist methodologies requires a distinction between two types of significance. We define an *ex ante* significance level, $\bar{\alpha} \in [0, 1]$, and an *ex post* significance level, $\alpha(x) \in [0, 1]$. The former represents the decision-maker's prior threshold for abandoning the judgmental decision $\tilde{a}$, while the latter is a data-dependent mapping from the observed *p-value* $\tilde{\alpha}(x)$ to the unit interval. While $\alpha(x)$ is computed only after seeing the data, the functional form of the mapping is a subjective choice made *ex ante*. This mapping represents the second dimension of conditioning: the Bayesian posterior does not merely update beliefs, but implicitly selects a significance level based on the realized strength of evidence.

Proposition 2.4 (**Representation in a Gaussian Quadratic Environment**)**.** *Suppose Assumption 2.1 holds. The Bayesian decision with prior distribution* $\pi(\theta) = \mathcal{N}(0, 1)$ *can be represented as a decision with judgment when the judgmental decision is* $\tilde{a} = 0$*, the* ex ante *level of significance is* $\bar{\alpha}^{\pi(\theta)} = 1$ *and the* ex post *level of significance is:*

$$\alpha^{\pi(\theta)}(x) = 2\Phi(-|x|/2), \tag{5}$$

or, writing $\tilde{\alpha}(x) = 2\Phi(-|x|)$ *for the* p-value*:*

$$\alpha^{\pi(\theta)}(x) = 2\Phi[\Phi^{-1}(\tilde{\alpha}(x)/2)/2]. \tag{6}$$

Proof. See Appendix. □

The *ex ante* significance $\bar{\alpha}^{\pi(\theta)} = 1$ implies that the Bayesian decision affords no protection to the judgmental anchor, as the Type I error of the associated test equals 1, characterizing Bayesian behavior as uncertainty neutral (Section 4).

2.4 Representation of Robust Bayesian Decisions

By treating the significance level as a data-dependent choice, we now analyze the structural differences between standard Bayesian updates, robust Bayesian decisions, and frequentist decisions with judgment. We specifically show that ambiguity aversion can be represented with the introduction of a frequentist inertia zone, a set of data realizations where the evidence is insufficient to warrant the abandonment of the judgmental anchor.

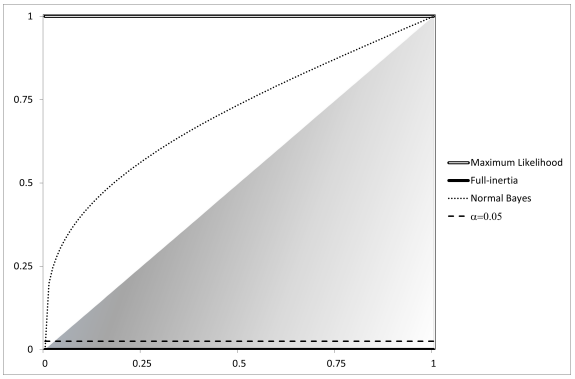
To visualize the connection among alternative decision rules, Figure 1 represents the $(\tilde{\alpha}(x), \alpha(x))$ space of *ex post* significance levels. This geometry allows us to interpret the "surprise" of the data (horizontal axis) against the decision-maker's *ex post* rejection threshold (vertical axis).

All statistical decision rules are characterized by a level of significance mapping that falls between polar decision behaviors. At one extreme lies the *full-inertia* decision ($\alpha(x) = 0$), which relies entirely on judgment and ignores the data. A decision-maker who assigns zero significance to any sample realization treats $\tilde{a}$ as unconditionally optimal. At the opposite extreme is the maximum likelihood decision ($\alpha(x) = 1$), which treats every sample realization as fully decisive.

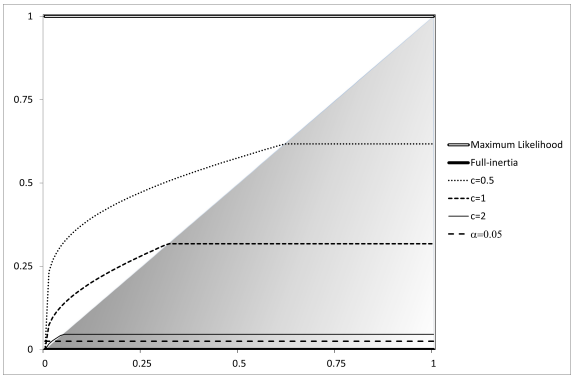
The shaded area below the 45° line represents the set of points where the *p-value* $\tilde{\alpha}(x)$ associated with the judgmental decision $\tilde{a}$ exceeds the chosen level of significance $\alpha(x)$. Within

Figure 1: *Ex post* levels of significance for alternative decision behaviors

Panel A: Bayesian Decisions



Panel B: Robust Bayesian Decisions



Note: The horizontal axis reports the *p-value* $\tilde{\alpha}(x)$ of the gradient evaluated at the observed realization x and at the judgmental decision $\tilde{a} = 0$. The vertical axis is the chosen level of significance $\alpha(x)$. The shaded area represents the points where the null hypothesis $H_0 : \theta = 0$ is not rejected, as in this area the *p-value* is greater than $\alpha(x)$. The figure plots five distinct decision behaviors. **1. Maximum Likelihood Decision**: $\alpha(x) = 1$, complete reliance on the data; **2. Full-inertia**: $\alpha(x) = 0$, complete reliance on judgment; **3. Decision with Judgment**: $\alpha(x) = 0.05$, a constant *ex post* level of significance; **4. Gaussian Bayesian Decision**: $\alpha^{\pi(\theta)}(x)$ determined by a $\mathcal{N}(0, 1)$ prior (Panel A); **5. Ambiguity-Averse Decision**: $\alpha^{\Gamma}(x)$ determined by the set of priors $\Gamma = \{\pi : \pi \text{ is } \mathcal{N}(\mu, c^{-1}), -1 \leq \mu \leq 1\}$ for different levels of precision $c \in \{0.5, 1, 2\}$ (Panel B).

this region, the null hypothesis $H_0 : \theta = \tilde{a}$ is not rejected, and the decision-maker retains the judgmental decision itself.

When the *ex post* level of significance is independent of the sample (represented in Figure 1 by the horizontal line at $\alpha(x) = 0.05$), the mapping represents the frequentist decision with judgment of Manganelli (2025). By the Probability Integral Transform theorem, the *ex ante* probability of observing a *p-value* smaller than the level of significance under the null hypothesis $H_0 : \theta = \tilde{a}$ is exactly equal to that threshold:

$$P_{\tilde{a}}^{X}\left(\tilde{\alpha}(X) < \bar{\alpha}^{A}\right) = \bar{\alpha}^{A}. \tag{7}$$

As shown in Proposition 2.4, the Bayesian decision always departs from the judgmental anchor $\tilde{a}$, meaning the null hypothesis $H_0 : \theta = \tilde{a}$ is rejected for every possible realization of the data. This implies that the mapping $\alpha^{\pi(\theta)}(x)$ determined by equation (6) lies strictly above the diagonal ($\alpha^{\pi(\theta)}(x) > \tilde{\alpha}(x)$) for all x, as reported in Panel A. Consequently, the *ex*

ante level of significance for the Bayesian rule is 1:

$$\bar{\alpha}^{\pi(\theta)} = P_{\tilde{a}}^{X}(\tilde{\alpha}(X) < \alpha^{\pi(\theta)}(X)) = 1. \tag{8}$$

Viewed through the frequentist lens, the Bayesian procedure exhibits three distinct structural features that contrast with standard decision-making heuristics. The first two properties hold generally, as will be established in Theorem 4.1 of Section 4.

1. **Absence of Inertia** – The *ex ante* level of significance equal to 1 implies that the judgmental decision is abandoned for any infinitesimal amount of evidence, regardless of the data realization.

2. **Confidence Reset** – The recursive nature of Bayesian updating introduces time variation in the decision-maker's significance levels. When a new sample is observed, the previous posterior is adopted as the new prior. While this new prior is centered on the recently updated decision, the *ex ante* significance level is simultaneously reset to 1. Consequently, the decision-maker's significance level for the judgmental decision undergoes a discrete jump: from a measured *ex post* significance strictly less than 1 (after the previous realization) to a state of certain rejection (before the next realization).

3. **Tail Paradox** – The functional form of the Bayesian mapping in Figure 1 reveals a specific relationship between the evidence and the level of significance. As the *p-value* $\tilde{\alpha}(x)$ approaches zero, indicating that the data is increasingly inconsistent with the judgmental anchor, the *ex post* level of significance $\alpha^{\pi(\theta)}(x)$ also decreases: the Bayesian decision-maker effectively lowers the significance threshold $\alpha^{\pi(\theta)}(x)$ required to justify further deviations from $\tilde{a}$ precisely when the data is most surprising. This behavior is specific to thin-tailed priors such as the normal distribution and may be difficult to reconcile with a frequentist interpretation of strengthening evidence.

The absence of an inertia zone in standard Bayesian updates suggests a need for a

framework that permits holding on to the judgmental decision in the absence of sufficient empirical evidence. Ambiguity aversion, which replaces the single subjective prior with a class of priors, provides this mechanism. Panel B of Figure 1 shows that this approach allows the mapping to cross the diagonal, creating the aforementioned *inertia zone*.

Gilboa and Schmeidler (1989) consider an ambiguity-averse decision-maker who is characterized by a set of priors Γ and evaluates actions by minimizing the expected loss under the worst-case prior within that set. By evaluating actions against the worst-case prior within Γ rather than a single subjective prior, this approach provides a robust foundation for decision-making that can accommodate the judgmental anchor $\tilde{a}$ without requiring its immediate abandonment. Within the frequentist framework, ambiguity aversion can be characterized as follows:

Proposition 2.5 (**Ambiguity**). *Consider the decision environment of Assumption 2.1, with loss function replaced by* $L(\theta, a) = -a\theta + \frac{1}{2}a^2$. *An ambiguity-averse decision-maker with the set of priors*

$$\Gamma = \{\pi : \pi \text{ is } \mathcal{N}(\mu, c^{-1}),\ c > 0,\ \ -1 \leq \mu \leq 1\}$$

over θ *chooses the action:*

$$\delta^{\Gamma}(x) = \begin{cases} (1+c)^{-1}(x+c) & \text{if } x < -c \\ 0 & \text{if } |x| \leq c \\ (1+c)^{-1}(x-c) & \text{if } x > c \end{cases} . \tag{9}$$

This action coincides with the decision with judgment when the judgmental decision is $\tilde{a} = 0$, *the* ex ante *level of significance is* $\bar{\alpha}^{\Gamma} = 2\Phi(-c)$ *and the* ex post *level of significance is:*

$$\alpha^{\Gamma}(x) = 2\Phi\left(-c(1+c)^{-1}(1 - \Phi^{-1}(\tilde{\alpha}(x)/2))\right), \tag{10}$$

where $\tilde{\alpha}(x)$ *is the* p-value *of the judgmental decision* $\tilde{a} = 0$ *given the realization* x *and* $\Phi(\cdot)$

denotes the cdf of the standard normal distribution.

