Perspectival Realism and Frequentist Statistics:
The Case of Jerzy Neyman's Methodology and Philosophy[*]

9 September 2024

1. Introduction

Perspectival realism ("PR" hereafter) is a developing trend in contemporary thought, which can be recognised as one of the post-Kuhnian theories of science. In these theories, significant emphasis is placed on the inseparable nature of cognitive and social dynamics within the cognitive act and the development of scientific knowledge (see Collins, Evans 2002). This sets it apart from a static approach of (logical) positivists. In particular, proponents of PR "share the general idea that there is no 'view from nowhere' and that scientific knowledge cannot transcend a human perspective" (premiss 1). This implies that the truth condition or justification of a hypothesis depends on an epistemic vantage point. However, "it is in part mind-independent facts that make our theories true or false" (premiss 2) (Ruyant 2020).

Scientific outcomes are contingent on the statistical methodology adopted. This non-physical statistical instrument, utilized for shaping data collection and drawing conclusions, is influenced by a scientist's viewpoint. Multiple valid statistical sampling and inference methods exist, requiring researchers to decide on specifics. Thus, it's scholarly justified to explore the relationship between statistics and PR. All the more that understanding of what might count as a 'perspective' is quite eclectic and sometimes underspecified in the philosophical literature.

It has been posited that PR aligns with various factual elements and methodological practices evident in the formulation and evolution of scientific theories (see, e.g., Massimi 2018b). While PR finds firm ground in scenarios involving concrete content within the exact sciences, its connection to statistical methodology remains underdeveloped. There are perspectival accounts of investigations of aspects of the process of scientific scrutiny that concentrate on data (see, e.g., Jacoby 2020) observational instruments (see, e.g., Creţu 2022), and the nature of numerical representations (see, e.g., Wolff 2019). Many authors, like Giere (2010), Rueger (2016), Massimi (2018c) or Potters (2020) advocate for perspectivism concerning scientific (including mathematical) models of experiments and data. However, they do so without delving into the specifics of statistical frameworks for sampling, inference,

---

[*] Forthcoming in *Synthese*.

and their interpretations. These analyses predominantly focus on the substantive aspects of models, neglecting the intricacies of the statistical procedures and tools underpinning them. Giere (1976) discusses frequentist hypothesis testing from a realist stance, albeit without a fledged perspectivist aspect.

A recent exploration into the Bayesian (alternative to frequentist) statistical approach within the framework of PR, has been conducted by Massimi (2021). However, the current body of literature lacks a comprehensive examination of how contemporary PR aligns with frequentist statistical methodologies for sampling and inference. Investigating the connections between PR and frequentism offers a novel vantage point for addressing inquiries about validity, universality, normativity, and the philosophical potential inherent in PR.

Frequentist statistics, also known as classical, orthodox, or error-based statistics, is characterised by its premise: the absence of inherent probabilistic statements (or choices) regarding probabilities that concern what the state of affairs actually is. Probabilities are introduced in the form of probabilities of detecting the true state of affairs (power function over the domain of possible point hypotheses), which pose the basis for knowing pre-observational relative error risks. The risk of false rejection of the hypothesis tested is fixed in advance. The risk of false acceptance of it, relative to a particular (point) alternative hypothesis, is equal to one minus the power corresponding to this particular alternative. Another distinctive aspect lies in the dependence of the inferential framework on assumed probabilities of unobserved data, guided by a model of sampling probability distribution.

Regarding PR's focus on the inseparability of social factors from cognitive processes, frequentist statistics stands out for the direct influence of social factors on sampling schemes (see, e.g., Kubiak, Kawalec 2022) and error risks (see Kubiak et al. 2021). This approach also allows for treating accepted statistical hypotheses as accurate descriptions of real physical systems, with objective assessment and control of the likelihood of accepting falsehoods, as discussed by Giere (1976) and Mayo (2018). This brings frequentist methods to be potentially highly interconnected to PR. However, frequentism varies in methodological and philosophical assumptions, as highlighted by Lenhard (2006). Therefore, when comparing PR to frequentism, it's essential to select a philosophically and methodologically comprehensive version of frequentism proposed by a specific author.

In the philosophy of statistics, a long-lasting disagreement exists between Bayesianism and frequentism (see Sprenger 2016). Among frequentists, Jerzy Neyman, with his sharp philosophical views (see, e.g., Neyman 1937, 344; 1957b), is one of the most historically recognisable opponents of Bayesian statistics. Neyman was a 20th-century statistician who is recognised as one of the co-founders of the frequentist statistical paradigm, which dominated the methodology of natural and social sciences in the 20th century (Lehmann 1985). Reevaluating Neyman’s methodology and philosophy of statistics from an unbiased philosophical standpoint, one not constrained by the Bayesian-frequentist controversy, shall bring about a new dimension to the debate on his conceptions. As the reader will see, some points in Neyman’s writing are balancing between realist and anti-realist (or perspectivist) statements, enhancing his relevance from the perspective of PR.

The aim of this article is to explore the compatibility of PR with frequentist statistics, focusing on Neyman’s approach. I seek to uncover the implications of this examination for PR.

The structure of the article runs as follows. Firstly, in Section 2, The PR assumptions are presented (2.1). Based on the problem of the optional stopping rule a motivating example of how PR can be integrated into frequentist statistics in general is offered (2.2). Next, the perspective is narrowed down. A reconstruction of Neyman’s conception of statistical inference with an emphasis on his philosophical views is presented and compared with PR. Aspects in which Neyman’s methodological and philosophical views aligns with realism (3.1) and perspectivism (3.2) are discussed. In Section 4, antirealistic (4.1) pragmatistic (4.2) and antipluralistic (4.3) elements of Neyman’s ideas are delved into. Section 5 explores the authenticity of perspectives (5.1) and provides some solutions (5.2-5.4) to the issues raised within the three aspects discussed in 4.1-4.3. Section 6 summarises the findings.

2. PR as Applied to Frequentist Statistics

2.1. Assumptions of PR

Perspectival realism serves as a middle ground between the extremes of objective realism and social constructivism. The perspectivist premiss (1), introduced in Sect. 1, implies that perspectival realism advocates for epistemic pluralism (Premiss I) according to which perspectival knowledge about mind-independent states of affairs viewed from different and sometimes incompatible angles is equally valid epistemically. This is because any knowledge of objectively existing facts concerning objects or processes can only be acquired from a perspective (see Massimi 2012). Furthermore, the fact that these perspective-relative claims are true concerning the same objectively existing state of affairs (as stated in the realist Premiss 2 from Section 1) implies that (Premiss II) these claims or their justifications retain, cross-perspectively, their ‘performance adequacy’ as evaluated from the points of view of the internal standards set by each of the perspectives (see Massimi 2018a, 172); this means the epistemic performance of a scientific claim or a justification (method) must be judged as adequate given standards set by a perspective by practitioners of different scientific perspectives (see Massimi 2018d, 354).

Three primary versions of PR can be distinguished. Two refer to the issue of truth-value of statements and in the working classification used in this paper, they are regarded as 'semantic-ontological'. The first asserts that scientific claims can be deemed true relative to a given perspective and "not true simpliciter" (see Crețu 2020, 1-2). This implies that although a scientific claim is either true or false as a claim framed within a perspective, the question of its truth-value outside that perspective remains meaningless. The second asserts that it can be said that "models are useful to get calculations done but their representational content should not be taken literally as giving us a true story about what the target system is like" because they are about "a modal aspect: it's about exploring and ruling out the space of possibilities in domains that are still very much open-ended for scientific discovery" (Massimi 2018c, 36-38). Model's "being about X is not purported to stand in any mapping relation to worldly-states-of-affairs (X) so as to fulfil the realist quest via a plurality of partially accurate models of X, each of which may give a partial, yet accurate, and veridical image of X." (Massimi 2018c, 38). The third variant shifts focus from a claim to its justification(s): "the truth makers of our beliefs are non-perspectival facts about nature, yet the justification of our beliefs is intrinsically perspectival and rooted in our epistemic perspectives as human agents" (Massimi 2012, 28); this version I call epistemic PR.

The three variants of PR are prone to criticism. The first type is vulnerable to charges of relativism: that there are no non-perspectival true claims entails that no non-perspectival facts are illuminated by their meaning, while realism appears to assume that science is (truly) telling what non-perspectival facts are (see Chakravartty 2010). The aim of this article does not allow for a consideration of versions of PR that are susceptible to objections of being non-realist positions. The second type directly contradicts the very notion of error by denying that the true value of a parameter serves as the truth-maker for a rightly asserted true hypothesis, which is presupposed by the idea of the probability of making a right/false assertion. Neyman presupposed the existence of the true value of a hypothesis' parameter(s) and of the risk of committing an error by accepting a false statement about this value. They were meant to be independent of a researcher's ignorance regarding the truthfulness/falsehood of the value in the case of the singular test (see Sect. 3.1). This presumes that some accepted statements about a parameter value will be in mapping relation to worldly states of affairs, rendering the second version of PR incompatible. This mentioned feature of Neyman's method also implies that the claim about parameter value encapsulated in a statistical hypothesis can be meaningfully thought of as possibly true non-perspectivally which is contradictory to what the first type of PR states. I find the above reasons sufficient not to consider the first two, ontological versions of PR in my analysis and to focus on the more balanced, and weakest, epistemic version of PR. The choice is also dictated by the concern to restrict the size of the article. However, the third version of PR is susceptible to the objection of being philosophically trivial, given its weaker nature. To address this, I consider the additional premise (III) that genuine, non-trivial perspectival justifications exist (see Massimi 2012).

The notion of a perspective in the literature is quite vague. It ranges from a broad type of perspectives like research traditions to narrow perspectives intricate theoretical frameworks or even individual attitudes of a scientist or group of scientists (see Crețu 2022). The perspectival aspects of statistical methodology, discussed in this paper, fall within both these broad and narrow categories.

