Prior Information in Frequentist Research Designs and Social (Non-epistemic) Influences: The Case of Neyman's Sampling Theory

Adam P. Kubiak and Paweł Kawalec
March 10, 2021

Abstract

We analyze the issue of using prior information in frequentist statistical inference by bringing out the sampling theory of Jerzy Neyman (a key figure in frequentist statistics), which has so far been largely ignored in philosophical discussions on frequentism. Our scrutiny of the different kinds of sampling designs supported by Neyman's theory reveals a variety of ways to explicitly, objectively engage with prior information. We argue that Neyman's approach to sampling enables researchers to let values classically classified as non-epistemic influence the procedure of collecting evidence and formulating statistical conclusions in order to not compromise the epistemic reliability of a procedure and even to improve upon procedures. With this, the discussed solutions of Neyman pose a methodological argument against the distinguishing of epistemic and non-epistemic values, and against the value-free ideal of scientific inference.

1. Introduction

Jerzy Neyman was a 20th-century statistician who is recognized 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). His main contributions to inferring 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 criticized as disadvantageous with regards to the Bayesian statistical paradigm[1] (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

[1] The debate between Bayesians and frequentists has been ongoing since the origins of the latter at the beginning of the 20th century.

largely neglected by philosophers of science,[2,3] even though his contribution to this field is significant (Little 2014) and still remains a standard element of present-day sampling frameworks (Srivastava 2016).[4] Bringing out 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 philosophical literature.

Only recently have some of Neyman's views on sampling been introduced to philosophical discussion (Zhao, 2020)—with Neyman being referred to as one of the representatives of the so-called design-based (as opposed to model-based) general approach to sampling.[5] Zhao referred to Neyman's statements concerning the general notion of a sample's representativeness and Neyman's critique of purposive sampling, but 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 randomization in which the use of prior information[6] concerning a population is minimized. This image of Neyman, as is shown in this paper, 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, with this being because Bayesianism explicitly encodes available prior information as a prior probability (e.g. Howson and Urbach 2006, 153–154). Furthermore, by leaving out priors, frequentism makes itself vulnerable to problems such as the base rate fallacy (Romeijn 2017).

On several occasions frequentism, and especially the approaches of Neyman and the Neyman-Pearson approach, was regarded as being poor at using prior information in the way it does. For example, it was stated that the frequentist procedure uses "implicit prior assumptions" (Sprenger 2009, 240); and that the frequentist inference assumptions that precede statistical inference, "are often hidden behind the curtain" while the Bayesian

[2] Even Mayo and Spanos, who covered a considerable part of Neyman's other methods and views in their works, did not touch upon this topic.
[3] Except perhaps for his conceptions of causal effect in a randomized experiment (Pearl 2009, 126-132).
[4] There are several other sampling plans in scientific literature that we do not discuss in this paper because we restrict our work to Neyman's contribution and the philosophical analysis thereof.
[5] Design-based inference is based on a pre-observational inference scheme that is dependent on sampling design but is independent from the values obtained from a possible probability sample, whilst the model-based inference scheme is based on post-observational considerations: it is conditioned by a model assumed on the basis of the one and only realized sample (see Särndal 2010, 116)
[6] 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. From this perspective it could also be termed "common knowledge" (cf. Lewis 1969, 53)

framework reveals such assumptions in more explicit way (Sprenger 2018). Bayesianism was thought to be superior when considering the "conventional" methods that are used in frequentist statistics, with this being because "conventional statistical methods are forced to ignore any relevant information other than that contained in the data" (McCarthy 2007, 2). This lack of sensitivity to context-specific prior information was ascribed to frequentist sampling design due to Neyman's ideas to the extent that the approach of Neyman was called "maximally uninformative" when discussing the use of prior information in sampling design (Zhao 2020). The approach of Neyman (and Pearson) to statistics was believed 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 contrasted with "building and adjusting a model to the data at hand and to the questions under discussion", which was thought to be a key feature of Fisherian's competing approach (Lenhard 2006, 84). Objections suggest that the use of prior information in frequentist statistical methodology, and especially in Neyman's methods is—to use terms from the above cited critical passages—„implicit", „hidden", „ignored" „uninformative", and „fixed" (not context-sensitive). This might make an overall impression that prior information is principally not utilized by Neyman's frequentist statistical methods in an objective and epistemically fruitful way. The important question is then whether the hypothesis of the uninformed and epistemically disadvantageous use of prior information in frequentism is strenghtened or weakened when it is considered from the perspective of Neyman's theory of sampling.

Our third source of motivation in analyzing Neyman's sampling designs is the debate concerning the role of non-epistemic values in science. 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 minimize the influence of these values on scientific reasoning (see e.g. Douglas 2009; Betz 2013); with non-epistemic values being called *contextual* values and including moral, political, and social values (Reiss, Sprenger 2017). Contrary to this stance, some authors (e.g. Steel 2010) argue that the influence of this type of values is 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. 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), but some authors (e.g. Rooney 1992) hold that distinguishing epistemic values from non-epistemic values has no unequivocal basis.

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). 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 in a process. The influence of premises (information) of an economic, cultural, moral, and other societal type 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.