Proof. See Appendix. □

Remark. The proposition uses the loss $L^l(\theta, a) \equiv -a\theta + \frac{1}{2}a^2$, which is linear in θ, rather than the quadratic loss $L^q(\theta, a) \equiv \frac{1}{2}(\theta - a)^2$ of Assumption 2.1. The two losses share the same gradient $\nabla_a L = a - \theta$ and are equivalent for all gradient-based decision rules, including the Bayesian decision and the decision with judgment. The distinction matters only for the Gilboa-Schmeidler minmax criterion, which takes the expectation of the full loss over θ. Under L^q, the posterior expected loss is quadratic in the prior mean μ, reducing the minmax problem to finding the point a that minimizes the maximum distance from the posterior mean interval $[(x - c)/(1 + c), (x + c)/(1 + c)]$. This problem has unique solution $a^* = x/(1 + c)$, the midpoint of the interval, for all x, producing no inertia zone. Under L^l, the posterior expected loss is linear in μ, so the worst-case prior is always a boundary solution $\mu^* = -\operatorname{sgn}(a)$, generating the inertia zone. The linearity of L^l in θ is therefore the key structural feature that generates inertia under the Gilboa-Schmeidler criterion, and explains why the gradient alone does not determine the behavior of the minmax problem. The absence of an inertia zone under L^q does not contradict the unifying framework of Section 4: it corresponds to the boundary case $\bar{\alpha}^\Gamma = 1$, in which the Gilboa–Schmeidler minmax agent is uncertainty neutral in the sense of Theorem 4.1(iii), and coincides with the Bayesian boundary. The loss function L^l is a concrete example where the minmax criterion generates interior uncertainty aversion, i.e. $\bar{\alpha}^\Gamma < 1$. This observation has a broader implication: within the Gilboa–Schmeidler minmax framework, a non-degenerate set of priors Γ is a necessary but not sufficient condition for the decision-maker to exhibit uncertainty aversion in the sense of $\bar{\alpha}^\Gamma < 1$. Necessity follows because a singleton set collapses the minmax criterion to standard Bayesian behavior, giving $\bar{\alpha} = 1$. Sufficiency fails, as the L^q case demonstrates: the structural feature that generates interior uncertainty aversion is not the set of priors alone but its interaction with the loss function.

Proposition 2.5 provides the formal bridge between the robust Bayesian and frequentist

frameworks. The resulting mapping, plotted for various precision levels c in Panel B, demonstrates that the ambiguity-averse agent only abandons the anchor $\tilde{a}$ when the *p-value* $\tilde{\alpha}(x)$ falls below the *ex post* significance level $\alpha^{\Gamma}(x)$ (equation (10)). When the sample realization falls within the interval $(-c, c)$, ambiguity-averse decision-makers retain their judgmental decision $\tilde{a} = 0$. This behavior mirrors the frequentist procedure in which the test fails to reject the null hypothesis $H_0 : \theta = \tilde{a}$, that is, when the *p-value* $\tilde{\alpha}(x)$ exceeds the significance threshold $\bar{\alpha}^{\Gamma}$.

In this inertia zone, the ambiguity-averse action coincides exactly with the judgmental decision. This result is consistent with the characterization in Bewley (2011), which demonstrates that frequentist confidence regions can be mapped to the set of prior means entertained by an uncertainty-averse agent. Unlike the standard Bayesian rule, which rejects the prior with probability 1, the ambiguity-averse mapping allows for a set of data realizations where judgment is preserved.

Formally, the null hypothesis $H_0 : \theta = \tilde{a}$ under the ambiguity-averse decision rule is rejected with an *ex ante* probability strictly less than 1:

$$\bar{\alpha}^{\Gamma} = P_{\tilde{a}}^{X}(\tilde{\alpha}(X) < \alpha^{\Gamma}(X)) < 1. \tag{11}$$

The precise value of $\bar{\alpha}^{\Gamma}$ depends on the precision parameter c, as established in Proposition 2.5. Geometrically, this threshold is identified in Panel B of Figure 1 at the intersection of the level of significance mappings and the diagonal of the unit square. Mathematically, the *ex ante* significance level $\bar{\alpha}^{\Gamma}$ is the fixed point of the mapping, where the curve crosses the 45° diagonal (i.e., where $\alpha^{\Gamma}(x) = \tilde{\alpha}(x)$). This intersection marks the boundary of the inertia zone, where the decision-maker transitions from retaining the judgmental decision to responding to the empirical sample. Increasing c expands the inertia zone by shifting the fixed point $\bar{\alpha}^{\Gamma}$ toward zero.

Since $\alpha^{\Gamma}(x) < \bar{\alpha}^{\Gamma}$ for $\tilde{\alpha}(x) < \bar{\alpha}^{\Gamma}$, the judgmental decision $\tilde{a}$ is rejected with Type I error

probability $\bar{\alpha}^{\Gamma}$. In the event of rejection, the ambiguity-averse decision-maker adopts the action at the boundary of the $(1-\alpha^{\Gamma}(x))$ confidence interval.

The analysis of this section illustrates three conceptual findings that motivate Sections 3 and 4. First, the *ex ante* level of significance serves as the index of uncertainty aversion: it equals 1 for the Bayesian decision-maker, lies strictly between 0 and 1 for the ambiguity-averse decision-maker, and equals 0 for the full-inertia decision-maker. Second, the *ex post* level of significance provides the formal mapping between the Bayesian and frequentist frameworks: for any Bayesian decision, there exists a unique data-dependent significance level that rationalizes it as a decision with judgment. Third, the figure geometry suggests that the three paradigms are not competing philosophies but specific points along a single continuum, unified by the $(\tilde{\alpha}(x), \alpha(x))$ space. The next sections formalize these insights in a multivariate setting under minimal assumptions.

3 Frequentist Representation of Bayesian Decisions

This section establishes the conditions for Bayesian decisions to be represented as multivariate decisions with judgment.

We begin by characterizing the decision environment:

Assumption 3.1 (**Decision Environment**). *(i) The random vector* $\mathbf{X}$ *takes values in a sample space* $\mathcal{X} \subseteq \mathbb{R}^n$ *with joint probability density function* $f(\mathbf{x}|\boldsymbol{\theta})$*, where* $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ *is an unknown parameter vector.*

(ii) Given the realization $\mathbf{x} \in \mathcal{X}$*, the decision-maker chooses an action* $\mathbf{a} \in \mathcal{A} = \mathbb{R}^k$*.*

(iii) The loss function is separable across components of $\mathbf{a}$*:* $L(\boldsymbol{\theta}, \mathbf{a}) = \sum_{i=1}^{k} L_i(\boldsymbol{\theta}, a_i)$*, where each* L_i *is twice continuously differentiable, strictly convex, and coercive in* a_i*, for all* $\boldsymbol{\theta} \in \Theta$*.*

Strict convexity and coercivity guarantee a unique optimal action for any $\boldsymbol{\theta} \in \Theta$. As will

be established in Section 3.2, the separability condition ensures that the multivariate decision problem decouples into independent component-wise tests, needed for frequentist calibration.

3.1 The Multivariate Bayesian Decision

In the Bayesian framework, judgment is incorporated with a prior distribution over the parameter space Θ, and the resulting action is an implicit consequence of how the likelihood updates those parameter beliefs. This contrasts with the decision with judgment introduced in Section 3.2, where judgment is placed on the decision variable $\mathbf{a}$.

Assumption 3.2 (**Bayesian Prior**). *The decision-maker's judgment is represented by a proper prior distribution $\pi(\boldsymbol{\theta})$ supported on Θ.*

To ensure the Bayesian decision is well-defined and characterizable via infinitesimal variations, we require the following regularity conditions:

Assumption 3.3 (**Bayesian Regularity**). *For all $\mathbf{x} \in \mathcal{X}$:*

(i) The marginal likelihood $m(\mathbf{x}) = \int_\Theta f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ satisfies $0 < m(\mathbf{x}) < \infty$, ensuring the posterior $\pi(\boldsymbol{\theta}|\mathbf{x})$ is well-defined.

(ii) There exists a function $G(\boldsymbol{\theta})$, independent of $\mathbf{a}$ and $\mathbf{x}$, such that $\|\nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \mathbf{a})\| \leq G(\boldsymbol{\theta})$ for all $\mathbf{a} \in \mathcal{A}$, and $\int_\Theta G(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} < \infty$ for all $\mathbf{x} \in \mathcal{X}$.

Theorem 3.1 (**Bayesian Decision**). *Under Assumptions 3.1, 3.2, and 3.3, the Bayesian decision $\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x})$ exists, is unique, and is the solution to the system of first-order conditions:*

$$\int_\Theta \nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x}))\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta} = \mathbf{0}. \tag{12}$$

Proof. See Appendix. □

3.2 The Multivariate Decision with Judgment

The decision with judgment is constructed from a judgmental decision and associated level of significance. Manganelli (2025) implements it by using a scalar level of significance. A more general approach defines the levels of significance as a k-vector of probabilities, to allow for component-wise hypothesis testing of the optimality of the judgmental decision.

Assumption 3.4 (**Judgment**). *The decision-maker's judgment is characterized by the pair* $A = \{\tilde{\mathbf{a}}, \boldsymbol{\alpha}\}$*, where* $\tilde{\mathbf{a}} \in \mathcal{A}$ *is the judgmental decision and* $\boldsymbol{\alpha} \in [0,1]^k$ *is the vector of levels of significance.*

To ensure the decision framework is operationally well-defined, we impose the following regularity on the estimator and the parameter space.

Assumption 3.5 (**Properties of the Maximum Likelihood Estimator**). *For any realization* $\mathbf{x} \in \mathcal{X}$*, the maximum likelihood estimate* $\hat{\boldsymbol{\theta}}(\mathbf{x})$ *exists and belongs to the interior of the parameter space* Θ*. Furthermore, the sampling distribution of the estimator* $\hat{\boldsymbol{\theta}}(\mathbf{X})$ *is absolutely continuous with respect to the Lebesgue measure on* Θ*.*

This high-level assumption ensures the existence of a valid estimator without referring to specific estimation frameworks. This allows the decision rule to be operationalized using either exact finite-sample distributions, when known, or standard asymptotic approximations.

Define the maximum likelihood and the shrinkage actions as follows:

Definition 3.1 (**Maximum Likelihood and Shrinkage Actions**). *Suppose Assumptions 3.1, 3.4 and 3.5 hold. The* ***maximum likelihood action*** *is*

$$\hat{\mathbf{a}}(\mathbf{x}) = \arg\min_{\mathbf{a} \in \mathcal{A}} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{a}). \tag{13}$$

The ***shrinkage action*** *is the component-wise convex combination*

$$\mathbf{a}(\boldsymbol{\lambda}; \mathbf{x}) = \boldsymbol{\lambda} \odot \hat{\mathbf{a}}(\mathbf{x}) + (\mathbf{1} - \boldsymbol{\lambda}) \odot \tilde{\mathbf{a}}, \tag{14}$$

where $\mathbf{1} = [1, \ldots, 1]^\top \in \mathbb{R}^k$, $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_k]^\top \in [0,1]^k$ *is the vector of shrinkage factors,* $\tilde{\mathbf{a}}$ *is the judgmental decision, and* $\odot$ *denotes the Hadamard (element-wise) product.*

Strict convexity and coercivity guarantee that the maximum likelihood action exists and is uniquely characterized by $\nabla_{\mathbf{a}} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \hat{\mathbf{a}}(\mathbf{x})) = \mathbf{0}$. The shrinkage action confines the feasible action space to an interpolation between the judgmental and maximum likelihood actions.

The action $\delta^A(\mathbf{x})$ is determined by testing the null hypothesis that the gradient of the loss function, evaluated component-wise at the judgmental decision $\tilde{\mathbf{a}}$, is zero: $H_0 : \nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \tilde{\mathbf{a}}) = \mathbf{0}$, or equivalently, $H_0 : \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{\theta}, \mathbf{a}(\mathbf{0}; \mathbf{x})) = \mathbf{0}$.

