

From Epistemic Justification to Methodological Decision: The Epistemicization of Hypotheses and the Status of H_0-H_1 in Scientific Research

Bülent Bayat

bbayat@nny.edu.tr

Nuh Naci Yazgan University

Systematic Review

Keywords: Neyman–Pearson decision theory, H_0-H_1 hypotheses, verification and falsification, Faulty use of hypotheses, Philosophy of science, Statistical methodology, epistemology–methodology distinction

Posted Date: October 26th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10632801/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

This article aims to examine the epistemological and methodological status of Neyman–Pearson decision theory by referencing the philosophy of science debates of the nineteenth and twentieth centuries and through the lens of what it evaluates as an “epistemic and methodological gap”, namely, the “research programs–hypotheses” relationship. The main argument of our article is that the epistemological approaches along the “verification–confirmation–falsification” axis, extending from Comte to Lakatos, while focusing on the problem of rational justification of scientific knowledge, have failed to directly address the increasingly evident methodological and functional needs of scientific research practices. This situation has created a distinct methodological/pragmatic gap, particularly in the processes related to the formulation, use, and testing of hypotheses.

Our study argues that this gap has been effectively filled by statistical approaches that began with Fisher and gained a systematic structure with Neyman–Pearson decision theory. Within the Neyman–Pearson framework, the H_0 and H_1 hypotheses are not constructed as symmetric propositions (each other's inverse/opposite) that represent epistemic truth claims, but rather as normative instruments that regulate rational decision-making under epistemic uncertainty in research. Concepts such as “Type I–Type II error, alpha–beta levels, and test power” in Neyman–Pearson theory are not criteria for epistemic justification; they are methodological parameters that ensure the long-term predictive performance and consistency of decision processes.

The article also interprets the relationship between verification and falsification approaches and Neyman–Pearson decision theory not as a direct theoretical continuity, but as a historical and conceptual background connection. In this context, it is argued that reading statistical tests as applications of epistemological theories is misleading and that interpreting the H_0 and H_1 structure in current research practices from an epistemic perspective lead to a systematic category error. Thus, the article addresses the interpretation of H_0 and H_1 hypothesis formulations from an epistemic perspective in current practice and the resulting faulty formulations.

In conclusion, the study proposes that a distinction between epistemology and methodology should be based not on hierarchy (order of priority) but on problem levels. It demonstrates that understanding Neyman–Pearson decision theory not as an epistemic foundation of scientific rationality but as a normative framework that regulates methodological operations provides a more consistent ground for evaluation, both from the perspective of the philosophy of science and statistical research practice.

Regarding the subject it addresses (the epistemic and methodological status of hypotheses), the article argues for a non-hierarchical, “problem-based” distinction between epistemology and methodology. Therefore, it considers it necessary to address the topic within a broad historical–epistemological context extending from early positivist epistemological approaches to Lakatos.

Introduction

One of the fundamental debates in the philosophy of science, shaped from the nineteenth century onward, has concentrated on the question of how the claims of universal validity (generalization) of scientific knowledge can be justified. Particularly within the positivist tradition, these debates have been formed by the quest for an epistemic ground that would secure the rational status of science, with the concepts of verification, confirmation, and falsification becoming the central tools of this quest. Nevertheless, despite the intensive theoretical debates carried out around these concepts throughout the twentieth century, the problem of the justification of scientific knowledge has continued to generate unresolved difficulties on both the philosophical and practical levels.

The verification argument of logical positivism has been subjected to serious epistemic criticisms due to the structural difficulties concerning the problem of induction (both Hempel and Popper agree on the impossibility of induction and verification), and these criticisms have made evident not only the limits of a particular philosophical approach but also a more general problem regarding the grounding of scientific knowledge. In response to this intractability, alternative epistemological approaches (confirmation theory, paradigm and research program models) have aimed not to entirely abandon the ideal of epistemic justification but to reformulate it in different ways or to stabilize it (dialectical stabilization). However, most of these attempts, while focusing heavily on the meaning and rational status of scientific knowledge, have not produced a direct response to the increasingly pronounced methodological needs of scientific practice.

Particularly, the processes of hypothesis formation, testing, and decision-making, which have become determinant in the experimental sciences and statistical research practices, have not fully coincided with the conceptual frameworks offered by epistemological discussions. While the debates on verification or falsification interrogated the epistemic status of hypotheses, research practice has increasingly been confronted with problems such as error, risk, sampling uncertainty, and the necessity of decision-making. Precisely at this point, the tension created by this uncertainty points to the widening distance between the theoretical level of the philosophy of science and the operational needs of statistical methods. This distance (or, in other words, the methodological gap not directly addressed by epistemological debates) has been *de facto* filled by statistical approaches that began with Fisher and were systematized with the Neyman–Pearson decision theory.

The Neyman–Pearson approach, whether intentionally or not, has not so much offered an epistemic solution to the "verification–falsification" problem but has rather suspended the uncertainty this problem creates in scientific practice (research programs) by establishing a normative and procedural structure. Today, the logical ground and manner of use of the H_0 and H_1 hypothesis structure, which is widely employed in research programs, has been largely shaped by the framework of this decision theory.

In this context, our article, on the one hand, foregrounds the historical and conceptual importance of the epistemological debates along the verification–falsification axis, while on the other hand, it seeks to understand why these debates have refrained from producing lasting solutions to the problems concerning "the operational use of hypotheses in research programs."

More importantly, it is observed that there is a widespread conceptual ambiguity in current research practice regarding the status and function of hypotheses. Meanings that have no counterpart either in the classical epistemological debates or in the methodological framework of statistical decision theory are attributed to hypotheses, and accordingly, H_0 and H_1 hypotheses are frequently equated with epistemic claims of truth, research hypotheses, or theoretical propositions. These modes of interpretation are repeated in textbooks, methodological sources, and empirical research articles indexed in international databases, and over time, they turn into established research practices. Although some methodology authors in recent years have pointed out the problems caused by these practices and have proposed alternative reporting or inferential approaches, they have not provided a systematic explanation concerning the epistemic status of hypotheses and hypothesis tests. Therefore, the present study aims not only to discuss the methodological problems in practice but also to reveal the epistemological ambiguity lying behind these problems and to reassess the actual function of H_0 and H_1 hypotheses in research programs.

The central thesis of this article is that the widespread epistemic interpretation of H_0 and H_1 represents a category error arising from the conflation of epistemological and methodological levels.

From Textbook Narratives to Research Practices: Illustrations of the Epistemological Confusion in the Status of Hypotheses

- **A Representative Example of Textbook Narrative: The Epistemic Reading of Hypotheses (Encyclopedia of Social Measurement, 2005)**

The entry titled "Hypothesis Tests and Evidence" in the *Encyclopedia of Social Measurement* (2005, pp. 265–270), published by ScienceDirect/Elsevier, summarizes the classic textbook-level presentation of NHST. This presentation serves as a representative example, as it clearly reflects certain assumptions that are increasingly questioned in the contemporary methodology literature and that constitute the very subject of discussion in our article. The encyclopedia entry reflects a common epistemological approach/understanding on four fundamental points;

First, the entry defines H_0 (null hypothesis) as "the hypothesis whose refutation by statistical means is taken as evidence supporting the alternative hypothesis"; as is evident, this is a language of epistemic acceptance. This interpretation construes the rejection of H_0 as evidence in favour of the alternative hypothesis, thereby establishing a direct epistemic relationship between the rejection of H_0 and the support of H_1 . However, in the Neyman–Pearson decision theory, the rejection of H_0 does not imply that H_1 is verified or epistemically supported. The decision rendered is merely a consequence of a behavioural rule to be applied under predetermined error probabilities. Thus, the rejection of H_0 and the epistemic support of H_1 are not the same conceptual operation.

