

# Prior Information in Frequentist Research Designs: The Case of Neyman's Sampling Theory\*

Adam P. Kubiak and Paweł Kawalec

May 17, 2022

## 1. Introduction

Jerzy 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 20<sup>th</sup> century (Lehmann 1985). His main contributions to inference from data (estimation, hypothesis evaluation; see, e.g., Neyman, Pearson 1928) and the process of interpreting the outcomes of experiments (philosophical assumptions and the goals of science; see, e.g., Neyman 1957) have been widely discussed by philosophers of science (e.g., Hacking 1965; Giere 1969; Mayo 1983; Mayo and Spanos 2006) and have often been criticised as disadvantageous with regards to the Bayesian statistical paradigm (see e.g., Romeijn 2017; Sprenger 2016, 2018) and the likelihoodist statistical paradigm (e.g. Royall 1997). However, Neyman's contribution to data collection and sampling designs has been, until recently (Zhao 2021), largely neglected by philosophers of science,<sup>1</sup> even though his contribution to this field is significant (Little 2014) and still remains a standard element of present-day sampling frameworks (Srivastava 2016).<sup>2</sup>

Highlighting the sampling theory of Jerzy Neyman is vital in light of the lack of self-standing and proper expositions of Neyman's views concerning sampling in the philosophical literature. Zhao (2021, section 2) depicted Neyman as one of the representatives of the so-called *design-based* (as opposed to *model-based*) general approach to sampling. In the design-based approach, the inference scheme and mathematical correctness of the estimation rely on the sampling design that determines selection probabilities ascribed to sampling units, while in the model-based approach, the inference scheme does not require a sampling design (assumptions regarding selection probabilities are not necessary<sup>3</sup>) (see, e.g. Särndal 1978).

---

\* This is a pre-copyedited author produced version of the article published in Journal for General Philosophy of Science. The final authenticated version is available online at: <https://doi.org/10.1007/s10838-022-09600-x>

<sup>1</sup> Except perhaps for his conceptions of causal effect in a randomised experiment (Pearl 2009, 126-132).

<sup>2</sup> There are several other sampling plans in the scientific literature that we do not discuss in this paper because we restrict our work to Neyman's contribution and the philosophical analysis thereof.

<sup>3</sup> The topic of making such assumptions is continued in Section 2.

Gregoire (1998) puts it this way: “[i]n the design-based framework, the probabilistic nature of the sampling designs is crucial [...] This is not the case in the model-based approach” (1431). While the inference scheme in the design-based approach is essentially pre-observational, the model-based inference scheme is essentially post-observational: “the model is fitted to sample data according to some criterion” and “inference in the model-based approach stems from the model, not from the sampling design” (Gregoire 1998, 1436). Zhao referred to Neyman’s statements concerning the general notion of a sample’s representativeness and Neyman’s critiques of sampling that rely on the researcher’s decisions instead of randomisation. Nonetheless, Neyman’s sampling designs are not fleshed out by this author. Moreover, in citing only selected fragments of Neyman’s views, Zhao depicted Neyman as a proponent of unrestricted randomisation in which the use of prior information<sup>4</sup> concerning a population is minimised. She presented design-based sampling as maximally uninformative. This image of Neyman’s view on sampling and of the design-based approach, as we show in this article, is misleading.

The second important reason to bring out Neyman’s original sampling theory regards the philosophical debate between frequentism and Bayesianism, in which Neyman’s sampling theory has been omitted. Many philosophers of the scientific method claim that Bayesianism provides a more adequate account of scientific inference than frequentism because Bayesianism explicitly encodes available prior information as a prior probability (e.g. Howson and Urbach 2006, 153–154, Romeijn 2017).

Frequentism, and especially Neyman-Pearson’s approach, is often regarded as unable to articulate the prior information it presupposes. For example, Sprenger (2009, 240) claims that the frequentist procedure uses “implicit prior assumptions”; and that the frequentist inference assumptions that precede statistical inference, “are often hidden behind the curtain”, while the Bayesian framework reveals such assumptions in a more explicit way (Sprenger 2018, 549, sect. 4). Bayesianism is regarded as superior to the “conventional” methods that are used in frequentist statistics because “conventional statistical methods are forced to ignore any relevant information other than that contained in the data” (McCarthy 2007, 2). This purported lack of sensitivity to context-specific prior information is expressed as “maximally uninformative” use of prior information in sampling design (Zhao 2021, 9101). The approach

---

<sup>4</sup> We use this term to denote a piece of information that is potentially or actually used in scientific inference as an element of a particular study and which is not a part of the observational data gathered when the study is conducted. Prior information is or can be shared and communicated as something that plays a role in drawing scientific conclusions.

of Neyman (and Pearson) to statistics is considered to “rely on a concept of model that includes much more preconditions, according to which much of the statistician’s method is already fixed” which contrast with “building and adjusting a model to the data at hand and to the questions under discussion”, which is thought to be a key feature of Fisher’s competing approach (Lenhard 2006, 84). These objections entail that prior information is in principle not utilised by Neyman’s frequentist statistical methods in an objective and epistemically fruitful way. The important question then is whether these objections stand when we consider the perspective of Neyman’s theory of sampling.

Our third source of motivation in analysing Neyman’s sampling designs is the debate concerning the role of non-epistemic values in science. Classically, social values, such as economic, ethical, cultural, political, and religious values, are understood in opposition to epistemic (cognitive) values (see e.g. Laudan 2004; Reiss, Sprenger 2020). The *value-free ideal of science* (VFI) assumes that collecting evidence and formulating scientific conclusions can be undertaken without making non-epistemic value judgments, and states that scientists should attempt to minimise the influence of these values on scientific reasoning (see e.g. Douglas 2009; Betz 2013). In frequentist statistics, the choice of a sampling scheme influences the process and outcome of statistical reasoning. This is accomplished by determining the mathematical model of the study design (see e.g. Lindley, Phillips 1976) and by the fact that the choice of sampling scheme influences sample composition. This prompts the question of whether, and how, an explicit influence of some social factors on the process of forming a scientific conclusion is present in Neyman’s sampling designs, and if so, whether the implementation of this type of prior information at the stage of designing a sampling scheme is adverse, neutral, or perhaps beneficial epistemically (with regards to estimation). Such a type of influence on a sampling scheme is different from the type of influence that has the form of the practical considerations that dictate the uneven setting of error rates in Neyman-Pearson’s theory of hypothesis testing. The latter has already been a subject of philosophical debate since long before (see e.g. Levi 1962) but the influence of practical, ethical, and societal considerations on the process of collecting evidence and formulating scientific conclusions with regards to Neyman’s sampling theory has not been philosophically elaborated. If it could be shown that allowing for the influence of some social values on sampling design is beneficial epistemically, then this would pose an argument against VFI, as it maintains that the influence of social (non-epistemic) values is epistemically adverse.

In this article, we analyse the use of prior information in Neyman’s sampling theory (Sect. 2). We show that in Neyman’s frequentism explicit and epistemically beneficial use of

manifold types of prior information is possible and of primary concern when designing the study. This is contrary to philosophers' statements like the ones mentioned above by Lenhard, Sprenger, or Zhao. We indicate that this applies not only to sampling in connection with estimation but also to testing hypotheses (Sect. 3.1). We refer to the outcomes of the analysis to support two philosophical-methodological conclusions. The first is weakened opposition between frequentist and Bayesian approaches to sampling and estimation (Sect. 3.2). The second is undermined VFI (Sect. 3.3).