Definition 3.2 (**Multivariate Decision with Judgment**)**.** *Let* $\mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})$ *be the shrinkage action of Definition 3.1, and let* $\mathbf{Y}$ *be an auxiliary random vector independent of* $\mathbf{X}$ *but with the same distribution. Define the component-wise standardized gradient:*

$$Z_i(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\lambda}; \mathbf{x}) \equiv \frac{\nabla_{\lambda_i} L(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \mathbf{a}(\boldsymbol{\lambda}; \mathbf{x}))}{\sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})}}, \quad i = 1, \ldots, k \tag{15}$$

where $\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})$ *is the* i*-th diagonal element of the variance-covariance matrix* $\Omega(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})$ *of the gradient vector* $\nabla_{\boldsymbol{\lambda}} L(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \mathbf{a}(\boldsymbol{\lambda}; \mathbf{x}))$*. The* ***realized standardized gradient*** *is defined as the value of* Z_i *when the auxiliary estimator is evaluated at the observed sample:*

$$\hat{z}_i(\boldsymbol{\lambda}; \mathbf{x}) \equiv Z_i(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda}; \mathbf{x}) = \frac{\nabla_{\lambda_i} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{a}(\boldsymbol{\lambda}; \mathbf{x}))}{\sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})}}. \tag{16}$$

For any scalar $\mu \in [0,1]$*, let* $(\mu, \hat{\boldsymbol{\lambda}}_{-i})$ *denote the vector* $\hat{\boldsymbol{\lambda}}$ *with the* i*-th component replaced by* μ*. The multivariate decision with judgment* $A = \{\tilde{\mathbf{a}}, \boldsymbol{\alpha}\}$ *is* $\delta^A(\mathbf{x}) = \mathbf{a}(\hat{\boldsymbol{\lambda}}; \mathbf{x})$*, where* $\hat{\boldsymbol{\lambda}} \in [0,1]^k$ *is a fixed point satisfying the following system of conditions for all* $i = 1, \ldots, k$*:*

$$\hat{\lambda}_i = \begin{cases} \lambda_i^*, & \text{if } \hat{z}_i((0, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) < q_{i,\alpha_i/2}, \\ 0, & \text{otherwise}, \end{cases} \tag{17}$$

where $q_{i,\alpha_i/2}$ *denotes the* $(\alpha_i/2)$*-quantile of the distribution of* $Z_i(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\lambda}; \mathbf{x})$*, under the null hypothesis* $H_0^{(0,\hat{\boldsymbol{\lambda}}_{-i})} : \nabla_{\lambda_i} L(\boldsymbol{\theta}, \mathbf{a}((0, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x})) = 0$*, and* λ_i^* *is a scalar satisfying the boundary condition* $\hat{z}_i((\lambda_i^*, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) = q_{i,\alpha_i/2}$ *if* $\hat{z}_i((1, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) \geq q_{i,\alpha_i/2}$*, and* $\lambda_i^* = 1$ *otherwise.*

Definition 3.2 characterizes the decision as a system of simultaneous fixed-point conditions. However, under the separability of Assumption 3.1(iii), the gradient in dimension i does not depend on $\boldsymbol{\lambda}_{-i}$. Consequently, the fixed-point system mathematically decouples into k independent scalar problems and all subsequent results are derived in this setting.

Remark (**Role of the auxiliary variable**)**.** The introduction of the auxiliary random vector $\mathbf{Y}$ enforces a concept central to this framework: statistical inference is conducted *conditional on the observed sample realization* $\mathbf{x}$. The fixed realization $\mathbf{x}$ plays a non-random role: it defines the coordinate system by anchoring the shrinkage path $\mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})$. The auxiliary vector $\mathbf{Y}$ provides the sampling distribution for the estimator $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ under the null hypothesis. The realized test statistic $\hat{z}_i$ is then obtained by the legitimate substitution $\mathbf{Y} = \mathbf{x}$ into this conditionally derived distribution. Attempting to use the data vector $\mathbf{X}$ to simultaneously determine the shrinkage path and supply the random estimator draw would conflate these two roles and yield an ill-formed sampling distribution. See also the discussion right after Proposition 2.3 and the argument used in the relative proof.

To show that the multivariate decision with judgment exists and is unique, the following additional regularity conditions are needed.

Assumption 3.6 (**Distributional Invariance**)**.** *For every* $\mathbf{x} \in \mathcal{X}$ *and for all* $i = 1, \ldots, k$*, the following properties hold:*

(i) (Nondegenerate variance) *The conditional variance*

$$\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda}) = \mathrm{Var}\Big[\nabla_{\lambda_i} L\Big(\hat{\boldsymbol{\theta}}(\mathbf{Y}),\, \mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})\Big)\Big]$$

is strictly positive.

(ii) (Distributional invariance) *The distribution of the standardized gradient* $Z_i(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\lambda}; \mathbf{x})$ *under the null hypothesis* $H_0^{(\boldsymbol{\lambda})} : \nabla_{\lambda_i} L(\boldsymbol{\theta}, \mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})) = 0$ *is invariant to* $\boldsymbol{\lambda} \in [0,1]^k$*, so that the quantiles* $q_{i,\alpha_i/2}$ *do not depend on* $\boldsymbol{\lambda}$*.*

(iii) (Variance invariance) *The normalizing variance* $\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})$ *does not depend on* $\boldsymbol{\lambda} \in [0,1]^k$*.*

(iv) (Symmetry) *Under the null hypothesis* $H_0^{(\boldsymbol{\lambda})}$*, the distribution of* $Z_i(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\lambda}; \mathbf{x})$ *is symmetric around zero, so that* $q_{i,1/2} = 0$*.*

Non-degeneracy (i) and symmetry (iv) are satisfied under standard asymptotic regularity conditions. Distributional invariance (ii) follows from separability asymptotically. Variance invariance (iii) is a separate maintained requirement that does not follow from (ii) alone, and is needed for the penalty term $\sum_{i=1}^k \lambda_i \, q_{i,\alpha_i/2} \sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})}$ in the auxiliary objective of the proof of Theorem 3.2 to be linear in $\boldsymbol{\lambda}$.

The following theorem establishes that the system of conditions in (17) admits a unique solution for any judgment $A = \{\tilde{\mathbf{a}}, \boldsymbol{\alpha}\}$ and any realization $\mathbf{x}$.

Theorem 3.2 (**Existence and Uniqueness of the Multivariate Decision with Judgment**)**.** *Suppose Assumptions 3.1, 3.4, 3.5 and 3.6 hold. For any judgment* $A = \{\tilde{\mathbf{a}}, \boldsymbol{\alpha}\}$ *and any realization* $\mathbf{x}$ *satisfying* $\hat{\mathbf{a}}(\mathbf{x}) \neq \tilde{\mathbf{a}}$ *component-wise, the system of conditions in* (17) *admits a unique solution vector* $\hat{\boldsymbol{\lambda}} \in [0,1]^k$*, which uniquely determines the decision* $\delta^A(\mathbf{x})$*.*

Proof. See Appendix. □

The following definition introduces the concept of *profile p-value*, which generalizes the scalar *p-value* of Section 2 to the multivariate setting by measuring the evidence against the optimality of $\tilde{a}_i$ in each dimension i separately, with the remaining shrinkage components held fixed.

Definition 3.3 (**Profile p-value**)**.** *Let* $\boldsymbol{\lambda}_{-i} \in [0,1]^{k-1}$ *be any fixed vector of shrinkage values for the components* $j \neq i$*. Let* F_i *denote the cdf of the standardized gradient* $Z_i(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\lambda}; \mathbf{x})$

*and let $\hat{z}_i$ denote its realized value, both as defined in Definition 3.2, under the null hypothesis $H_0^{(0,\boldsymbol{\lambda}_{-i})} : \nabla_{\lambda_i} L(\boldsymbol{\theta}, \mathbf{a}((0, \boldsymbol{\lambda}_{-i}); \mathbf{x})) = 0$. The **profile p-value** of the judgmental decision $\tilde{\mathbf{a}}$ in dimension i at realization $\mathbf{x}$, given $\boldsymbol{\lambda}_{-i}$, is:*

$$\tilde{\alpha}_i(\boldsymbol{\lambda}_{-i}; \mathbf{x}) \equiv 2F_i(\hat{z}_i((0, \boldsymbol{\lambda}_{-i}); \mathbf{x}))\,, \quad i = 1, \ldots, k. \tag{18}$$

In the multivariate case, a naive approach would evaluate the evidence against $\tilde{a}_i$ by evaluating all components of the gradient simultaneously at the judgmental anchor ($\boldsymbol{\lambda} = \mathbf{0}$). The *profile p-value* $\tilde{\alpha}_i(\boldsymbol{\lambda}_{-i}; \mathbf{x})$ corrects for potential cross-dimensional dependencies by conditioning on the jointly determined $\boldsymbol{\lambda}_{-i}$. As noted previously, while this conditioning accommodates general loss functions, under the separability of Assumption 3.1(iii), the conditioning becomes mathematically vacuous, and the evidence against $\tilde{a}_i$ collapses to the marginal p-value.

Theorem 3.3 (**Existence of Representative Significance Levels**). *Suppose Assumptions 3.1, 3.4, 3.5 and 3.6 hold, and $\hat{\mathbf{a}}(\mathbf{x}) \neq \tilde{\mathbf{a}}$ component-wise. Then, for any shrinkage vector $\boldsymbol{\lambda}^* \in [0, 1]^k$, there exists a vector of data-dependent significance levels $\boldsymbol{\alpha}(\mathbf{x}) \in [0, 1]^k$ such that the decision rule (17) yields $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^*$, and consequently $\delta^A(\mathbf{x}) = \mathbf{a}(\boldsymbol{\lambda}^*; \mathbf{x})$. For each component i with $\lambda_i^* > 0$, the significance level $\alpha_i(\mathbf{x})$ is uniquely determined by the boundary condition $\hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x}) = q_{i,\alpha_i(\mathbf{x})/2}$. For each component i with $\lambda_i^* = 0$, any value $\alpha_i(\mathbf{x}) \leq \tilde{\alpha}_i(\boldsymbol{\lambda}_{-i}^*; \mathbf{x})$ rationalizes $\hat{\lambda}_i = 0$; the canonical choice $\alpha_i(\mathbf{x}) = \tilde{\alpha}_i(\boldsymbol{\lambda}_{-i}^*; \mathbf{x})$ is adopted by convention.*

Proof. See Appendix. □

3.3 Bayesian Representation

To establish the Bayesian representation result, we must ensure that the Bayesian updating process is structurally compatible with the interpolative nature of the decision with judgment.

Assumption 3.7 (**Component-wise Shrinkage Representation**). *For a given prior*

$\pi(\boldsymbol{\theta})$ and all realizations $\mathbf{x} \in \mathcal{X}$, there exists a vector $\boldsymbol{\lambda}^B \in [0,1]^k$ such that:

$$\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x}) = \boldsymbol{\lambda}^B \odot \hat{\mathbf{a}}(\mathbf{x}) + (\mathbf{1} - \boldsymbol{\lambda}^B) \odot \delta^{\pi(\boldsymbol{\theta})}, \tag{19}$$

where $\delta^{\pi(\boldsymbol{\theta})}$ is the prior action satisfying $\mathbb{E}_{\pi(\boldsymbol{\theta})}[\nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \delta^{\pi(\boldsymbol{\theta})})] = \mathbf{0}$, $\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x})$ is the posterior action satisfying $\mathbb{E}_{\pi(\boldsymbol{\theta}|\mathbf{x})}[\nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x}))] = \mathbf{0}$, and $\hat{\mathbf{a}}(\mathbf{x})$ is the maximum likelihood action satisfying $\nabla_{\mathbf{a}} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \hat{\mathbf{a}}(\mathbf{x})) = \mathbf{0}$.

This assumption restricts attention to Bayesian models in which the posterior action is a component-wise interpolation between the prior action and the maximum likelihood action. A Bayesian update that moves in the opposite direction ($\lambda_i < 0$ for some $i \in \{1, \ldots, k\}$) or overshoots the empirical evidence ($\lambda_i > 1$) is incompatible with the interpolative logic of the decision with judgment, and arises only when the posterior mean is not a convex combination of the prior mean and the maximum likelihood action, as with multimodal or heavily asymmetric priors. The assumption is satisfied by exponential families with conjugate priors: in these models the posterior mean is by construction an affine combination of the prior mean and the maximum likelihood estimator, as verified explicitly in the Gaussian-quadratic environment of Section 2, where $\lambda^B = 1/2 \in [0,1]$ for all x. Verification of the assumption in the multivariate case is model-specific.

We next generalize the scalar significance level of Assumption 3.4 to allow for data-dependence.