Second, in the glossary section of the same encyclopedia entry, the alternative hypothesis is defined as: "The belief the investigator holds." This definition equates the alternative hypothesis with the investigator's pre-existing scientific belief. Yet, particularly in the Neyman–Pearson approach, H_1 does

not represent the investigator's psychological belief; it merely denotes the alternative decision state constructed solely for the purpose of defining error probabilities and operating the decision procedure. Therefore, the use of the term "belief" leads to a blurring of the conceptual boundary between scientific expectation and statistical hypothesis.

Third, the introductory statement in the encyclopedia entry, "A hypothesis about an underlying population is a statement that may be true or false," presents statistical hypotheses as if they were scientific propositions carrying truth values. Undoubtedly, in a mathematical sense, H_0 is either true or false; however, the aim of the Neyman–Pearson system is not to determine whether H_0 or H_1 is true. The purpose of this approach is to establish a decision procedure that regulates which decision will be made under predetermined Type I and Type II error probabilities. Consequently, the language of truth and the language of decision should not be evaluated on the same epistemological plane. The truth of scientific hypotheses and the function of statistical hypotheses in the decision process belong to distinct conceptual categories.

Finally, the purpose of NHST is stated in the entry as follows: "NHST allows the investigator to test the probability of a null hypothesis against a research or alternative hypothesis." This statement mis defines the probability that NHST computes. Classical NHST does not test the probability that H_0 is true, nor does it calculate the probability that H_1 is true. The quantity computed is the probability of obtaining the observed or more extreme data, given that H_0 is assumed to be true [$P(\text{Data}|H_0)$]. In other words, NHST is concerned not with $P(H_0|\text{Data})$ or $P(H_1|\text{Data})$, but only with $P(\text{Data}|H_0)$.

Hence, the p-value provides information not about the probability of hypotheses, but about the extent to which the observed data are expectable under H_0 .

These four examples should not be regarded as isolated terminological problems. On the contrary, they collectively represent different manifestations of a common epistemological assumption. This shared assumption is the notion that statistical hypotheses possess the same epistemic status as scientific claims, and that NHST produces direct epistemic evidence about scientific theories. The central claim of our article is precisely the opposite: in the Neyman–Pearson system, H_0 and H_1 are not scientific propositions carrying truth claims, but rather methodological decision tools that regulate behaviour under error rates. Therefore, the results produced by NHST should be interpreted not as epistemic verification or falsification, but as the application of pre-defined decision rules.

- **The Ritualization of Hypothesis Labels**

The transition from textbooks to empirical research reveals how this pedagogical convention has become ritualized. Contemporary articles routinely formulate multiple hypotheses labelled H_1 , H_2 , H_3 , H_4 , and beyond. Each label is presented as though it represents an independent statistical hypothesis to be accepted or rejected.

The article titled "The influence mechanism analysis on the farmers' intention to adopt Internet of Things based on UTAUT TOE model," published on the www.nature.com/scientificreports webpage on July 1, 2024 (an applied structural equation modelling/SEM study), formulated a series of hypotheses numbered sequentially from H_1 to H_{14} and tested these hypotheses. Thus, the systematic formulation of hypotheses numbered from H_1 to H_{14} constitutes a typical example of how the epistemological confusion discussed in this article is reflected in research practice.

Moreover, in the Neyman–Pearson decision theory, there is no conceptual structure involving independent statistical hypotheses such as H_2 , H_3 , H_4 ...; consequently, labelling hypotheses in the form of H_1 – H_{14} does not reflect the original logic of hypothesis testing within the Neyman–Pearson decision theory. Furthermore, the results table of this article also employs epistemic evidence language, such as "supported–not supported." Both examples represent the problematic issues we criticize in our article, since the reader may draw an epistemic inference that H_{11} is verified or that H_3 is falsified. Hence, in the article, structural equation modelling/SEM is used as an epistemic decision mechanism, applying Neyman–Pearson logic to SEM and employing a "statistics–decision–interpretation" mechanism. However, as we contend in our article, equating "decision" with "scientific verification" is a problematic assumption.

- **Contemporary Methodological Reform Initiatives**

The article titled "When Null Hypothesis Significance Testing Is Unsuitable for Research; A Reassessment," published on August 3, 2017 (Denes Szucs and John P. A. Ioannidis), argues that NHST should cease to be the default method in scientific research.

This article, which has significant influence in the contemporary methodology literature, puts forward the following claims; NHST is one of the important causes of the replication crisis and should no longer be the default method, different inference methods should be used for different research questions, and power analysis, effect size, preregistration, and open data should become mandatory.

These recommendations can be considered important methodological reforms aimed at improving the quality of statistical inference. However, the article does not address the question of in what sense scientific hypotheses are verified, supported, or rejected. Consequently, the gap between methodological reform proposals and epistemological interpretation persists. This article by Szucs and Ioannidis makes the following significant determinations;

- Fisher and Neyman–Pearson are not the same method.
- Neyman–Pearson is a long-term behavioural decision procedure.
- NHST does not calculate the probability that H_0 or H_1 is true.
- The inference " H_0 is rejected and therefore H_1 is true" is logically invalid.

The authors criticize strongly the misapplication of NHST, conceptual confusions, and the misinterpretation of the p-value. However, as a solution, they again propose better statistical inference (Bayesian approaches, prior probabilities, FPR/TPR, power analysis, etc.). Nevertheless, these proposed reforms still address the evaluation of scientific hypotheses within the framework of statistical inference, thus leaving open the question of how the epistemic status of scientific hypotheses can be grounded independently of statistical methods. At the same time, the following question remains in the void; Is determining the epistemic status of scientific hypotheses truly the task of statistical procedures?

Another illustrative example is the article titled "Hypothesis Testing Requires Reliable Data: A Simulation Approach to Inferential Statistics and the Defense of a Research Program Strategy" (Antonia Krefeld-Schwalb, Erich H. Witte, and Frank Zenker), published in *Frontiers in Psychology/Quantitative Psychology and Measurement* on April 24, 2018.

The most important claims of the article are, in summary; NHST is not actually hypothesis testing, it is only concerned with $p(D|H_0)$. In contrast, genuine hypothesis testing should evaluate $L(H|D)$, that is, the extent to which the hypothesis is supported by the data. In summary, the article argues that since NHST is only a data probability and not genuine hypothesis testing, we should develop a "Research Program Strategy (RPS)." The aim of the proposed RPS is explicitly stated as "statistical H_1 verification." However, as we will address in our article, the foundational literature on hypotheses, namely the Neyman–Pearson decision theory, adopts an approach that is distant from the language and claim of "verification." In the article by Krefeld-Schwalb, Witte, and Zenker, this language is maintained, and terms such as *preliminary verification*, *substantial verification*, and *H_1 verification* are frequently used; indeed, the final aim of the six-stage RPS is explicitly defined as "substantial H_1 verification." This verification language places an unnecessary epistemic burden on "statistical processes." While the foundational theory (Neyman–Pearson) does not make any claim of verification, the reform literature reproduces the terminology of verification.