In frequentist statistics, the choice of a sampling scheme is something that influences the process and outcome of statistical reasoning. This is done by determining how an estimator is mathematically defined (see e.g. Lindley, Phillips 1976) and by the obvious fact that a choice 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 also present in Neyman's sampling designs, and if 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 for a long time (see e.g. Levi 1962). 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 the influence of some social values on sampling design is beneficial epistemically, then we would have a methodological argument against the general distinction of social (non-epistemic) values and epistemic (not social) values: allowing social values to influence the process of drawing scientific conclusions would mean the simultaneous realization of some epistemic values. This in turn would pose an argument against VFI, as the ideal taken jointly with the classic distinction of social values as belonging to the non-epistemic type assumes that the influence of social (non-epistemic) values is epistemically adverse.

As the reader may see, the introduction of Neyman's ignored sampling theory into philosophical debate may contribute to the history and philosophy of statistics by filling the gap found in philosophical debate concerning Neyman's conceptions. This may also shed a new light onto the criticism of the lack of use or misuse of prior information by frequentist methodologies. Finally, our introduction may have consequences for the general philosophy of the scientific method, especially for the debate regarding the presence and role of non-epistemic values in scientific inference. This is why the goal of this paper is to analyze Neyman's framework for sampling with a focus on prior information in research designs in order to see how different types of this aforementioned information are used in Neyman's sampling designs. Our main contribution is to explicate the use of prior information in Neyman's sampling theory, and to show the consequences thereof for distinguishing epistemic and non-epistemic values in science and VFI. The goal of this paper is not to make comparative analyzes of the use of prior information in Neyman's approach and how prior information is used in other competing approaches. Neverthless, findings from our analysis might serve as a new premise for further reconsideration of the fine-grained outcomes of comparisons between competing paradigms thoroughly discussed in relevant literature (see e.g. Romeijn 2017; Mayo 2018).

This paper is structured in the following way. In the last paragraph of this section we explicate our understanding of the most basic concepts used in our analysis. In Section 2 we reconstruct Neyman's ideas regarding sampling from the perspective of the use of prior information, including information of social factors. Then, in Section 3, we draw conclusions regarding the twofold goal of our investigation. Firstly, we answer the question concerning the objectivity and epistemic import of the use of prior information in Neyman's sampling designs. Secondly, we show the consequences for distinguishing epistemic and non-epistemic values in science and VFI. In Section 4 we address some issues that can be raised in reference to classical problems of frequentist statistics. We end by presenting overall conclusions in Section 5.

In this paper we use the term *objectivity* (*objective*) 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 by means of the shared standards of its expression and use. The second is that procedures are not contingent on non-epistemic, including social, factors that would negatively influence the epistemic valour of procedures (see Reiss, Sprenger 2017). By the term *epistemic value* we understand a value whose influence poses a positive contribution towards 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. Firstly, (a) to be able to generate statistically reliable conclusions and to have control over the nominal level of false conclusions. Secondly, (b) to increase the accuracy of true conclusions. This means (a) 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 (b) maximize the accuracy[7] of an interval estimator[8] (minimize length of possible intervals) (see Neyman 1937). When we speak of the influence of social values on statistical inference we think of letting prior information of social factors to be implemented in sampling design and thus influence the process (and effect) of estimation in respect to aspects (a) and (b). The influences of social factors considered by us are the influences on reaching the epistemic goal understood as indicated in (a) and (b). Realization of the epistemic goal in its two described aspects can be understood as the realization of the 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.[9]

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

[7] In Neyman's terminology, the value of variance is the indicator of an estimate's accuracy (Neyman 1938a).

[8] The random variable (a function that assigns a real number to each possible outcome of a random phenomenon), which is used to generate estimates of a sought-out population parameter, is called an estimator. To give a simple example: an estimator of the population mean is a random variable $\bar{X}$ that is a function of random variables that refer to possible outcomes of particular trials (related to selections of one sampling unit), while the numerical value $\bar{x}$, which is a function of the observed values, is an estimate. The variance of an estimator is the expected size of the squared deviation of a random variable (an estimator) from its expected value; to simplify it can be said that variance is a measure of the expected departure of an estimator's value from the true one. An interval estimator generates numerical values in the form of intervals with known pre-observational probability of the generated interval will cover the true value of the population parameter.

[9] 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.

latter[10] (cf. Kuusela 2011, 91-93). In his groundbreaking paper (1934) Neyman compared two "very broad" (559) groups of sampling techniques that presuppose taking representative samples from finite populations—*random* sampling—in its special form of 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 the second broad group of sampling called 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, 567-68) including simple random sampling with or without replacement, and stratified and cluster sampling (discussed by us below in this paper), among others. The meaning of random sampling can be rephrased with more recent terms as follows: which of the possible sets of n sampling units are to be selected is not determined by a researcher's decision and each such set has a definite nonzero probability to be selected. The main tenet of assurance of a random selection of units, is, due to Neyman, that such a selection allows for the use of probability calculus: the conducting of interval estimation and the calculating of error probability, which, in Neyman's view, is not feasible in the case of purposive selection (1934, 559, 572, 586).[11] Therefore, the meaning of purposive selection might be rephrased as follows: which of the possible sets of sampling units is selected is determined by a researcher's decision and it is either impossible to ascribe probabilities to the selection of a particular possible set, or these probabilities are a priori known to be either 0 or 1.