On one hand, this methodology encompasses principles or assumptions that form part of the working stance of a scientist, which is classified as a narrow perspective (see Crețu 2022, 522-523). On the other hand, these methodological attitudes are “second-order (methodological-epistemic) principles that can *justify* the scientific knowledge claims advanced” (Massimi 2019, 3), which is classified as a broad perspective (see Crețu 2022, 522). A narrow perspective might entail specific assumptions about, for example, the mathematical model governing probability distribution in Bayesianism or the model defining the probability distribution of sampled units in frequentism. In contrast, a broad perspective, within the context of research traditions, could involve the statistical methodology paradigm adopted, such as Bayesian, frequentist, likelihoodist, or Akaikean.

This paper critically examines the perspectival nature of the frequentist sampling scheme and inferential pattern, both at a general level (e.g., frequentist vs. Bayesian methodological traditions or approaches) and a detailed level (particularly in establishing error risk levels or the specifics of observational patterns).

For the sake of clarity of the following considerations and due to the diverse interpretations of ‘perspective’ in philosophical discussions, it’s worthwhile to clarify how perspective might be understood in the simplest sense. Thus, a working definition of perspective in PR is proposed. A perspective (in PR as applied to statistics) is an assumption included in the statistical inference/justification (e.g. about the data collection protocol, about some feature of a statistical model, or inferential procedure) which is made based on either subjective or non-epistemic[1] reasons relative to a particular research community.[2]

2.2. The Optional Stopping Case-study

An interesting illustration of the PR's capacity to encompass the frequentist statistical methodology can be found in the examination of the issue of *optional stopping rules*[3] (see, e.g., Savage 1962; Lindley, Phillips 1976). The problem can be illustrated using an example where a hypothesis about the sex ratio of koala pouch young in poor physical conditions is tested (see McCarthy 2007, 31-33).

Assume that the ecological substantive hypothesis in question posits that the proportion of female pouch young koalas in the population is 50% (the number of males and females is equal). The reasonable alternative hypothesis states that the proportion exceeds 50%, indicating female dominance.

This thought experiment involves a researcher following a specific path in the field and encountering 12 koala mothers, each with a pouch young. Consider the raw observational data, or empirical phenomena pertaining to the tested hypothesis, which the researcher experiences, as observing a distinct sequence of 12 pouch young koalas, either female (f) or male (m). Let's consider that the observed sequence was, for instance: m, f, f, f, f, f, f, f, f, m, f, m (I refer to this definite observational data as D for brevity). Prior to embarking on the predefined route, which includes 12 koala mothers with their offspring, the researcher had to make a critical decision regarding the experimental protocol. This decision entailed establishing the procedural rule governing the termination of data collection. There are at least two potential

---

[1] The terms 'subjective' and 'non-epistemic' are not synonymous, and they do not necessarily exclude each other. For instance, in frequentist statistics, the choice of specific error risk levels can have objective justifications, such as adhering to societal standards related to utilities or risks. This type of justification is pragmatic rather than epistemic. Conversely, a Bayesian's subjective justification for assigning a prior probability to a hypothesis based on their individual experience and knowledge related to the subject matter is both subjective and epistemic in nature. Furthermore, when a researcher rounds up data for personal convenience or includes fabricated data in statistical inference to enhance their chances of publication, these actions are both subjective and non-epistemic

[2] A research community can be defined at various levels, depending on the context. This can range from a collective of researchers sharing a common approach to an extreme case where a single individual forms their research community. For instance, consider the subjective Bayesian approach. It can serve as a reason for a research team, or for an individual researcher working independently, to choose to assign a particular prior probability to the hypothesis under investigation, based on their adherence to the subjective Bayesian framework.

[3] The issue of stopping rules has been a subject of extensive discussion among philosophers and methodologists, approached from various perspectives. In this subsection, I use a specific case to illustrate how frequentist methodology operates and relate it to PR. It's important to clarify that the purpose of this paper is not to analyse how this exemplified frequentist feature compares to alternative methods that adhere to the likelihood principle, which is insensitive to stopping rules. Nor does it aim to engage in a discussion of whether the frequentist violation of that principle is right or wrong. In essence, this article's primary focus is to analyse the interplay between PR and Neyman's frequentism, rather than delving into methodological problems or the pros and cons of specific aspects of frequentist statistics in comparison to alternative approaches.

scenarios concerning the protocol a researcher employed, under the assumption of which D could have been observed. The data could have been observed either by terminating sampling after recording the 12th individual ($S1$), or by terminating data collection when the 3rd male was recorded ($S2$). Let's consider that once the researcher had selected one of the protocols and encountered D during their observation, they subsequently commissioned a statistical analysis. It's assumed that the analyst is aware of whether the researcher had assumed protocol $S1$ or $S2$ and that the empirical data collected in a subsequent observation was D. There are two potential statistical inferences based on what the researcher could have conveyed to the analyst (statistician):

If the researcher's report was $S1$ and D, the sampling follows a binomial distribution. This distribution models the probabilities of obtaining n female samples until reaching the fixed total of 12 trials. The cumulative probability P_1 of observing the data (with 9 females) and more extreme outcomes (in this case of having 10, 11, or 12 females in the sample) equals 0.073. Consequently, given 0.05 error rate threshold, the observed female ratio in the sample (0.75) is not significantly different from (greater than) the hypothesised population ratio (0.5). Therefore the conclusion of the test is not to reject the hypothesis.

If the report were $S2$ and D, the sampling scheme would be described by a different model, specifically the negative binomial distribution. This distribution accounts for the probability of collecting a certain number of females before reaching a fixed total of 3 males in the sample. The corresponding p-value P_2 includes the probability of observing 9 females and less likely scenarios: 10 female records, $11, 12, 13, 14,$ and so on. In this case, the p-value amounts to 0.033, which is significantly below the conventional 0.05 error threshold. Consequently, the conclusion is to reject the hypothesis that the population ratio is 0.5.[4]

Herein lies an epistemic anomaly: two distinct sampling strategies, linked to different statistical experiment models, can yield disparate verdicts regarding the acceptance or rejection of the same substantive hypothesis., in the light of, allegedly, "identical" (McCarthy 2007, 33) sequence of empirical data in both cases, specifically the aforementioned D. Presented below is an illustration of how this particular frequentist methodological issue can exemplify PR.

Considering that perspectives can be principles or assumptions that form part of the

---

[4] Although the two results are not identical, p-values of 0.073 and 0.033 are not very different in terms of the strength of evidence in the data against the tested hypothesis. This suggests that the orthodox use of 'accept vs. reject' without grading may not always be practically advisable.

working stance of a scientist, $S1$ and $S2$ can be viewed as distinct methodological perspectives. These perspectives diverge regarding their knowledge claims due to their assumption or derivation of varying statistical hypotheses, specifically differing statements concerning probability distributions. This distinction arises from the disparate sampling spaces and models used to formulate the tested statistical hypotheses in each case. The perspectival evidence considered in these two hypothetical cases also differs. This discrepancy arises from the adoption of distinct sets of relevant information (evidence) for inferential purposes, even if the raw observational data D (a definite sequence of male and female pouch young) remains the same in both scenarios. The experimental protocol is defined here as the pre-observational 'rule' that guides a scientist's termination of collecting observations. It's therefore distinct from the (information about) the observed states of affairs D but serves as the perspectival standard for discerning and determining which information from the observed states of affairs qualifies as evidence. In $S1$, the evidence considered can be expressed by the proposition: 'exactly three males and nine females were recorded in the sample until (and including) the twelfth trial'. In sampling framework $S2$ (and its related statistical model), partial information about the order of males is utilized, and the evidence can be expressed as: 'exactly three males were recorded in the sample until (and including) the twelfth trial, and the twelfth trial recorded in the sample was male'. It is noticeable that the second set of evidence implies the first, but not vice versa, indicating that the evidence considered is not equivalent in both cases (see Kubiak 2014, 138-139).

Despite both possible observational viewpoints are yielding distinct perspectival claims and justifications for the conclusions drawn, which substantiates Premiss 1, the same substantive hypothesis is statistically defined and examined in both instances. Both statistical hypotheses describe a mind-independent state of affairs, an objectively existing population characteristic—the proportion of pouch young females. A shared realist element is one of the model's parameter value—probability of a female in a trial equal to 0.5, which signifies the population ratio. This aligns with Premiss 2.

Epistemic PR focuses on the aspect of justification. A statistical procedure/framework serves as the justification for a conclusion drawn with the use of it (obviously, including the empirical evidence obtained). Therefore, in asking about the performance adequacy of the discussed perspectival statistical set-ups/procedures it's important to know their performance in being epistemically successful inferential, and by that justificatory, tools. In both $S1$ and $S2$ the method assumes definite performance adequacy in them being able to yield a true conclusion if the state of the world, as described in the substantive hypothesis, is indeed true.

While the example considered resulted in different conclusions in both cases, this discrepancy is not at odds with the principles of PR. The considered method operates under the assumption that if the proportion of males in the pouch young population is 0.5, then the conclusion drawn from $S2$, over multiple iterations, will retain high performance adequacy. This expectation arises from the fact that if observations using the $S2$ sampling strategy were repeated numerous times, the method would reliably lead to the acceptance of the true hypothesis that the population ratio is 0.5, closely aligning with the predefined standards (with an error risk close to 5%). The same principle applies to sampling strategy $S1$. Therefore, if the male proportion in the pouch young population is 0.5, then both distributions that express the hypothesis tested are true. And if the value 0.5 of the parameter p is true in both cases of application of different stopping rules, both perspectival protocols will nominally retain their epistemic performance adequately to standards set for both models; this is a cross-perspectivally recognizable methodological fact. Thereby Premiss II of PR is fulfilled. It also seems that the perspectival assumptions $S1$ and $S1$, at least in the considered case where the number of trials and the observed sex ratio are the same, hold equal epistemic validity—there doesn’t appear to be any clear and objective reason why the researcher who makes the assumption should give epistemic preference to one over the other (although pragmatic reason may come into play). This means the requirement of epistemic pluralism (Premiss I) is satisfied too.