We use the term *objective* (*objectivity*) in the sense of *process objectivity*, meaning the objectivity of scientific procedures. Of the possible facets of objectivity, we concentrate on two. The first is that the prior information on which an outcome is contingent is explicitly and unequivocally stated, and thus knowledge is intersubjectively communicable and controllable through the shared standards of its expression and use. The second is that the procedures are not contingent on non-epistemic factors, including social ones, that would negatively influence the epistemic value of those procedures (see Reiss, Sprenger 2020). By the term *epistemic value*, we understand a value that positively contributes to reaching the epistemic goal of the assertion of new theses that are close to the truth and the avoidance of the assertion of theses that are far from the truth (see David 2001). In the case considered by Neyman, desired properties of the method of statistical estimation from a sample oriented towards the aforementioned general goal translate into two more specific goals: (I) to be able to generate statistically reliable conclusions and to have control over the nominal frequency of false conclusions, and (II) to increase the accuracy of true conclusions. More precisely, these goals are (I) being able to carry out an unbiased statistical interval estimation of a sought after quantity and to calculate error probability in the first place and—once such estimation is achievable—to (II) maximise the accuracy of an interval estimator<sup>5</sup> (minimise the length of possible intervals) (see Neyman 1937, 371). When we speak of the influence of social values on statistical inference we think of letting prior information of social factors be implemented in the sampling design and thus influence the process (and effect) of estimation in respect to

---

<sup>5</sup> A function that assigns a real number to each possible outcome of particular trial (related to a random selection of one sampling unit) or to a set of such outcomes is a random variable. An estimator is a random variable used to generate estimates of a sought-out population parameter. For example, an estimator of the population mean is a random variable  $\bar{X}$  that is a function of random variables  $X_1 \dots X_n$  that refer to possible outcomes of consecutive selections numbered from 1 to  $n$ . An estimate of the population mean is the numerical value  $\bar{x}$ , which is a function of the observed values  $x_1 \dots x_n$ . The variance of an estimator is the expected size of the squared deviation of the estimator from its expected value. An interval estimator of a parameter generates numerical values in the form of intervals with some pre-observational probability that a generated interval will cover the true value of the parameter.

aspects (I) and (II). Realisation of the epistemic goal in its two described aspects can be understood as the realisation of two epistemic values respectively: the value of achieving statistical reliability in the method of estimation (which, as we present later in the text, is called *consistency* by Neyman), and the value of increasing the accuracy of estimation methods.<sup>6</sup>

## 2. The Use of Prior Information in Neyman's Theory of Sampling Designs

In this section, we refer to Neyman's contributions to the methodology of sampling (in connection with estimation) in order to reveal that his framework aims at the explicit incorporation of the diverse types of prior information that are available in different research designs.

Historically, the challenge of drawing inferences from a sample rather than from a whole population was tantamount to ascertaining that the former is a *representation* of the latter<sup>7</sup> (cf. Kuusela 2011, 91-93). In his groundbreaking paper, Neyman (1934) compared two "very broad" (559) groups of sampling techniques that presuppose taking representative samples from finite populations: *random* sampling in its special form, so-called stratified sampling and *purposive* selection (sampling). What was, for Neyman, distinctive for random sampling was that there was some randomness present in the selection process, as opposed to purposive selection, where there is no randomness in the selection process. It follows from his paper that the method of random sampling may be of "several types" (Neyman 1934, 567-568), including simple random sampling with or without replacement, and stratified and cluster sampling (discussed by us below in this article), among others. The meaning of random sampling can be rephrased in more recent terms as *probability sampling*. In probability sampling, each unit in a finite population of interest has a definite non-zero chance of selection (Steel 2011, 896-898). This chance does not need to be equal for every unit. Neyman's rationale for random selection is that it enables the application of the probability calculus to interval estimation and the calculation of error probability, which, in Neyman's

---

<sup>6</sup> The necessity of narrowing down possible topics of our investigation due to paper size limitations and our goal of taking a closer look at Neyman's sampling theory led us to consider only those epistemic aspects of sampling techniques that were considered by Neyman himself.

<sup>7</sup> One of the still most commonly used descriptions of a *representative sample* is that such a sample is a *miniature of the population* (cf. Domicic 2011, 1222-1224).

view, is not feasible in the case of purposive selection (1934, 559, 572, 586).<sup>8</sup> Purposive selection means that the selection of sampling units is determined by a researcher's arbitrary, non-random choice and it is either impossible to ascribe probabilities to the selection of a particular possible set, or these probabilities are ex ante known to be either 0 or 1.

## **2.1. Stratified and Cluster Sampling**

Stratified sampling is a kind of probability sampling in which, before drawing a random sample, a population is divided into several, mutually exclusive and exhaustive groups called *strata*, from which the name of the approach derives. Next, the sample is divided into partial samples, each being randomly drawn from the strata. Stratified sampling is often a more convenient, economical way of sampling, e.g., in a survey about support for a new presidential candidate conducted separately in each province of a country where, roughly speaking, a province corresponds to a stratum. Citizens in such a case are not randomly selected from the population of the country's inhabitants as a whole but from each stratum separately. If the ratio of each stratum sample size to the size of the stratum's population is the same for each province, then every inhabitant of the country has the same chance of being included in the survey.

This form of stratified sampling prevailed at the time of the publication of Neyman's classic paper in 1934. A simple example can help to understand the core idea. Imagine a country with three provinces with 25, 10, and 5 inhabitants, respectively. If the stratified sample includes 8 inhabitants, then the sizes of corresponding subsamples must be 5, 2, and 1, accordingly. This is to assure that none of the strata will be under- or overrepresented and for the whole sample to remain representative of the relative proportions of the population. Stratified sampling is particularly useful when the variability of the investigated characteristic is known to be in some way dependent on an auxiliary factor. Strata should then be determined to represent the ranges of values of such a factor—we discuss this later in this section.

Sometimes the characteristics of a population or its environment makes it difficult to sample individual units of a population. The cost or inconvenience of sampling units is simply too high compared to its benefits, all things considered, as in the case of investigating per

---

<sup>8</sup> Neyman recognised R.A. Fisher as the one who introduced to sampling and experimentation the principle that to control errors and produce a rigorous measure of uncertainty, it is necessary that a sample be collected by random selection and not through arbitrary choice (Neyman 1950, 292; Neyman 1977, 110; cf. Marks 2003, 933).

capita food expenditure. It is much easier to get to know what the monthly food expenditure of a household is with a known number of members than it is to draw at random a particular citizen and to determine how much she spends per month. This is because food for all members of a household is usually bought jointly and shared without discriminating how much of a product was bought or used by an individual member. The investigated state of affairs regarding individuals exists and relevant data could theoretically be obtained—individuals might be randomly selected and asked to record their expenditures or consumption—but this would be very inconvenient for the individuals and require high compensation for their agreement to participate in the survey. One approach to preserving convenience and thriftiness is to randomly draw and investigate *clusters* (new sampling units of a higher order), like households, rather than the units themselves. In other cases, cultural conventions or moral considerations might be worth taking into account, such as in the case of determining the value of weekly church donations per person in a particular city. Imagine no public data is available and you want to estimate it based on a random sample. In some countries, the amount donated is not formally predetermined and some parishioners may believe that the amount of an individual's donation should remain undisclosed.<sup>9</sup> In this case, a possible way of data collection that preserves the indicated people's moral values would be to treat parishes with a known number of parishioners and a known total sum of donations as sampling units—clusters.<sup>10</sup>

Thus, this type of sampling for Neyman consists of treating groups of individuals as units of sampling. Clusters as groups are collectives of units that are always taken into consideration together: first, some of the clusters are selected at random, and then all members of the selected clusters are included in the sample. Strata, in contrast, are conceived as subsets of a population, and from every stratum, some units are drawn at random. For example, if a country's districts were treated as clusters, rather than strata, then random drawing would apply to districts themselves: some districts would be randomly selected and then all the citizens from the selected districts would be subjected to the questionnaire. Sometimes the attributes of a cluster's elements are measured separately for each element and generalised, while in other cases, a generalised measure is immediately available (being

---

<sup>9</sup> For example, such beliefs are held by members of the so-called Neocatechumenal Way communities operating in parishes of the Catholic Church.

<sup>10</sup> One could decide not to take the indicated value into account and sample individuals instead. This might be associated with a lack of response from some of the sampled units, the effects of which can be remedied using Neyman's double sampling and optimum allocation methods (see Hansen, Hurwitz 1946, 518), which are discussed later in this article.

unique). This second case would be the just mentioned examples of parishes and households, where measures of an element's attributes are not available. A clear advantage of cluster sampling is that it seems to naturally capture the structure of many studied populations, and so it may be the only reasonable sampling scheme in the socio-economic realm, for "human populations are rarely spread in single individuals. Mostly they are grouped" (Neyman 1934, 568). This type of sampling was later classified as one-stage cluster sampling. This type is distinguished from the multi-stage type, in which clusters are randomly selected in the first stage but random selection is continued in the follow-up stage(s) within the selected clusters (see Levy, Lemeshow 2008, 225).