2.1. Stratified Sampling

Stratified sampling is a kind of probability sampling[12] in which, prior to drawing a random sample, a population is divided into several, mutually exclusive and exhaustive groups (groups are 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 natural way of sampling, e.g., in a survey about support for a new presidential

[10] One of the still most commonly used descriptions of a *representative sample* is that such a sample is a *miniature of the population* (cf. Dumicic 2011, 1222-1224).

[11] Neyman recognized 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).

[12] Probability sampling is the case in which 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. In some cases random sampling can be pathological, with some units having zero chance of selection (like in nonresponse sampling). Neyman did not offer any systematic mathematical solutions to pathological sampling but such solutions emerged later.

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 sub-populations of a strata. If the number of persons interrogated differs among the strata in a way that the proportion of a sub-sample size to the size of a sub-population will be the same for each province, then every inhabitant of the country will have the same chance of being included in the survey. Such a form of stratified sampling, where inclusion probabilities are equal for every member of a population, prevailed at the time of the publication of Neyman's classic paper in 1934. A simple example can help to understand this idea. Imagine a country has three provinces. Imagine the number of inhabitants in each province is 25, 10, and 5 respectively. The sample size is to be comprised of 8 trials, thus 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 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 this factor—we discuss this in later parts of this section.

2.2. Cluster Sampling

It is often very difficult to sample individual units of a population. Sometimes the characteristics of a population or its environment make this impossible in principle, such as, for example, in the case of determining the value of weekly church donations per person in a particular city. Such information is in principle a secret for many parishioners. In this case, a possible way of data collection would be to treat parishes with a known number of parishioners and a known sum of donations as a *cluster*—a new sampling unit of a higher order. In other cases, the cost of sampling units is simply too high compared to its benefits, all things considered, as in the case of investigating per capita food expenditure. It is much easier to get to know what the monthly food expenditure of a household is with the known number of members of a household than it is to draw a particular member of a household and to know 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 used by an individual member. One of the approaches in such a case is to select clusters of units, like households, as the units of investigation, instead of units themselves.

Thus, strictly speaking, this type of sampling, as understood by Neyman, consists in treating groups of individuals as units of sampling. Clusters as groups are collectives of units that are always taken into consideration together: some groups are randomly selected and all members of the selected group are included in the sample. Strata as groups are conceived as subsets of a population of sampling units and from every distinguished subset (stratum), some units are drawn at random. For example, if a country's districts were treated not as strata, but as clusters, then random drawing would not apply to units inside a district, but 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 generalized, while in other cases, a generalized measure is immediately available (being unique). This second case would be the just mentioned examples of parishes and households, where measures of an element's attributes are not available in turn. Cluster sampling seems to correspond to the natural structure of many studied populations, and sometimes such sampling may be socio-economically the only reasonable sampling scheme, especially in regard to humans, as "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 opposed to the multi-stage type, in which clusters are randomly selected in the first stage but random selection is continued in follow-up stage(es) 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 reconceptualized as a population of clusters, and stratification thus performed on the reconceptualized population of clusters. This was the case that Neyman used for his analyses in his 1934 paper. This does not preclude the consideration of assumptions, roles, and consequences separately for clustering and stratification, with this also being the case in Neyman's analysis.

2.3. Prior Information in Stratification/Clustering, and Consistent Estimation

What Neyman has mathematically shown is that the information on how a population is organized and socio-economic factors like those above described can be objectively applied in the process of scientific investigation at the stage of designing the sampling scheme. He has shown how to let these factors influence the process of statistical inference—thus how to let social values influence statistical inference in order to enable statistically reliable conclusions and for there to be control over the nominal level of false conclusions; this means to reach the epistemic goal in its (a) aspect.

Even the arbitrariness of stratification and/or clustering does not rule out the feasibility of an estimation that will satisfy the (a) aspect of the epistemic goal by the proper use of Neyman's concept of the best estimator, today called the best[13] linear unbiased estimator (*B.L.U.E.*),[14] which was presented in Neyman's 1934 paper and denoted the linear unbiased[15] estimator of minimal variance (Neyman 1934, 563-567). In Neyman's terminology, the value of variance was the indicator of an estimate's *accuracy* (Neyman 1938a). 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 organized into clusters (Neyman1934, 579). An increase in accuracy of estimation means *shorter* confidence intervals (see Neyman 1937, 371).

Neyman's analysis of stratified and clustered sampling designs[16] indicates how to properly implement information—known prior to the research process—concerning how a population is organized and what are the socio-economic factors like those described earlier. He has mathematically argued that information that represents the influence of these factors can be implemented in an explicit, objective way—without a negative influence on the correctness of an estimation's procedure. This means without obstructing consistent estimation. As we argued, in some cases, letting these factors influence a procedure is even the only way to undergo a consistent estimation, which would be impossible otherwise (see. e.g. the example of households). In other cases this is necessary if one wants to comply with social standards, like in the example of parishes, in which, as we have already explained, ethical standards would require not to sample individuals.