- **Widespread Assumptions and Usage Patterns**

Whether in educational processes, scientific research practices, or studies in the methodology and statistics literature, traditional rituals concerning the status of hypotheses are widely maintained. This mode of use reproduces meanings that are not actually present in the epistemic and methodological statuses of hypotheses; in particular, it transfers the concluding language of rejection, acceptance, or verification of hypotheses into scientific practices, thereby perpetuating it across generations. Consequently, the conceptual boundaries between scientific hypotheses and statistical decision procedures become progressively blurred, and methodological decisions are often interpreted as epistemic conclusions, while statistical outputs are interpreted as scientific verifications. This situation is not merely a terminological preference but a fundamental methodological problem that directly affects the nature of scientific inference, the ontological and epistemic statuses of hypotheses, and the role of statistical methods in the production of scientific knowledge. In this article, the problem will be addressed, first, on the epistemological plane, subsequently within the framework of the ontological

status of hypotheses, and finally in terms of the methodological limits of statistical decision theory. Our main thesis is that many of the epistemic meanings attributed to hypotheses arise not from their constitutive function but from methodological rituals that have become entrenched over time, and therefore, the discussion must be conducted on epistemological foundations prior to statistical techniques.

Although the historical conditions of this transformation (publishing practices, institutional incentives, etc.) are beyond the scope of this study, the emphasis here is that this transformation can in no way be derived from the internal logic of the N–P theory and is based entirely on a ritualized misreading.

The Historical Roots of the Epistemic Interpretation of Hypotheses

The examples discussed in the previous section demonstrate that the epistemic interpretation of hypotheses is not an isolated terminological habit but a widespread feature of contemporary scientific practice. To understand why this interpretation has become so persistent, it is necessary to examine the historical development of the search for epistemic justification in the philosophy of science. The following section therefore traces this development from early positivism to Lakatos, not because these traditions directly generated contemporary hypothesis testing practices, but because they constitute the historical background against which the methodological role of hypotheses gradually emerged.

- **From Comte to Lakatos: The Limits of the Search for Epistemic Justification**

In the early forms of the positivist conception of science, the rationality of scientific knowledge was largely founded on the idea of the verification of observable regularities. Auguste Comte argued that the fundamental criterion distinguishing scientific knowledge from metaphysical speculation is factual verification, and this approach conceptualized science as a kind of aggregate of verified regularities. However, this framework proved insufficient in clarifying the question of how universal laws (generalizations) can be legitimately derived from singular observations.

The confirmation theory developed by Carl Gustav Hempel in the first half of the twentieth century aimed to move the idea of verification into a logically more sophisticated domain. Nevertheless, rather than solving the problem of induction and verification, the Hempelian approach reformulated it in a more technical manner through confirmation theory, yet the question of by which criteria universal generalizations are "sufficiently" confirmed remained unanswered. This situation led to the gradual increase of epistemic expectations regarding the justification of scientific knowledge.

One of the most radical objections directed at these expectations finds expression in Karl Popper's falsificationist approach. Popper, arguing that scientific theories cannot be verified but can be falsified, placed the concepts of "risk and criticizability" at the center of epistemic rationality. Yet, while falsificationism explains how propositions are to be rejected, it has provided limited answers to the question of under what conditions propositions (hypotheses) can be tentatively accepted and used in scientific practice (at the research level).

At this point, Thomas Kuhn's concept of the paradigm (which is not detailed here as it is not directly related to the subject of this article) and Imre Lakatos's research programs model proposed new frameworks for explaining scientific continuity. Both approaches have shown that scientific theories are evaluated not only through singular experiments but also within broader historical and methodological contexts. However, these models also did not directly place at their center the questions of how hypotheses are structured, how they are tested, and by which criteria decisions are made in the scientific research process. Although these epistemological approaches offer strong analyses on the rational meaning of scientific knowledge, they have left a distinct gap concerning the operational function of hypotheses. In other words, the debates along the "verification–falsification" axis have interrogated what hypotheses "mean," but have largely relegated the question of what hypotheses "are good for" in scientific practice to methodology (this is not a conscious abandonment). This very point constitutes the ground for the historical and conceptual divergence between epistemology and statistical decision theory (methodology/practice).

- **Transition from the Epistemic Gap (Justification) to Methodological Need**

Although the epistemological debates conducted along the verification–falsification axis produced deep and sophisticated analyses in terms of explaining the rational status of scientific knowledge, by the mid-twentieth century they struggled to provide direct answers to certain increasingly pronounced demands of scientific research practice. These demands focused not so much on the question of whether propositions (hypotheses) are true, but rather on the problem of how to act under conditions of uncertainty (the ambiguity between rejection and acceptance). In other words, while epistemology predominantly discussed the question "What should we believe?", scientific practice (research) increasingly imposed, with growing urgency, the question "What decision should we make under conditions where the obtained results are contestable?"

The decision-related uncertainty created by this ambiguity became concretized especially in experimental and statistical research. In these situations of uncertainty/ambiguity generated by limited samples, measurement errors, randomness, and replicability issues, researchers "who inevitably had to make decisions" often required criteria pertaining to keeping error probabilities at acceptable levels, managing risks, and ensuring the comparability of results, rather than relying on full epistemic justification. As we have previously asserted, neither verification nor falsification approaches proposed conceptual tools that would systematically regulate such operational decision processes.

The problem that becomes apparent at this point is that, although epistemological theories are strong in explaining the meaning of scientific knowledge, they fail to provide a normative framework for the conduct of scientific activity. While verification–falsification debates discuss under what conditions hypotheses can be considered closer to the truth, research practice has been confronted with the question of under what conditions a hypothesis will be tentatively rejected or accepted. This distinction between the two levels of problems has often not been explicitly and conceptually clarified and has been left to the implicit assumptions of epistemic discussions.

In this context, it can be said that, rather than condemning scientific practice to deadlock, the problem of epistemological justification triggered a new orientation toward practice. Epistemic uncertainty did not lead to a postponement of scientific decision-making; on the contrary, it led to the development of methods for working *with* this uncertainty. This transformation marks a shift from epistemological projects that attempt to determine the truth conditions of knowledge, toward methodological approaches that aim to regulate rational decision-action under uncertainty.

The distinguishing feature of this shift is that the center of scientific rationality moves from epistemic grounding toward normative regulation. Here, normativity refers to the rules that determine which procedures should be followed at the decision stage, rather than the extent to which beliefs (researchers' predictions) are justified. In this context, hypotheses begin to be treated not primarily as epistemic propositions representing truth, but as methodological tools that function within the framework of certain error rates and decision thresholds.

This transformation does not mean, under any circumstances, that epistemological debates have become entirely invalid. On the contrary, the conceptual incompatibilities revealed by verification and falsification approaches constitute the historical and intellectual background of this methodological orientation. However, this background functions as a ground that explains *why* these methods became necessary, rather than determining the logical structure of statistical methods. In other words, the verification–falsification debates are not the source of decision theories, but rather the practical-epistemic justification behind this attitude.