Sampling of clusters can be combined with stratified sampling. If prior information prompts one towards sample clusters instead of the original units of the population, then the original population can be reconceptualised as a population of clusters, and stratification can thus be performed on the reconceptualised population of clusters. Neyman used this approach in his 1934 paper. Still, the assumptions, roles, and consequences may be examined separately for clustering and stratification, as exemplified by Neyman.

We turn now to the epistemic consequences of the use of prior information by means of stratification and clustering. Neyman has mathematically demonstrated that the information on how a population is organised and socio-economic factors like those mentioned above can be objectively applied in the process of scientific investigation at the stage of designing the sampling scheme with the use of stratification and clustering. He has shown how these factors influence the process of statistical inference—thus how social values of convenience, thriftiness, or abidance of cultural norms can influence statistical inference and enable statistically reliable conclusions and for there to be control over the nominal level of false conclusions, as a means to reach the epistemic goal in aspect (*I*).

Even when stratification and/or clustering is arbitrary it does not rule out the feasibility of an estimation (aspect (*I*) of the epistemic goal) that will use the *best linear unbiased estimator (B.L.U.E.)*,<sup>11</sup> the conception of which was introduced in Neyman's 1934 paper and meant the linear unbiased estimator of minimal variance (Neyman 1934, 563-567).

---

<sup>11</sup> The term "best" means: of minimal variance among estimators of the type considered and under the condition of no prior assumption of the probability (density) function of data. Neyman (1934, 564-565) stressed that linear estimators are not the best in an "unequivocal" sense and argued for this type of estimator to be an element of his theory based on certain "important advantages" of linear estimators, discussion of which is beyond the scope of this paper. An unbiased estimator is an estimator the expected value of which is equal to the true value of the parameter being estimated (in opposition to biased estimator, for which the expected value is not equal to the true value).



In Neyman's terminology, the value of the variance of an estimator is inversely proportional to its *accuracy* (Neyman 1938a). An increase in the accuracy of estimation means *shorter* confidence intervals (see Neyman 1937, 371).<sup>12</sup> That a method of sampling is *representative* means that it enables *consistent* estimation of a research variable and of the accuracy of an estimate (see Neyman 1934, 587-88). Consistency of the method of estimation means, in Neyman's theory, that interval estimation with a predefined confidence level can be ascribed to every sample irrespective of the unknown properties of a population (Neyman 1934, 586). Consistent estimation can be achieved regardless of the variation of the research variable within a particular strata, the way a population is divided into strata and the primary entities organised into clusters (Neyman 1934, 579).

Neyman's analysis of stratified and clustered sampling designs<sup>13</sup> indicate how to properly implement information available prior to the onset of the research process concerning how a population is organised and its relevant socio-economic factors. He has mathematically shown that information representing the influence of these factors on sampling and estimation can be implemented in an explicit, objective way without obstructing consistent estimation.

## **2.2. Purposive Selection and Optimum Allocation Sampling**

In contrast to the method of stratified sampling (or, more generally, the method of random sampling), purposive selection aims not at random selection, but at the maximal representativeness of a sample by intentional (purposive) selection of certain groups of entities. This selection is based on an investigator's expert knowledge of general facts about the population in question or her own experience concerning the results of previous investigations. This kind of approach may sometimes appear natural to a researcher. For example, consider an ecologist who wants to assess the difference in blooming periods of certain herb species from two large forest complexes exposed to different climatic conditions. If an investigator knows about the presence of a certain factor of secondary interest and its influence on the abnormal disturbance of the selected species' blooming, she might tend to

---

<sup>12</sup> The discussed properties relate directly to the method (estimator—mathematical construct), not to the estimate—single, particular outcome based on the observed data—or calculated interval yielded by the use of this construct. The interpretation of the connection between the methodological optimisation and outcomes is that, on average, an estimate's (data) variance will be smaller and the calculated confidence interval will be shorter (more accurate). It could also be said that an increase of the accuracy of the estimator yields that the expected variance of an estimate is smaller, and length of an interval is shorter.

<sup>13</sup> The mathematical background for Neyman's statements concerning (stratified) cluster sampling can be found in full detail in Neyman (1933, 33-69).

exclude sampling from those forest sites (and thus those individuals of the herb) that are to a large extent subject to the local extreme (abnormal) disturbances of the aforementioned factor. This can be explained as an attempt to minimise the risk of a random drawing of an ‘extreme’ sample whose observational mean would be very distant from the population mean of the blooming period. It seems reasonable in such a case to *purposively select* specimens growing in sites that represent normal conditions with regards to this factor. By avoiding the risk of selecting an extreme sample, a more representative sample will be selected which, ideally, should lead to better accuracy of the assessment of the relevant characteristic of the population.

According to Neyman, the basic assumption underlying purposive selection was that the values of an investigated quantity (ascribed to particular units of the investigated population from which a sample is to be taken) are correlated with the auxiliary variable and that the regression of these values on the values of this same auxiliary variable is linear (Neyman 1934, 571). Neyman stated that if one assumes that the above hypothesis is true, then successful purposive selection must sample units of the population for which the mean value of the auxiliary variable will have the same value, or at least as close as possible to the value for the whole population (see Neyman 1934, 571). This can be motivated by the following simple example: supposing that the quantity of an average weekly income from donations is positively correlated with the mean age of the members of a parish, then, if most of the parishes from the investigated population were “senior” (in terms of the average age of members), the sample should include an adequately larger number of “senior” parishes than “younger” ones so that the mean “age” of a parish in a sample is close to the mean age of a parish from the whole population of parishes.

As mentioned earlier, purposive selection originally concerned non-probabilistic sampling. Neyman later modified the concept of purposive selection so that it became a special case of random sampling. What was assumed, to differentiate random sampling from purposive selection before Neyman’s paper, was first that “the unit is an aggregate, such as a whole district, and the sample is an aggregate of these aggregates” (1934, 570). Neyman has shown that the fact that “elements of sampling are [...] groups of [...] individuals, does not necessarily involve a negation of the randomness of the sampling”. We discussed this in Subsection 2.2 under the label of cluster sampling, as it is called nowadays. Thus, “the nature of the elements of sampling”, whether the unit of sample is an individual, or a cluster (a group of individuals), should not be considered as “constituting any essential difference between random sampling and purposive selection” (1934, 571).

Second, it was assumed by the time of Neyman's analysis that "the fact that the selection is purposive very generally involves intentional dependence on correlation, the correlation between the quantity sought and one or more known quantities" (1934, 570-571). Neyman has shown that this dependence can be reformulated as a special case of stratified sampling, which was by then regarded to be a type of random sampling. The effect of joining these two facts was as follows: "the method of stratified sampling by groups (clusters) includes as a special case the method of purposive selection" (1934, 570). Neyman stressed that this reconceptualised purposive sampling can be applied without difficulties only in exceptional cases. As an improved alternative to the method of purposive selection, but also to the method of simple random sampling and the method of stratified sampling with sample sizes for strata being proportionate to the sizes of the strata from which they are drawn, Neyman (1934) offered a method that is today called *optimum allocation sampling*.