Consider the epistemic goal in its (a) aspect, i.e. in being able to generate statistically reliable conclusions and to have control over the nominal level of false conclusions. Properly

---

[13] 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.

[14] Neyman (1934, 564-565) stressed that he didn't state that linear estimators are the best in an "unequivocable" sense and argued for the choice of 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.

[15] 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 the biased estimator, for which the expected value is not equal to the true value).

[16] The mathematical background for Neyman's statements concerning (stratified) cluster sampling can be found in full detail in Neyman (1933, 33-69).

designed cases of stratified and clustered sampling can become cases of the influence of social (non-epistemic) values on the process of collecting evidence and formulating scientific conclusions in which influence not only does not obstruct the epistemic goal, but can contribute to the realization of this goal by making it possible to generate statistically reliable conclusions and to have control over the nominal level of false conclusions when this would be difficult to achieve otherwise.

2.4. Purposive Selection and Optimum Allocation Sampling

In contrast to the method of stratified sampling (or, more generally, the method of random sampling), purposive selection, since Neyman's paper from 1934, has been aimed not at random selection, but at the maximal representativeness of a sample obtained by means of 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.[17] This kind of approach may sometimes appear like quite a natural attitude for a researcher. For example, consider an ecologist who wants to assess the difference in blooming periods of certain herb species, say, *Hepatica Nobilis*, 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 exclude from sampling from those forest sites (and thus those individuals of the herb) that are to a large extent subjected to the local extreme (abnormal) disturbances of the aforementioned factor. This can be explained as an attempt to minimize 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. The avoidance of the risk of selecting an extreme sample entails there being a more representative sample, which ideally should secure the better accuracy of the assessment of the population's characteristic in question.

Due to Neyman, the basic assumption of the method of purposive selection was that the values of an investigated quantity (ascribed to particular elements of the investigated population from which a sample is to be taken) are correlated with the auxiliary variable and

[17] Purposive sampling was anticipated by the *monograph* method, in which entities typical to a population were selected based on the concept in which a population as a whole "had a reality external and superior to individuals (…) whereas individuals were simply contingent manifestations of it" (Desrosières 1998/1993, 214).

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, the selection method that will supply a very representative sample must comprise of the purposive selection of such elements of the population for which the mean value of the auxiliary variable will have the same value, or at least as nearly the same as is possible, as it has for the whole population (see Neyman 1934, 571). This can be motivated by the following simple example. If the quantity of a parish's average weekly income from donations is positively correlated with the mean age of its members, then, if most of the parishes from the investigated population were "old" (in terms of the average age of members), the sample should include an adequately larger number of old 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 we already mentioned, before Neyman's reconsideration of the conception of purposive selection and the proper mathematical explication thereof, purposive selection concerned the idea of non-probability sampling. Neyman modified the then adopted concept of purposive selection, with this becoming a special case of random sampling. What was assumed to differentiate random sampling from purposive selection before Neyman's paper, was, firstly, 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). Secondly, 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 in the form of 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) and thus the concept of purposive selection became a special case of random sampling that takes the form of stratified sampling of clusters. The consistency of estimation does not mean that its accuracy cannot be further improved by an appropriate sampling scheme. Neyman stressed that reconceptualized 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 an improved method that is today called *optimum allocation sampling*.

Neyman showed in his analysis of how to minimize the variance 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 size of the partial 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).[18] The variance of an estimator is proportional to the variability of the research variable within a strata. Therefore, to minimize the variance of the estimator by optimal sample allocation, the sample size for a stratum should be proportionate to the quotient of the size of a stratum and 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 strata and subsequently a more accurate estimation (see Neyman 1933, 41, 89; 1934, 579-580). In the case of the absence of any ready data, estimation of the variability of the investigated quantity within strata requires preliminary research; the result of such an initial trial may subsequently be reused as a part of the actual (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).

2.5. Prior Information in Stratification/Clustering and Accuracy of Estimation

The methodological ideas proposed by Neyman that we presented in Subsection 2.4. 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, an outcome (its accuracy) can depend on the correlation of an investigated quantity with another

[18] What was overlooked by Neyman is that Tschuprow (1923) also derived this rule of optimal sample allocation. Neyman acknowledged Tschuprow's priority in later years (Neyman 1952b).

quantity known prior to sampling in order to yield an epistemic profit of an estimation's increased accuracy. The same holds for implementing prior information about the estimated variability of an investigated property.[19]

If clusters are the elements of sampling, minimizing their size also increases the accuracy of an estimator (Neyman 1934, 582). Making clusters comprised of the same number of entities gives the same effect (Neyman 1933, 90). What was not addressed by Neyman is that more internally heterogeneous clusters also increase the accuracy of an estimation. An important upshot of the above is that pre-study information concerning some social factors that govern 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 in statistical procedures in an objective way to increase the accuracy of estimation. This means the epistemic benefit as seen from the (b) aspect of the epistemic goal.