Definition 3.4 (**Significance Level Vectors**). *Given a judgment $A = \{\tilde{\mathbf{a}}, \cdot\}$ as in Assumption 3.4, the levels of significance are characterized by the pair $\{\bar{\boldsymbol{\alpha}}, \boldsymbol{\alpha}(\mathbf{x})\}$, where:*

(i) $\bar{\boldsymbol{\alpha}} \in [0,1]^k$ *is a vector of constants, the* **ex ante *significance vector***, *fixed prior to observing the data.*

(ii) $\boldsymbol{\alpha}(\mathbf{x}) = [\alpha_1(\mathbf{x}), \ldots, \alpha_k(\mathbf{x})]^\top \in [0,1]^k$, *for $\mathbf{x} \in \mathcal{X}$, is a data-dependent vector, the* **ex post *significance vector***, *determined upon observing $\mathbf{x}$.*

Definition 3.4 extends the judgment of Assumption 3.4 by replacing the fixed significance vector $\boldsymbol{\alpha}$ with the pair $\{\bar{\boldsymbol{\alpha}}, \boldsymbol{\alpha}(\mathbf{x})\}$. Henceforth, $A = \{\tilde{\mathbf{a}}, \bar{\boldsymbol{\alpha}}, \boldsymbol{\alpha}(\mathbf{x})\}$ denotes the extended judgment. Assumption 3.4 is recovered as the special case in which $\boldsymbol{\alpha}(\mathbf{x}) = \bar{\boldsymbol{\alpha}}$ for all $\mathbf{x} \in \mathcal{X}$, i.e., the *ex post* significance vector is constant and coincides with the *ex ante* vector.

The *ex ante* vector $\bar{\boldsymbol{\alpha}}$ governs whether the judgmental decision $\tilde{\mathbf{a}}$ is abandoned at all, while the *ex post* vector $\boldsymbol{\alpha}(\mathbf{x})$ determines the magnitude of the departure along the shrinkage path. The central result of this section shows that the Bayesian update corresponds to $\bar{\boldsymbol{\alpha}} = [1, \ldots, 1]^\top$ and a specific data-dependent choice of $\boldsymbol{\alpha}(\mathbf{x})$.

The following theorem establishes that the Bayesian update is a boundary case of the decision with judgment.

Theorem 3.4 (**Frequentist Representation of Bayesian Decisions**). *Suppose the Assumptions 3.1–3.7 are satisfied. For any Bayesian decision rule $\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x})$ and any realization $\mathbf{x} \in \mathcal{X}$ satisfying $\hat{\mathbf{a}}(\mathbf{x}) \neq \tilde{\mathbf{a}}$ component-wise, there exists a frequentist representation with judgment $A = \{\tilde{\mathbf{a}}, \bar{\boldsymbol{\alpha}}, \boldsymbol{\alpha}(\mathbf{x})\}$, in the sense of Definition 3.4, such that $\delta^A(\mathbf{x}) = \delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x})$. The judgment A is characterized by:*

- *a judgmental decision $\tilde{\mathbf{a}}$ implicitly defined by $\mathbb{E}_{\pi(\boldsymbol{\theta})}[\nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \tilde{\mathbf{a}})] = \mathbf{0}$;*
- *a vector of* ex ante *significance levels $\bar{\boldsymbol{\alpha}} = [1, \ldots, 1]^\top$;*
- *a vector of data-dependent,* ex post *significance levels $\boldsymbol{\alpha}(\mathbf{x}) = [\alpha_1(\mathbf{x}), \ldots, \alpha_k(\mathbf{x})]^\top$, where $\alpha_i(\mathbf{x}) = 2F_i\Big(\hat{z}_i(\boldsymbol{\lambda}^B; \mathbf{x})\Big)$, for $i = 1, \ldots, k$, and $\boldsymbol{\lambda}^B$ is the shrinkage vector identifying the Bayesian action guaranteed by Assumption 3.7.*

Proof. See Appendix. □

Theorem 3.4 is the logical culmination of the double conditioning mechanism: the first dimension anchors the test to a data-dependent shrinkage path $\mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})$, while the second dimension maps the observed evidence to a specific *ex post* significance level $\alpha_i(\mathbf{x})$. The result that $\bar{\boldsymbol{\alpha}} = \mathbf{1}$ characterizes the Bayesian decision-maker as uncertainty neutral: the judgmental

anchor is abandoned for any $\mathbf{x}$, regardless of the strength of evidence. This provides the frequentist benchmark against which the ambiguity-averse inertia of Section 4 is measured.

3.4 Two Alternative Strategies to Incorporate Judgment

Wald's statistical decision problem is fully specified by three primitives: the statistical model $f(\mathbf{x}|\boldsymbol{\theta})$, the action space $\mathcal{A}$, and the loss function $L(\boldsymbol{\theta}, \mathbf{a})$. The loss function is the common ground, as both Bayesian and frequentist traditions accept it as the object to be minimized. The Bayesian-frequentist tension reduces to a single question: on which of the remaining two primitives is judgment imposed? The two frameworks give opposite answers:

- **The Bayesian Framework** encodes judgment on the parameter space Θ.
- **The Frequentist Framework** encodes judgment on the decision space $\mathcal{A}$.

Theorem 3.4 shows that these are not competing approaches but alternative representations of the same decision: any prior distribution over $\boldsymbol{\theta}$ implies a judgmental decision $\tilde{\mathbf{a}}$ and a data-dependent significance mapping $\boldsymbol{\alpha}(\mathbf{x})$. The two frameworks are related within the class of models satisfying Assumption 3.7, where judgment and data operate as complementary rather than contradictory forces.

This relationship is asymmetric. Theorem 3.4 maps Bayesian decisions to decisions with judgment, but the converse does not hold in general. Because Bayesian decisions necessarily imply $\bar{\boldsymbol{\alpha}} = \mathbf{1}$, any decision rule exhibiting inertia ($\bar{\alpha}_i < 1$ for some i) is fundamentally non-Bayesian. This identifies the *ex ante* significance vector as the primitive that distinguishes standard Bayesian behavior from rules that protect the judgmental anchor.

The following section formalizes this by establishing $\bar{\boldsymbol{\alpha}}$ as the primitive indexing a continuum from full inertia to uncertainty neutrality.

4 Frequentist Representation of Uncertainty

This section generalizes the representation of Section 3 to the full continuum of uncertainty aversion. We establish that uncertainty-averse rules are represented by an *ex ante* significance vector with $\bar{\alpha}_i < 1$ for some $i \in \{1, \ldots, k\}$, while Bayesian uncertainty neutrality remains the boundary case $\bar{\boldsymbol{\alpha}} = \mathbf{1}$.

4.1 Uncertainty Representation

The general representation relies on the formalization of the induced shrinkage path and a behavioral threshold for inertia.

Definition 4.1 (**Induced Shrinkage Vector and Equilibrium Profile p-value**). *Let $\delta(\cdot) : \mathcal{X} \to \mathcal{A}$ be a decision rule satisfying Assumption 3.7. The **induced shrinkage vector** $\boldsymbol{\lambda}(\mathbf{x}) \in [0,1]^k$ is the unique vector satisfying:*

$$\delta(\mathbf{x}) = \boldsymbol{\lambda}(\mathbf{x}) \odot \hat{\mathbf{a}}(\mathbf{x}) + (\mathbf{1} - \boldsymbol{\lambda}(\mathbf{x})) \odot \tilde{\mathbf{a}}, \quad \mathbf{x} \in \mathcal{X}, \tag{20}$$

*where uniqueness follows from $\hat{\mathbf{a}}(\mathbf{x}) \neq \tilde{\mathbf{a}}$ component-wise. The **equilibrium profile p-value** in dimension i is:*

$$\tilde{\alpha}_i(\mathbf{x}) \equiv \tilde{\alpha}_i(\boldsymbol{\lambda}_{-i}(\mathbf{x}); \mathbf{x}), \quad i = 1, \ldots, k, \tag{21}$$

where $\tilde{\alpha}_i(\boldsymbol{\lambda}_{-i}; \mathbf{x})$ is the profile p-value *of Definition 3.3 evaluated at $\boldsymbol{\lambda}_{-i} = \boldsymbol{\lambda}_{-i}(\mathbf{x})$.*

The induced shrinkage vector expresses any interpolating decision rule in terms of its component-wise weights between the judgmental anchor and the maximum likelihood action, while the equilibrium profile *p-value* specializes the profile *p-value* of Definition 3.3 to the shrinkage values simultaneously determined by the decision rule.

To establish the general representation result, the following assumption is needed.

Assumption 4.1 (**Threshold Inertia**). *For each $i = 1, \ldots, k$, there exists a threshold*

$\bar{\alpha}_i \in [0,1]$, *fixed prior to observing the data, such that:*

$$\lambda_i(\mathbf{x}) = 0 \iff \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i, \quad \mathbf{x} \in \mathcal{X}.$$

Assumption 4.1 imposes a precise behavioral restriction: in each dimension i, the decision to retain or abandon the judgmental anchor is governed exclusively by the statistical evidence against it, as measured by the *equilibrium profile p-value* $\tilde{\alpha}_i(\mathbf{x})$, and by a fixed pre-committed threshold $\bar{\alpha}_i$. Any interpolating decision rule partitions the sample space into an inertia zone $\{\lambda_i(\mathbf{x}) = 0\}$ and a rejection zone $\{\lambda_i(\mathbf{x}) > 0\}$: the assumption requires that this partition be determined by statistical evidence alone, not by factors extraneous to the hypothesis being tested. This is precisely the frequentist definition of a significance level as a pre-committed statistical threshold, now applied directly to the decision problem.

The assumption is satisfied by all decision frameworks considered in this paper. The frequentist decision with judgment satisfies it by construction. The ambiguity-averse decision of Proposition 2.5 satisfies it with threshold $\bar{\alpha}^{\Gamma} = 2\Phi(-c)$: the inertia zone $\{|x| \leq c\}$ coincides exactly with $\{\tilde{\alpha}(x) \geq 2\Phi(-c)\}$. The Bayesian decision satisfies it with threshold $\bar{\alpha}_i = 1$, since the inertia zone is empty.

The assumption excludes rules that condition the inertia decision on features of the data that are extraneous to the hypothesis being tested, such as the sign of x_i independently of its *equilibrium p-value*, the magnitude of components unrelated to the null, or institutional and behavioral factors outside the statistical framework. It also excludes rules whose inertia zone is not an upper level set of the *equilibrium p-value* in each dimension i. One such example is a rule that retains the anchor when the *equilibrium p-value* is either very small or very large but rejects it for intermediate values, producing a non-connected inertia zone that cannot be characterized by a single threshold. In the next section, we show that decisions associated with non-connected inertia zones are suboptimal.

Definition 4.2 (**Uncertainty Aversion**). *A decision rule $\delta(\cdot)$ satisfying Assumption 4.1 is*

uncertainty averse *if* $\bar{\alpha}_i < 1$ *for some* $i \in \{1, \ldots, k\}$*, and* ***uncertainty neutral*** *if* $\bar{\alpha}_i = 1$ *for all* $i = 1, \ldots, k$*.*

The next theorem establishes that within this class the threshold $\bar{\alpha}_i$ is uniquely identified by the decision rule as its *ex ante* rejection probability, and that it orders inertia zones monotonically across the uncertainty continuum.