Neyman showed in his analysis of how to minimise the length of an estimator in the case of stratified sampling design that the size of the stratum is not the only factor that should be taken into account when determining the needed size of a sample of a stratum. It is more optimal for an estimate's accuracy to also take into account estimates of the standard deviation of the research variable in strata (Neyman 1933, 92).<sup>14</sup> The variance of an estimator of a quantity is proportional to the variability of the research variable within strata. Therefore, to minimise the variance of the estimator by optimal sample allocation, the sample size for a stratum should be proportional to the quotient of the size of a stratum with the variability of the research variable in a stratum (Neyman 1933, 64; 1934, 577-580). If the variability of an auxiliary characteristic is known to be correlated with the variability of the research variable, one can use this information to divide the population into more homogenous strata with regards to the auxiliary variable, which will result in smaller (estimated) variances of the research variable within a stratum and subsequently a more accurate estimation (Neyman 1933, 41, 89; 1934, 579-580). Neyman stated that "There is no essential difference between cases where the number of controls is one or more" (Neyman 1934, 571), and if there is more than one known correlation, then one can implement all the relevant knowledge about manifold existing correlations using the "weighted regression" of the variable of interest upon multiple controls (see Neyman 1934, 574-575). In the case of the absence of any ready data, estimation of the variability of the investigated quantity within strata requires preliminary

---

<sup>14</sup> Neyman overlooked that Tschuprow (1923) also derived this rule of optimal sample allocation. Neyman acknowledged Tschuprow's priority in later years (Neyman 1952b).

research; the result of such an initial trial may subsequently be reused as a part of the main trial (Neyman 1933, 43-44). When one cannot make any specific assumption about the shape of the regression line of the research variable on the auxiliary variable, “The best we can do is to sample proportionately to the sizes of strata” (Neyman, 1934, 581-83). It is important to note that Neyman’s idea of optimum allocation sampling implies unequal inclusion probabilities (Kuusela 2011, 164)—sampling units that belong to strata with greater variability of the research variable will have higher inclusion probability. Methodological ideas proposed are clear cases of the direct, objective methodological inclusion of prior information of relationships between the sought after characteristics of the investigated population and some other auxiliary characteristics. These ideas demonstrate how sampling design and, eventually, the accuracy of an outcome can depend on the correlation of an investigated quantity with another quantity. If such information is known prior to sampling, it can increase estimation’s accuracy. The same holds for implementing prior information about the estimated variability of an investigated property.<sup>15</sup>

If clusters are the elements of sampling, minimising their size also increases the accuracy of an estimator (Neyman 1934, 582). Making clusters comprised of the same number of entities also increases the accuracy (Neyman 1933, 90). What was not addressed by Neyman is that more internally heterogeneous clusters also increase the accuracy of an estimation. So, pre-study information concerning some social factors in how a human population is structured in terms of the research variable can serve to devise smaller, or more internally varied, clusters so as to increase accuracy.

These facts about stratification and clustering indicate that via the use of Neyman’s theory of sampling and estimation, prior information about the changeability of an investigated property, about the dependence of the research variable on auxiliary factors, and about contextual social factors can be implemented using statistical procedures in an objective way to increase the accuracy of estimation. This yields the epistemic benefit of aspect (II) of the epistemic goal.

### **2.3. Double Sampling**

---

<sup>15</sup> Except for improving the accuracy of confidence intervals, stratified sampling has other epistemic advantages that we could consider if there was no limit to the size of this article. For example, this type of sampling can provide information for optimising estimators in so-called model-assisted estimation techniques (see, e.g., Royall, Henson, 1973), that are exploited, for example, in small area estimation.

Now we turn to aspects of Neyman's sampling design that concern a factor that inevitably and essentially influences the processes of collecting evidence and of formulating conclusions, namely the prior information regarding the costs of research.

It is taken for granted in statistics that Neyman "invented" (Singh 2003, 529) or "developed" (Breslow 2005, 1) a method called *double sampling* (Neyman 1938a) or *two-phase sampling* (Legg, Fuller 2009). Neyman, in his analysis of stratified sampling (1934), proved that if a certain auxiliary characteristic is well known for the population, one can use it to divide the whole population into strata and undertake optimum allocation sampling to improve the accuracy of the original estimate. The problem of double sampling refers in turn to the situation in which there is no means of obtaining a large sample which would give a result with sufficient accuracy because sampling the variable of interest is very expensive and because knowledge of an auxiliary variable, which could improve the estimate's accuracy, is not yet available. The first step of the sampling procedure, in this case, is to secure data for the auxiliary variable only from a relatively large random sample of the population in order to obtain an accurate estimate of the distribution of this auxiliary character. The second step is to divide this population, as in stratified sampling, into strata according to the value of the auxiliary variable and to draw at random from each of the strata a small sample to secure data regarding the research variable (Neyman 1938a, 101-102). Neyman intended this second stage to follow the optimum allocation principle (Neyman 1938b, 153).<sup>16</sup>

The main problem in double sampling is how to rationally allocate the total expenditure between the two samplings so that the sizes of the first large sample and the second small sample, as well as sizes of samples drawn from particular strata, are optimal from the perspective of the accuracy of estimation (Neyman 1938b, 155). For example, suppose that the average value of food expenditure per family in a certain district is to be determined. Because the cost of ascertaining the value of this research variable for one sampling unit is very high, limited research funds only allow one to take quite a small sample. However, the attribute in question is correlated with another attribute, for example, a family's income, whose per-unit sampling cost is relatively low. An estimate of the original attribute can be obtained for a given expenditure either by a direct random sample of the attribute or by arranging the sampling of the population in the two steps as described above.

---

<sup>16</sup> It is important not to confuse 2-phase sampling with 2-stage sampling. In the first case both samples are drawn from the same population, but with regards to different variables, whilst in the second case a sample is taken from the population studied and a second sample is taken from a subpopulation comprised only of entities that belong to the sample obtained at the first stage.

Neyman provided formulas for the allocation of funds in double sampling that yield greater accuracy of estimation compared to estimation calculated from data obtained in one-step sampling – both having the same budget. Nevertheless, in certain circumstances, double sampling will lead to less accurate results. Neyman indicated that certain preliminary information must be available in order to verify whether the sampling pattern will lead to better or worse accuracy and to know how to allocate funds (Neyman 1938a, 112-115). So, double sampling requires prior estimates of the following characteristics: the proportion of individuals belonging to first-stage strata, the standard deviation of the research variable within strata, the mean values of the research variable in strata, and, obviously, the costs of gathering data for the auxiliary variable and research variable per sampling unit (see Neyman 1938a, 115).<sup>17</sup> To increase the efficiency of estimation by using double sampling, both types of costs must differ enough, and the between-stratum variance of the research variable must be sufficiently large when compared to the within-stratum variance (Neyman 1938a, 112-115). Thus, to evaluate which of the two methods might be more efficient, prior information concerning the above-indicated properties of the population sampled is required. It is also needed to approximately determine the optimal size of the samples (Neyman 1938a, 115).

What we have shown is that the method of double sampling articulates the rules of using the prior information concerning the structure of a population (with regard to an auxiliary variable interrelated with a research variable), the information about the estimated values of a research variable, its variability, as well as typical economic factors: the costs of different types of data collection and available research funds. Those rigid rules determine the estimation procedure and its effects in an objective manner. More importantly, this method guides a researcher towards the realisation of the second (*II*) aspect of the epistemic goal: the correct use of these types of information can increase the accuracy of estimation.

### **3. Methodological and Philosophical Consequences**

Manifold types of prior information are used at the stage of planning and executing the collection of evidence. Neyman's method uses not only prior information relating directly to a sought quantity but also related indirectly to it and also information concerning non-cognitive

---

<sup>17</sup> Given the need for some previous knowledge or preliminary estimation of these quantities, Neyman ultimately labeled the method “triple sampling” (Neyman 1938b, 150).

factors that can influence a given outcome. All these types of information available prior to conducting the research process can be regarded as originating from different research contexts in which new research is being carried out. Thus, three main types of prior information used in Neyman's sampling designs can be distinguished:

- 1) prior estimates of the research variable and its variability within the population,
- 2) correlations between other characteristics of the studied population (auxiliary variables) and research variable(s), and
- 3) social factors: the technical convenience and availability of research objects (which depend on known characteristics of the population), financial factors—costs of the manifold ways of gathering data and available funds—and moral considerations.

These indicated types of information are used in an explicit and unequivocal way: they are encapsulated in the form of definite mathematical constructs for sampling designs or in the definite values of these constructs' parameters. Therefore, their use is objective and coherent from the perspective of the statistical framework adopted by Neyman. This use of a vast spectrum of prior information in designing the study can have a positive epistemic influence on scientific inference and conclusions derived (as shortening a confidence interval means changing the contents of a conclusion).

In what follows we analyse Neyman's use of prior information in study design from the perspectives of the frequentism vs. Bayesianism controversy (3.1-3.2) and the debate on the role of non-epistemic values in science (3.3).

