ZAGADNIENIA
NAUKOZNAWSTWA
2 (200), 2014
PL ISSN 0044 – 1619

*Adam P. Kubiak*
Faculty of Philosophy
John Paul II Catholic University of Lublin

# A Frequentist Solution to Lindley & Phillips' *Stopping Rule* Problem in Ecological Realm

**Abstract.** In this paper I provide a frequentist philosophical-methodological solution for the stopping rule problem presented by Lindley & Phillips in 1976, which is settled in the ecological realm of testing koalas' sex ratio. I deliver criteria for discerning a stopping rule, an evidence and a model that are epistemically more appropriate for testing the hypothesis of the case studied, by appealing to physical notion of probability and by analyzing the content of possible formulations of evidence, assumptions of models and meaning of the ecological hypothesis. First, I show the difference in the evidence taken into account in different frequentist sampling procedures presented in the problem. Next, I discuss the inapplicability of the Carnapian principle of total evidence in deciding which formulation of evidence associated with a given sampling procedure and statistical model is epistemically more appropriate for testing the hypothesis in question. Then I propose a double-perspective (evidence and model) frequentist solution based on the choice of evidence which better corresponds to the investigated ecological hypothesis, as well as on the choice of a model that embraces less unrealistic ontological assumptions. Finally, I discuss two perspectives of the stopping rule dependence.
**Keywords:** sampling, principle of total evidence, binomial distribution, significance test, point null hypothesis, objective probability, accuracy, data, relativity, arbitrariness

## Częstościowe rozwiązanie Lindleya i Phillipsa problemu reguły stopu na przykładzie z dyscypliny ekologii

**Abstrakt.** W niniejszym artykule przedstawiam cząstkową propozycję filozoficzno-metodologicznej obrony klasycznej metodologii testowania hipotez statystycznych. Dokonuję tego poprzez próbę odpowiedzi na zarzuty wynikające z problemu stopu przedstawionego przez Lindleya i Phillipsa w 1976 roku, a osadzonego w realiach hipotezy ekologicznej związanej z pytaniem o stosunek płci u młodych osobników gatunku koala. Rozważania rozpoczynam od przedstawienia problemu, następnie omawiam aplikowalność Carnapa zasady całkowitej obserwacji i związek możliwych form obserwacji oraz modeli z hipotezą ekologiczną. W konsekwencji przedstawiam dwa rozwiązania przedstawionego problemu: pierwsze z perspektywy porównywania alternatywnych form formułowania obserwacji, drugie z perspektywy porównywania modeli. Obydwa odwołują się do ontologicznych konsekwencji klasycznej obiektywnej interpretacji prawdopodobieństwa oraz postawionej hipotezy ekologicznej. Argument kończy przedstawienie dwóch perspektyw zależności wyniku testowania hipotezy od wyboru reguły stopu.
**Słowa kluczowe:** próbkowanie, zasada całkowitej obserwacji, rozkład dwumianowy, test istotności, punktowa hipoteza zerowa, prawdopodobieństwo obiektywne, trafność, dane, relatywizm, arbitralność

## 1. Introduction

Debates on Bayesian *versus* classical (orthodox) statistics are present in multiple disciplines, including ecology, within which diverse Bayesian-oriented textbooks have been recently published (Link & Barker 2010; Kéry & Schaub 2012). Yet, the Bayesian and the orthodox methodologies of scientific inference are, among some statisticians as well as some philosophers, acknowledged as desirably compatible (Efron 2005) or even inevitably complementary (Williamson 2013). Methodological discussion in ecology goes beyond these two frameworks and also embraces other approaches like information theoretic methods and effect size statistics (Stephens et al. 2006).

One of the objections to classical hypothesis testing is a renowned problem known as Lindley's paradox (Lindley 1957). An account of the standard Bayesian critique that relates to this paradox can be found in McCarthy's handbook *Bayesian Methods for Ecology* (2007) where the paradox is embodied into the ecological realm of testing the sex ratio of pouch young of koalas' mothers in poor physical conditions. McCarthy's main objection is based on Lindley & Phillips' (1976) 'stopping rule problem', which directly relates to the paradox.