At this stage, in the context of the transition from epistemic justification to methodological need, Deborah Mayo's (1996) "Error Statistics" approach can be briefly addressed. Through her concept of the "severe test," Mayo aims not to confirm the truth of hypotheses, but to evaluate the extent to which the test procedures used are sensitive to error. According to her, a test is considered severe if it is capable of detecting, with high probability, a result concerning a hypothesis that is false in a particular respect. In this approach, "evidence" is based not on demonstrating that the hypothesis is true, but on the serious elimination of possibilities of error. Thus, Mayo's approach reformulates the Popperian ideal of falsification as a reference for methodological legitimacy through statistical error control, without transforming it into an epistemic claim of truth. This approach makes it possible to read the emergence of statistical tests and hypothesis structures as an indirect but effective response to epistemological problems. In Mayo, scientific research ceases to be merely an activity that discusses the truth of propositions and is redefined as an activity aimed at making decisions under uncertainty, distributing risks, and stabilizing collective research practice.

One recent debate on theoretical stability in epistemic realism (Nyström, 2025) emphasizes that stability should not be confused with epistemic reality but rather should be understood as a product of methodological and institutional research practices. According to Nyström, the assumption that theoretical stability naturally carries an epistemic meaning is a problematic assumption, and this assumption is shaped around the question of how stability is to be explained within the realist–

antirealist dispute. That is, the meaning of stability is not uncontested, natural, or automatically accepted; it is a subject of debate.

The claim defended in our article is that a structurally similar confusion emerges at the level of hypothesis testing, where methodological decision mechanisms are mistakenly identified with epistemic justification. As we have addressed and emphasized, in the Neyman–Pearson decision theory, the H_0/H_1 framework functions as a normatively defined decision mechanism aimed at stabilizing research practice under uncertainty. Interpreting test results as epistemic confirmations or refutations of hypotheses reproduces, at the micro-methodological level, the same category error that Nyström identifies in macro-level realism debates; that is, methodological stability is conflated with epistemic justification.

The continuing debates over the foundations and interpretation of statistical inference, extending from Birnbaum's (1962) attempt to establish a unified framework for statistical inference to more recent discussions by Lakens and colleagues (2018), demonstrate that no consensus has emerged regarding the inferential meaning of statistical procedures. Within this broader debate, Lakens and colleagues focus specifically on the inferential limitations of traditional NHST and propose methodological alternatives rather than epistemological solutions.

A similar point was made by Senn (2007), who argued that many of the practical difficulties associated with hypothesis testing stem not from the statistical framework itself but from persistent misunderstandings of its interpretation.

Lakens et al. (2018) argue that the traditional structure of NHST is inadequate, both logically and statistically, for grounding claims of a "null effect" or "no difference". Failure to reject a hypothesis does not mean it is accepted (failure to reject $\neq$ acceptance). Therefore, the widely used "accept/reject" language in frequentist statistics is misleading. Lakens and colleagues argue that this problem can be overcome through Bayesian approaches and equivalence tests, by generating positive evidence in favour of H_0 . In brief, Lakens and colleagues state; if you want to claim that there is no effect, do not merely test H_0 ; define H_1 as well and test it, or demonstrate that H_0 falls within a specified (small) effect range.

Nevertheless, from the perspective of this study, we can state the following; Bayesian and equivalence-based tests, which aim to remedy the inadequacy of traditional NHST in producing a claim of no difference, function not so much to produce epistemic claims of truth, but "similarly to the Neyman–Pearson framework" as methodological tools that regulate decision-making under uncertainty. In conclusion, even if the claims of Lakens and colleagues are accepted "that is, even if these approaches produce pragmatically more refined decisions" the epistemic status of the results produced cannot be extended beyond methodological decision processes. In other words, these alternatives do not produce anything epistemically different from Neyman–Pearson.

This conclusion is broadly consistent with the argument of Gelman and Shalizi (2013), who likewise treat statistical inference as a methodological process of model criticism and scientific learning rather than a

procedure for establishing the truth of hypotheses.

- **Methodological Silence and Empirical Practice in Popperian Epistemology**

At this point, a discussion regarding the pragmatic necessity (knowingly or unknowingly) left incomplete by the Popperian epistemological perspective becomes unavoidable. The Popperian perspective makes a scientific "claim" (although Popper himself does not use such a label, this claim now corresponds to the H_1 hypothesis) epistemically visible, while the reference state (the H_0 hypothesis) that this claim suspends effectively emerges but remains epistemically silent within the theory. In terms of the burden of proof/demonstration responsibility, Popper implies the following: if you assert the contrary of an existing state/status (*status quo*), you must falsify this apparent existing state/status and (a priori) objective reality (Popper, 1963).

In short, falsifiability is epistemically silent in Popper, yet it produces necessary epistemic consequences empirically/in practice. That is, if the H_1 hypothesis is a claim that suspends the current state within the reference framework it defines (through variables), this "current state" cannot be merely a practical starting point and must also be defined epistemically. For the methodological structure (H_0-H_1) necessarily established in the practice of empirical research must effectively generate an epistemic status in a domain about which Popper theoretically remains silent. Yet Popper suspends the claim but does not name the suspended entity itself and remains silent.

If H_1 is a claim "suspending the current state," the current state itself (H_0) cannot be left undefined for H_1 to have epistemic meaning. In other words, in empirical research, H_1 is the researcher's claim, and this claim is an intervention that suspends the current state within the reference framework. Consequently, H_0 , as its alternative, must be a hypothesis that rejects this claim and defines the current state of the variables (*status quo*).

In conclusion, within Popper's falsificationist framework, the only visible element of scientific knowledge is the proposed claim, and this claim acquires epistemic status provided it is falsifiable. However, this claim "taking a risk" requires the effective suspension of "a reference state." Popper does not define this reference state as an epistemic object (there is no such definition/attempt in Popper); he leaves it silent as a background assumption. In methodological practice, however, this background takes the form of a counterhypothesis that rejects the claim and defines the current state. Therefore, H_0 emerges as the epistemic reference framework necessitated by empirical practice, in a situation where Popper remains theoretically silent.

Finally, on this point, we can state the following: H_0 has no independent epistemic status within the Popperian framework. However, in empirical research practice, it is necessary to define H_0 with specific content as a prerequisite for the testability of H_1 . Although this definition is not formulated as an epistemic claim at the theoretical level, it implicitly entails an epistemic acceptance through measurement instruments, observational procedures, and confidence in background assumptions. Hence, although H_0 appears to be theoretically devoid of epistemic status, in practice it functions as an

epistemically loaded reference point. This necessary yet theoretically unprocessed relationship points to a systematic silence within the Popperian framework.

Returning to our discussion, the first systematic articulations of the transformation concerning the methodological demand left unaddressed by the verification–falsification debates, which we have examined in detail in the previous sections, will first become apparent in the work of Ronald Fisher, and subsequently acquire a clear and consistent normative framework with the Neyman–Pearson decision theory.

- **Ronald A. Fisher: Significance, Testing, and Epistemic Differentiation**

The increasing centrality of statistical methods in scientific research practice was achieved not through a "direct philosophical answer" to the problem of epistemic justification, but rather through operational solutions developed for the practical problems encountered by researchers. One early and decisive example of this transformation is Ronald Fisher's approach, shaped around significance tests. Fisher's work can be read as a precursor to a new methodological sensibility aimed at rendering scientific research practicable under conditions of uncertainty.

At the core of Fisher's understanding of statistics lies the epistemic differentiation between the inherently random structure of sample data and scientific inferences. In this context, the "significance test" he developed aims not to establish the truth of a hypothesis, but to demonstrate the extent to which the observed data can be considered "unusual/deviant" under a specific assumption (theory). In this vein, "significance/p-value" is designed and used as a tool that quantitatively expresses this unusualness. However, this tool is less an epistemic criterion of verification and more a statistical indicator that regulates the researcher's threshold of doubt (the probability of making an erroneous decision) concerning the correctness of the results. This situation indicates that the relationship Fisher's approach establishes with epistemology is neither direct nor unproblematic. That is, on the one hand, Fisher's significance tests are regarded as methods that contribute to the rational evaluation of scientific evidence, while on the other hand, it is assumed that the results offered by these methods do not carry claims of epistemic certainty.