### **3.1. Sensitivity of Study Design to Prior Information and Transparency of its Use in Hypothesis Tests**

It is taken for granted that Bayesian procedures are more transparent than frequentist ones thanks to explicitly included prior information encapsulated in prior probability distributions (Sprenger 2018). Sprenger points out that the outcome of a frequentist test is sensitive to issues such as how one defines the hypothesis and the plausible alternative, or whether a test is one- or two-tailed, and that it is hard to imagine frequentist consideration of these type of assumptions without a fair amount of adhocery. In the same article Sprenger also objects to frequentists' ignoring the issue of scientifically meaningful effect size or prior plausibility of a hypothesis (2018, Section 4). These types of prior inferential assumptions are thus thought not to be explicitly and objectively considered by frequentists.

Conversely, Neyman argues that these types of test features can and must be tailored to a particular research problem in reference to prior knowledge (see Neyman 1950, 277-291). For example, Neyman (278-79) insists that the effect size of substantial relevance should be clearly set and explicitly considered in setting experimental design. The same stands for the decision of whether a test is to be one- or two-sided, which itself should be subject to experimental verification (282-285).<sup>18</sup> Neyman and Pearson (1928, 178, 186) also admit that there is usually a prior expectation in regard to the truth-value of an investigated hypothesis.<sup>19</sup> Even though this information is not used as a premise in frequentist inferential procedures, it can be referred to in determining the statistical design of research and ultimately influence the outcome.

An example of how this fact could function in practice can be shown in reference to McCarthy's (2007, 4-13) simplified example. McCarthy recalls a case of detecting the presence of a frog species in a pond. He assumes the probability of positive detection in case the species is present to be 0.8 and the probability of no positive detection in case it is absent to be 1. He rightly states that the outcome of Bayesian reasoning could be sensitive to the knowledge of which type of pond a researcher comes across: whether this would be a type of pond in which this species almost always occurs (perfect habitat), or in which it almost never occurs (unwelcome habitat). Noting absence would not make the researcher believe the frog was absent in the case of perfect habitat, but could suffice to conclude so in the case of unwelcome habitat. McCarthy indicates that the influence of this type of prior information on the outcome is a key feature of Bayesianism, which the frequentist approach is lacking.

Nonetheless, the knowledge concerning the type of pond can play a role in frequentism at the stage of construing research design. Following Neyman and Pearson (1928, 178, 186) one could assert that a researcher usually has prior information that prompts them to believe that the hypothesis tested is true. If the pond to be examined exemplifies the frog's natural habitat so that they expect the frog to occupy it, this assumption could be used to define the hypothesis to be tested as the statement that the species is present. The effect of the application of Neyman's testing scheme (acceptance or rejection based on the  $p$ -value), under the conditions assumed by McCarthy, would be acceptance of the statement that the species is present. Analogically, in the case of unwelcome habitat, the hypothesis to be tested would

---

<sup>18</sup> In practice, this means that subject matter knowledge about the possible direction of the effect should be available or obtained and considered first.

<sup>19</sup> An analysis of a possible epistemic import of this information when frequentist hypothesis testing is considered can be found in (Kubiak et al. 2021).



state that the frog is absent, and the lack of observation would make the researcher accept that it is absent. Therefore, the prior information about the type of pond can possibly be utilised by a frequentist at the stage of designing the statistical model (of the hypothesis to be tested in this case) and influence the outcome of the investigation.

The above exemplary considerations regarding hypothesis testing are consistent with the methodological conclusion from the analysis of Neyman's sampling designs. They both show that in Neyman's frequentism it is the study design where taking into account various types of prior information is possible and of primary epistemic concern. An interesting question for future research could be to investigate, based on case studies, whether and how some assumptions concerning study designs in frequentist hypothesis testing play a role analogous to the role of inferential assumptions in Bayesianism. This type of investigation would be in line with a recent statement that the best choice of one of the two—Bayesianism or frequentism (that follow from Neyman and Pearson's perspective)—depends on the case considered (see Lakens et al. 2020).

### **3.2. Reconciliation of Bayesian and Frequentist Approaches to Sampling and Estimation**

Zhao (2021) distinguished two senses of sample representation: “the design-based approach where a representative sample is one drawn randomly and the model-based approach where a representative sample is balanced on all features relevant to the research target” (9111). Zhao suggested that the core of the first approach was maximally uninformative randomization: “[r]andom selection is, at its core, a maximally uninformative selection procedure.” (9101) She stressed that “maximal noninformation precludes outside factors from systematically affecting (‘informing’) a sample's composition” (9101) whereas the key feature of the model-based approach is that “model-based inference in sampling relies on assumptions concerning the relationship between control and target variables” (9110). She pointed out Neyman as the representative of the design-based approach and outlined some of his basic statements that indicate the importance of randomization (9099-9101).

It may be taken for granted that Neyman is believed to be a co-founder of the design-based approach to sampling and estimation (Sterba 2009, 713; Särndal 2010, 114) but what we have shown in Section 2 is that in the design-based approach outside factors can well affect a sample's composition in a very informed way. In particular, the information about the regression of research variable on auxiliary variable(s) can be implemented through stratified random sampling that enables a more balanced sample in the sense adopted (following Royall and Herson) by Zhao (2021, 9108). Therefore, Zhao's assertion that an informed sampling

based on the use of prior information to balance the sample on auxiliary factor is (an advantageous) special feature of the model-based approach that distinguishes it from the design-based approach was far-fetched. Depicting Neyman as the proponent of unrestricted randomization with equal inclusion probabilities (see Zhao 2021, 9101) is also misleading. Our conclusions may lead one to wonder whether it is necessary to regard the design-based and model-based approaches as contradictory.

It is also believed that an inference pattern in the design-based approach is conditional on sampling design established prior to sampling whilst the model-based approach is conditional on an actual sample obtained (Särndal 2010, 116; Royall, Herson 1973, 883). Bayesian modelling requires specification of the prior distribution for investigated quantities whilst the design-based conception assumes that the investigated quantities are fixed and unknown values exist independently of the observer (Little 2004, 547-548).<sup>20</sup> The above can be encapsulated by the statement that “Design-based inference is inherently frequentist, and the purest form of model-based inference is Bayes” (Little 2014, 417). In both conceptions, prior information plays a role in construing models that affects a sample composition and the outcome of estimation, although in each of them it is used differently. Below we argue that juxtaposition of Neyman’s design-based conception with the Bayesian model-based one reveals that they are complementary or even analogical in certain respects.

Both approaches to sampling and estimation have deficiencies. Shortcomings of the design-based approach are mainly the limited guidance in the case of small samples and inapplicability when randomisation is highly corrupted (Zhao 2021). The major weakness of the model-based approach is that it can lead to much worse inferences than the design-based approach when the model is seriously misspecified (Little, 2004). These deficiencies can be diminished by granting the complementarity of the two approaches.

Firstly, they are complementary in having strengths in different circumstances. There are cases in which one of them is more effective than the other and therefore the status of the universal superiority of either of the two approaches depends on the context of research (Samaniego, Reneau 1994).<sup>21</sup>

Secondly, the complementariness stems from the fact that “[t]here are certain statistical scenarios in which a joint frequentist-Bayesian approach is arguably required”

---

<sup>20</sup> There exists also a frequentist modelling (of a Fisherian-type) in which investigated quantity is a random sample from a “superpopulation” but the argument we present in this article does not require reference to this conception.

<sup>21</sup> Interestingly, it is known that sometimes the two frameworks coincide in terms of the numerical outcomes (see Tillé, Wilhelm 2017, 183).

(Bayarri, Berger 2004, 59) as each method can be improved when supported by elements of the other one: In the design-based approach, crude design-based estimators can be post-observationally refined in reference to values estimated by the model; the design-based estimation with this kind of refinement stemming from the model-based approach is called model-assisted design-based estimation (Stahl et al. 2016, 3). The model-based approach, in turn, can be assisted by a design-based sampling technique: balanced, design-based random sampling allows a researcher to find better-specified and more robust models to be used for inference (Särndal 1978, 35; Little 2012, 316; Williamson 2013). This suggests that the two approaches are complementary rather than exclusive.