2.6. Double Sampling

Thus far, we have considered prior information, such as the structure of a population, the variability of the research variable, and the research variable's regression on other variables, which could in many cases be classified as belonging to the category of information concerning the natural characteristics of studied phenomenon. Some of this information may be said to be partially of a social or economic character (see e.g. the case of households, parishes, and provinces from the previous subsection of this section). Now we turn to Neyman's sampling design that explicitly embodies the eminently social factor that appears to be extremely conventional on the one hand while, on the other, inevitably and essentially influences the process of collecting evidence and formulating conclusions—the prior information regarding the costs of research.

It is established in statistics that Neyman "invented" (Singh 2003, 529) or "developed" (Breslow 2005, 1) a method called *double sampling* (Neyman 1938a) or *two-phase sampling*

---

[19] Except for improving the accuracy of CI, there exist other aspects of the epistemic advantages of stratified sampling 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 optimizing estimators in so-called model-assisted estimation techniques (see e.g. Royall, Henson, 1973), that are exploited, for example, in small area estimation.

(Legg, Fuller 2009). Neyman, in his analysis of stratified sampling (1934), proved that if a certain auxiliary characteristic is well known for the population, we can use it to divide the whole population into strata and undertake optimum allocation sampling which would result in the improving of 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 a variable of interest is very expensive and because preliminary thorough knowledge of an auxiliary variable, which could improve the estimate's accuracy, is not yet available. The sampling procedure is as follows: the first step 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 sample, 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 be made according to the optimum allocation principle (Neyman 1938b, 153).[20]

The main problem in double sampling is how to rationally allocate the total expenditure thereof into 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, the average value of food expenditure per family in a certain district is to be determined. Because the cost of ascertaining for one sampling unit the value of this research variable is very high, limited research funds allow for only the collection of a quite small sample. However, the attribute in question is correlated with another attribute, for example, a family's income, of which the per-unit cost of obtaining this datum 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 described above.

Neyman provided formulas for the allocation of funds in double sampling that yield greater accuracy of estimation compared to when estimation is calculated from data obtained in one-step sampling and with the same amount of available funds. Nevertheless, in certain circumstances, double sampling will help to avoid the unnecessary loss of the accuracy of an estimate, while in others, it will lead to less accurate results. Neyman indicated that certain

[20] 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 comprising only of entities that belong to the sample obtained at the first stage.

preliminary information must be available in order to verify whether the sampling pattern will lead to a better or worse accuracy and to know how to allocate money (Neyman 1938a, 112-115). Because of this, double sampling requires prior estimates of the following characteristics: the proportion of individuals belonging to a first-stage strata, the standard deviation of the research variable within a strata, the mean values of the research variable in strata, and, obviously, the costs of gathering datum of the auxiliary variable and research variable per sampling unit (see Neyman 1938a, 115).[21]

To increase the efficiency of estimation 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, one requires some prior information concerning the above-indicated properties of the population sampled. This is also needed to approximately determine the optimal size of samples (see Neyman 1938a, 115).

2.7. Prior Information in Double Sampling and Accuracy of Estimation

In the conclusions found in Subsection 2.5., we stressed that using information about the relations between a research variable and other known characteristics may serve to plan a sampling pattern that allows for one to make a more accurate estimation. Double sampling instructs the researcher on how to obtain such information as to improve estimation in a resource-limited context. Some knowledge of the value of the preliminary estimate of the research variable and of its variance is essential in regards to the technique of double sampling. This technique allows one to also include typical economic factors: the costs of different types of data collection and available research funds.

This method shows how to use prior information concerning the structure of a population (in the aspect of an auxiliary variable interrelated with a research variable), information about the estimated values of a research variable, its variability, and information concerning financial possibilities and limitations. This rigid use influences the estimation procedure and its effects in an objective way. More importantly, this method guides a researcher towards the realization of the second (b) aspect of the epistemic goal: the correct use of these types of information can increase the accuracy of estimation.

[21] Given the need for some previous knowledge or preliminary estimation of these quantities, Neyman ultimately labeled the method “triple sampling” (Neyman 1938b, 150).

3. Methodological and Philosophical Consequences

3.1. The Objective and Epistemically profitable Use of a Vast Spectrum of Prior Information

As argued persuasively by Johann Lenhard (2006), one of the advantages of Neyman's approach is that he proposes an integrative mathematical model of the research process, which itself becomes "an object of mathematical argument" (Lenhard 2006, 84). In terms of this approach, it is on mathematical grounds that the most efficient course of action can be determined. Kino Zhao (2020) suggested, that, for Neyman, the core of this mathematical design was maximally uninformative randomization based on the presumption that "maximal noninformation precludes outside factors from systematically affecting ("informing") a sample's composition". What we have shown in Section 2 is that this is the other way around. In Neyman's approach, outside factors affect a sample's composition in a very informed way. This is done by implementing into mathematical sampling designs a possibly vast scope of available prior information regarding the domain in question: prior estimates regarding the research variable, correlations between the sought-after quantities and collateral factors, and the social factors that accompany research.