Theorem 4.1 (**Representation of Decision Rules Along the Uncertainty Continuum**)**.** *Suppose Assumptions 3.1, 3.4, 3.5, and 3.6 hold, and that* $\hat{\mathbf{a}}(\mathbf{x}) \neq \tilde{\mathbf{a}}$ *component-wise. Let* $\delta(\cdot) : \mathcal{X} \to \mathcal{A}$ *be a decision rule satisfying Assumptions 3.7 and 4.1, with induced shrinkage vector* $\boldsymbol{\lambda}(\mathbf{x})$ *as defined in* (20)*. Within this class, the following holds:*

(i) (Representation) *The rule* $\delta(\cdot)$ *can be represented as a decision with judgment* $A = \{\tilde{\mathbf{a}}, \bar{\boldsymbol{\alpha}}, \boldsymbol{\alpha}(\cdot)\}$*, where the* ex ante *significance vector* $\bar{\boldsymbol{\alpha}} = [\bar{\alpha}_1, \ldots, \bar{\alpha}_k]^\top$ *is the threshold vector of Assumption 4.1 and the* ex post *significance levels are given at each realization* $\mathbf{x} \in \mathcal{X}$ *by:*

$$\alpha_i(\mathbf{x}) = \begin{cases} 2F_i(\hat{z}_i(\boldsymbol{\lambda}(\mathbf{x}); \mathbf{x})) & \text{if } \lambda_i(\mathbf{x}) > 0, \\ \tilde{\alpha}_i(\mathbf{x}) & \text{if } \lambda_i(\mathbf{x}) = 0, \end{cases} \qquad i = 1, \ldots, k.$$

(ii) (*Ex ante* characterization) *The* ex ante *significance vector is uniquely determined by the rule as:*

$$\bar{\alpha}_i = P_{\tilde{\mathbf{a}}}\Big(\lambda_i(\mathbf{X}) > 0\Big) = P_{\tilde{\mathbf{a}}}\Big(\tilde{\alpha}_i(\mathbf{X}) < \alpha_i(\mathbf{X})\Big), \quad i = 1, \ldots, k. \tag{22}$$

(iii) (Uncertainty neutrality) *If* $\delta(\cdot)$ *can be represented as a Bayesian decision, then* $\bar{\alpha}_i = 1$ *for all* $i = 1, \ldots, k$*. The converse holds within the class of decision rules admitting a prior representation: if* $\bar{\alpha}_i = 1$ *for all* i *and* $\delta(\cdot)$ *is rationalized by posterior expected loss minimization under some proper prior* $\pi(\boldsymbol{\theta})$*, then* $\delta(\cdot)$ *can be represented as a Bayesian decision.*

(iv) (Monotone uncertainty aversion) *In dimension i, the inertia zone* $\{\mathbf{x} \in \mathcal{X} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i\}$ *is non-decreasing in the set-inclusion sense as* $\bar{\alpha}_i$ *decreases: for any* $\bar{\alpha}_i^{(1)} \leq \bar{\alpha}_i^{(2)}$,

$$\left\{\mathbf{x} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i^{(2)}\right\} \subseteq \left\{\mathbf{x} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i^{(1)}\right\},$$

so that a smaller $\bar{\alpha}_i$ *corresponds to a larger inertia zone and a greater reluctance to depart from the judgmental anchor* $\tilde{\mathbf{a}}$.

Proof. See Appendix. □

Remark. The *ex post* significance level $\alpha_i(\mathbf{x})$ is uniquely identified by the rule $\delta(\cdot)$ only in the rejection zone $\{\mathbf{x} : \lambda_i(\mathbf{x}) > 0\}$, where the formula $\alpha_i(\mathbf{x}) = 2F_i(\hat{z}_i(\boldsymbol{\lambda}(\mathbf{x}); \mathbf{x}))$ yields a value strictly above $\tilde{\alpha}_i(\mathbf{x})$, as established in Step 1 of the proof. In the inertia zone $\{\mathbf{x} : \lambda_i(\mathbf{x}) = 0\}$, any value $\alpha_i(\mathbf{x}) \leq \tilde{\alpha}_i(\mathbf{x})$ produces the same decision. The uniquely identified primitive of the representation is the threshold $\bar{\alpha}_i$, characterized in part (ii).

Theorem 4.1, by establishing that every interpolating decision rule satisfying Assumption 4.1 can be represented as a decision with judgment, reveals that the *ex ante* significance vector $\bar{\boldsymbol{\alpha}}$ is uniquely identified from observed behavior as the *ex ante* rejection probability, making it an empirically measurable index of a decision-maker's position on the uncertainty continuum. The Bayesian, frequentist and robust Bayesian paradigms are not competing philosophical positions but specific points along a single continuum, indexed by $\bar{\boldsymbol{\alpha}}$. At one extreme, the Bayesian decision-maker sets $\bar{\alpha}_i = 1$ in every dimension, committing to abandon the judgmental anchor for any infinitesimal amount of evidence. At the other extreme, the full-inertia decision-maker sets $\bar{\alpha}_i = 0$, retaining the anchor regardless of the data. Every decision rule between these poles corresponds to an intermediate value of $\bar{\boldsymbol{\alpha}}$, with a larger inertia zone and greater reluctance to depart from the anchor as $\bar{\boldsymbol{\alpha}}$ decreases.

The scope of the theorem is determined by Assumption 4.1, which requires that the inertia decision be governed exclusively by statistical evidence against the judgmental anchor, as measured by the equilibrium profile *p-value*. This is a behavioral restriction on the

decision rule rather than a restriction on the underlying belief structure: it is satisfied by the frequentist decision with judgment by construction, by the Bayesian decision with $\bar{\alpha}_i = 1$, and, as established in Proposition 2.5, by the Gilboa-Schmeidler minmax decision in the Gaussian-quadratic environment, where the inertia zone $\{|x| \leq c\}$ coincides exactly with the level set $\{\tilde{\alpha}(x) \geq 2\Phi(-c)\}$. Verification that a specific robust Bayesian model satisfies Assumption 4.1 is model-specific and depends on the structure of the set of priors Γ.

4.2 Alternative Foundations for Threshold Inertia

Threshold inertia is the unique optimal accept-or-reject rule under two independent traditions: the axiomatic approach of Gilboa and Schmeidler (1989) and the error-control framework of Neyman and Pearson (1933). While the former generates inertia through preference axioms over acts, the latter generates it via statistical power maximization under a Type I error constraint.

The result requires one additional distributional condition on the test statistic.

Assumption 4.2 (**Monotone Likelihood Ratio**). *For each $i = 1, \ldots, k$ and each fixed $\mathbf{x} \in \mathcal{X}$, define the population gradient $\eta_i(\boldsymbol{\theta}; \mathbf{x}) \equiv \nabla_{\lambda_i} L(\boldsymbol{\theta}, \mathbf{a}((0, \boldsymbol{\lambda}_{-i}(\mathbf{x})); \mathbf{x}))$. The family of distributions of $Z_i(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\lambda}(\mathbf{x}); \mathbf{x})$, indexed by the scalar $\eta_i(\boldsymbol{\theta}; \mathbf{x})$ for fixed $\mathbf{x}$, has the monotone likelihood ratio (MLR) property in Z_i.*

Assumption 4.2 is satisfied in all standard parametric families used in econometrics. In natural exponential families the sufficient statistic has the MLR property by construction, and the gradient-based test statistic inherits it. The Gaussian-quadratic environment of Section 2 satisfies this condition, as do t-distributed and chi-squared test statistics.

Proposition 4.2 (**Neyman–Pearson Optimality of Threshold Inertia**). *Suppose Assumptions 3.1, 3.4, 3.5, 3.6, and 4.2 hold. Fix any $\bar{\alpha}_i \in [0, 1]$ and any dimension $i \in \{1, \ldots, k\}$. Let $\mathcal{C}(\bar{\alpha}_i)$ denote the class of all decision rules $\delta(\cdot)$ satisfying the Type I error*

constraint under $H_0 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0$*:*

$$P_{\boldsymbol{\theta}}(\lambda_i(\mathbf{X}) > 0) = \bar{\alpha}_i \quad \textit{for all } \boldsymbol{\theta} \textit{ such that } \eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0.$$

Within this class, the threshold inertia rule $\delta^*(\cdot)$ *with inertia zone* $\mathcal{I}_i^* = \{\mathbf{x} \in \mathcal{X} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i\}$ *is the uniformly most powerful test of* $H_0 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0$ *against* $H_1 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) < 0$ *at size* $\bar{\alpha}_i$*: for any other rule* $\delta'(\cdot) \in \mathcal{C}(\bar{\alpha}_i)$ *with inertia zone* $\mathcal{I}' \neq \mathcal{I}_i^*$*,*

$$P_{\boldsymbol{\theta}}^{\delta^*}(\lambda_i(\mathbf{X}) > 0) \geq P_{\boldsymbol{\theta}}^{\delta'}(\lambda_i(\mathbf{X}) > 0)$$

for all $\boldsymbol{\theta}$ *with* $\eta_i(\boldsymbol{\theta}; \mathbf{x}) < 0$*, with strict inequality for some such* $\boldsymbol{\theta}$*.*

Proof. See Appendix. □

Remark. Two features of the hypothesis testing framework deserve explicit clarification. First, although $\eta_i(\boldsymbol{\theta}; \mathbf{x}) \equiv \nabla_{\lambda_i} L(\boldsymbol{\theta}, \mathbf{a}((0, \boldsymbol{\lambda}_{-i}(\mathbf{x})); \mathbf{x}))$ involves the sample realization $\mathbf{x}$, the null hypothesis $H_0 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0$ is independent of $\mathbf{x}$ under separability (Assumption 3.1(iii)): by separability, $\eta_i(\boldsymbol{\theta}; \mathbf{x}) = \nabla_{a_i} L(\boldsymbol{\theta}, \tilde{a}_i)$, which depends only on $\boldsymbol{\theta}$ and the judgmental anchor $\tilde{a}_i$. Second, the one-sided direction of the alternative $H_1 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) < 0$ is structurally determined rather than inferred from the data. By strict convexity of L_i in a_i (Assumption 3.1(iii)) and the first-order characterization of $\hat{a}_i(\mathbf{x})$, the gradient $\nabla_{\lambda_i} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{a}((0, \boldsymbol{\lambda}_{-i}(\mathbf{x})); \mathbf{x}))$ is non-positive for all $\boldsymbol{\lambda}^* \in [0, 1]^k$, as established in the proof of Theorem 3.3. The empirically relevant alternative is therefore always in the direction toward the maximum likelihood action, independently of the sample realization. The Karlin–Rubin theorem accordingly applies without qualification.

Proposition 4.2 establishes that within the class of accept-or-reject rules controlling Type I error at $\bar{\alpha}_i$, the threshold rule is not merely adequate but uniquely optimal: any rule that conditions the inertia decision on features of the data extraneous to $\tilde{\alpha}_i(\mathbf{x})$ (producing, for instance, a non-connected inertia zone) is inadmissible within this class. In environments

where Assumption 4.2 does not hold, Assumption 4.1 is maintained directly and can be verified case by case, as illustrated by the Gaussian-quadratic environment of Section 2.

Proposition 4.2 and Proposition 2.5 together establish that the threshold inertia structure arises as the unique optimal accept-or-reject rule under two theoretically independent frameworks. The Gilboa-Schmeidler minmax criterion generates it through preference axioms over acts: the agent retains the anchor whenever no act dominates it under the worst-case prior. The Neyman-Pearson framework generates it through statistical optimality: the threshold rule is UMP among all rules controlling Type I error at $\bar{\alpha}_i$. Neither derivation depends on the other, yet both converge on the same behavioral prediction, with $\bar{\alpha}_i$ as the single identifiable parameter. This convergence suggests that $\bar{\alpha}_i$ is a theory-robust primitive: a decision-maker need not commit to either the belief-based framework of the ambiguity aversion literature or the error-control framework of frequentist statistics to operationalize it. Its behavioral content, the pre-committed probability of departing from the judgmental anchor, is invariant to the underlying theoretical motivation.

4.3 Eliciting the *Ex Ante* Significance Level

Theorem 4.1 establishes that $\bar{\boldsymbol{\alpha}}$ is the unique identifiable primitive of uncertainty aversion: it is recovered from the decision rule as the *ex ante* rejection probability (Definition 4.2). Unlike Bayesian priors, which require elicitation of beliefs over the full parameter space Θ, this primitive admits a direct behavioral interpretation. The experiment described in this section makes that interpretation operational.