Tillé and Wilhelm argue that in current practices the idea of random sampling interplays with informed sampling via two principles, the principle of restriction—the idea of avoiding extreme samples by balancing on auxiliary variables—and the principle of giving higher probability inclusion for units that contribute more to the variability of the estimator (2017, 179-181). Zhao (2021) finds randomisation distinctive to Neyman’s design-based approach and informed sampling to be specific for the model-based approach. As we have shown this is not true because informing the sample by means of adequate stratification with unequal inclusion probabilities is an important element of Neyman’s sampling theory. This means the distinction, as drawn by Zhao, dissolves when Neyman’s theory is considered. Neyman’s theory is an example of frequentist joint use of randomisation and informed sampling. This means the Bayesian model-based approach is not the only one which can rely on prior information to perform more informed sampling. The functional analogy between Bayesian model-based and Neyman’s design-based approach becomes more perspicuous as the influence of the information about the actual sample on the quality of estimation is considered. In some cases, it is more optimal from the perspective of the accuracy of the outcome to balance sample on auxiliary variable(s) based on the design of adequate stratified sampling with respect to this variable(s) (Neyman 1933, 41, 89; Neyman 1934, 574-575). With a lack of adequate prior information, a preliminary trial may be required in order to establish the sampling design, and the result of such an initial trial may subsequently be reused as a part of the actual (main) trial (Neyman 1933, 43-44). This means that Neyman allows for the actual sample to influence the quality of the estimation procedure in a systematic way, whereas Zhao (2021) claims this type of feature to be specific to the model-based approach: “the design-based framework does not provide guidance for how sample

composition should be analyzed” (9103).<sup>22</sup> “Functional analogy” in this context means analogy of epistemic function or role that the use of prior information eventually plays in estimation. Although the information is in both cases employed by different means, it leads to the epistemic effect of improvement of the accuracy of estimation. This analogy of the two methodologies can be compared to analogous organs in biology, like lungs and gills, by which oxygen is taken into the body in different ways, enabling cellular respiration. That there is a functional analogy does not erase the distinction between the two types of organs, or the two types of methods.

In conclusion, Bayesian model-based and Neyman’s design-based approaches to sampling and estimation, while remaining methodologically distinct approaches, can be complementary and are in part functionally analogical with respect to the use of prior information and the use of information about the actual sample for the sake of epistemic profit. This supports the idea of reconciliation in the frequentism vs. Bayesianism debate. The idea is to leave aside overly discussed interpretative issues and to turn to—best by a joint eclectic approach—the real issue to be solved, which is the gap between assumed probabilistic models and reality; this is the common ground for the two paradigms to meet (Kass 2011). In the model-based approach, the model in question is the model of the probability distribution of the outcomes that may be far from the truth with respect to the reality of population values. This model can be refined thanks to design-based sampling. In the design-based approach in turn it is the model of the probability distribution of the sampled units (the model of research design) that may not fully meet the reality of research conditions. Unfavourable effects of this can be levelled by refining a design-based estimator thanks to the assistance of a model for outcome’s distribution.

### **3.3. The Role of Social Values in Research Design**

One widely held view among scientists and philosophers regarding scientific objectivity is their “freedom from personal or cultural bias” (Feigl 1949, 369). Thus, to ensure the objectivity of scientific procedures and outcomes, the research process should be robust with regards to personal subjective values as well as independent from the social and economic contexts of scientific research. One way to accomplish this value-free ideal of science is to

---

<sup>22</sup> A reader may note that the (partial) influence of the actual sample on the design of the sampling scheme and on the accuracy of an estimator does not make double sampling an example of model-based inference. The inference scheme still relies on the (probability) sampling design, which is not the case in the model-based approach.

ignore these contexts of research activities and exclusively “focus on the logic of science, divorced from scientific practice and social realities” (Douglas 2009, 48). As we indicated in the introductory section, the VFI states that the process of collecting evidence and formulating scientific conclusions can proceed without the influence of these type of values, and that these influences should be avoided. Contrary to this stance, some authors (e.g. Steel 2010) argue that the influence of this type of values are inseparable and/or does not need to have an adverse effect on scientific cognition. Others (e.g. Elliott, McKaughan 2014) state that VFI is inconsistent with the actual goals of scientists which are a mixture of epistemic and non-epistemic considerations.

The influence of social values on the scientific research process and its outcomes is well illustrated by a number of recently debated research areas, most notably climate change (for an overview of which see Elliott 2017), where the focus of research is determined by value-laden prior information. As succinctly expressed by Baumgaertner and Holthuijzen (2016, 51), who advance an analogous point for conservation biology, “The research is guided by what is deemed important; however, that ends up being measured (e.g., by an anthropocentric perspective or an ecocentric approach). That means that the areas of research that are focused on are selected by nonepistemic values.” An apt example of this is the relativity of an outcome of vegetation classification: the choice of different ontologies and thus the choice of how data is presented to a computer program that performs the vegetation classification may depend on the practical purpose for which the classification is being made (Kubiak, Wodzisz 2012).

The influence of non-epistemic factors is present in frequentist statistical methodology. Neyman and E. Pearson’s conception of hypothesis testing includes the explicit influence of factors of a societal type upon the process of the formation of scientific conclusions (see e.g. Neyman 1952a). As we already said in the introductory section, this is done by relying on practical factors in the uneven setting of error risks. Knowledge of these factors, which is available prior to sampling, once included can be regarded as the implementation of a special type of prior information. The influence of premises (information) of economic, cultural, moral, and other societal types on the process of collecting evidence and formulating scientific conclusions can be understood as the influence of social values on this aforementioned process. This is a violation of the VFI.

The realm of non-epistemic values influencing the discussed research procedures and outcomes can be contested by the suspicion that all that has been shown is that certain social facts, or factors, play a role in sampling and estimation. How could this entail an influence of

non-epistemic values? Indeed, a social state of affairs, like an economic, political, or moral circumstance encountered by a researcher can be considered a social factor. These could be, for example, political or moral expectations or beliefs (e.g. moral/religious values of anonymity of church donation), the way people organise themselves in social structures (subgroups), or prices of products or services established by the society's economic interactions. The existence of different social factors is an acknowledged fact but it is a researcher who decides (not) to let a factor influence the research process and its outcome—for example, by letting the Marxist-Leninist politics in the Soviet Union influence the practice and outcomes of biological research (the historical phenomenon known as Lysenkoism; see e.g. Soyfer 1994). In the case of statistical methods and pragmatic, economic, and moral factors discussed by us, this would take place in the form of deciding not to implement knowledge of the discussed factors in research design by choosing an uninformed sampling scheme, like simple random sampling, instead of using stratification, clustering, or other methodological tools discussed. As we tried to argue, such implementations are not inevitable, and the motives to use particular solutions can be non-epistemic. A value can be understood as “[a] fundamental standard to which one holds the behavior of self and Others” (Lacey 1999, 24). Letting different social factors, like the above indicated, affect the research scheme and outcome, can be understood as the behavior of following important political, moral, or pragmatic standards a researcher sticks to. This means proceeding in accord with the value of satisfying political ideas, respecting moral standards/beliefs/expectations, or maintaining practical convenience or thriftiness, respectively. Such values can be regarded as non-epistemic values. Proceeding in accord with such values when deciding on the sampling scheme will mean letting non-epistemic value judgments influence the scientific process of collecting evidence and drawing conclusions.

By now it is evident that an influence of non-epistemic values is actually present in some disciplines and in the Neyman-Pearson statistical methodology of testing hypotheses. This does not necessarily seriously undermine the VFI as some could argue that these disciplines do not fully realise the ideal of scientificness (when they are compared to, for example, physics or chemistry), and this methodology is undesired and replaceable by an alternative one. One way to rebut this would be to show that the impact of non-epistemic values can be neutral or even beneficial epistemically. As far as the mentioned impact on methodology of testing hypotheses is concerned, the issue turns out to be multifaceted and the jury is out. The epistemic import of the impact of non-epistemic values on setting error risks, which is an element of research design, may be positive or negative depending on the case

considered (Kubiak et al. 2021). It depends also on the aspect considered. For example, it may differ depending on whether outcome replicability or experiment replicability is studied (see Kubiak, Kawalec 2021).