Consequently, in Fisher, a significance test does not assert that a hypothesis is true (Cox, 1958; Cox, 2006); it merely indicates the extent to which the obtained data are compatible with the hypothesis in question (the hypothesis being tested). This leads to a blurring of the boundary between claims of epistemic truth and criteria of methodological evaluation.

Another notable issue in Fisher's approach is the ambiguous status of hypotheses. In Fisher, the hypothesis is neither explicitly defined as a normative basis for decision-making nor conclusively established as a claim of epistemic truth. This intermediate position occupied by hypotheses (the ambiguity of their epistemic status) allows Fisher's methods to maintain their connections with epistemology, while simultaneously enabling a distance from the classical conception of verification.

However, this intermediate position has been the source of enduring debates concerning the interpretation of Fisherian statistics.

In this context, Fisher's significance tests function as a tool that enables scientific research to operate under uncertainty, without entirely severing it from the ideal of epistemic certainty. Nevertheless, this function leaves open the question of by which criteria scientific decisions should be made. In the theory, whether the p-value is "small" or "large" does not by itself provide a decision rule; questions such as which thresholds are acceptable and which errors are to be tolerated cannot be systematically answered within the Fisherian framework.

This situation leads to the clear perception of a gap/ambiguity between epistemic justification and methodological decision-making in Fisher's approach. While Fisher views statistical tests as part of scientific reasoning, he deliberately leaves unclear how these tests will yield binding decisions. Thus, the researcher largely bears the responsibility of interpreting the obtained statistical results. The method does not necessitate a decision; it merely guides it (Cox, 1958; Cox, 2006).

At this point, the limits of the Fisherian approach become apparent. In a context where scientific practice is becoming increasingly institutionalized and collective research programs are gaining prominence, statistical interpretations left to individual judgment begin to be deemed insufficient. Questions such as which errors are acceptable, which results will be reported, and how research findings will be compared have come to demand a clearer and more normative explication. Fisher's approach addressed this demand intuitively but did not develop a systematic decision theory. That is to say, in essence, "Fisher is aware of the situation but has not developed a decision theory."

Therefore, Fisher's understanding of statistics can be evaluated as "an intermediate phase" in the transition from epistemological disagreement to methodological transformation. Fisherian significance tests are not a direct solution to the verification–falsification debates, but they constitute one of the first strong examples demonstrating that it is possible to work with the uncertainty revealed by these debates. Without entirely abandoning the ideal of epistemic truth, Fisher initiates a methodological orientation aimed at rendering scientific practice operable under uncertainty. It is with the Neyman–Pearson decision theory that this orientation attains a systematic and normative decision structure.

- **Neyman–Pearson: Normative Decision Theory and the Redefinition of Hypotheses**

Although statistical methods surpassed a significant threshold with Fisher's significance tests in terms of gaining a central role in scientific research practice, this approach left open the question of how scientific decisions should be made. This gap was systematically addressed by the framework of decision theory developed by Jerzy Neyman and Egon Pearson. The Neyman–Pearson approach positions "statistical tests" not as "tools of evaluation" concerning epistemic truth claims, but as a normative structure that systematizes the ability of research programs to make rational decisions under uncertainty. The distinctive feature of this approach is its radical redefinition of the methodological status of hypotheses. Within the Neyman–Pearson framework, H_0 and H_1 (null and test/alternative

hypotheses) are not treated as symmetric epistemic propositions (symmetry such as if H_0 is true, H_1 is false) or as competing truth claims. On the contrary, these hypotheses are defined as methodological solutions that determine which decision will be made under specific error probabilities. The meaning of hypotheses is derived not from the "truth" they represent, but from their roles within decision processes.

This normative structure becomes operational through the explicit definition of statistical errors. Type I error denotes the probability of rejecting a true null hypothesis (H_0), while Type II error denotes the probability of failing to reject (erroneously accepting) a false null hypothesis (H_0). In the Neyman–Pearson approach, these error types represent the risks to which scientific research is exposed under unavoidable conditions of uncertainty. Hence, error probabilities are not flaws to be eliminated (the N–P theory is aware that this is impossible), but methodological parameters to be predetermined and controlled. Within this framework, the α (alpha) level defines the acceptable upper limit for Type I error. Alpha is not an epistemic degree of confidence or intensity of belief, but merely an a priori decision parameter concerning the extent to which a false rejection occurring under specific conditions will be tolerated. Similarly, the β (beta) level and the correspondingly defined test power ($1-\beta$) express the research's capacity to detect a false null hypothesis. The concept of power is not a measure of epistemic truth, but an operational characteristic describing the performance of a specific decision rule. That is, statistical tests produce decisions, not epistemology.

This structure of the Neyman–Pearson decision theory shifts the center of scientific rationality from epistemic justification toward methodological regulation. Here, rationality is associated not with an answer to the question of which hypothesis is "truer," but with which decision process produces more predictive/stable and replicable results in the long run. Therefore, the Neyman–Pearson approach focuses less on the justification of individual beliefs (researchers' predictions) concerning research findings, and more on the sustainability of collective research practice. This normative orientation makes it possible to apply statistical tests in a context-independent manner. Researchers produce results within the framework of predetermined error rates and decision rules, independent of their personal convictions or intuitive assessments. While this enhances the comparability of scientific findings across different studies, it also largely purifies decision-making processes from personal interpretations.

An important point that needs emphasizing here is that the Neyman–Pearson decision theory does not entirely sever its relationship with epistemological debates. The uncertainties and the emphasis on risk revealed by verification and falsification approaches constitute the historical background of this decision theory. Nevertheless, rather than providing direct answers to these epistemological problems, the Neyman–Pearson framework offers a procedural solution that secures the functioning of scientific practice. Epistemic problems are not solved, but they are rendered manageable through methodological tools.

Therefore, the central position acquired by the H_0 and H_1 structure in contemporary research programs should be understood not as a direct extension of epistemological theories, but as the institutionalization of a decision-based normative regulation. Hypotheses are no longer primarily

statements representing truth (epistemic); they become tools that stabilize scientific activity through error rates, power, and decision rules. This redefinition constitutes one of the cornerstones of the normative architecture of modern scientific research.

- **The Indirect Relationship Between Verification-Falsification and Decision Theory**

The normative structure of the Neyman–Pearson decision theory has often, and erroneously, been interpreted as a direct continuation of verification or falsification approaches. This interpretation is nourished by the widespread linguistic usage suggesting that statistical tests "falsify" or "confirm" scientific hypotheses.

As Hacking (1965) emphasized, statistical inference should not be interpreted as a direct theory of scientific knowledge but as a distinctive form of inferential reasoning. This perspective reinforces the view defended here that the relationship between statistical decision procedures and epistemological theories is one of conceptual interaction rather than theoretical identity.