The researcher’s choice of $S1$ or $S2$ before their observation and the subsequent analysis by the statistician doesn’t undermine the perspectival nature of the analyst’s justification for their conclusion. The perspectival nature of the analyst’s justification doesn’t involve any new decisions regarding how to interpret D after the observation during the analysis stage. The experimental protocol adopted by the researcher predetermines what conclusion an analyst will draw (or how they will justify their conclusion) if D is observed by the researcher. The analyst’s justification becomes perspectival due to their reliance on the experimental protocol, an assumption that, as I have previously discussed, is perspectival in nature. If one of the premises that forms the basis for the conclusion, such as $S1$ or $S2$ assumed to be the experimental protocol, is perspectival, then such justification can be considered perspectival. The temporal precedence of the perspectival factor aligns with the conventional approach to perspectivism (see, e.g. Giere’s 2006 discussion on the perspectival nature of colour vision).

In the following two sections of this article, I investigate Neyman’s frequentist methodology taken alongside his philosophical interpretation. This investigation allows for a comparison with PR and an examination of both consistencies (Section 3) and inconsistencies

(Section 4) between Neyman's ideas and PR.

3. Neyman's Theory—Elements Coherent with PR

Jerzy Neyman was not a professional philosopher and didn't employ the typical terminology found in philosophical discussions when expressing his philosophical views. Nevertheless, some aspects of his philosophical stance have been examined and debated in philosophical literature (e.g., Hacking 1965; Mayo, Spanos 2006). However, these discussions have not directly considered his ideas from the perspective of PR. In this section, I will appropriately frame and discuss those elements of his methodological and philosophical conceptions that share similarities with both realism and perspectivism.

3.1. Neyman's Views and Realism

Some of Neyman's basic methodological and meta-methodological conceptions align with realist ideas. To begin with, Neyman did not dismiss the assumption of the existence of an independent reality, which is an ontological aspect of realism. He asserted that each study involves a "true state of nature" that is unknown (Neyman 1971, 2). This is represented by the 'true value' of the hypothesis' parameters. According to Neyman, the values of the hypothesis' parameters that a researcher asks about, were "generally unknown constants" (Neyman 1937, 343). The constant value of the statistical model's parameter is, as such, a mathematical concept. However, Neyman asserts that "there are real objects that correspond to these abstract concepts in a certain sense" (Neyman 1952, 24). Consequently, the truthfulness of the hypothesis parameter's value implies that this value, in some way, corresponds to, or denotes an unknown, but independently existing state of affairs in the real world. In this regard, Neyman appears to be endorsing at least a form of ontological realism.

The fundamental concept behind the application of statistical methodologies to experiments and observations is to "assume that the real value of the sought-after quantity exists […] and—based on laws of large numbers—to seek for calculable measurement results' functions[5] that can be considered approximations of the 'true value' and mean error" (Neyman 1923a[1], 19, auth. transl.). Therefore, it becomes evident that Neyman's ideal is to arrive at conclusions represented by these functions, where "numerical values of mathematical

---

[5] A *function* of the results of measurements means a value that summarises the data obtained from all trials. It's standardly called *statistic*. For example, the numerical value $\bar{x}$, an estimate of the population mean, is a function of the observed values $x_1 \dots x_n$.

formulas more or less agree with the results of the actual measurements” (Neyman 1952, 24). The values of these functions of actual measurements are expected to be approximately the same as the “real values” that exist independently in the real world, which assumes an epistemic realist stance. Notably, Neyman employs the plural form when discussing “functions”, hinting at the potential use of various functions to derive results from the empirical evidence obtained. Nevertheless, all these potentially distinct outcomes are expected to align with the evidence to a certain degree and provide approximations of the objective truth.

Next, the conception of the method’s reliability is rooted in the concepts of the two types of error probabilities: the probability of rejecting the tested hypothesis if it’s true and the probability of accepting it when it’s false (Neyman 1952, 55). A true hypothesis is characterized by having its specified parameter range encompassing this unknown, real value. The method’s reliability is contingent on its performance in producing true conclusions in the long run. Consequently, a form of epistemic realism appears to underpin the method’s long-term reliability.

Finally, Neyman required research schemes to account for real-world factors that exist objectively and independently of the research scheme itself. Neglecting these factors could potentially impact the alignment between a physical (substantial) hypothesis and a statistical hypothesis. Neyman’s illustration of this matter draws from Fisher's well-known hypothetical scenario involving a tea-tasting lady, who is challenged to determine whether the tea or milk was poured into a cup first based solely on the taste of the tea; the substantive hypothesis is that she is not able to distinguish it. An independent factor in this scenario could be, for instance, the lady’s ability to associate her impression of a specific pouring sequence with the thickness of the cup, which she can perceive through touch. If the experimental design fails to consider this factor and one pouring method is predominantly used with thinner cups while the other is associated with thicker cups, then the substantive hypothesis of the lady lacking the ability may be true whereas the corresponding statistical hypothesis, which involves the distribution of probabilities for potential experimental outcomes under the assumption of her lacking this ability, would be false (Neyman 1950, 282-291). Hence, it becomes evident that the adequacy of the model (or statistical hypothesis) in describing independently existing real-world characteristics is paramount when assessing the truthfulness of that model.

In conclusion, it can be inferred that several fundamental methodological and meta-methodological concepts proposed by Neyman align with realist ideas. Hypotheses in his framework refer to independently existing realities, and are categorized as either true or false objectively (ontically). Furthermore, the entire research framework is anticipated to be congruent with independently existing, genuine factors. This reaffirms the presence of ontological and epistemic realist ideas in Neyman's thought.

3.2. Neyman's Views and Perspectivism

While a true statistical hypothesis encapsulates a value of an independently existing real quantity, it does so by attributing empirical meaning to this value. A statistical hypothesis constitutes a statement concerning the probability (density) distribution of a random variable, with this random variable being a function of a collection of random phenomena derived during the course of a random experiment. This implies that both the distribution and, consequently, the hypothesis are somewhat shaped by the particulars of the observational (experimental) set-up.

Specifically, the concept of probability within the realm of statistical hypotheses pertains not to physical entities or the characteristics of physical entities but rather to the attributes of physical events aligned with an observational arrangement. In simpler terms, probability is attributed not to objects themselves but to the events associated with an observational setup, as elucidated by Neyman (1952, 10-12). This distinction becomes evident in Neyman's response to Jeffrey's hypothetical scenario involving two boxes. One box contains one white and one black ball, whereas the other contains one white and two black balls. Initially, a box is chosen randomly, followed by the random selection of a ball from the chosen box. Consider Neyman's definition of probability "the probability $P(B|A)$, of an object A having the property B will be defined as the ratio $P(B|A) = m(B)/m(A)$" (Neyman 1937, 337). When applied to this toy example, it's not telling about the probability of the selected ball having the property of being white:

> "the objects are obviously not balls, but pairs of random selections, the first of a box, and the second of a ball [thus], the probability sought is that of a pair of selections ending with a white ball" (Neyman 1952, 11).

Thus, in Neyman's perspective, probabilities are directly associated with characteristics of observational designs or procedures. Similarly, statistical hypotheses, which are statements concerning probability distributions, are also relative to these designs. Even when they are meant to describe substantive hypotheses about the mechanisms or other attributes of an objectively existing reality, they do so exclusively through the lens of experimental constructs that dictate what can be observed and which data functions (test statistics) are utilized. This Neymanian perspective aligns with the implications of the stopping rule problem discussed in Section 2.

The perspectival aspect of frequentist inference is also evident in the socially guided differentiated validity of two types of error (false rejection of a hypothesis called the first type of error and false acceptance of a hypothesis dependent on the power to detect the true alternative): "[…] with rare exceptions, the importance of the two errors is different, and this difference must be taken into consideration when selecting the appropriate test" (Neyman, 1950, 261).[6] From one perspective, it might be more important to prevent a false rejection error, while from another "(…) the desirable property of the test of H is as high a power as practicable, perhaps with some neglect of the probability of rejecting H when true" (Neyman, 1971, 4). Consider, for example, a lady who claims to have the ability to distinguish by taste whether milk or tea was first poured into a cup. From the perspective of a commission responsible for evaluating the lady's claim and potentially granting her recognition or rewards for this skill, the more significant error to avoid might be mistakenly granting the claim when it is, in fact, false. However, from the lady's viewpoint, the error of falsely asserting that she lacks this ability might be more critical to avoid (Neyman, 1950, 274).

Therefore, depending on the perspective adopted, the error rates might be set differently and ultimately lead to different decisions regarding whether to reject or accept the hypothesis based on the same statistical test (understood as function of the data) and the same data. All other factors being equal, when one choose to decrease the risk of one type of error, it results in an increase in the risk of the other type. The different risk settings in the two pragmatic contexts described above don't make the error risk rates inherently incommensurable—they are expressed using the same scale, i.e., the measure of probability. What makes these two cases perspectival is that, while different choices can lead to different testing outcomes, the method itself doesn't favour one risk setting over the other from an epistemic perspective. The choice of a specific risk balance between the two types of error is purely pragmatic.

The conclusion from sections 3.1 and 3.2 is that on Neyman's account statistical hypotheses, justifications and conclusions based on the use of statistical tools are always relative to the perspectival idealised assumptions, experimental constructs and pragmatic reasons. However, they simultaneously refer to the perspective-independent, true states of affairs: real parameter values and real experimental circumstances (see, e.g., Neyman 1934).