A researcher could collect the data with respect to either a fixed number of trials, or a fixed number of successes (males). In the case of Bayesian analysis, the result of testing will be insensitive to how a researcher decides to stop the sampling. In the case of the point null hypothesis significance test (NHST), according to McCarthy, it may be the case that "two different stopping rules for the sampling strategies lead to different conclusions about the null hypothesis, even though the actual data are identical" (McCarthy 2007, 33). This statement is challenging for the proponent of the frequentist classical framework, unless she will show that there is no arbitrariness in her choice of the stopping rule. To do this she should provide a criterion for identifying which formula is epistemically favorable to be applied in the case of the specific hypothesis.

Although a philosophical defense of the physical interpretation of probability (Strevens 2013) and NHST (e.g. Mayo 1996) is present in the literature, there is no paper which provides an explicit frequentist philosophical answer to stopping rule problem as formulated by Lindley & Phillips. Also ecologists compare Bayesian statistics with NHST framework, but are usually proponents of the first and present at best moderate acceptance of the latter. Frequentists' response to Bayesians' critique seems to be neglected in this field of life sciences – only one small contribution (Dennis 1996) clearly advocates the orthodox method, yet it does not refer to Lindley & Phillips' stopping rule problem.

The aim of my argument is to provide a frequentist philosophical-methodological solution for Lindley & Phillips' stopping rule problem in the ecological realm of testing koalas' sex ratio.

I deliver criteria for discerning a stopping rule, an evidence and a model that are epistemically more appropriate for testing the hypothesis of the case studied, by appealing to the physical notion of the probability as a fraction (LaPlace 1812) and frequency (von Mises 1919) and by analyzing the content of possible formulations of evidence, assumptions of models and meaning of the ecological hypothesis. Bayesianism has analogical problem concerning arbitrariness in the choice of the *prior* probability (Bernard 1996), but the general discussion whether Bayesians are superior to frequentists because of the stopping rule problem, or any other difficulty, is beyond the scope of this paper.

The structure of the paper is as follows. First, I present the stopping rule problem. Then I show the difference in the evidence taken into account in different frequentist sampling procedures presented in the problem. Next, I discuss the inapplicability of the Carnapian (1947; 1950) principle of total evidence in deciding which formulation of evidence associated with a given sampling procedure and statistical model is epistemically more appropriate for testing the hypothesis in question. Then I propose a double-perspective (evidence and model) frequentist solution based on the choice of a formulation of evidence that better corresponds to the investigated ecological hypothesis, as well as on choice of a model that embraces less unrealistic ontological assumptions. Finally, I discuss two perspectives of stopping rule dependence.

## 2. An outline of the problem: different outcomes of significance test

The ecological hypothesis in question (H) states that the proportion of males in the population of pouch young is 0.5 (number of males and females is equal), and the reasonable alternative hypothesis states that the proportion is less than 0.5 (females prevail).

The researcher surveyed 12 females each with an offspring in its pouch – three of the offspring were males and nine were females. The data could be obtained in at least two ways: the researcher could sample until the 12th individual is recorded ($S_1$), or until the 3rd male is recorded ($S_2$). Regardless of the sampling strategy the data seem to be equivalent and the two alternative inferences are as follows (McCarthy, 31-33).

Sampling in the case $S_1$ is modeled by the binomial distribution that represents probabilities of collecting *n* number of males until a number of trials in a sample reaches a fixed value; the sum of probability of the observed datum $Pr_1$(males = 3) and more extreme data (in this case $Pr_1$ of having 2, 1, and 0 males in the sample) equals 0.073, thus the observed male ratio, given 0.05 cut-off error rate, is not significantly less than the hypothesized ratio 0.5.

Sampling in the case $S_2$ is modeled by the negative binomial distribution that represents probabilities of collecting *n* number of females until a number of males in a sample reaches a fixed value. The *p*-value in this case is a sum of the $Pr_2$(females = 9) and less probable outcomes: having 10 female records, 11, 12, and so on. The *p*-value equals 0.033 in this case, so with the conventional 0.05 error rate it is significantly lower than the expected value of 0.5 and the conclusion is to reject the null hypothesis.