However, to the extent that such a reading fails to sufficiently observe the distinction between epistemological justification and methodological decision-making, it risks conflating two different levels of problems. Verification and falsification approaches deal with the epistemic status of scientific theories. The questions of under what conditions a theory can be rationally accepted, what kinds of experimental risks it should take, or in the face of which observations it should be abandoned are at the center of these debates. For example, in Karl R. Popper's falsificationist approach, "the value of theories/hypotheses/scientific claims" is measured by their falsifiability and their exposure to severe tests. The emphasis here on "being falsifiable" concerns "the openness of truth claims to critical scrutiny."

In contrast, decision theory determines which procedures should be followed under conditions of uncertainty, rather than grounding the truth of scientific theories. In the Neyman–Pearson framework, tests do not produce an epistemic verdict in the sense of "falsifying" a hypothesis; they merely enable the selection of a particular action (rejection or non-rejection) as a result of operating decision rules for which specific error rates have been pre-accepted. This difference leads to the same concepts (e.g., "rejection") carrying different meanings in two different contexts (epistemic and methodological).

At this point, we must state that it would be more appropriate to understand the relationship between the verification–falsification debates and decision theory as a historical and conceptual background connection, rather than as a substantive continuity (from the epistemic to the methodological). Epistemological debates have rendered visible that scientific knowledge cannot be definitively justified and always contains a certain degree of risk and uncertainty. This awareness has made the question of how scientific research can be sustained in practice even more urgent. Decision theory is precisely a product of this urgency and, rather than solving epistemological problems, regulates how to work with them.

This distinction can be seen more clearly, particularly through the concept of "falsification." In the Popperian context, falsification is a strong claim concerning the epistemic status of a proposition, and a falsified proposition (in connection with a specific observation) must be rationally abandoned. In the Neyman–Pearson context, however, the decision to reject is a methodological action taken under an accepted probability of error and is not necessarily the expression of the conclusion that the theory (proposition) is false. The fact that the same term undertakes such significantly different functions in the two contexts makes interpreting statistical tests (and their results) through an epistemological reading attractive, but at the same time misleading.

A similar situation applies to discussions of verification and confirmation. While confirmation theory attempts to explain the extent to which observational evidence supports a hypothesis, decision theory does not treat such degrees of support as a central concept. In the Neyman–Pearson approach, what matters is not "how supportive" the evidence is, but what error profile/probability a particular decision rule possesses in the long run. Thus, while the concept of confirmation occupies a central position in the epistemological context, it remains secondary, and often even external, in the theoretical context of "decision."

In this framework, the relationship between the verification–falsification debates and decision theory can be described as a kind of indirect interaction. Epistemological approaches have rendered visible the fragility of scientific knowledge and its intertwined structure with uncertainty, and decision theory, moving from this visibility, has developed a normative regulation concerning how scientific activity can be sustained in practice. However, this regulation consciously refrains from reconstructing epistemic truth claims.

Therefore, it is misleading to view statistical tests and the H_0 and H_1 structure as the "application" of epistemological theories. Decision theory is not a solution to epistemological problems; it is a procedural framework that prevents these problems from paralyzing scientific practice (research programs). The relationship of this framework with epistemology is contextual, not resolutive.

The failure to make this distinction sufficiently clear constitutes the source of the conceptual confusions frequently encountered in contemporary educational programs and research practices, as exemplified in the earlier sections of this article. Even some reformist attempts have focused on statistical procedures and remained ambiguous regarding the epistemological foundations of hypothesis tests. The reading of hypotheses as epistemic truth claims, the reporting of statistical decisions in the language of "evidence," and the presentation of methodological tools as if they produce philosophical conclusions, stem largely from the articulation of these two levels. As will be discussed in the next section, this situation is not merely a terminological problem but points to a deeper category error that leads to the mispositioning of scientific rationality.

Category Error in Current Research Practices: The Epistemicisation of H_0 and H_1 Hypotheses

Criticisms regarding this issue, concerning the reading of statistical norms as epistemic criteria, are not limited merely to the philosophy of statistics. In his articles in the decision theory literature, Gigerenzer (2004) has objected to the use of normative procedures as context-independent epistemic truth standards, arguing that this tendency may lead to category errors. Although Gigerenzer's criticisms are primarily directed at the evaluation of human reasoning, they bear a structural similarity to the problem discussed here. In both cases, methodological or normative decision rules are interpreted as producing epistemic content. From this perspective, interpreting the failure to reject H_0 as the "absence of an effect" means transforming a decision-theoretic procedure into an epistemic claim. However, in the Neyman–Pearson framework, the test result is merely an expression of a decision made under specific error rates and does not provide direct knowledge about the ontological or epistemic status of the research object.

As Gigerenzer emphasizes from a different perspective, the widespread adoption of the N-P statistical decision theory in research practices has provided significant standardization in terms of the production and reporting of scientific results. Nevertheless, the failure to adequately distinguish this theoretical framework from epistemological debates has brought with it a persistent conceptual confusion regarding the status of hypotheses. Expressions frequently encountered in the current research literature, such as "the research hypothesis was confirmed," " H_1 was accepted," or " H_0 was falsified," are the most visible indicators of this confusion.

Although this linguistic practice may appear to be an innocent abbreviation at first glance, upon closer examination it points to a systematic category error based on reading a methodological tool as if it were an epistemic claim. In the Neyman–Pearson framework, the H_0 and H_1 structure is a normative mechanism that regulates which decision is to be made under specific error rates. The output of this structure is not an epistemic judgment about the truth of a hypothesis, but an action taken as a result of operating pre-defined decision rules. Nevertheless, in research reports, these actions are frequently translated into the language of truth and imbued with epistemic meaning.

An important source of this category error is the disregard for the difference between the everyday and philosophical usages of the terms "rejection" and "acceptance" and their decision-theoretical usage. In decision theory, rejection means "deciding in favor of H_1 against H_0 under the condition that a certain alpha level is exceeded." This decision is not evidence that H_0 is false; it merely indicates that a certain error risk has been accepted (a decision in favor of H_1 has been made, accepting the possibility that H_0 might be false) and that an alternative action has been chosen. In contrast, in the epistemological context, rejection often means the justified demonstration of the falsity of a proposition (a fully Popperian acceptance). The different functions that the same term undertakes in these two contexts prepare the ground for reading decision outputs as if they were epistemic verdicts.

A similar problem arises in the use of the concepts of "support" and "evidence." Statistical tests indicate the extent to which the data are compatible with a specific model, or which action is chosen under a specific decision rule. However, these indicators are often presented as proof of the truth of the

hypothesis. Thus, methodological performance criteria replace epistemic justification criteria. Consequently, research hypotheses cease to be tools that regulate decision processes and are positioned (through a flawed positioning) as epistemic claims expected to be confirmed or falsified.

This tendency, by obscuring the fact that H_1 is a legitimate epistemic proposition containing the researcher's original theoretical claim, often places it in an epistemically secondary position vis-à-vis H_0 . However, in the Neyman–Pearson framework, these roles correspond not to an ontological or epistemic distinction, but to a methodological division of labor. H_0 is chosen as the reference point for error control, and H_1 is defined in line with a specific effect size and power target. Each of these choices is aimed at optimizing decision performance rather than making truth claims.

This misreading creates cascading effects in the interpretation of research results. Once a decision is made, the idea that the hypothesis has been "proven" becomes established, leading to the derivation of strong epistemic conclusions (such as generalizations) from singular studies/observations. Yet the fundamental assumption of decision theory is to ensure the statistical properties of long-term and repeated applications, not of singular tests. In other words, decision theory regulates the error profile of collective research practice, not individual truth claims (beliefs about the correctness of research findings).