---

[6] The idea of error acceptance depending on the social context is classically referred to as epistemic risk in the literature (see Rudner 1953).

Neyman considered some of these conclusions to be true (in the classical sense) to the extent defined by the error risks. As a result, both premisses 1 and 2 are satisfied.

Furthermore, Neyman seemed to acknowledge the potential for equally valid perspectives regarding the fundamental assumptions of scientific methodology:

> "[…] in theoretical work, the choice between several equally legitimate theories is a matter of personal taste. In problems of application, the personal taste is again the decisive moment, but it is certainly influenced by considerations of relative convenience and empirical facts" (Neyman 1937, 336 footnote *).

Here, the term 'theories' pertains to methodological frameworks. If these frameworks can be considered 'equally legitimate' it suggests the existence of epistemic pluralism in perspectives, which aligns with an aspect of PR (Premiss I).

The aspects of Neyman's ideas that can be seen as having similarities to both realism and perspectivism align quite well with perspectival realism up to this point. However, Neyman's approach also incorporates other crucial elements that appear to clash with PR and introduce internal inconsistencies within his statements. These elements pertain to the notion of scientific concepts being fictional, the pragmatic (non-epistemic) interpretation of scientific assertions, and the concept of normative anti-pluralism. I will explore these three topics in the subsequent section.

4. Neyman's Theory—Elements Potentially Inconsistent with PR

4.1. Fictional Nature of Scientific Concepts

According to Neyman, statistical hypotheses are formulated based on idealised assumptions that do not hold true regarding the real world and empirical evidence:

> "The objects in the real world, or rather our sensations connected with them, are always more or less vague and since the time of Kant, it has been realized that no general statement concerning them is possible. The human mind grew tired of this vagueness and constructed a science from which everything that is vague is excluded—this is mathematics. […] there are many mathematical theories that are successfully applied to practical problems. However, this does not mean that these theories deal with real objects [...] the theory [of mathematical statistics] itself deals with abstract concepts not existing in the real world" (Neyman, 1952, 23-24).

The issue of correspondence between the real world, scientific statements, and evidence is a challenge that exists at the level of an individual trial and the empirical evidence derived from it. This can be illustrated by referencing one of Neyman's early papers (1923a^1)[7], in which he introduced a general design for a field experiment aimed at comparing various crop types based on their potential yields.

In that paper, he explored the experimental design involving the random allocation of seeds to plots within a test field. Each seed sowing was associated with what he referred to as the 'true yield'. Nevertheless, the result obtained from measuring the yield of specific yeast varieties at a particular plot did not represent the true yield of that variety at that plot. The true yield, in this context, was an unknown, fixed value (Neyman 1923a^1, 465-67).

This discrepancy arises from a technical error in the measurement process. The concept of true yield is, in fact, an idealised abstraction, representing the average value derived from an infinite number of measurements where all conditions are identical except for the variation introduced by random technical errors, resulting in measurement inaccuracies. This type of error is distinct from the errors involved in statistical inference about hypotheses, and it is important to note that "no statistical methods can improve the accuracy of the experiment beyond the limits fixed by the technical random error" (Neyman et al. 1935, 110).

Hence, there exists no direct correspondence between the 'true yield' derived from a specific trial (a scientific concept) and the observed yield obtained from that trial. Additionally, the existence of random technical errors implies that there is no direct equivalence between the actual yield present at a particular plot (an ontological fact) and the observed yield recorded at that plot (an empirical fact). The disparities between these two notions become quite apparent when one recognizes that the 'true yield' at a specific trial (plot) is essentially an a priori counterfactual state of affairs. This is due to the infinite number of unrealized counterfactual measurements inherent in the concept of the true yield (see Rubin 1990).

To emphasise the absence of equivalence between a scientific concept and observational facts, Neyman discerned two distinct meanings of terms like 'yield' when employed in two different facets of the scientific process: one for describing empirical data (referred to by Neyman as 'pure empiricism') and another for making inferences within a scientific framework (Neyman 1923a^1, 18). In the first case, one can speak of the result(s) of empirical observations (measurements), whereas in the second—of scientific concepts that put

---

[7] Originally published in Polish. In this article, reference to the Polish original will take the form of "1923a^1" while the form "1923a^2" will refer to the fragment translated in 1990 by D.M. Dąbrowska and T.P. Speed.

these observations into more general frames. What distinguishes the use of a term in the context of a scientific concept is that "all scientific terms, which are defining properties and relations between investigated objects, are fictions" (Neyman 1923a[1], 18). Neyman's notion of the 'true yield' serves as an illustration of such a scientific, and thus fictitious, term. The scientific concept of a yield at a specific plot is fictional since it does not align with the real actual state of affairs, which is finite.[8] It's fictitious concerning a yield at a particular plot because it does not align with the potential observational situation either. This is because an infinite series of measurements, which the concept is assumed to represent, is impossible to actualise. The notion that scientific concepts are fictional somewhat contradicts Neyman's realistic views as presented in section 3.1 and Premiss 2. Issues presented and discussed in this subsection are interconnected with those presented in the following subsection.

4.2. The Pragmatistic (Non-epistemic) and Long-run-based Interpretation of Reliability of a Statistical Procedure

The second element of Neyman's theory that could be inconsistent with PR is his position that accepting a scientific statement does not result in any belief regarding the truthfulness of that specific scientific statement:

> "The terms 'accepting' and 'rejecting' are very convenient and are well-established. It is important, however, to keep their exact meaning in mind, and to discard various additional implications which may be suggested by intuition. Thus, to accept [or reject respectively] a hypothesis H means only to decide to take action A rather than action B. This does not mean that we necessarily believe that the hypothesis is true [or false respectively]" (Neyman 1950, 259).

This agrees with Neyman's moral-like postulate: "The beliefs of particular scientists are a very personal matter and it's useless to attempt to norm them by any dogmatic formula" (Neyman 1957b, 16).

Epistemic accounts of a hypothesis' confirmation (or disconfirmation) that do not rely on probability or degrees of belief are conceivable. However, Neyman and Pearson emphasised that accepting or rejecting a specific hypothesis based on a statistical test could not, for methodological reasons, be seen as having epistemic justification. In other words, a

---

[8] The concept was said to represent a statement about empirical realisation in an infinite string of measurements.

single statistical test of a hypothesis does not provide any measure of the degree of confirmation or disconfirmation of that hypothesis:

> "[…] as far as a particular hypothesis is concerned, no test based upon the theory of probability can by itself provide any valuable evidence of the truth or falsehood of […] hypothesis. However, we may look at the purpose of tests from another viewpoint. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour concerning them, in following which we insure that, in the long run of experience, we shall not be too often wrong" (Neyman, Pearson 1933, 291).

Almost twenty-five years later, Neyman continued to assert that statistical tests are decision rules with relative-frequency (long-run) performance. Error probabilities associated with such rules address questions like, "[h]ow frequently will the contemplated rule prescribe mass application of a given vaccine when, in fact, this vaccine is dangerously toxic?" (Neyman 1957b, 18). The assumption that the method's reliability applies not to a single testing situation but to a long sequence of tests implies that it does not serve the purpose of (dis)confirming a particular hypothesis in a single testing situation. The assumed absence of an epistemic interpretation for a single application of frequentist procedures is reflected in Neyman's pragmatic interpretation of the goal of the scientific investigation method using statistical tools. Although the "[…] theory was born and constructed with the view of diminishing the relative frequency of errors, particularly of 'important' errors" (Neyman 1977, 108), accepting a hypothesis is an act of will, a decision to behave as if the hypothesis were true. This decision is based on the assumption that the method we are employing is reliable enough not to lead us away from the truth in a sufficiently large fraction of practically important cases. This is why the final stage of accepting a hypothesis:

> "[…] amounts to taking a 'calculated risk', to an act of will to behave in the future (perhaps until new experiments are performed) in a particular manner, conforming with the outcome of the experiment. It is this act of adjusting our behaviour to the results of observations, that is the overlooked element of the final stages in scientific research and that is covered by the term 'inductive behaviour'" (Neyman 1957b, 12).

Neyman equated his concept of 'inductive behaviour' with the decision-theoretic idea of 'statistical decision-making' introduced by Wald (1950). In this view, one evaluates the expected loss associated from possible decisions over probability distribution of data for a specific hypothesis, where the hypothesis cannot itself be treated as a random variable (see Neyman 1957b; Neyman 1937, 343-344). The selection of the optimal decision rule relies on the analysis of these values for all possible hypotheses. The information necessary to achieve this is contained in decision rule’s ‘performance characteristic’. To clarify, consider a function of which the value represents the probability of a specific decision (e.g., accepting the tested hypothesis), and the domain comprises all possible hypotheses (hypothesis’ parameter values). The performance characteristic encompasses a set of such functions for all possible decisions. This approach involves describing the expected losses across all possible states of affairs, with the primary goal being to identify the rule that minimises the expected loss in the worst-case scenario(s). Subsequently, efforts are made to minimise the expected loss for, ideally, every other possible scenario (Neyman 1950, 1-14; 1957b, 18). Hence, the crucial aspect of the described feature of the method is that it’s pre-observational, encompassing all potential evidential outcomes. As a result, it doesn’t provide an epistemic interpretation of specific post-observational evidential conditions. Consequently, it doesn’t directly offer a post-observational degree of confirmation or epistemic justification for accepting a particular hypothesis. Additionally, it fundamentally operates on pragmatic grounds, further complicating its epistemic status. Scientific realism, on the other hand, appears to assume that scientific conclusions (results of specific research) are considered to be either true or very close to the truth, relying on statistical justifications that have provided them with a substantial degree of confirmation.

4.3. Anti-pluralistic Elements in Neyman’s Conception

In Section 3.2, I pointed out Neyman’s declaration that appeared to endorse methodological pluralism. However, in the subsequent part of the paper, there is an indication that his methodological solutions and views partially challenge this presumed pluralism.