As a result, the two different sampling strategies lead to different conclusions about the null hypothesis based on the same set of raw data. What Lindley and Phillips (1976, 114) stresses is that this is due to the fact that "(…) the significance (in a technical sense) to be associated with the hypothesis of equal chances depends heavily on what other results could have been achieved". In practice it means that "The usual statistical significance test requires the sample space, or alternatively, the stopping rule to be specified". In consequence, a frequentist methodology may be accused of yielding two epistemically equally valid outcomes that are contradictory. Such an objection would appeal to arbitrariness of the choice between the two stopping rules or models. As I argue below, this choice is not arbitrary. I provide, with respect to the ecological hypothesis in question, two simple methodological solutions for discriminating the better one of the two – at first sight equivalent – sampling strategies or models.

## Response to the problem. The issue of data (evidence) equivalence

First, I will present an explication of models and stopping rules by means of their evidential correlates. It is obvious that in practice the raw data collected during scientific investigation is in many cases richer in information than the data that we abstract from the original collection for the purposes of transformation and inference: a simple illustration is to neglect extreme (unreliable) data, or irrelevant data (standard elements of measurement data or additional data recognized for other reasons as potentially valuable). Irrelevant data in the discussed case would include, for instance, the exact place, date and time of the particular report or the relevant ordering of males and females records.

In the case of sampling framework $S_1$, connected to a binomial distribution, evidence that was taken into account ($E_1$) can be expressed by means of a proposition: "Exactly three males and nine females were recorded in the sample until (and including) the twelfth trial was recorded in the sample".

In the sampling framework $S_2$ bound up to negative binomial distribution the evidence considered ($E_2$) can be expressed in the proposition: "Exactly two males had been recorded in the sample until (and including) the eleventh trial and the twelfth trial recorded in the sample was male".

It can be seen that $E_2$ embodies some information about the order of the outcomes, while $E_1$ does not. McCarthy claims that "Regardless of the sampling strategy, the data are equivalent" (2007, 31) and "(…) the only difference is the stopping rule for the researcher's sampling strategy". But this, as I argue, is not the case. After unpacking the evidence, the latter shows that $E_2$ implies $E_1$, but not vice versa. For these two cases different information is taken into account. Even if the raw data at hand might be equivalent in both cases, the data taken into consideration are different: $E_2$ is more informative than $E_1$ by saying something more about the order of the outcomes, and $E_2$ implies $E_1$.

One might still think that McCarthy – by saying 'equivalent' – meant that in both cases the information sublimed in a form of E was taken from the same raw data. But imagine a case in which for some reason one evidence from the data ($E_a$) was to be an information about the place and time, and another ($E_b$) – an information about sex. Saying that in these cases $E_a$ and $E_b$ are equivalent would be bizarre. McCarthy could also have in mind that 'equivalent' means 'taken from the same raw data and applied to the same H'. Yet I will argue further that the ecological hypotheses tested by means of $E_1$ and $E_2$ are not one and the same (see Section 3).

## 3. Choice based on evidence comparison

Imagine the situation in which a frequentist ecologist who wants to make use of a data gathered by someone else does not know whether it was collected with respect to a fixed number of trials or a fixed number of successes and the data allows it to be interpreted in both ways.

Among the two alternative bodies of evidence[1], purportedly appropriate to be used for inferences about H, $E_2$ is informatively richer than (and entails) $E_1$. For this reason a researcher might intuitively appeal to the advantage of using all the available information and choose more informative evidence $E_2$. Especially that, recalling the Monty Hall problem (Gorroochurn 2012, 264-270), the situation in choosing between $E_1$ and $E_2$ at first sight seems to be analogous to the situation of choosing the evidence in Monty Hall's game. The fact that Monty opens a particular door entails the fact that there is no car behind that door, but not vice versa. In accordance to the 'principle of total evidence' the first (richer) fact should be then taken into account as evidence used for proper evaluation of probability of a hypothesis concerning the presence of a car in one of the remaining doors (Bovens and Ferreira 2010, 480). However, the requirement, that the researcher should take as evidence "(…) his total knowledge of the results of his observation" (Carnap 1947, 138-139) is not applicable in the analyzed case and cannot indicate a proper E. The requirement as stated by Carnap, is meant to be a "(…) basis for determining the degree of confirmation" (Carnap 1950, 211), whereas in NHST the measured value is not the degree of confirmation of H.