In this context, the epistemitization of H_0 and H_1 is not merely a terminological problem but also means the misplacement of epistemic rationality at an incorrect level. When the concepts of epistemological debates are conflated with the outputs of methodological tools, both epistemic burdens are demanded from statistical tests that they were not designed to bear, and problems that epistemology should solve are relegated to statistical procedures. The resulting picture produces neither a sound epistemic justification nor a consistent methodological application. Therefore, what is needed in contemporary research practices is not to redefine the epistemological status of the H_0 – H_1 structure, but to render visible the methodological and normative status it already possesses. When hypotheses are treated not as propositions representing truth, but as tools that make rational decision-making under uncertainty possible, both the limits of statistical methods become clearer, and the domain of epistemological debates is properly delineated.

The Methodological Status of H_0 and H_1

Although hypotheses in statistical research are often used intertwined with epistemological concepts, within the context of the Neyman–Pearson decision theory, the primary function of these hypotheses is not to prove truth but to regulate decision-making under uncertainty. Therefore, the meaning of H_0 and H_1 hypotheses differs from the meanings attributed to them in everyday language or philosophical debates. The misunderstandings frequently encountered in research practices largely arise from the disregard of this difference. Regarding the status of hypotheses within the Neyman–Pearson decision theory framework, the following can be stated:

- **H₀ Hypothesis: A Hypothesis of "No Difference," Not "Absence."** One of the most common errors in practice is the designation of the H₀ hypothesis as the "null hypothesis" (or "absence hypothesis"). This labelling creates the impression that H₀ is an ontological proposition claiming that no effect or relationship exists. However, within the Neyman–Pearson framework, the H₀ hypothesis is not an ontological claim of absence, but rather an assumption of no difference (or no effect) chosen as the starting point of comparison.

As Gerd Gigerenzer argues in *Statistical Rituals* (2018) and related works, using NHST as if it were an epistemological confirmation theory leads to a serious category error. According to Gigerenzer, the Neyman–Pearson approach is historically a normative procedure built on the logic of decision-making; however, in practice, H₀ is often interpreted as an ontological claim of "absence," and non-significant results are read at the epistemic level as "the absence of an effect." Gigerenzer maintains that this usage produces a ritualized interpretive practice that transforms statistical decision rules into epistemic truth criteria.

Consequently, the primary function of H₀ is to establish a reference state within which error rates can be defined and controlled in statistical testing. Therefore, H₀ should be understood not as a proposition that the researcher believes or does not believe, but as a methodologically fixed state of comparison. That is, the rejection of H₀ does not mean "the refutation of no difference" or "the proof of the existence of an effect of the selected independent variable"; it merely indicates that, under the chosen alpha (significance) level, a result incompatible with the assumption of no difference has been obtained.

- **H₁ Hypothesis: A Test Hypothesis, Not an Epistemic Claim.** Similarly, the H₁ hypothesis is also frequently and erroneously presented as the "research hypothesis" or "the proposition to be proven." This usage turns H₁ into an epistemic truth claim and places upon it a philosophical and epistemic burden that cannot be expected from statistical tests. However, in the Neyman–Pearson approach, H₁ is not a proposition assumed to be true, but a definition of an alternative state intended to be tested. The H₁ hypothesis is typically defined along with a specific effect size and direction. The purpose of this definition is to determine under which conditions the test will have sufficient power, and in which cases the decision to reject H₀ will be considered meaningful. Hence, the role of H₁ is to regulate the performance characteristics of the test, rather than to represent the researcher's belief.
- **Rejection and Acceptance: Decisions at the Research Level, Not Epistemic Verdicts.** Another point that particularly needs emphasis here is the meaning of the term's "rejection" and "acceptance." In the Neyman–Pearson framework, these terms do not refer to epistemic verdicts concerning whether a hypothesis is true or false, but to actions selected as a result of applying pre-determined decision rules (which entail a specific error assumption). Therefore, the statement "H₀ was rejected" does not mean "H₀ is false"; it merely means "a decision has been made in favour of H₁ by accepting a certain error risk."

When this distinction is not observed, the results of statistical tests are reported in the language of epistemic certainty, and overgeneralizations are made from singular research findings. Yet the logic of decision theory is aimed at ensuring the error profiles of long-term and repeated applications, rather than singular results. The correct positioning of hypotheses can be summarized as follows; **H₀**: Not a claim representing absence, but a methodological reference of no difference. **H₁**: Not a proposition of truth intended to be proven, but a test hypothesis. **Rejection/Acceptance**: Not epistemic judgments, but decision actions at the research level.

Explicitly observing this distinction both clarifies the limits of statistical tests and prevents the epistemological over-interpretation of scientific results. The relationship of statistical methods with epistemology should be evaluated not through the question of what these methods prove, but through the question of how they render scientific practice operable under uncertainty.

Hypothesis Labels Beyond H₀ and H₁: An Epistemic and Methodological Critique

In positivist epistemology, the status of the hypothesis is characterized by indeterminacy. Within positivist epistemology (particularly in logical positivism), the concept of hypothesis does not possess a normative status; hypotheses are not explicitly defined, are not symbolically standardized, and play an instrumental role in the verification process, but their epistemic status is never definitively established.

In positivist epistemology, the primary interest is focused on "verifiable propositions," and the formal structure of hypotheses remains a secondary, technical issue. Consequently, there is no form defining "the modes of formulating hypotheses" in positivist epistemology. This situation renders the status of hypotheses ambiguous from the perspective of positivist epistemology.

With the Neyman–Pearson decision theory, hypotheses acquire a status, but this status is limited. The N–P decision theory grants a status to hypotheses; however, the status defined by decision theory is not an epistemic position but is situated in a "decision-theoretic" position. That is, in the N–P decision theory, hypotheses are decision alternatives and do not carry claims of truth. They are merely either rejected or not rejected. Moreover, in the theory, this structure is deliberately limited to two symbols: H₀, representing the no-difference/reference state, and H₁, the alternative to be tested. This limitation is not accidental; it derives from the binary nature of decision theory.

Consequently, symbolic usages such as H₂, H₃, H₄, and/or H_{0.1}, H_{0.2}, H_{1.1}, H_{1.2}, have "no defined place within the binary decision structure of the N–P framework" and their methodological statuses are ambiguous. This situation arises from the internal logic of the theory itself. Gigerenzer's critique of "statistical rituals" is directed at the mechanical and reflexive application of hypothesis tests, detached from theoretical content. In this context, the widespread use of indexed hypothesis structures (H₂, H₃, H₁₁, H₁₂, etc.) in the literature, often merely duplicated as testable units and reduced to reporting convenience, although not directly overlapping with the "lack of epistemic and methodological status" argument advanced in this study, shows a strong parallelism.

The Disconnect in Practice and Symbolic Inflation in Hypothesis Formulations

It is frequently observed today that researchers employ symbols such as H_2 , H_3 , H_4 , and/or $H_{0.1}$, $H_{0.2}$, $H_{1.1}$, $H_{1.2}$ in their hypothesis formulations. Yet none of these symbols possesses a defined status either in the epistemological literature or in statistical theory (methodologically). For they are neither epistemic claims, nor independent decision alternatives, nor, ultimately, theoretical necessities; they are merely reporting conventions/schemas, and they cannot be considered "hypotheses"; they are "designations/labels."