Neyman’s thinking exhibits two anti-pluralistic dimensions, both reflecting a somewhat ‘God's-eye’ viewpoint. One can be described as the ‘in-theory’ perspective, while the other pertains to justifying the theory from a meta-level standpoint. The ‘in-theory’ perspective can be further categorized into two forms of anti-pluralism: bottom-up, which concerns the epistemic adequacy of models and setups, and top-down, which deals with the epistemic effectiveness of statistical inference.

The bottom-up perspective focuses on identifying and selecting the experiment model that aligns optimally with physical reality (as seen in Neyman 1950, 282-291). Neyman illustrated this by using the example of the tea-tasting lady (as discussed in 3.1). In conducting the experiment, it’s essential to define a suitable set of permissible hypotheses. For instance, one might question whether the alternative hypothesis to the lady lacking the ability (making random guesses) indicates a perfect guess or perfect misguidance. The lady might possess the ability to differentiate between pouring methods but tends to confuse one method with the other. Another aspect to consider is whether the question pertains to the lady’s capacity to distinguish between the two methods or to identify each method separately. In the latter scenario, the cups should be evaluated independently, not by comparing them within a pair. However, what if she can identify one of the methods but remains uncertain about the other? Does she have knowledge of the number of cups created using one of the methods that she will receive? If that’s the case, then the trials should be treated as dependent. Finally, it’s crucial to establish an appropriate technique for a random experiment, wherein any factor that could influence the alignment between a physical and a statistical hypothesis is mitigated through randomisation. An example of this, as mentioned in 3.1, pertains to the order of pouring. In summary, considering numerous factors encourages the pursuit of the most effective experimental setup rather than treating various potential setups as equally valid.

Another instance of Neyman’s bottom-up anti-pluralist approach is his theory concerning the use of sampling design to achieve optimal estimation (1934; 1938). In this context, discussing estimation is just as relevant as discussing hypothesis tests, given the duality between hypothesis tests and Neyman’s interval estimation technique. An estimation technique is essentially akin to conducting a series of hypothesis tests (see Neyman 1937, 372; Lehmann, Romano 2005,164-168). Neyman developed techniques for maximising the accuracy of estimation by considering additional (auxiliary) factors related to the population’s structure. This technique involves evaluating several mathematically equivalent methods for drawing a sample from the population and selecting the one that, based on this knowledge, provides the most accurate sampling design. In cases where the epistemic adequacy of certain assumptions in models or setups cannot be verified due to insufficient data (such as assumptions about distribution shape or sample dependence), it becomes challenging to make definitive choices. Neyman’s approach in such situations was to seek models or setups that minimise the need to make unverifiable assumptions. For example, when comparing alternative sampling schemes like purposive and stratified sampling, Neyman argued that it’s often uncertain whether hypotheses regarding the relationship between the research variable and an auxiliary variable are met. In such cases, it’s preferable to select a simpler, less accurate but ‘safer’ sampling scheme (see Neyman 1934). Neyman’s aversion to adopting a Bayesian methodological perspective was also driven by a similar desire to avoid relying on unverified assumptions (see the last paragraph of this section). All of the above indicates that Neyman advocated achieving optimal adequacy between theoretical models of observation and all known aspects of the investigated reality by fulfilling specific conditions like those mentioned earlier. This implies that he promoted a principled approach to narrowing down the possible observational perspectives from which a hypothesis could be tested to the one that best corresponds to reality.

The top-down type of Neyman's anti-pluralistic view on the choice of research perspective is perhaps best exemplified by the normative requirement to use a test whereby the probability of correctly rejecting a hypothesis would be maximal for a preassigned error of the first type:

> "if two different critical regions w_1 and w_2 are suggested, both insuring the same probability of error of the first kind, then the choice between these regions depends on their effectiveness in controlling the error of the second kind" (Neyman 1950, 304).

Originally, the rule was presented as applied to choosing among several test statistics (see Neyman and Pearson, 1933). However, this idea of minimising the error of the second kind can also be applied when choosing between different protocols for collecting data. This is because the critical regions w_1 and w_2, represent two distinct sets of possible observation outcomes that result in the rejection of a tested hypothesis, assuming the same risk of the first type of error in both cases. In protocols $S1$ and S2, these sets of outcomes that lead to rejection are different.

From this point of view, the two alternative perspectives adopted in the discussed example of testing the hypothesis that the number of males and females of pouch young is even will not be equally valid epistemically. Imagine a test is conducted in both cases using the so-called likelihood ratio test statistic, which Neyman and Pearson (1933, 298-301) found to yield greater power than any other test statistic in the case of point hypothesis testing against another point alternative when all other factors are equal.

Imagine the hypothesis (as discussed in 2.2) that the population ratio is 0.5 is tested against the alternative hypothesis that the ratio in question is 0.75. If the consideration of a test's power function "seems to be the proper rational basis for choosing the test" (Neyman, 1952, 58), then the perspective of sampling design related to $S2$ is preferable. This is because the likelihood ratio test devised based on sampling distribution $S2$ has higher power (equal to 0.46) to detect the true alternative compared to the power equal to 0.39 for detecting it in the case of the likelihood ratio test devised based on sampling distribution $S1$.

Neyman's meta-methodological views also seem to be in contrast with the idea of plurality of perspectives, as he suggested that, in principle, some general methodologies would be better suited for particular cases than others. Notably, he acknowledged that while the Bayesian methodological framework for testing or estimation can be mathematically valid, it could only be applied in practice under quite exceptional circumstances, specifically in cases where there's a lack of evidence to support assumptions about the probabilities of hypotheses: "Even if the parameters to be estimated, $\theta_1, \theta_2 \dots \theta_l$ could be considered as random variables, the elementary probability law *a priori*, $p(\theta_1, \theta_2 \dots \theta_l)$, is usually unknown, and hence the [Bayesian formula] cannot be used because of the lack of the necessary data" (1937, 343).

5. Philosophical Consequences

In this paper, I argued that PR can be applied to frequentist statistical methodology and shed light on some of its unique characteristics. However, when we consider Neyman's frequentism, we see that PR is only partially consistent with it. Neyman's views seem to balance between compatibility and incompatibility with PR, depending on the specific aspect of the methods under discussion. Moreover, the concerns raised in section 2.1 about the genuineness of perspectives remain unaddressed.

In this section, I will delve into the question of real and substantial presence of perspectives (5.1). Furthermore, I will investigate the extent to which it's feasible to resolve the challenges that have emerged from the juxtaposition of PR with Neyman's conceptions (5.2-5.4). Consequently, this section aims to demonstrate that by interpreting Neyman's statements and the characteristics of his methods appropriately, without necessarily conflicting with his views, it's possible to resolve the philosophical inconsistencies highlighted earlier. This will illustrate that Neyman's methodology and views can, under these interpretations or adjustments, be compatible with PR. Simultaneously, I will emphasise that it might be more suitable to embrace a weaker version of PR. In this interpretation, while recognising the existence of epistemic pluralism in perspectives to some degree, it's not deemed irremovable in principle.

5.1. Perspectives' Authenticity

Epistemic Perspectival Realism, like Perspectival Realism in general, could conceivably be reduced to a form of dispositionalism. This argument posits that perspectives can be elucidated by invoking the multifaceted dispositional nature of the causal properties inherent to the target system. On this account, dispositions are understood as genuinely occurring real properties that "[d]ispose the systems that have them to behave in particular ways in specific circumstances." (Chakravartty 2010, 409). Dispositions "[…] are non-perspectival facts: they are true whatever perspective one takes. One must take a perspective in order to investigate it, of course; that is, one must view the phenomena from a particular vantage point, use a particular sort of instrument, or perform a particular kind of experiment, to determine how a disposition manifests itself in that particular interaction. However, the facts produced by these investigations are perfectly non-perspectival ones" (Chakravartty 2010, 409). The crucial point to grasp is that the invocation of various dispositions, in the broadest sense, results in the observation of fundamentally distinct properties. This is exemplified in investigations regarding the corpuscular and wave aspects of light (Chakravartty 2010, 410-411). Likewise, inquiries into the hardness versus conductivity of a material, or a hornet species' carbon dioxide resistance versus its venomousness, would necessitate distinct experimental approaches and considerations of different properties.[9] Examining these dispositions as either solely relative to perspective or solely dependent on knowledge doesn't appear to offer any profound or thought-provoking philosophical insights.

The question arises of whether this argument applies to the features of frequentist statistics. When we consider the example of Koalas, we are essentially observing the same property (the sex of pouch young of randomly selected Koala females). This suggests there may be no causal evocation of different dispositional facts about the studied population. It would be even more unusual to suggest that two different dispositions are evoked in $S1$ and $S2$ when examining the same observations (including their order) in both cases. To be more precise, apart from the differences in knowledge claims and the evidence considered, the two aspects are genuinely perspectival because there is no causal evocation of different dispositions of the studied population implied by $S1$ and $S2$; observations in both cases could be assumed to be an identical empirical event (i.e. an identical sequence of trial outcomes). Even if one assumes that the evidence considered in both cases is different, including some information about the order in $S2$, it can only be interpreted as a result of the researcher's decision influenced by the choice of the observational protocol, rather than indicating any fundamental ontic difference in the nature of the observed phenomenon. However, one might

---

[9] For this reason, regarding the choice of a hypothesis as genuinely perspectival would be too weak a perspectivism because of merely focusing on different dispositions that one is interested in when deciding on what to investigate about the object of research.

contend that the justification in both cases also involves reference to sets of counterfactual possible absolute (non-relative) frequencies of outcomes, which do differ between $S1$ and $S2$. Similarly, the argument could be made that depending on the adopted sampling rule, one might encounter different observational data than the considered one that would be unattainable using the other rule. For instance, in the case of $S2$, the number of investigated koalas could have been either larger or smaller. Firstly, it's important to note that different sets of counterfactual unobserved outcomes should not be considered as representing two distinct 'facts produced' by different dispositional properties because they do not correspond to actually produced (and observed) different facts. Secondly, the difference in possible observations is quantitative rather than qualitative. It involves variations in the number of observed koala mothers, while the fundamental feature under scrutiny in both cases remains conceptually the same. The property under analysis could be framed as dispositional (the disposition/propensity of koalas to conceive male or female offspring) but there is no basis for recognising two essentially different dispositional properties being involved in $S1$ and $S2$.