Proposition 4.3 (**Loss Bound**). *Suppose Assumptions 3.1, 3.4, 3.5, and 3.6 hold, and let $\bar{\alpha}_i \in [0,1]$ be fixed. Under Assumption 4.1, the decision rule $\delta^A(\mathbf{x})$ representable as a decision with judgment $A = \{\tilde{\mathbf{a}}, \bar{\boldsymbol{\alpha}}, \boldsymbol{\alpha}(\cdot)\}$ in the sense of Theorem 4.1 incurs a higher loss than the judgmental anchor $\tilde{\mathbf{a}}$ with probability at most $\bar{\alpha}_i$:*

$$P_{\boldsymbol{\theta}}\Big(L_i(\boldsymbol{\theta}, \delta^A(\mathbf{X})) > L_i(\boldsymbol{\theta}, \tilde{\mathbf{a}})\Big) \leq \bar{\alpha}_i, \tag{23}$$

for all $\boldsymbol{\theta} \in \Theta$ *such that* $\eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0$.

Proof. See Appendix. □

Proposition 4.3 provides the decision-theoretic interpretation of $\bar{\alpha}_i$: it is the maximum probability a decision-maker tolerates of performing worse than their benchmark when that benchmark is optimal. This suggests an Ellsberg-style experiment to elicit $\bar{\alpha}_i$ directly, as illustrated in the following table.

Table 1: Elicitation of $\bar{\alpha}_i$ via Urn Composition

	Urn 1 (H_0: Anchor is optimal)	**Urn 2 (H_1: Anchor is not optimal)**
Action: Reject	Black balls (Loss) $= 100\bar{\alpha}_i$	Red balls (Gain) $\geq 100\bar{\alpha}_i$
Action: Accept	White balls (Neutral) $= 100(1 - \bar{\alpha}_i)$	White balls (Missed Gain) $\leq 100(1 - \bar{\alpha}_i)$

Decision-makers choose $\bar{\alpha}_i \in \{0.01, \ldots, 0.99, 1\}$, which fixes the probability of a loss (number of black balls) from Urn 1. By the unbiasedness property of UMP tests, this choice simultaneously guarantees a probability of gain (number of red balls) from Urn 2 of at least $\bar{\alpha}_i$. Decision-makers can also choose $\bar{\alpha}_i = 0$, corresponding to the trivial rule that always retains the anchor: the composition of both urns is known, containing exactly zero black and zero red balls. The decision-maker faces both urns simultaneously without knowing from which urn a ball will be drawn.

The setup is a variant of the classic Ellsberg (1961) experiment, featuring two layers of uncertainty. The first is the irreducible uncertainty regarding which urn is selected, corresponding to the unknown true hypothesis, determined by the state of nature. The second is the residual uncertainty regarding the composition of Urn 2, which represents the test's power against an unspecified alternative. Unlike the selection of the urn, this second source is partially determined by the decision-maker's selection of $\bar{\alpha}_i$: by adjusting their tolerance for Type I error, they endogenously determine the floor for the test's power, thereby exercising a degree of control over the ambiguity they are willing to face.

A decision-maker who sets $\bar{\alpha}_i = 0$ opts for full inertia: no black balls in Urn 1, but also no red balls in Urn 2, and the anchor is retained regardless of the evidence. A decision-maker

who sets $\bar{\alpha}_i = 1$ is uncertainty-neutral in the sense of Theorem 4.1(iii): the judgmental anchor is abandoned for any realization of the data, imposing no constraint on Type I error. Intermediate choices of $\bar{\alpha}_i$ correspond to intermediate degrees of uncertainty aversion along the continuum identified by the representation theorem.

The elicitation experiment provides the empirical discipline that identifies the degree of uncertainty aversion without committing to any specific theoretical framework.

5 Conclusion

This paper identifies the Bayesian-frequentist tension as a divergence in the assignment of judgment to the primitives of Wald's framework. While the Bayesian tradition encodes judgment as a prior distribution over statistical parameters, the frequentist tradition encodes it as a point in the action space, via a judgmental anchor and a significance level designed to limit Type I errors. We have shown that the frequentist double conditioning mechanism subsumes the Bayesian approach within the class of interpolating decision rules. Under this unified structure, the Bayesian posterior emerges as the boundary case of zero structural inertia, where the judgmental anchor is abandoned in favor of data-driven actions for any empirical evidence.

The central finding is that the *ex ante* significance level is a theory-robust behavioral primitive that indexes a continuum of uncertainty attitudes. This primitive places the Bayesian, frequentist, and robust Bayesian paradigms along a single axis, from the uncertainty-neutral Bayesian ($\bar{\alpha} = 1$) to the full-inertia decision-maker ($\bar{\alpha} = 0$). The robustness of this primitive is evidenced by the convergence of two independent traditions: the Neyman-Pearson framework identifies the threshold rule as uniquely optimal under the monotone likelihood ratio condition, while the Gilboa-Schmeidler minmax framework generates the same behavioral signature from preference axioms. This convergence suggests that the significance threshold is a fundamental feature of optimal decision-making under uncertainty, regardless of the

underlying theoretical motivation.

This framework provides a decision-theoretic foundation for the standard significance thresholds $(1\%, 5\%, 10\%)$ common in empirical research. Rather than a mere statistical convention, the *ex ante* significance level can be identified from observed behavior as the *ex ante* rejection probability, providing an empirically measurable index of ambiguity aversion or institutional inertia. Because it caps the probability of performing worse than a judgmental decision when that decision is optimal, the threshold is an elicitable index of uncertainty aversion. By providing a unified language to reconcile subjective judgment with objective evidence, this representation places statistical practice and decision theory on a common, theory-robust foundation.

Proofs

Proof of Proposition 2.1 (Bayesian Decision in the Simplified Environment). The posterior distribution is $\theta|x \sim \mathcal{N}(x/2, 1/2)$. The Bayesian optimal action minimizes the posterior expected loss $\mathbb{E}_{\pi(\theta|x)}[L(\theta, a)]$. Setting the first order conditions equal to zero yields:

$$\mathbb{E}_{\pi(\theta|x)}[a - \theta] = 0 \implies a = \mathbb{E}_{\pi(\theta|x)}[\theta] = x/2.$$

□

Proof of Proposition 2.2 (Decision with Judgment in the Simplified Environment). The result follows from application of Theorem 3.2 of Manganelli (2025). □

Proof of Proposition 2.3 (p-value). Let $a(\lambda; x) = \lambda x + (1 - \lambda)\tilde{a}$, for $\lambda \in [0, 1]$, so that the decision variable becomes λ. Write $\nabla_\lambda L(\theta, \lambda)$ as shorthand for $\nabla_\lambda L(\theta, a(\lambda; x))$. The first derivative of the loss function w.r.t. λ is:

$$\nabla_\lambda L(\theta, \lambda) = ((\lambda x + (1 - \lambda)\tilde{a}) - \theta)(x - \tilde{a}).$$

Optimality of $\tilde{a}$ is tested at $\lambda = 0$. When $\tilde{a} = 0$, the null hypothesis becomes:

$$H_0 : \nabla_\lambda L(\theta, 0) = -\theta x = 0.$$

To construct the test statistic, the population parameter θ in the null hypothesis must be replaced by its maximum likelihood estimator $\hat{\theta}(X) = X$. Since X is already used to denote the random variable whose realization x defines the action being tested, we introduce the auxiliary random variable $Y \sim \mathcal{N}(\theta, 1)$, independent of X but identically distributed, and replace $\hat{\theta}(X)$ with $\hat{\theta}(Y) = Y$. Under $H_0 : \theta = 0$, the test statistic is:

$$\nabla_\lambda L(\hat{\theta}(Y), 0) = -xY \sim \mathcal{N}(0, x^2). \tag{24}$$

Substituting $\hat{\theta}(y) = y = x$, the realized value is $-x \cdot x = -x^2 \leq 0$, confirming the shrinkage direction. The *p-value*, conditioning on the realized sign of the test statistic, is:

$$\begin{aligned} P_{\theta=0}(-xY < -x^2 | -xY \leq 0) &= P_{\theta=0}(-xY \leq 0 | -xY < -x^2) \frac{P_{\theta=0}(-xY < -x^2)}{P_{\theta=0}(-xY \leq 0)} \\ &= 2P_{\theta=0}(-xY < -x^2) \\ &= 2\Phi(-|x|). \end{aligned}$$

□

Proof of Proposition 2.4 (Representation in a Gaussian Quadratic Environment). We show that the *ex ante* level of significance is $\bar{\alpha}^{\pi(\theta)} = 1$. The Bayesian decision is $\delta^{\pi(\theta|x)}(x) = x/2$, while the judgmental decision associated with the prior distribution is $\delta^{\pi(\theta)} = 0$, defined as the action minimizing the prior expected loss. If $x = 0$, the two decisions coincide for any $\bar{\alpha}^{\pi(\theta)} \in [0, 1]$. The value $\bar{\alpha}^{\pi(\theta)} = 1$ is consistent with $x = 0$ even though it is not uniquely identified in this case. If $x \neq 0$, however, the Bayesian decision satisfies $\delta^{\pi(\theta|x)}(x) = x/2 \neq 0$, so observational equivalence requires that the decision with judgment always rejects the

judgmental decision $\delta^{\pi(\theta)} = \tilde{a} = 0$. This occurs only if the judgmental decision is never retained, which requires the *ex ante* level of significance $\bar{\alpha}^{\pi(\theta)} = 1$.

Given $\bar{\alpha}^{\pi(\theta)} = 1$, the decision with judgment always selects an interior shrinkage action of the form $\delta^A(x) = \lambda^* x$. The observational equivalence is obtained by equating the Bayesian decision and the decision with judgment: $\delta^{\pi(\theta|x)}(x) = \delta^A(x)$. If $x = 0$, the result is immediate. If $x \neq 0$, we get $x/2 = \lambda^* x$, which implies $\lambda^* = 1/2$.

Hence, solving for the *ex post* significance level $\alpha(x) \equiv \alpha^{\pi(\theta)}(x)$:

$$\frac{1}{2} = 1 + \frac{c_{\alpha(x)/2}}{|x|} \quad \Rightarrow \quad c_{\alpha(x)/2} = -\frac{|x|}{2}.$$

Since $c_{\alpha/2} = \Phi^{-1}(\alpha/2)$, this yields $\alpha^{\pi(\theta)}(x) = 2\Phi(-|x|/2)$. □

Proof of Proposition 2.5 (Ambiguity). An ambiguity-averse decision-maker chooses the action:

$$\arg\min_a \max_{\pi \in \Gamma} \int L(\theta, a) dF^{\pi(\theta|x)}(\theta),$$

where $F^{\pi(\theta|x)}(\theta)$ denotes the posterior distribution of a given $\pi \in \Gamma$.

For a prior $\mathcal{N}(\mu, c^{-1})$, the posterior mean of θ is $(1+c)^{-1}(x + c\mu)$. The loss becomes:

$$\max_{\mu \in [-1,1]} \left[-a \frac{x + c\mu}{1 + c} + 0.5a^2 \right].$$

Due to the linearity of the posterior mean in μ, the maximum occurs at the boundaries: if $a > 0$, the loss is maximized by $\mu = -1$, and if $a < 0$ by $\mu = 1$. Therefore, the worst-case prior mean is $\mu^* = -\operatorname{sgn}(a)$, and the objective function becomes:

$$\min_a \left[-a \frac{x - c \operatorname{sgn}(a)}{1 + c} + 0.5a^2 \right],$$

Taking the derivative with respect to a (for $a \neq 0$):

$$a - \frac{x - c\,\mathrm{sgn}(a)}{1+c} = 0 \implies a = \frac{x - c\,\mathrm{sgn}(a)}{1+c}.$$

Solving for a yields the following piecewise rule:

1. If $x > c$: The only solution is $a > 0$, giving $a = \frac{x-c}{1+c}$.

2. If $x < -c$: The only solution is $a < 0$, giving $a = \frac{x+c}{1+c}$.

3. If $|x| \leq c$: the objective is non-smooth at $a = 0$. The right derivative is $-(x-c)/(1+c) \geq 0$ and the left derivative is $-(x+c)/(1+c) \leq 0$, so $a = 0$ satisfies the subgradient optimality condition and is the unique minimizer.