It can be said, that the choice of the primary unit of the population studied and the division of the population onto strata can be done by taking into account information concerning the existing population structure (e.g. in the case of separated provinces), economics (e.g. the household as units of investigation of people's food expenditure), and by respecting moral/social norms (e.g. when one asks about the average donation of a church member). A different choice of the primary unit of selection or a different stratification is objectively dependent on this information and the realization of the epistemic goal in its (a) aspect is not hindered by this dependence. Moreover, if auxiliary factors (like a natural factor that is disturbing a species' blooming, or a social factor in the form of income influencing expenditure) are known to be correlated with the quantity in question, and this knowledge is implemented in the form of appropriate sampling design, then the estimation procedure becomes epistemically more reliable— the accuracy of estimation increases, which means that confidence intervals become shorter. This is possible thanks to redefined purposive selection (which was regarded by Neyman as very rarely applicable), and optimum allocation sampling. If a thorough preliminary knowledge of an auxiliary variable, which could improve an estimate's accuracy, is not yet available, one can use the double sampling scheme. This design guides a researcher on how the accuracy of an estimator can be improved in a resource-limited situation and when knowledge regarding manifold costs and available funds is implemented in the research process. These manifold types of prior information are used at the stage of planning and executing the collection of evidence. Their implementation not only does not obstruct the correctness of statistical estimation, but can even be beneficial to it in two ways: by enabling estimation which would be hard to execute otherwise (like in the case of the investigation of food expenditure per capita), and by increasing an estimator's accuracy. We refer to all of the indicated information that can be used in this way as *prior contextual information*. This phrase indicates that Neyman's method makes use not only of prior information relating directly to a sought quantity but also of other types of knowledge that relate to a sought quantity indirectly and of those that tell us about other, non-cognitive 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, we propose to distinguish the three main types of prior contextual information used in Neyman's sampling designs:

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 has an influence on scientific inference and conclusions derived—shortening a confidence interval means changing the contents of a conclusion. As argued in the previous section, in many cases this influence will have a positive impact on the realization of the epistemic goal. This influence can contribute to the realization of the epistemic goal by making it possible to generate statistically reliable conclusions and to have control over the nominal level of false conclusions—the (a) aspect of the goal, with this being difficult otherwise (see 2.3). This can also contribute to the realization of the (b) aspect of the epistemic goal by increasing the accuracy of estimation (see 2.5 and 2.7).

3.2. The Role of Social Values in Collecting Evidence and Formulating Scientific Conclusions

As we argued in Section 1, the existence of the methodological influence of social factors represents an embodiment of social values, which are classically understood as being in opposition to epistemic values. As we have shown in Section 2, the influence of social values takes place in the form of sampling design driven by such values as cost-effectiveness, practical convenience, or compliance with social standards. This influence takes effect in the form of social-value-laden appropriate clustering, stratification, and sample allocation.

The fact, that the influence of social values can have a positive epistemic influence on estimation in either the (a) and/or (b) aspect of the epistemic goal as earlier described, reveals that the classic dichotomous distinction between epistemic and non-epistemic factors/values become pointless as one moves towards a serious and engaged consideration of a given domain, which also includes the available results of measurements; this is what currently makes Neyman's approach especially topical. This point is well illustrated by a number of recently debated research areas, most notably climate change (for an overview, 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 (see Kubiak, Wodzisz 2012).

The relaxation of the discussed dichotomies entails a general philosophical conclusion concerning the recent revival of the debate about the value-ladenness of science. 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 ignore these contexts of research activities and exclusively "focus on the logic of science, divorced from scientific practice and social realities" (Douglas 2009, 48). What we have concluded is that Neyman's sampling method covers common non-epistemic factors such as financial factors, technical convenience, and moral considerations. Admittedly, these do not exhaust all possible factors, but still embrace 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 realization of the epistemic goal in the two aspects discussed in this paper. If these social values can have a positive epistemic import, then the classic dichotomy between epistemic and non-epistemic (social) values in the scientific processes is undermined. The value-free ideal of science assumes such a dichotomy, therefore it is also undermined. What we have indicated in the introductory section is that the VFI states that contextual, social values are non-epistemic values and should not influence the process of collecting evidence and formulating scientific conclusions. Our analysis has shown that realization of the considered social values by letting them influence this process and these conclusions simultaneously becomes a realization of the epistemic values of 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 the sake of epistemic profits. In the following section we comment on how our argumentation relates to some issues that can be raised in reference to some classical problems of frequentist statistics.