The presence of perspectivism in Neyman's (and Pearson's) conception of hypothesis testing is even more evident. Although the sampling scheme and raw data may remain the same, various factors can change, such as the levels of error risks, data transformations (test statistics), or specific and necessary model assumptions. An epistemic justification for the choice or assumption of a particular methodological setting of this kind over other possible choices may not always be available. In such cases, the alternative choices can be considered equally valid from an epistemic standpoint (see section 5.4 for further discussion). The justification for these choices may then rely solely on pragmatic considerations. These varying assumptions, such as those concerning the rates of the two types of error risk, can lead to different justifications for conclusions and potentially different conclusions drawn from the same observed raw data.

It should be noted that the outcomes of random sampling, even under the same model, are inherently unique as they do not have to include the same individuals each time. Additionally, one researcher or team may not have access to samples drawn by another. From this point of view, any specific observation made by a scientist or team could be seen as a unique perspective. Some might view this understanding of perspective as stemming from the randomness of the sampling process, which might appear too trivial to be considered genuinely perspectival in a philosophically significant sense. The perspectival differences discussed so far are grounded in disparities in methodological or theoretical assumptions. The 'perspectival' difference in random samples doesn't arise from the application of distinct observational, inferential, or interpretative assumptions. When examining scientific methodologies, perspectives understood as ‘human vantage points’ entail certain methodological distinctions driven by human choices and decisions. Generating different outcomes solely due to the randomness of data collection processes might not be considered sufficiently human-driven in this context. The distinctions between cases like $S1$ and $S2$, as well as differences arising from the adoption of various testing procedures, especially regarding error rates, are perspectival in the sense that they originate from different methodological and pragmatic vantage points. They are non-trivially human-driven. These considerations affirm Premiss III, which asserts the authenticity and non-trivial nature of perspectival justifications evident in frequentist statistical methodology.

5.2. Overcoming the Unclear Status of Scientific Concepts from Realist Perspective

As mentioned in 4.1, Neyman emphasised the fictional nature of hypotheses (scientific models) and the tenuous connection between evidence (data) and statistical models. However, this doesn’t imply that models (and consequently statistical hypotheses) are entirely fictitious in their relation to an independent reality (referred to as Neyman’s ‘real world’). This is because the real world cannot be reduced solely to the realm of empirical data (actual, or potential), although it’s “inhabited by data” (Kass, 2011, 2). Neyman only indicated that fictional nature of scientific statements becomes evident when they are contrasted with ‘empiricism’. In philosophy, empiricism posits that scientific knowledge reduces to, or is equivalent to actual, or potential empirical evidence. Neyman seemed sceptical of accepting ‘pure empiricism’ which suggests that he acknowledged that a scientific model conveys more than just information about actual or potential observational outcomes. It can be true in terms of it representing the independent reality of mechanisms, propensities or static features of populations. These elements, when investigated through experiments or observations, can serve as (probabilistic) causes for obtaining specific outcomes. Certainly, Neyman was referring to an actual, albeit unknown, characteristic of the studied population and a hypothesis that posits a statement about this characteristic which can either be true or false. Hence, Neyman’s assertion regarding the fictitious nature of all scientific terms should be interpreted in the context of the correspondence between weak (actual or potential) evidence and models, where evidence is assumed to be something different than the real world being modelled.

The understanding of the probability-empirical data relationship that allows for sidestepping the anti-realist inclinations of the infinitesimal approach is the concept according to which "[…] probabilities are understood as mathematically convenient approximations to long-run relative frequencies" (Pitman 1997, 11). This can be further distilled into a more explicit understanding: "[…] probability of an event A is the expected or estimated relative frequency of A in a large number of trials" (Pitman 1997, 12). This viewpoint, even though it reflects the prevailing practical stance of contemporary frequentists on probability, should not be categorized as a strict definition or elucidation of probability. Instead, it should be seen as a demonstration of how theoretical probability, as defined by figures like Neyman (1937, 336-337) that does not refer to, nor entail, frequency interpretation—can be practically applied through supplementary mathematical theorems to yield estimates of frequencies.

The neutrality of the theoretical probability definition permits its flexible utilization in various contexts, contingent upon the specific aspect of the model under consideration. This versatility enables us to connect different facets of reality to which it pertains, encompassing both epistemic, related to perspectivism, and ontic, related to realism. Therefore, one could conceptualise a statistical model as providing insights into unobserved ontic dispositions or propensities of objects or collectives of objects, akin to mechanisms (see, e.g., Peirce 1910/1932, 2.664; Popper 1959, 37). For example, the parameter $p = 0.75$ can be interpreted as indicating the propensity of two koalas to conceive female rather than male, which is estimated at 0.75. Similarly, the unknown true yield at a specific plot could represent the propensity of physical (system of) objects to exhibit certain observable behaviours under repeated observations in certain conditions. Even though obtaining an exact empirical realization of this hypothesized behaviour is impossible, the accepted scientific statement can still be considered approximately true with respect to the unobserved actual or dispositional features, as well as the ontic mechanisms of the real world.

In 4.1, Neyman's perspective is that the real world is inherently 'vague' while the mathematical models used to describe these ontic states are 'fictional' idealisations that deviate from these ontic states. For instance, consider a hypothetical example of a wheel of fortune that was intended by its designer to be perfectly fair (although this is known to be rather uncommon in practice). The actual ontic mechanism or nature of its propensity must inherently differ from the mathematical model that describes it, as there will never be such a perfect symmetry in the real-world mechanism as envisioned by the designer's model. Nonetheless, idealisation doesn't necessarily negate the approximate truthfulness of such models, therefore, realism still applies. Ultimately, it's understood that it's impossible to have models that are entirely adequate, which is why Neyman stops short of asserting that a model must be 'found satisfactory' solely in terms of its empirical adequacy (1952, 27).

5.3. Overcoming The lack of Epistemic Interpretation of a Single Outcome

Neyman’s stance on realism appears to be somewhat ambiguous. He emphasized that a statistical method serves as a tool for making practical decisions rather than obtaining true beliefs. However, at the same time, he also assumed that the method's trustworthiness in providing these pragmatically valuable conclusions, over the long term, relies on the method’s ability to yield conclusions that are frequently true in a realistic sense, as discussed in section 3.1.

Referencing a single-case probabilistic measure of the degree of belief in a hypothesis is not the only way to render to a scientific method epistemic realist characteristics. Epistemic realism can also be applied to a set of assertions as an outcome of the application of N-P methods. Neyman himself acknowledged this when he stated that “in the long run of experience, we shall not be too often wrong” (Neyman, Pearson 1933, 291). Furthermore, Neyman suggests that a procedure can be considered reliable when it’s consistently applied across various research contexts, involving different hypotheses and error rates set at varying levels. According to Neyman, the Central Limit Theorem (he likely meant the Law Of Large Numbers, which is implied by the Central Limit Theorem) enables us to infer that the relative frequency of errors will tend to be close to the arithmetic mean of the errors, regardless of the specific research context (Neyman 1977, 108-109). In essence, this implies that the average error serves as an indicator of the proportion of assertions within a body of outcomes from statistical tests that are true, although the question of the truthfulness of any particular one must be abandoned.

Hence, a distinctive form of realism can be applied as an interpretation of Neyman’s methodology. In this context, this specific form of realism entails assessing a collection of outcomes that, when combined, constitute a body of scientific knowledge, where a significant portion is considered true. The general epistemic credibility of the method can also be based on the concept of pre-observational total probability that the method will lead to acceptance of a true statement (see Kubiak et al. 2022).

Another approach to support an epistemic and realistic interpretation of N-P, the single-case application of the method in this case, is to demonstrate that other measures of single-case confirmation for a specific hypothesis, beyond those based on the concept of degree or belief strength, can be applied within frequentist methodology. An example of such a frequentist measure, which involves an analysis of statistical power, is the concept of ‘severity’ proposed by Mayo and Spanos (see, e.g., 2006): the post-observation evaluation of how convincingly an accepted hypothesis has withstood the test in comparison to specific alternatives. It can be considered a measure of the strength of evidential support for a hypothesis within a particular observational context. Interestingly, Mayo and Spanos noted that Neyman himself drew attention to the concept of "post-data use of power" (2006, 334). Mayo and Spanos pointed out instances where Neyman wrote ambiguously about the confirmation of a hypothesis in relation to a high probability of detecting the alternative hypothesis if it were true.[10] This perspective supports the notion of a single-case epistemic interpretation of Neyman's method. It's crucial to understand that the concept of severity doesn't measure the confirmation of the original hypothesis in isolation but also considers nearby alternatives. Passing a severe test implies that reality is not significantly distant from the tested hypothesis, suggesting its approximate truthfulness or proximity to the truth. This interpretation aligns Neyman's method more closely with epistemic realism and, consequently, with PR.

5.4. Finding Perspectival Pluralist Aspects in Neyman's Conceptions

In the previously cited footnote from the 1937 paper, Neyman suggests that when it comes to selecting statistical theories or methodological frameworks for practical applications, 'personal taste' remains 'decisive'. In one of his later works, he goes even further by asserting that there should be no 'dogmatism' regarding application aspects:

> "What I am opposed to is the *dogmatism* which is occasionally apparent in the application of Bayes' formula when the probabilities *a priori* are not implied by the problem treated, and the author attempts to impose on the consumer of statistical methods the particular *a priori* probabilities invented by himself for this particular purpose" (Neyman 1957b, 19).