If a symbolic formulation/label has no epistemic or methodological counterpart, that is, if their statuses are epistemically and methodologically ambiguous, then such formulations (i.e., symbols such as H_2 , H_3 , H_4 , and/or $H_{0.1}$, $H_{0.2}$, $H_{1.1}$, $H_{1.2}$) do not carry scientific status. We can formulate this situation in terms of epistemic and methodological logic as follows: only H_0 – H_1 are methodological hypothesis formulation symbols with defined statuses. In contrast, hypothesis formulations such as H_2 , H_3 , H_4 ; $H_{0.1}$, $H_{0.2}$; $H_{1.1}$, $H_{1.2}$ are symbolic labels/conventions with ambiguous statuses.

However, the logical (epistemic and methodological) framework we have established so far does not imply that such hypothesis symbols cannot be used. It merely means that the epistemic and methodological statuses of such hypothesis symbol usages are ambiguous and that they lack (and cannot be defended on) scientific grounds.

Conclusion

This study has aimed to demonstrate that the relationship between the philosophy of science and statistical methodology possesses a more indirect and asymmetric structure than is often assumed. Although the epistemological debates along the "verification–confirmation–falsification" axis from Comte to Lakatos have produced powerful conceptual analyses concerning the rational status of scientific knowledge, they have not provided direct answers to the increasingly pronounced methodological requirements of scientific research practice. This situation should be evaluated not as a failure of epistemological justification projects, but as a differentiation of problem domains.

The central argument defended throughout this article is that this differentiation becomes particularly visible through the status of hypotheses. While epistemological approaches have largely treated hypotheses as propositions carrying truth claims, statistical decision theory has structured hypotheses as normative tools that regulate rational decision-making under uncertainty. This methodological orientation, which began with Fisher and acquired systematic form with the Neyman–Pearson decision theory, has developed not a solution to epistemological problems, but a procedural framework that prevents these problems from paralyzing scientific practice.

Within this framework, the central position acquired by the H_0 – H_1 structure in contemporary research programs should be understood not as a direct extension of epistemological theories, but as the

institutionalization of a decision-theoretic regulation. Concepts such as Type I–Type II error, alpha–beta, and power do not express degrees of epistemic truth; they are normative parameters that secure the long-term performance of specific decision rules. Therefore, interpreting the outputs of statistical tests as if they were epistemic truth claims leads to a systematic categorization error, which amounts to transforming a methodological tool into an epistemic object.

This mispositioning is not merely a terminological issue arising in reporting language. The epistemicisation of hypotheses leads to demanding from statistical methods philosophical burdens they were not designed to bear, and to relegating problems that epistemology should solve to methodology. Consequently, statistical tests neither produce sound epistemic justifications, nor can epistemological analyses meet the operational needs of research practice.

The solution proposed by this study is not to establish a hierarchical relationship between epistemology and methodology, but to clearly distinguish the problem levels of these two domains. Epistemology may continue to reflect on the meaning, scope, and rational status of scientific knowledge, while methodology may continue to develop normative frameworks that regulate how decisions are to be made under uncertainty. Positioning the H_0 – H_1 hypothesis structure at this second level both clarifies the limits of statistical methods and prevents epistemological debates from being directed to the wrong addresses.

Thus, it becomes evident that scientific rationality cannot be defined along a single axis (neither solely through epistemic justification nor solely through methodological functioning). Modern scientific practice places at its center not the elimination of epistemic uncertainty, but the institutional and procedural arrangements that make it possible to work with this uncertainty. The Neyman–Pearson decision theory is one of the clearest examples of such arrangements. Understanding this approach within its own normative context, without epistemicisation it, provides a more consistent ground for evaluation both for the philosophy of science and for statistical methodology.

Declarations

Author Contribution

The entire article was written by a single author.

References

1. Birnbaum, A.: On the foundations of statistical inference. *J. Am. Stat. Assoc.* **57**(298), 269–306 (1962). <https://doi.org/10.1080/01621459.1962.10480679>
2. Comte, A.: (1830–1842). *Cours de philosophie positive* (Vols. 1–6). Paris, France: Bachelier
3. Cox, D.R.: *Principles of statistical inference*. Cambridge University Press (2006)
4. Fisher, R.A.: *Statistical methods for research workers*. Oliver & Boyd (1925)

5. Fisher, R.A.: The design of experiments. Oliver & Boyd (1935)
6. Gelman, A., Shalizi, C.R.: Philosophy and the practice of Bayesian statistics. *Br. J. Math. Stat. Psychol.* **66**(1), 8–38 (2013). <https://doi.org/10.1111/j.2044-8317.2011.02037.x>
7. Gigerenzer, G.: Mindless statistics. *J. Socio-Econ.* **33**(5), 587–606 (2004). <https://doi.org/10.1016/j.socec.2004.09.033>
8. Gigerenzer, G.: Statistical rituals: The replication delusion and how we got there. *Adv. Methods Practices Psychol. Sci.* **1**(2), 198–218 (2018). <https://doi.org/10.1177/2515245918771329>
9. Gigerenzer, G.: The legacy of Daniel Kahneman: A personal view. *Erasmus J. Philos. Econ.* **18**(1), 28–61 (2025). <https://doi.org/10.23941/ejpe.v18i1.1075>
10. Hacking, I.: Logic of statistical inference. Cambridge University Press (1965)
11. Hempel, C.G.: *Aspects of scientific explanation and other essays in the philosophy of science*. Free Press. (See especially *Studies in the Logic of Confirmation*, pp. 3–46.) (1965)
12. Lakatos, I.: Falsification and the methodology of scientific research programmes. In: Lakatos, I., Musgrave, A. (eds.) *Criticism and the growth of knowledge*, pp. 91–196. Cambridge University Press (1970)
13. Lakens, D., McLatchie, N., Isager, P.M., Scheel, A.M., Dienes, Z.: Improving inferences about null effects with Bayes factors and equivalence tests. *Journals Gerontology: Ser. B: Psychol. Sci. Social Sci.* **75**(1), 45–57 (2018). <https://doi.org/10.1093/geronb/gby065>
14. Mayo, D.G.: Error and the growth of experimental knowledge. University of Chicago Press (1996)
15. Neyman, J.: *Lectures and conferences on mathematical statistics and probability*. U.S. Department of Agriculture. (1952)
16. Neyman, J., Pearson, E.S.: On the problem of the most efficient tests of statistical hypotheses. *Philosophical Trans. Royal Soc. Lond. Ser. Containing Papers Math. or. Phys. Character.* **231**(694–706), 289–337 (1933). <https://doi.org/10.1098/rsta.1933.0009>
17. Nyström, J.: Theoretical stability and scientific realism. *Philos. Sci.* **92**, 900–921 (2025). <https://doi.org/10.1017/psa.2025.22>
18. Pearson, E.S.: Statistical concepts in their relation to reality. *J. Am. Stat. Assoc.* **50**(272), 204–207 (1955)
19. Popper, K.R.: The logic of scientific discovery. Routledge (1959)
20. Popper, K.R.: Conjectures and refutations: The growth of scientific knowledge. Routledge (1963)
21. Senn, S.: Statistical issues in drug development, 2nd edn. Wiley (2007)
22. Szucs, D., Ioannidis, J.P.A.: When null hypothesis significance testing is unsuitable for research: A reassessment. *Front. Hum. Neurosci.* **11** (2017). Article 390. <https://doi.org/10.3389/fnhum.2017.00390> [reference:0][reference:1]