4. Discussion

We have argued that although in the case of Neyman's frequentism prior information related to an investigated quantity is implemented differently than in the case of Bayesianism, the implementation thereof remains objective. Still, one might then object that what makes the use of prior information unjustified is that it is not used in a Bayesian manner, that is, in the form of the prior probability of a sought quantity. However, evaluating whether prior information is implemented in a right or wrong way from a perspective outside of the framework of a given paradigm in which an investigation is done shifts the discussion to the issue of the fundamental conceptual and metaphysical assumptions of both paradigms. Neyman himself had strong metaphysical and conceptual reasons for discarding the attribution of probability (prior or posterior) to sought-out quantities, one of which was that he interpreted the investigated parameter value as an unknown constant and not as a random variable while ascribing probability to the investigated parameter value would be to speak of it as a random variable and would eventually—in terms of the way he was connecting probabilities to empirical frequencies—violate the so-called empirical law of large numbers[22] (Neyman 1937, 340-345; cf. Neyman 1952a, 18, 27). Nevertheless, the fundamental philosophical topic of interpreting issues surrounding the concept of probability itself is beyond the scope of this paper; in particular the controversy regarding the epistemic interpretation of particular outcomes that engages some fundamental issues concerning the interpretation of probability (cf. e.g. Hájek 2012).[23] The problem with epistemically interpreting a particular outcome is a classic (see Spielman 1973) problem, but it does not affect our conclusions. The concept of the epistemic goal as explicated in Section 1 refers to the epistemic reliability of a procedure. One can speak of the long-run epistemic reliability of an estimation procedure without assuming any kind of epistemic interpretation of a particular outcome. One element of this reliability is the probability that the estimator will cover the true value, and the other is how close on average the estimator will approximate the true value in the case of covering the true value, which—in Neyman's terminology—means how accurate this estimator will be on average (what its width will be). The first aspect of the procedure's epistemic reliability can be assessed based on the nominal value of the confidence level, and the second on the basis of the Monte Carlo simulation, which is a standard way of examining the properties of estimators. Input variance for this simulation can be estimated, for example, on the basis of its prior estimation or current sample variance. Technical issues such as the topic of the relation of nominal reliability to the actual reliability of confidence intervals are beyond the scope and goal of our paper. We only mention that Bayesian convergence theorems are also limiting properties, and may therefore face similar problems from such a perspective.

To sum up, it can be stated that a method's pre-observational propensity to derive true and accurate conclusions can be given the status of the epistemic characteristic of a procedure. This is plausible even if this epistemic feature of a method is different from the type of epistemic aspect that might be of Bayesians' major interest, namely the post-observational credibility (probability) of an outcome (conclusion).[24] These pre-observational and post-

[22]It is a statement that many stable empirical frequency distributions have been observed in the world that are numerically close to those theoretical distributions derived by mathematical models and the laws of statistics. For example, that the mean value obtained from fairly symmetric die throws converges to the expected value calculated for such a die is enormously well empirically confirmed (see e.g. Freudenthal 1972). This is why using mathematical laws to approximate relative frequencies of observations is very well justified (see Neyman 1952a, 19-27; Mayo 2018, 254-258).

[23] Neither do we aim at a resolution of the long-lasting disagreement between Neyman and his prominent adversary, Sir Ronald Fisher. The intricacies of this debate are investigated in greater detail in e.g. (Lehmann 2011; Sabbaghi and Rubin 2014).

[24] The conflation of pre-observational and post-observational perspectives is a classic example of misplaced critique of the epistemic reliability of frequentist methods (see Graves 1978).

observational aspects of the epistemic credibility of a procedure and of an outcome respectively may be conceived as complementary, and not as competing aspects (see Rochefort-Maranda 2017). Similarly, both Bayesianism and Neyman's frequentism use a priori information in an essentially different way but in an internally justified manner. The status of the universal superiority of any of the two approaches to estimation is inconclusive and preferability depends on the ccircumstance of research (Samaniego, Reneau 1994) Also, "There are certain statistical scenarios in which a joint frequentist-Bayesian approach is arguably required" (Bayarri, Berger 2004, 59). Neyman's frequentist sampling designs as connected with his estimation theory can be complementary to Bayesianism and could even be perceived as superior in some respects, as Neyman's designs explicitly makes use of a very wide spectrum of types of relevant auxiliary information, and not only the prior information (preliminary estimate) of an estimated quantity. Moreover, adequate use of prior information in sampling designs can give epistemic profit. Perhaps some Bayesians might object that this is false because the proper aspect of evaluating the use of prior information for the sake of epistemic profit should be post-observational, not pre-observational. But, if the Bayesian perspective is one possibility, not universally superior, and sometimes complementary to the frequentist perspective, such an objection would not pose a valid counterargument. Neyman's frequentist use of prior information in sampling and estimation is objective and epistamically profitable on its own way. This way can be regarded equally valid and relevant in practice, as the Bayesian way of using prior information is.

A different angle from which Neyman could be criticized is the accusation of his not proposing a method for incorporating prior information into scientific research that would be sufficiently generalized, and his offering of a bunch of local solutions instead; high generality is in turn what the Bayesian approach appears to offer. Indeed, Neyman proposed to differentiate the method of dependent sampling on some assumptions regarding prior information, but such an approach cannot be regarded as a drawback in light of what he proved, namely that applying the same sampling template under the circumstances of prior information of a different type can be a less effective approach in definite cases. Additionally, the lack of generalized approach is not what can be fully attributed to his solutions: the unifying element thereof is the criterion of proper stratification (which, for obvious reasons, prescribes a different stratification given different prior information) and the optimum allocation principle. A more general criterion could be said to be the principle of constructing sampling designs that will allow the use of the B.L.U.E. estimator in a given context of available prior information.