Methodological choices can indeed have a significant impact on the outcomes of applying statistical procedures.[11] However, the act of 'inventing' a prior also shapes the outcome and can be seen as a methodological decision based on 'personal taste'. Why influential personal 'choice', 'taste', 'invention', is acceptable when Neyman speaks of the statistical methodology adopted by him and is not acceptable when Neyman speaks of the Bayesian statistical

---

[10] However, he didn't say that the power under consideration was calculated in respect to post-observational p-value, except for one instance (1957a, 13) where he spoke of the situation of 'large p-value' being obtained.

[11] I exposed this in Section 2.

methodology lacks explanation and appears inconsistent.

As demonstrated in section 4.3, a tension exists between the notion of the importance of ‘personal taste’ in selecting a mathematical framework, which implies an acceptance of the plurality of perspectives, and the inclination to eliminate the equivalence of perspectives at both the methodological and meta-methodological levels. Section 4.3 suggests that Neyman’s views imply that the most epistemically favourable perspective should be determined by the objective epistemic context of the research, such as prior knowledge of the studied population or the experimental conditions. The anti-pluralistic meta-methodological inclination is particularly evident in the passage (paradoxically, from the same paper where the aforementioned footnote is found) cited at the very end of section 4.3. In this passage, Neyman suggests that the choice of a statistical framework should be grounded in physical reality. One way to interpret Neyman’s position is to argue that his dissatisfaction with Bayesian priors while accepting a role for preferences in error risk management, could be this. Bayesian priors are not inherently value-laden since they are degrees of credence, and since epistemic preferences are not allowed by Neyman. In contrast, error risk considerations can be value-laden and subject to preferences that are of pragmatic, not epistemic nature. Another way of putting this would be to say that both general methodologies can be considered equally valid epistemically, provided that the internal epistemic standards they impose are met. Neyman seemed to suggest that the epistemic standard required for choosing the Bayesian method is the presence of epistemically objective evidential basis for setting the priors, which he believed was usually lacking in practice. On the contrary, when it comes to adopting the frequentist method, no specific epistemic standard is mandated concerning the trade-off between error risks. The choice is considered epistemically neutral, but it requires pragmatic justification in turn. Another approach to understanding Neyman’s seemingly inconsistent expressions could be to suggest that the issue of aligning with PR primarily emerges at the lower methodological level. Neyman might have advocated a form of meta-methodological pluralism at a higher, more abstract level, while concurrently endorsing anti-pluralism at the lower methodological level.[12] Despite Neyman’s assertion that “this subjective element [choices concerning error risks] lies outside of the theory of statistics” (Neyman 1950, 263) it determines an important part of the statistical procedure and results. Therefore Neyman’s method can be viewed as endorsing perspectival pluralism in respect of the pragmatic-value driven differentiation of error risks.

It can be argued that this perspectival aspect opens the door to different perspectival statistical justifications, each potentially carrying equal epistemic credibility. Let’s consider

---

[12] I thank the anonymous Reviewers for drawing attention to these tactics.

two hypothetical scenarios involving error risk settings for a research question.

In the first scenario, the risk of falsely rejecting the tested hypothesis is set at 5%, while the risk of falsely accepting it's set at 20%. In the second scenario, the risk of falsely rejecting the hypothesis is 20% while the risk of falsely accepting 5%. In some cases, these two different test settings would produce different results. When the hypothesis being tested is true, the first procedure will have a higher nominal, long-term epistemic credibility in leading to true conclusions compared to the second procedure. However, if the alternative hypothesis is true, the epistemic credibility of the second procedure will be greater by the same amount. As previously argued, the default Neymanian approach is to view the reliability of the procedure as a pre-observational feature, where both types of errors (Type I and Type II) are relevant for its assessment. Additionally, since there's no prior information available about the likelihood of any hypothesis being true, both types of risk avoidance must be considered equally relevant from an epistemic perspective. Consequently, when solely considering the epistemic aspect of the specified error risks, without taking into account their pragmatic rationale, both of the described testing settings must be deemed equally epistemically credible. Therefore, Neyman's method implies partial epistemic pluralism (Premiss I is partially satisfied), despite Neyman's explicit interpretative assertions and his legitimate intention to eliminate the subjective, epistemically pluralistic element from his methodology.

Another example of the challenges in completely avoiding some level of pluralism in perspectives arises in the quest to find a test for a hypothesis that possesses greater power than any other conceivable test, regardless of the specific admissible alternative point hypothesis. This type of test has been referred to as the *uniformly most powerful test* (UMPT) (see Neyman 1950, 324-326). In numerous scenarios, the existence of such tests is simply not possible,[13] and the minimal epistemic requirement becomes that a test should possess power against any alternative hypothesis that is, at the very least, as high as the specified value for the risk of Type I error (see Neyman, Pearson 1933). This permits the utilisation of various tests, which might lead to potentially distinct conclusions, for analysing identical data when the experimental protocol remains the same. The selection depends on the specific alternative parameter value against which one aims to test hypothesis H. However, in certain situations, there might be no epistemically objective rationale to favor certain potential alternative hypotheses over others, given the absence of prior information that would enable the assessment of their chance of being

---

[13] Such a situation arises, for example, when a point hypothesis about the mean of the normal distribution with known standard deviation is tested against the alternative of the form $H'\colon \theta \neq \theta_0$: a UMPT exists against the alternatives which state that the mean is greater than the one purported by the hypothesis tested, and a different UMPT against the alternatives that state the mean is smaller than the assumed (Neyman, Pearson 1933, 319-321).

true.

In conclusion, despite Neyman's strong emphasis on the requirement for having ample data to substantiate the empirical adequacy of the underlying assumptions of the adopted model (as discussed in Section 4.3), there are situations where this may not be entirely feasible. This potential pluralism arises in Neyman's discussions of certain aspects or assumptions of frequentist models, which remain unknown and cannot be directly assessed through observation.[14] For instance, in (1923a^2, 470-471) Neyman delved into the problem of having to specify a parameter that couldn't be empirically identifiable from the observed data. This parameter related to the correlation between reactions to soil conditions among different varieties of crops of which difference in yield was meant to be estimated. In a work written with Scott (1948) he explored the issue of 'incidental' parameters, which cannot be consistently estimated but are known to be factors that are present in research and whose number increases with the size of the sample. While Neyman did attempt to address the challenges posed by factors that are necessary to consider but not directly observable, his proposed solutions often came with certain restrictions and were not applicable in all situations.[15]

In summary, despite Neyman's considerable efforts to eliminate epistemic pluralism in various instances of perspective at different levels, epistemic pluralism of perspectives continues to persist in the aspects of frequentist methodology described in this section. Considering the partial ambivalence in Neyman's statements, his efforts to minimise pluralism whenever feasible, and the presence of pluralistic aspects in frequentism, Neyman could be interpreted as acknowledging pluralism but not in the sense of inalienable epistemic necessity. He appeared to be a methodological anti-pluralist, asserting that although we cannot currently epistemically discriminate between genuinely existing perspectives in every possible case, we should not assume that we are forever bound to this state of affairs. In this context, epistemic pluralism of perspectives seems potentially reducible in the future as statistical methodology and empirical knowledge continue to develop.

6. Conclusions

In this article, the fundamental principles of PR within the realm of frequentist statistics have been elucidated. Various forms of PR have been categorized into three types, and a minimal proposal of definition of 'perspective' has been provided. It has been demonstrated

---

[14] I thank the anonymous Reviewer for highlighting this issue.

[15] See Hennig (2023) for a recent framing of the general problem.

that PR could serve as a viable framework for frequentism, with the use of stopping rules as an illustrative example. Following on from the aforementioned matters an extensive analysis of how Neyman’s frequentism relates to the set of PR’s basic assumptions was conducted. It was found that Neyman's perspectives on this matter are somewhat ambivalent, shifting between compatibility and incompatibility with PR, contingent on the specific aspect of the methods under consideration. Additionally, there exists a duality between what Neyman explicitly articulates in his philosophical-methodological reflections and what is implicitly implied by his methodology. These two aspects do not align seamlessly.

Nonetheless, it was feasible to propose certain solutions to the consistency issues by introducing interpretative refinements of Neyman’s frequentism, which render epistemic PR reasonably consistent with the methodology. It has also been proposed that within Neyman’s frequentist methodology, one encounters genuine and non-trivial perspectives, some of which are not equally valid epistemically (like in the case of stopping rules), whereas others might be considered to be (like some error risk settings).

Neyman does not advocate the complete elimination of all forms of plurality but aims to reduce the number of potential perspectives as much as possible. His theory and views suggest the existence of some residual epistemic pluralism that is challenging to eliminate. This is especially relevant in situations involving different error risk settings, scenarios where optimal tests do not exist, and hardly empirically distinguishable parameters.

Neyman’s approach of seeking the most optimal methodological framings whenever possible, rather than advocating methodological pluralism, suggests an inclination to view PR as a descriptive rather than a normative position. It also indicates that PR is a case (or aspect)-dependent instead of universal, absolute, or binding stance.

Despite the attempts in this article to address the issues of inconsistency between PR and Neyman’s thoughts, one might conclude that, given the indicated inconsistencies it’s not more, neither less in line with PR than with major alternatives. However, Neyman’s frequentism appears to strike a balance between pure realism and constructivism, which arguably positions it closer to a more balanced PR position. Given Neyman’s assertion that scientific conceptions encompass hypothetical and idealised elements, are intertwined with experimental constructs, and are partially influenced by practical considerations regarding errors, it would be inaccurate to categorise him as a strict realist. Similarly, associating Neyman with pure constructivism would be a mismatch, considering his acknowledgment of the existence of a reality approximated by scientific concepts and their reliable long-term assertions. Alternatively, adopting a more moderate stance, such as perspectival realist, appears to align better with these statements.