The judgmental decision $\tilde{a} = 0$ is obtained by minimizing the maximum expected loss over the set of priors Γ prior to observing the sample realization. Given symmetry of the interval $\mu \in [-1, 1]$, the *ex ante* optimal action is

$$\tilde{a} = \arg\min_{a} \max_{\mu \in [-1,1]} \left[-a\mu + 0.5a^2 \right] = 0.$$

To establish the link with the decision with judgment, we equate $\delta^{\Gamma}(x)$ to $\delta^{A}(x)$:

$$\delta^{\Gamma}(x) = I(|x| > c)(x - \mathrm{sgn}(x)c)/(1+c)$$
$$\delta^{A}(x) = I(|x| > -c_{\bar{\alpha}/2})(x + \mathrm{sgn}(x)c_{\alpha(x)/2}).$$

The two decisions are equal when $c = -c_{\bar{\alpha}/2}$ and $(x - \mathrm{sgn}(x)c)/(1+c) = x + \mathrm{sgn}(x)c_{\alpha/2}$. Inverting the first condition gives the *ex ante* level of significance $\bar{\alpha}^{\Gamma}$. Solving the second condition for $\alpha^{\Gamma}(x)$ gives the *ex post* level of significance. □

Proof of Theorem 3.1 (Bayesian Decision). Strict convexity and coercivity of $L(\boldsymbol{\theta}, \cdot)$ (Assumption 3.1(iii)) are preserved under the expectation operator, so the posterior expected

loss $\rho(\mathbf{a}, \mathbf{x}) = \int_\Theta L(\boldsymbol{\theta}, \mathbf{a})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$ is strictly convex and coercive in $\mathbf{a}$. By Assumption 3.3(ii), the domination condition permits interchange of gradient and integral by the dominated convergence theorem, yielding $\nabla_{\mathbf{a}}\rho(\mathbf{a}, \mathbf{x}) = \int_\Theta \nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \mathbf{a})\pi(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}$. The unique minimizer $\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x}) = \arg\min_{\mathbf{a}} \rho(\mathbf{a}, \mathbf{x})$ is therefore characterized by $\nabla_{\mathbf{a}}\rho(\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x}), \mathbf{x}) = \mathbf{0}$. □

Proof of Theorem 3.2 (Existence and Uniqueness of the Multivariate Decision with Judgment). Consider the auxiliary objective function

$$J(\boldsymbol{\lambda}) = L(\hat{\boldsymbol{\theta}}(\mathbf{x}),\, \mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})) - \sum_{i=1}^{k} \lambda_i\, q_{i,\alpha_i/2}\, \sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})}.$$

By Assumption 3.1(iii), L is strictly convex in $\mathbf{a}$. Since $\hat{\mathbf{a}}(\mathbf{x}) \neq \tilde{\mathbf{a}}$ component-wise, $\mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})$ is an affine transformation of $\boldsymbol{\lambda}$ and by Assumption 3.6(iii) $q_{i,\alpha_i/2}\sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})}$ does not depend on $\boldsymbol{\lambda}$, $J(\boldsymbol{\lambda})$ is continuous and strictly convex on the compact convex set $[0, 1]^k$.

Since J is continuous on the compact set $[0, 1]^k$, a global minimizer $\hat{\boldsymbol{\lambda}}$ exists by the extreme value theorem. Since J is strictly convex, this minimizer is unique.

The gradient of J with respect to λ_i is obtained by the chain rule:

$$\begin{aligned}\frac{\partial J}{\partial \lambda_i} &= \nabla_{a_i} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})) \cdot (\hat{a}_i(\mathbf{x}) - \tilde{a}_i) - q_{i,\alpha_i/2}\sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})} \\ &= \left(\hat{z}_i((\lambda_i, \boldsymbol{\lambda}_{-i}); \mathbf{x}) - q_{i,\alpha_i/2}\right) \sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})},\end{aligned}$$

where the second equality uses the definition of $\hat{z}_i$ in (16). By strict convexity of L in a_i (Assumption 3.1(iii)), $\partial^2 L/\partial a_i^2 > 0$, and since $\hat{a}_i(\mathbf{x}) \neq \tilde{a}_i$ by assumption, $(\hat{a}_i(\mathbf{x}) - \tilde{a}_i)^2 > 0$. Therefore:

$$\frac{\partial \hat{z}_i}{\partial \lambda_i} = \frac{\partial^2 L}{\partial a_i^2} \cdot \frac{(\hat{a}_i(\mathbf{x}) - \tilde{a}_i)^2}{\sqrt{\Omega_{ii}(\hat{\boldsymbol{\theta}}(\mathbf{x}), \boldsymbol{\lambda})}} > 0,$$

so $\hat{z}_i((\lambda_i, \boldsymbol{\lambda}_{-i}); \mathbf{x})$ is strictly increasing in λ_i, and consequently $\partial J/\partial \lambda_i$ is strictly increasing in λ_i.

The Karush-Kuhn-Tucker conditions for the box-constrained problem $\boldsymbol{\lambda} \in [0, 1]^k$ yield three cases for each i:

1. *Lower bound active*: $\hat{\lambda}_i = 0$, when $\partial J/\partial \lambda_i\big|_{\lambda_i=0} \geq 0$, i.e., $\hat{z}_i((0, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) \geq q_{i,\alpha_i/2}$.

2. *Interior solution*: $\hat{\lambda}_i = \lambda_i^* \in (0,1)$, solving $\hat{z}_i((\lambda_i^*, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) = q_{i,\alpha_i/2}$.

3. *Upper bound active*: $\hat{\lambda}_i = 1$, when $\partial J/\partial \lambda_i\big|_{\lambda_i=1} \leq 0$, i.e., $\hat{z}_i((1, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) \leq q_{i,\alpha_i/2}$. This case splits into two sub-cases. If $\hat{z}_i((1, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) = q_{i,\alpha_i/2}$, then $\lambda_i^* = 1$ satisfies the boundary condition and this case is subsumed into case 2. If $\hat{z}_i((1, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) < q_{i,\alpha_i/2}$, then since $\hat{z}_i$ is strictly increasing in λ_i, the gradient $\partial J/\partial \lambda_i$ is strictly negative throughout $[0,1]$, so J is strictly decreasing in λ_i over the entire feasible range and the constrained minimum is at $\hat{\lambda}_i = \lambda_i^* = 1$, consistently with the definition of λ_i^*.

Cases 2 and 3 are therefore unified as $\hat{\lambda}_i = \lambda_i^* \in (0,1]$, where λ_i^* either solves the boundary condition $\hat{z}_i((\lambda_i^*, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) = q_{i,\alpha_i/2}$ or equals 1 when no such solution exists in $(0,1]$. These cases coincide exactly with the system of conditions in (17).

The uniqueness of $\delta^A(\mathbf{x})$ follows directly from the functional form of $\mathbf{a}(\boldsymbol{\lambda}; \mathbf{x})$. □

Proof of Theorem 3.3 (Existence of Equivalent Significance Levels). Fix an arbitrary shrinkage vector $\boldsymbol{\lambda}^* \in [0,1]^k$ and a data realization $\mathbf{x}$. To show that $\boldsymbol{\lambda}^*$ can be rationalized as a decision with judgment, we must find a vector $\boldsymbol{\alpha}$ such that the system of equations in (17) is satisfied at $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^*$.

By the absolute continuity of the estimator's distribution (Assumption 3.5) and the non-degeneracy of the variance (Assumption 3.6(i)), the cdf F_i of $Z_i(\hat{\boldsymbol{\theta}}(\mathbf{Y}), \boldsymbol{\lambda}; \mathbf{x})$ is continuous and strictly increasing. Define $\alpha_i(\mathbf{x}) = 2F_i(\hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x}))$, so that $q_{i,\alpha_i(\mathbf{x})/2} = \hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x})$ by construction. When $\lambda_i^* = 0$, this reduces to $\alpha_i(\mathbf{x}) = \tilde{\alpha}_i(\boldsymbol{\lambda}^*_{-i}; \mathbf{x})$ by Definition 3.3.

Since $a_i(\lambda_i^*; \mathbf{x})$ lies between $\tilde{a}_i$ and $\hat{a}_i(\mathbf{x})$ for $\lambda_i^* \in [0,1]$, and $\nabla_{a_i} L$ is strictly increasing in a_i by strict convexity with $\nabla_{a_i} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \hat{\mathbf{a}}(\mathbf{x})) = 0$, the factor $\nabla_{a_i} L(\hat{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{a}(\boldsymbol{\lambda}^*; \mathbf{x}))(\hat{a}_i(\mathbf{x}) - \tilde{a}_i)$ is non-positive, giving $\hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x}) \leq 0$ for all $\boldsymbol{\lambda}^* \in [0,1]^k$. Therefore $F_i(\hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x})) \leq F_i(0) = 1/2$ by Assumption 3.6(iii), giving $\alpha_i(\mathbf{x}) = 2F_i(\hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x})) \in [0,1]$ for all i, confirming $\boldsymbol{\alpha}(\mathbf{x}) \in [0,1]^k$.

When $\lambda_i^* \in (0,1]$, the constructed $\alpha_i(\mathbf{x})$ satisfies $\hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x}) = q_{i,\alpha_i(\mathbf{x})/2}$ by construction, so the boundary condition in (17) is satisfied at $\hat{\lambda}_i = \lambda_i^*$, covering both the interior case

$\lambda_i^* \in (0,1)$ and the upper bound case $\lambda_i^* = 1$. Uniqueness follows from the strict monotonicity of F_i: the boundary condition $q_{i,\alpha_i(\mathbf{x})/2} = \hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x})$ uniquely determines $\alpha_i(\mathbf{x}) = 2F_i(\hat{z}_i(\boldsymbol{\lambda}^*; \mathbf{x}))$.

When $\lambda_i^* = 0$, the "otherwise" branch of (17) requires $\hat{z}_i((0, \hat{\boldsymbol{\lambda}}_{-i}); \mathbf{x}) \geq q_{i,\alpha_i(\mathbf{x})/2}$. Since $\hat{\boldsymbol{\lambda}}_{-i} = \boldsymbol{\lambda}^*_{-i}$ at the solution and $q_{i,\alpha_i(\mathbf{x})/2} = \hat{z}_i((0, \boldsymbol{\lambda}^*_{-i}); \mathbf{x})$ by construction, both sides are equal and the condition holds trivially. When $\lambda_i^* = 0$, the decision $\hat{\lambda}_i = 0$ is rationalized by any $\alpha_i(\mathbf{x}) \leq \tilde{\alpha}_i(\boldsymbol{\lambda}^*_{-i}; \mathbf{x})$; the canonical choice $\alpha_i(\mathbf{x}) = \tilde{\alpha}_i(\boldsymbol{\lambda}^*_{-i}; \mathbf{x})$ is adopted by convention. □

Proof of Theorem 3.4 (Frequentist Representation of Bayesian Decisions). Fix a data realization $\mathbf{x}$. Under the stated assumptions, the Bayesian action $\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x})$ is uniquely defined. By Assumption 3.7, there exists a shrinkage vector $\boldsymbol{\lambda}^B \in [0,1]^k$ such that $\delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x}) = \mathbf{a}(\boldsymbol{\lambda}^B; \mathbf{x})$, where the path is anchored at the prior action $\delta^{\pi(\boldsymbol{\theta})}$.