Another issue, somewhat independent from the Bayesian critique, is if whether the prior contextual information proposed to be used by Neyman is itself reliable. In particular, Neyman's designs seem to presuppose, for example, that the collection of data is randomized not only theoretically but practically when prior estimation was conducted, that the population is indeed clustered, that the costs of particular types of observation are truly what we claim to be, that the data at hand is not simply an effect of fraud, etc. Similar concerns may apply to the manifold sources of error in actual research, for example the technical problems of gathering specific types of evidence, the influence of unknown factors on observational outcomes, and the precision and bias of measurement tools.[25] All these aspects are subjects of vivid philosophical and methodological discussions but are not concerned with the question of whether the logic of statistical design is correct, but rather concern if whether the design is correctly implemented or whether its premises are satisfied in a particular case. Analogous problems affect the Bayesian concept.

Although the reliability of prior information is not a problem that might threaten our statements presented in Section 3, a serious concern might be the possible subjectivity of stratification—what if there is more than one possible correlation known based on which different stratifications could be made? Neyman replies 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 by means of the "weighted regression" of the variable of interest upon multiple controls (cf. Neyman 1934, 574-575). In the cited paper, Neyman did not intend to develop an ultimate technical solution explaining how to best merge information about the regression of several variables, with this requiring one to include the regression of one particular auxiliary variable on another, and so on. Still, this is quite a technical and diverse issue discussed in statistical science (it has also been discussed by Neyman in the context of his "potential yield" experimental design, which is not the topic of our paper) and does not influence our general conclusions. This, again, is because it pertains to the reliability of prior information, which is unrelated to the goals of our investigation of the reliability of the methods of making inferences with the use of premises concerning prior information, and not to the reliability of the premises themselves.

[25] This is a physical/technical property that should not be conflated with an estimator's bias, which is a mathematical property.

5. Conclusion

Analyses of Neyman's sampling designs have been largely ignored in philosophical debates. We presented a self-standing reconstruction of the designs thereof. As a result of this, we have shown that Neyman's sampling designs enable a statistical estimation that would be hard to execute otherwise and minimize the variance of an estimator thanks to the use of a vast spectrum of prior contextual information about the presence of natural mechanisms, about the attributes of investigated populations, and socio-economic contexts.

The methodological result stemming from our investigation of Neyman's sampling designs weakens possible Bayesian methodological superiority in that it is opposed to the alleged frequentists' adhockery in using prior information, their lack of objectivity of the use thereof, the lack of clarity in how prior knowledge is used, the inconsistency of that use, etc.

The second, more philosophical conclusion entailed by our analysis concerns the recent revival of the debate concerning the value-ladenness of science. Neyman's sampling theory, which is a standard element of present-day sampling framework, allows for a systematic study of the complementarities between different kinds of values—epistemic, and non-epistemic—implemented in statistical methodology, with there being a possible positive effect of non-epistemic values on the method's epistemic reliability. This undermines a simplistic view of the research process which is presumed by the dichotomic epistemic vs. non-epistemic influence of values in the scientific process. By that, this undermines the value-free ideal of science.

6. References

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

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*. Cambrigde: Harvard University Press.

Douglas, Heather E. 2009. *Science, Policy and the Value-Free Ideal.* Pittsburgh: University of Pittsburgh Press.

Douglas, Heather E. The Value of Cognitive Values. *Philosophy of Science* 80: 796-806.

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: p. 47-60.

Freudenthal, Hans. 1972. “The 'Empirical Law of Large Numbers' or 'The Stability of Frequencies'” *Educational Studies in Mathematics* 4 (4): 484-490.

Giere, Ronald N. 1969. Bayesian Statistics and Biased Procedures. *Synthese 20*: 371-387.

Hacking, Ian. 1965. *Logic of Statistical Inference*. London: Cambridge University Press.

Hájek Alan. 2012. Interpretations of Probability. *The Stanford Encyclopedia of Philosophy,* (Winter 2012 Edition), ed. Edward N. Zalta.

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.

Kneeland, Hildegarde, 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., Wodzisz, 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.

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. 2011. *Fisher, Neyman, and the creation of classical statistics*. New York: Springer.

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.

Lewis, David. 1969. *Convention: A Philosophical Study*, Cambridge, MA: Harvard University Press.

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. 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. 2018. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge: Cambridge University Press.

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. 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, R. 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 2017 Edition)*, ed. Edward N. Zalta

Rochefort-Maranda, Guillaume. 2017. Statistical Power and P-values: An Epistemic Interpretation Without Power Approach Paradoxes. (Preprint) http://philsci-archive.pitt.edu/id/eprint/14220

Romeijn, Jan-Willem. 2017. Philosophy of Statistics. In *The Stanford Encyclopedia of Philosophy (Spring 2017 Edition)*, ed. Edward N. Zalta.

Rooney, Phyllis. 1992. On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful? In *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1: 13–22.

Sabbaghi, Arman, and Donald B. Rubin. 2014. Comments on the Neyman-Fisher Controversy and Its Consequences. *Statistical Science* 29: 267–284.

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. 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.

Spielman, Stephen. 1973. A refutation of the Neyman-Pearson theory of testing. *British Journal for the Philosophy of Science* 24(3): 201–222

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.

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.

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.

Zhao, Kino. 2020. Sample representation in the social sciences. *Synthese*. https://doi.org/10.1007/s11229-020-02621-3