An exciting avenue for future research is to explore the extent to which various types of perspectives in statistics are (in)commensurable. Currently, there are methods available for integrating evidence from different perspectival studies on the same subject of interest (see, e.g., Mikołajewicz, Komarowa 2019). Neyman himself embraced the concept of average errors as a way to summarise various configurations of error risks in numerous tests, regardless of the specific context and assumptions of the models used in each case. Furthermore, even when the Bayesian and frequentist statistical paradigms appear to produce contradictory conclusions from the same data (see, e.g., Wagenmakers, Ly 2023) conciliatory practices are indeed possible (see e.g. Bayarri, Berger, 2004) as long as we temporarily set aside overly discussed metaphysical interpretative issues (c.f. Kass 2011). All this raises an intriguing question for PR when viewed from the perspective of statistical methodology, as it allows for the possibility of merging perspectives and potentially enables epistemic discrimination between perspectives that currently seem equally valid.

References

Bayarri, M. Jesús, and James O. Berger. 2004. The Interplay of Bayesian and Frequentist Analysis. *Statistical Science* 19 (1): 58–80.

Chakravartty, Anjan. 2010. “Perspectivism, inconsistent models, and contrastive explanation” *Studies in History and Philosophy of Science Part A*, Volume 41 (4), 405-412.

Collins, Harry M., Evans, Robert. 2002. “The third wave of science studies: Studies of expertise and experience.” *Social Studies of Science* 32: 235–296.

Crețu, Ana-Maria. 2020. “Perspectival Realism.” In *Encyclopedia of Educational Philosophy and Theory*, ed. M. A. Peters. Singapore: Springer.

——— 2022. “Perspectival Instruments.” *Philosophy of Science*, 89(3), 521-541.

Fisher, Ronald A. 1956. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Giere, Ronald N. 1976. Empirical Probability, Objective Statistical Methods, and Scientific Inquiry. In Harper, W.L., Hooker, C.A. (eds) *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science. The University of Western Ontario Series in Philosophy of Science*, vol 6b. Springer, Dordrecht.

——— 2006. Scientific perspectivism. University of Chicago press, Chicago

——— 2010. “An agent-based conception of models and scientific representation.” *Synthese* 172: 269-281.Hacking, Ian. 1965. *Logic of Statistical Inference*. London: Cambridge University Press.

Hennig, C, 2023. “Parameters not empirically identifiable or distinguishable, including correlation between Gaussian observations.” *Stat Papers* https://doi.org/10.1007/s00362-023-01414-3

Jacoby, Franklin. 2020. “Data identity and perspectivism.” *Synthese* (2020).

https://doi.org/10.1007/s11229-020-02824-8

Kass, Robert E. 2011. “Statistical Inference: The Big Picture.” *Statist. Sci.* 26 (1): 1-9.

Kubiak, Adam P. 2014. “A Frequentist Solution to Lindley & Phillips’ Stopping Rule Problem in Ecological Realm.” *Zagadnienia Naukoznawstwa* 50 (200): 135-145.

——— and Kawalec, Paweł. 2022. Prior Information in Frequentist Research Designs: The Case of Neyman’s Sampling Theory. *J Gen Philos Sci* 53, 381–402.

——— Kawalec, Paweł and Kiersztyn, Adam. 2022. „Neyman-Pearson Hypothesis Testing, Epistemic Reliability and Pragmatic Value-Laden Asymmetric Error Risks.” Axiomathes 32, 585–604.

Lehmann, Erich L. 1985. “The Neyman-Pearson Theory After Fifty Years.” In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Volume 1*, ed. L.M. Le Cam, 1047–1060. Michigan: Wadsworth Advanced Books & Software.

Lehmann, Erich, L., Romano, Joseph, P. 2005. *Testing Statistical Hypotheses*.Third Edition. New York: Springer-Verlag.

Lenhard, Johannes. 2006. “Models and Statistical Inference: The Controversy between Fisher and Neyman–Pearson.” *The British Journal for the Philosophy of Science* 57: 69–91.

Lindley, Dennis V., Phillips, Lawrence D. 1976. “Inference for a Bernoulli Process (a Bayesian View).” *The American Statistician* 30 (3): 112–119.

Massimi, Michela. 2021. “Cosmic Bayes. Datasets and priors in the hunt for dark energy.” *Euro Jnl Phil Sci*, 11, 29 (2021). https://doi.org/10.1007/s13194-020-00338-1

——— 2019. “Realism, perspectivism, and disagreement in science.” *Synthese* (2019). https://doi.org/10.1007/s11229-019-02500-6

——— 2018a. “Perspectivism.” In *The Routledge Handbook of Scientific Realism*, ed. J. Saatsi, 164-175. Oxford: Routledge.

——— 2018b. “Three problems about multiscale modeling in cosmology” *Studies in History and Philosophy of Modern Physics*, 64**:**26–38.

——— 2018c. “Perspectival Modeling.” *Philosophy of Science* 85(3): 335–359.

——— 2018d. “Four kinds of perspectival truth.” *Philosophy and Phenomenological Research*, 96(2), 342–359.

——— 2012. “Scientific perspectivism and its foes.” *Philosophica* 84(1): 25–52.

Mayo, Deborah. 2018. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics*

Wars. Cambridge: Cambridge University Press.

——— and Spanos, Aris. 2006. “Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction.” *The British Journal for Philosophy of Science,* 57 (2): 323–357.

McCarthy, Michael A. 2007. *Bayesian Methods for Ecology*. Cambridge: Cambridge University Press.

Mikolajewicz Nicholas, Komarova Svetlana V. 2019. Meta-Analytic Methodology for Basic Research: A Practical Guide. Frontiers in Psychology 10. DOI=10.3389/fphys.2019.00203

Neyman, Jerzy. Spława-Neyman, Jerzy. 1923a^1. “Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych.” Roczniki Nauk Rolniczych I Leśnych 5: 1–51.

——— Spława-Neyman, Jerzy. 1923a^2. “On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.” Part of (Neyman 1923a^1) transl. by T. M. Dabrowska, T. P. Speed. Statistical Science 5: 465–472.

——— 1934. “On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection.” *Journal of the Royal Statistical Society* 97 (4): 558–625.

——— 1937. “Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability.” *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236 (767): 333–380.

——— 1938. “Contribution to the Theory of Sampling Human Populations.” *Journal of the American Statistical Association* 33:101–116.

——— 1950. *First Course in Probability and Statistics*. New York: Henry Holt and Co.

——— 1952. *Lectures and conferences on mathematical statistics and probability*. Washington: U.S. Department of Agriculture.

——— 1957a. “The Use of the Concept of Power in Agricultural Experimentation.” *Journal of the Indian Society of Agricultural Statistics* 9: 9–17.

——— 1957b. “‘Inductive Behavior’ as a Basic Concept of Philosophy of Science.” *Revue De L'Institut International De Statistique* 25 (1/3): 7–22.

——— 1971. “Foundations of Behavioral Statistics.” In *Foundations of Statistical Inference,* ed. V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart and Winston.

——— 1977. “Frequentist probability and frequentist statistics.” *Synthese*36 (1): 97–131.

——— and Pearson, Egon, S., 1933. “On the Problem of the Most Efficient Tests of Statistical Hypotheses.” Philosophical Transactions of the Royal Society of London. Series A 231:

289-337.

——— and Scott, E.L., 1948. “Consistent estimation from partially consistent observations.” *Econometrika* 16: 1–32.

Peirce, Charles S. 1910/1932. *The Collected Papers of Charles Sanders Peirce, Vol. II: Elements of Logic.* eds. C. Hartshorne, and P. Weiss. Cambridge: Harvard University Press.

Pitman, Jim. 1997. *Probability*. New York: Springer

Popper, Karl. 1959. *The Logic of Scientific Discovery*, Hutchinson & Co.

Potters, Jan. 2020. “Perspectivism and the epistemology of experimentation.” Euro Jnl Phil Sci 10, 24. https://doi.org/10.1007/s13194-020-00289-7.

Sprenger, Jan. 2016. “Bayesianism vs. Frequentism in Statistical Inference. In The *Oxford Handbook of Probability and Philosophy*, ed. A. Hájek and C. Hitchcock, 382-405. Oxford: Oxford University Press

Rubin, Mark. “Repeated sampling from the same population?” A critique of Neyman and Pearson’s response to Fisher. *Euro Jnl Phil Sci* 10, 42 (2020). https://doi.org/10.1007/s13194-020-00309-6

Rueger, Alexander. “Perspectival Realism and Incompatible Models.” *Axiomathes* 26: 401–410.

Rudner, R. 1953. “The scientist qua scientist makes value judgments.” *Philosophy of Science*, 20, 1–6.

Ruyant, Quentin. 2020. Perspectival realism and norms of scientific representation. Euro *Jnl Phil Sci* 10, 20 (2020). https://doi.org/10.1007/s13194-020-00285-x

Savage, Leonard. ed. 1962. *The foundations of statistical inference: A discussion.* London: Methuen.

Wagenmakers, EJ., Ly, A. History and nature of the Jeffreys–Lindley paradox. *Arch. Hist. Exact Sci.* 77, 25–72.

Wald, Abraham. 1950. *Statistical Decision Functions*. New York: John Wiley and Sons.

Wolff, J.E. 2019. Representationalism in Measurement Theory. Structuralism or Perspectivalism? In *Understanding Perspectivism: Scientific Challenges and Methodological Prospects,* ed. M. Massimi & C. D. McCoy, 109-126. London: Routledge.