To establish observational equivalence, first set the judgmental decision $\tilde{\mathbf{a}} = \delta^{\pi(\boldsymbol{\theta})}$. Second, set the *ex ante* significance levels $\bar{\alpha}_i = 1$ for all i, so that the corresponding quantile threshold is $q_{i,\bar{\alpha}_i/2} = q_{i,1/2} = 0$ by Assumption 3.6(iv). Since $\hat{\mathbf{a}}(\mathbf{x}) \neq \tilde{\mathbf{a}}$ component-wise by assumption, the realized standardized gradient satisfies $\hat{z}_i(\mathbf{0};\, \mathbf{x}) < 0 = q_{i,1/2}$ for all i, as established in the proof of Theorem 3.3. The lower bound $\hat{\lambda}_i = 0$ is therefore ruled out by the KKT conditions, so the unique solution satisfies $\hat{\boldsymbol{\lambda}} \in (0,1]^k$ and can be identified with $\boldsymbol{\lambda}^B$. By Theorem 3.3, there exists a unique vector of *ex post* significance levels $\boldsymbol{\alpha}(\mathbf{x})$ such that the boundary conditions $\hat{z}_i(\boldsymbol{\lambda}^B;\, \mathbf{x}) = q_{i,\alpha_i(\mathbf{x})/2}$ hold for all $i = 1, \ldots, k$. Substituting this $\boldsymbol{\alpha}(\mathbf{x})$ into the decision rule (17) yields $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^B$. Consequently, $\delta^A(\mathbf{x}) = \mathbf{a}(\boldsymbol{\lambda}^B; \mathbf{x}) = \delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x})$, establishing observational equivalence. □

Proof of Theorem 4.1 (Representation of Decision Rules Along the Uncertainty Continuum). The proof proceeds in three steps.

Step 1: Point-wise representation. Fix any realization $\mathbf{x} \in \mathcal{X}$. By Assumption 3.7, the induced shrinkage vector $\boldsymbol{\lambda}(\mathbf{x}) \in [0,1]^k$ is well-defined. By Theorem 3.3, there exists a vector

of *ex post* significance levels $\boldsymbol{\alpha}(\mathbf{x}) \in [0,1]^k$ rationalizing $\boldsymbol{\lambda}(\mathbf{x})$, with canonical choice:

$$\alpha_i(\mathbf{x}) = \begin{cases} 2F_i(\hat{z}_i(\boldsymbol{\lambda}(\mathbf{x});\, \mathbf{x})) & \text{if } \lambda_i(\mathbf{x}) > 0, \\ \tilde{\alpha}_i(\mathbf{x}) & \text{if } \lambda_i(\mathbf{x}) = 0, \end{cases}$$

where $\tilde{\alpha}_i(\mathbf{x})$ is the equilibrium profile *p-value* of Definition 4.1. When $\lambda_i(\mathbf{x}) > 0$, strict monotonicity of $\hat{z}_i$ in λ_i (Theorem 3.2) and strict monotonicity of F_i (Assumption 3.6(i)) give:

$$\alpha_i(\mathbf{x}) = 2F_i(\hat{z}_i(\boldsymbol{\lambda}(\mathbf{x}); \mathbf{x})) > 2F_i(\hat{z}_i((0, \boldsymbol{\lambda}_{-i}(\mathbf{x})); \mathbf{x})) = \tilde{\alpha}_i(\mathbf{x}),$$

so $\alpha_i(\mathbf{x}) > \tilde{\alpha}_i(\mathbf{x})$ in the rejection zone, and $\alpha_i(\mathbf{x}) = \tilde{\alpha}_i(\mathbf{x})$ in the inertia zone by convention. Since $\mathbf{x}$ was arbitrary, this establishes the point-wise representation in part (i).

Step 2: Ex ante *characterization.* By Assumption 4.1, the rejection zone satisfies:

$$\{\lambda_i(\mathbf{x}) > 0\} = \{\tilde{\alpha}_i(\mathbf{x}) < \bar{\alpha}_i\}.$$

By the piecewise definition of part (i), $\{\lambda_i(\mathbf{x}) > 0\} = \{\alpha_i(\mathbf{x}) > \tilde{\alpha}_i(\mathbf{x})\}$ by construction.

By separability (Assumption 3.1(iii)), $\nabla_{\lambda_i} L$ does not depend on $\boldsymbol{\lambda}_{-i}$, so the equilibrium profile *p-value* $\tilde{\alpha}_i(\mathbf{X})$ coincides with the marginal *p-value*, which is uniform on $[0,1]$ under $H_0 : \boldsymbol{\theta} = \tilde{\mathbf{a}}$ by the Probability Integral Transform. This implies that $P_{\tilde{\mathbf{a}}}(\tilde{\alpha}_i(\mathbf{X}) < \bar{\alpha}_i) = \bar{\alpha}_i$. Therefore:

$$P_{\tilde{\mathbf{a}}}(\lambda_i(\mathbf{X}) > 0) = P_{\tilde{\mathbf{a}}}(\tilde{\alpha}_i(\mathbf{X}) < \bar{\alpha}_i) = \bar{\alpha}_i = P_{\tilde{\mathbf{a}}}(\tilde{\alpha}_i(\mathbf{X}) < \alpha_i(\mathbf{X}))\,,$$

where the last equality uses the equivalence $\{\alpha_i(\mathbf{x}) > \tilde{\alpha}_i(\mathbf{x})\} = \{\tilde{\alpha}_i(\mathbf{x}) < \bar{\alpha}_i\}$ established in this step. This establishes part (ii).

Step 3: Boundary cases. For part (iii): if $\delta(\cdot)$ is observationally equivalent to a Bayesian decision, then $\bar{\alpha}_i = 1$ for all i by Theorem 3.4, since the Bayesian decision abandons the judgmental anchor for every data realization with probability 1. The conditional converse follows directly from the same theorem: if $\bar{\alpha}_i = 1$ for all i and $\delta(\cdot)$ is rationalized by posterior

expected loss minimization under a proper prior $\pi(\boldsymbol{\theta})$, then the conditions of Theorem 3.4 are satisfied and $\delta(\cdot) = \delta^{\pi(\boldsymbol{\theta}|\mathbf{x})}(\mathbf{x})$.

For part (iv): $\bar{\alpha}_i < 1$ implies, by Assumption 4.1, that the inertia zone $\{\mathbf{x} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i\}$ has strictly positive probability $1 - \bar{\alpha}_i > 0$ under H_0. Since the inertia zone is the level set $\{\tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i\}$, it is non-decreasing in the set-inclusion sense as $\bar{\alpha}_i$ decreases: for $\bar{\alpha}_i^{(1)} \leq \bar{\alpha}_i^{(2)}$,

$$\left\{\mathbf{x} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i^{(2)}\right\} \subseteq \left\{\mathbf{x} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i^{(1)}\right\},$$

so that a smaller $\bar{\alpha}_i$ corresponds to a larger inertia zone and a greater reluctance to depart from the judgmental anchor. □

Proof of Proposition 4.2 (Neyman–Pearson Optimality of Threshold Inertia). By separability (Assumption 3.1(iii)), $\nabla_{\lambda_i} L$ does not depend on $\boldsymbol{\lambda}_{-i}$, the equilibrium profile *p-value* $\tilde{\alpha}_i(\mathbf{X})$ coincides with the marginal *p-value*, and therefore by the Probability Integral Transform, $\tilde{\alpha}_i(\mathbf{X})$ is uniformly distributed on $[0, 1]$ under $H_0 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0$. The threshold rule therefore satisfies the Type I error constraint with equality: $P_{\boldsymbol{\theta}}(\tilde{\alpha}_i(\mathbf{X}) < \bar{\alpha}_i) = \bar{\alpha}_i$ under H_0, so it belongs to $\mathcal{C}(\bar{\alpha}_i)$.

Under Assumption 4.2, the standardized gradient Z_i has the MLR property with respect to $\eta_i(\boldsymbol{\theta}; \mathbf{x})$. For the one-sided hypotheses $H_0 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0$ against $H_1 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) < 0$, the MLR property implies that the distribution of Z_i stochastically decreases as η_i decreases below zero. By the Karlin–Rubin theorem (Lehmann and Romano, 2005, Theorem 3.4.1), the test that rejects H_0 when $Z_i < c$, equivalently when $\tilde{\alpha}_i(\mathbf{x}) < \bar{\alpha}_i$, is uniformly most powerful of size $\bar{\alpha}_i$ for all $\boldsymbol{\theta}$ such that $\eta_i(\boldsymbol{\theta}; \mathbf{x}) < 0$. Since the inertia zone of this rule is $\mathcal{I}_i^* = \{\mathbf{x} : \tilde{\alpha}_i(\mathbf{x}) \geq \bar{\alpha}_i\}$, any other rule $\delta' \in \mathcal{C}(\bar{\alpha}_i)$ with $\mathcal{I}' \neq \mathcal{I}_i^*$ has strictly lower power at some $\boldsymbol{\theta}$ with $\eta_i(\boldsymbol{\theta}; \mathbf{x}) < 0$. □

Proof of Proposition 4.3 (Loss Bound). By Assumption 4.1, the decision rule departs from the anchor $\tilde{\mathbf{a}}$ if and only if $\tilde{\alpha}_i(\mathbf{x}) < \bar{\alpha}_i$. Under $H_0 : \eta_i(\boldsymbol{\theta}; \mathbf{x}) = 0$, by separability (Assumption 3.1(iii)), $\nabla_{\lambda_i} L$ does not depend on $\boldsymbol{\lambda}_{-i}$, so the equilibrium profile p-value $\tilde{\alpha}_i(\mathbf{X})$ coincides

with the marginal p-value and is uniformly distributed on $[0,1]$ by the Probability Integral Transform (Assumptions 3.5 and 3.6(i)); consequently, $P_{\boldsymbol{\theta}}(\tilde{\alpha}_i(\mathbf{X}) < \bar{\alpha}_i) = \bar{\alpha}_i$. Under H_0, the anchor $\tilde{\mathbf{a}}$ satisfies $\nabla_{\mathbf{a}} L(\boldsymbol{\theta}, \tilde{\mathbf{a}}) = \mathbf{0}$, so by strict convexity and separability (Assumption 3.1(iii)), $L_i(\boldsymbol{\theta}, \mathbf{a}) > L_i(\boldsymbol{\theta}, \tilde{\mathbf{a}})$ for all i and all $\mathbf{a} \neq \tilde{\mathbf{a}}$. □

References

Bewley, T. (2011). Knightian decision theory and econometric inferences. *Journal of Economic Theory 146*, 1134–1147.

Brown, L. D. (1981). A complete class theorem for statistical problems with finite sample spaces. *Annals of Statistics 9*(6), 1289–1300.

Efron, B. and C. Morris (1973). Stein's estimation rule and its competitors — an empirical Bayes approach. *Journal of the American Statistical Association 68*(341), 117–130.

Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *The Quarterly Journal of Economics 75*(4), 643–669.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*(2), 141–153.

Hansen, L. P. and T. J. Sargent (2024). Risk, ambiguity, and misspecification: Decision theory, robust control, and statistics. *Journal of Applied Econometrics 39*(6), 969–999.

Ilut, C. and M. Schneider (2023). Chapter 24 - Ambiguity. In R. Bachmann, G. Topa, and W. van der Klaauw (Eds.), *Handbook of Economic Expectations*, pp. 749–777. Academic Press.

James, W. and C. Stein (1961). Estimation with quadratic loss. In J. Neyman (Ed.), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, pp. 361–379. Berkeley: University of California Press.

Lehmann, E. and J. Romano (2005). *Testing Statistical Hypotheses*. Springer.

Manganelli, S. (2025). Statistical decision functions with judgment. *Journal of Economic Theory 223*, 105940.

Manski, C. F. (2021). Econometrics for decision making: Building foundations sketched by Haavelmo and Wald. *Econometrica 89*(6), 2827–2853.

Neyman, J. and E. S. Pearson (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society A 231*(694–706), 289–337.

Savage, L. (1954). *The Foundations of Statistics*. New York, John Wiley & Sons.

Wald, A. (1950). *Statistical Decision Functions*. New York, John Wiley & Sons.