What is the impact of non-epistemic values when Neyman's theory of sampling is examined in turn? As we have shown in Section 2, the influence of non-epistemic premises regarding the process of collecting evidence and the shape of conclusions can rationally inform sampling design. What we have concluded is that Neyman's sampling method can include common non-epistemic factors such as financial factors, technical convenience, and moral considerations. Admittedly, these do not exhaust all possible factors, but still include the most pertinent ones. We also argued that this means that the influence of social values like cost-effectiveness, practical convenience, or compliance with social (e.g. ethical) standards on collecting evidence and formulating scientific conclusions can positively contribute to the realisation of the epistemic goal in the two aspects discussed in this article, what Neyman called the consistency and accuracy of estimation. Therefore, contrary to what VFI postulates, certain types of social values can, and sometimes even should influence the scientific process for epistemic benefit. Possible epistemic neutrality or even profitability of the influence of non-epistemic values on the process of sampling and estimation weakens the version of the VFI presented in the Introduction. Even if value-ladenness could be systematically avoided by a change of methodology, like it is proposed by Betz (2013), the rationale for doing so becomes unclear if value-ladenness is not always epistemically adverse and is profitable epistemically in some cases. Obviously, there are perspectives in light of which value-ladenness is unfavorable, just to mention the infamous Lysenkoism case. Our investigation is limited to the analysis of sampling methodology and some aspects of possible value-ladenness. It only shows that VFI as a generalized principle is too strong a statement. Remarkably, a similar conclusion has recently been delivered concerning the epistemic import of the value-ladenness of Neyman-Pearson hypothesis testing (Kubiak, Kawalec 2021). Owing to this, the case of Neyman's statistical methodology motivates the adoption of a more balanced, less principled position.

#### **4. Conclusion**

We presented a self-standing reconstruction of Neyman's theory of sampling designs that has been largely ignored in philosophical debates, except for its recent depiction by Zhao (2021), which is misleading. Zhao mischaracterized Neyman's theory and the design-based approach

by identifying them with maximally uninformed sampling while presenting balanced sampling as a distinguishing feature of the model-based approach.

Lenhard (2006, 84) claimed that adjusting a model to the question under discussion, and also to the data at hand, is not compatible with Neyman's approach. We have proven that this is not fully justified. For Neyman, it is on the model of study design where a great emphasis is put to implement prior information for epistemic benefits. This includes prior estimates about the research variable and the inclusion of information about an actual sample.

We also showed that Neyman's approach gives the possibility of objective inclusion of prior information in the study design not only for the purpose of better estimation but also to make better-informed hypothesis testing. We believe that statements reoccurring in philosophical debates about the uninformed use of prior information in frequentism, like e.g. Sprenger's (2018), rather refer to scientists' malpractices than to the conception itself, at least when Neyman's conception is concerned. This is perhaps because of the neglect of Neyman's crucial views regarding the use of prior information in the study design, especially his ideas regarding sampling designs.

In reference to the debate on the design-based vs. model-based approach to sampling and estimation, it can be concluded that the Neymanian way of informed sampling is different than, but not necessarily functionally contrary to the Bayesian way. They are complementary approaches, which strengthen the conciliatory approach to frequentist and Bayesian statistics.

Neyman's sampling designs enable consistent statistical estimation and can minimise the variance of an estimator along with an objective use of a vast spectrum of prior information about the presence of natural mechanisms, the attributes of investigated populations, and socio-economic contexts.

The specificity of the last type of prior information possible to be used in Neyman's sampling theory reveals that Neyman's methods let non-epistemic values influence the study design and outcome with potential epistemic profit. This methodological fact disconfirms the generalized version of the VFI and suggests that it should be further reconsidered from the perspective of specific statistical methodologies.

## References

Baumgaertner ,Bert, Holthuijzen, Wieteke. 2016. On nonepistemic values in conservation biology. *Conservation Biology* 31: 48–55.



- Bayarri, M. Jesús, and Berger, James, O. 2004. The Interplay of Bayesian and Frequentist Analysis. *Statistical Science* 19 (1): 58–80.
- Betz, Gregor. 2013. In defence of the value-free ideal. *European Journal for the Philosophy of Science* 2: 207–220.
- Bowley, Arthur L. 1926. Measurement of Precision attained in Sampling, *Bulletin de l'Institut International de Statistique* 22: 1-62.
- Breslow, Norman E. 2005. Case–Control Study, Two-phase. In *Encyclopedia of Biostatistics*, ed. Peter Armitage and Theodore Colton. Chichester: Wiley.
- Collins, Harry M. and Robert Evans. 2002. The third wave of science studies: Studies of expertise and experience. *Social Studies of Science* 32: 235–296.
- David, Marian. 2001. “Truth as the Epistemic Goal.” In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. M. Steup, 151–169. Oxford: Oxford University Press.
- Desrosières, Alain. 1998/1993. *The Politics of Large Numbers. The History of Statistical Reasoning*. Cambridge: Harvard University Press.
- Douglas, Heather E. 2009. *Science, Policy and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
- Dumicic, Ksenija. 2011. Representative Samples. In *International Encyclopedia of Statistical Science*, ed. Miodrag Lovric, 1222-1224. Berlin: Springer.
- Elliott, Kevin C., ed. 2017. *Exploring inductive risk: case studies of values in science*. New York: Oxford University Press.
- Elliott, Kevin C., & McKaughan, Daniel J. 2014. Nonepistemic Values and the Multiple Goals of Science. *Philosophy of Science*, 81(1), 1-21. doi:10.1086/674345
- Feigl, Herbert. 1949. Naturalism and Humanism: An Essay on Some Issues of General Education and a Critique of Current Misconceptions Regarding Scientific Method and the Scientific Outlook in Philosophy. *American Quarterly* 1: 135-148, reprinted in *Herbert Feigl. Inquiries and Provocations. Selected Writings 1929-1974* ed. R.S. Cohen, 366-377.
- Fienberg, Stephen E. and Judith M. Tanur. 1995. Reconsidering Neyman on Experimentation and Sampling: Controversies and Fundamental Contributions. *Probability and Mathematical Statistics* 15: 47-60.
- Giere, Ronald N. 1969. Bayesian Statistics and Biased Procedures. *Synthese* 20: 371-387.
- Gregoire, Timothy G. 1998. Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*. 28(10): 1429-1447.
- Hacking, Ian. 1965. *Logic of Statistical Inference*. London: Cambridge University Press.

- Hansen, Morris H., Hurwitz, William N. 1946. The Problem of Non-Response in Sample Surveys. *Journal of the American Statistical Association* 41 (236): 517-529.
- Hessels, Laurens K., Harro van Lente and Ruud Smits. 2009. In search of relevance: The changing contract between science and society. *Science and Public Policy* 36: 387–401.
- Howson, Colin, Urbach, Peter. 2006. *Scientific Reasoning. The Bayesian Approach*. Chicago: Open Court.
- Robert E. Kass. 2011. Statistical Inference: The Big Picture. *Statist. Sci.* 26(1): 1-9
- Kneeland, Hildegard, Erika H. Schoenberg and Milton Friedman. 1936. Plans for a Study of the Consumption of Goods and Services by American Families. *Journal of the American Statistical Association* 31: 135-140.
- Kubiak, Adam P., Kawalec, Paweł .2021. The Epistemic Consequences of Pragmatic Value-Laden Scientific Inference. *Euro Jnl Phil Sci* 11, 52.
- Kubiak, Adam P., Kawalec, Paweł, Kiersztyn, Adam. 2021. Neyman-Pearson Hypothesis Testing, Epistemic Reliability and Pragmatic Value-Laden Asymmetric Error Risks. *Axiomathes* <https://doi.org/10.1007/s10516-021-09541-y>
- Kubiak, Adam P., Wodzis, Rafał. R. 2012. Scientific essentialism in the light of classification practice in biology—a case study of phytosociology. *Zagadnienia Naukoznawstwa* 194 (4): 231-250.
- Kuusela, Vesa. 2011. *Paradigms in Statistical Inference for Finite Populations Up to the 1950s* Research Report 257, Statistics Finland.
- Lacey, Hugh. 1999. *Is Science Value Free?* London: Routledge.
- Lakens, Daniël, Neil McLatchie, Peder M. Isager, Anne M. Scheel, Zoltan Dienes. 2020. Improving Inferences About Null Effects With Bayes Factors and Equivalence Tests. *The Journals of Gerontology, Series B*, 75 (1): 45–57.
- Laudan, Larry. 2004. The Epistemic, the Cognitive, and the Social. In *Science, Values, and Objectivity*, ed. Peter Machamer and Gereon Wolters, 14–23. Pittsburgh: University of Pittsburgh Press.
- 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, vol. 1*, eds. L.M. Le Cam, R.A. Olshen, 1047–1060. Wadsworth Advanced Books & Software.
- Legg, Jason, C. and Fuller, Wayne, A. 2009. Two-Phase Sampling. In *Handbook of Statistics. Sample Surveys: Design, Methods and Applications, vol. 29, part A*, ed. C.R. Rao, 55-70. Elsevier.
- 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.

- Levi, Isaac. 1962. On the Seriousness of Mistakes. *Philosophy of Science* 29(1): 47-65.
- Levy, Paul S., and Lemeshow, Stanley. 1999. *Sampling of Populations: Methods and Applications* (4rd ed.), New York: John Wiley & Sons.
- Lindley D. V., Phillips L. D., 1976. Inference for a Bernoulli Process. *The American Statistician* 30: 112-119.
- Little, Roderick J.A. 2004. To Model or Not to Model? Competing Modes of Inference for Finite Population Sampling. *Journal of the American Statistical Association*, 99(466): 546-556.
- Little, Roderick J.A. 2012. Calibrated Bayes, an Alternative Inferential Paradigm for Official Statistics. *Journal of Official Statistics*, 28 (3): 309–334.
- Little, Roderick J.A. 2014. Survey sampling: Past controversies, current orthodoxy, and future paradigms. In *Past, present, and future of statistical science*, ed. Xihong Lin, 413–428. Boca Raton: CRC Press, Taylor & Francis Group.
- McCarthy Michael A. 2007. *Bayesian Methods for Ecology*, Cambridge: Cambridge University Press.
- Marks, Harry M. 2003. Rigorous uncertainty: why RA Fisher is important. *International Journal of Epidemiology* 32: 932–937.
- Mayo, Deborah G. 1983. An Objective Theory of Statistical Testing. *Synthese* 57: 297–340.
- Mayo, Deborah G. and Aris Spanos. 2006. Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction. *The British Journal of Philosophy of Science* 57: 323-357.
- Neyman, Jerzy and Egon S. Pearson. 1928. On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part II. *Biometrika* 20A: 263-294.
- Neyman, Jerzy. 1933. *Zarys teorii i praktyki badania struktury ludności metodą reprezentacyjną*. Warszawa: Instytut spraw społecznych.
- Neyman, Jerzy. 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: 558-625.
- Neyman, Jerzy. 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: 333-380.
- Neyman, Jerzy. 1938a. Contribution to the Theory of Sampling Human Populations. *Journal of the American Statistical Association*. 33: 101-116.
- Neyman, Jerzy. 1938b. O sposobie potrójnego losowania przy badaniach ludności metodą reprezentacyjną. *Przegląd statystyczny* 1: 150-160.

- Neyman, Jerzy. 1950. *First Course in Probability and Statistics*. New York: Henry Holt and Co.
- Neyman, Jerzy. 1952a. *Lectures and conferences on mathematical statistics and probability*. Washington: U.S. Department of Agriculture.
- Neyman, Jerzy. 1952b. Recognition of priority. *Jour. Roy. Stat. Soc.* 115: 602.
- Neyman, Jerzy. 1957. 'Inductive Behavior' as a Basic Concept of Philosophy of Science. *Revue De L'Institut International De Statistique* 25: 7-22.
- Neyman, Jerzy. 1977. Frequentist probability and frequentist statistics. *Synthese* 36: 97-131.
- Pearl, Judea. 2009. Causal inference in statistics: An overview. *Statistics Surveys* 3: 96–146.
- Royall, Richard. 1997. *Statistical evidence: A likelihood paradigm*. CRC Press.
- Royall, Richard M., & Herson, J. 1973. Robust estimation in finite populations. *Journal of the American Statistical Association*, 68(344), 880–893.
- Reid, Constance. 1998. *Neyman—from life*. New York: Springer-Verlag.
- Reiss, Julian and Sprenger, Jan, Scientific Objectivity. In *The Stanford Encyclopedia of Philosophy (Winter 2020 Edition)*, ed. Edward N. Zalta
- Romeijn, Jan-Willem. 2017. Philosophy of Statistics. In *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)*, ed. Edward N. Zalta.
- Samaniego, Francisco J. and Dana M. Reneau. 1994. Toward a Reconciliation of the Bayesian and Frequentist Approaches to Point Estimation. *Journal of the American Statistical Association* 89 (427): 947-957.
- Särndal, Carl-Eric. 1978. Design-based and model-based inference in survey sampling. *Scand. J. Statist.*, 5: 27-52.
- Särndal, Carl-Eric. 2010. Models in survey sampling. In: *Official Statistics Methodology and Applications in Honor of Daniel Thorburn*, eds M. Carlson, H. Nyquist, M. Villan 15–27. Stockholm: Stockholm University.
- Seng, You Poh. 1951. Historical Survey of the Development of Sampling Theories and Practice. *Journal of the Royal Statistical Society. Series A (General)* 114: 214–231.
- Singh, Sarjinder. 2003. *Advanced Sampling Theory with Applications. How Michael 'selected' Amy Volume I*. The Netherlands : Kluwer Academic Publisher.
- Smith, Fred T.M. 1976. The foundations of survey sampling. *Journal of the Royal Statistical Society. Series A (General)* 139, Part 2, 183-204.
- Soyfer, Valery N. 1994. *Lysenko and the tragedy of soviet science*. New York: Rutgers University Press.

- Sprenger, Jan. 2009. Statistics between Inductive Logic and Empirical Science. *Journal of Applied Logic* 7: 239-250.
- Sprenger, Jan. 2016. Bayesianism vs. Frequentism in Statistical Inference. In *The Oxford Handbook of Probability and Philosophy*, 382-405. Oxford: Oxford University Press.
- Sprenger, Jan. 2018. The objectivity of Subjective Bayesianism. *Euro Jnl Phil Sci*.  
<https://doi.org/10.1007/s13194-018-0200-1>
- Srivastava, A. K. 2016. Historical Perspective and Some Recent Trends in Sample Survey Applications. *Statistics and Applications* 14: 131–143.
- Ståhl, Göran., Svetlana Saarela, Sebastian Schnell, Sören Holm, Johannes Breidenbach, Sean P. Healey, Paul L. Patterson, Steen Magnussen, Erik Næsset, Ronald E. McRoberts & Timothy G. Gregoire. 2016. Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. *For. Ecosyst.* 3, 5.  
<https://doi.org/10.1186/s40663-016-0064-9>
- Steel, Daniel. 2010. Epistemic Values and the Argument from Inductive Risk. *Philosophy of Science* 77 (2010):14–34.
- Steel, David. 2011. Multistage Sampling. In *International Encyclopedia of Statistical Science*, ed. Miodrag Lovric, 896-898. Berlin: Springer.
- Sterba, Sonya K. 2009. Alternative model-based and design-based frameworks for inference from samples to populations: From polarization to integration. *Multivariate Behavioral Research* 44: 711–740.
- Tschuprow, Aleksandr A. 1923. On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron* 2: 461-493,646-683.
- Tillé, Yves, and Matthieu Wilhelm. 2017. Probability Sampling Designs: Principles for Choice of Design and Balancing. *Statistical Science* 32, no. 2: 176-89.
- Williamson, Jon. 2013. Why Frequentists and Bayesians Need Each Other. *Erkenntnis* 78 (2): 293–318.
- Zhao, Kino. 2021. Sample representation in the social sciences. *Synthese* 198, 9097–9115.