The measured value is the probability of the evidence with respect to the hypothesis, which does not even constitute a conditional probability, as there is no such thing as *prior* probability of the hypothesis in the frequentist sense. But what

[1] There could be plenty of other ways of formulating evidence depending on how much raw data there is, but for the sake of simplicity I take into account the two that stands for the two stopping rules presented by Lindley & Phillips (1976).

if one formulates the principle in a more general intuitive form – as a postulate to choose the richest evidence among those which are available and plausible for inference about a hypothesis? The principle in such a form will not help to make the choice in question, because $E_1$ and $E_2$ stand for different hypotheses. In the simple Monty Hall case, shifting the evidence does not impose a change of the hypothesis. However, the case of $E_1$ and $E_2$ is more complicated, due to $E_1$ and $E_2$ are considered through more complex probability models. The factual hypothesis that is under scrutiny in each case is a statistical hypothesis in a form of a statement about the sum of certain values of a particular probability function – in case $E_1$ and $E_2$ these are different functions. Hence, there is no way of considering $E_1$ and $E_2$ in relation to one and the same statistical hypothesis[2]. Still, one might argue, that what makes both $E_1$ and $E_2$ refer to the same physical (f.e. ecological) hypothesis is the same probability parameter $p = 0.5$ in both models which is the real thing reflecting the same ecological hypothesis and the difference between $E_1$ and $E_2$ is merely difference in which the hypothesized state of affairs may appear to the researcher. I will reject this objection in the further part of the argument (Section 3).

What could be then an evidence comparison-based criterion for making the right choice? Lindley & Phillips (1976) state that in their example information about the order of the failures and successes has no relevance, as "(…) technically, 9 and 3 are jointly sufficient" (p. 113). As opposed to this, Carnap (1947, 139-141) argues that the information about the (temporal) order in the evidence should not be neglected since it may reveal a periodical pattern in the evidence occurrence that might play a role in the prediction of future events. However, this might be discussed as eventually valid for single case hypotheses (like: what will be the next outcome). The ecological hypothesis that a researcher is concerned with is meant to be either a generalized statement about a fraction in population, or a kind of a law of nature which states something about the long run or large scale tendency. For this reason, the information about order is irrelevant in considerations of either type of hypothesis.

Regarding above, if the order were to be irrelevant, the sufficiency criterion could be used for making a choice between $E_1$ and $E_2$ – but this would be a pragmatic rather than an epistemic norm. Yet, irrelevance enables a researcher to epistemically discriminate between $E_1$ and $E_2$ if additional information that should be irrelevant makes a difference in the outcomes. Additional information in $E_2$, which is an information about the order, is the one which is responsible for causing the difference in the outcome, however – for this particular hypothesis – it should be

[2] When the evidence would relate to one fixed hypothesis, the principle in its intuitive form might apply. This could be the case, for instance, when one takes into consideration the binomial model and evidence comprised of six records against evidence comprised of twelve records. Yet, in practice it happens that such a requirement of total evidence is deliberately and legitimately violated when a researcher conceptually finds a greater number of trials to be of too little epistemic utility in comparison with the cost of collecting it.

irrelevant. This is the reason for dropping $E_2$ as inadequate for testing the given hypothesis.

Now, consider another possible approach for discerning between $E_1$ and $E_2$: maybe E that is more strongly (logically) connected to H should be preferred? Due to the fact, that in analyzed frequentist statistical framework there is no division on P(H) and P(H|E), nor P(E) and P(E|H), the eventual decision would have to rely exclusively on the comparison of $P(E_1)$ with $P(E_2)$. One might stipulate that the more likely the event is to occur in the assumed world of a true null hypothesis, the better it represents the hypothesis and thus there can be more confidence of the eventual rejection of the hypothesis. Considering the character of a relation of $E_1$ to $E_2$ it can be said that if $E_2$ implies $E_1$, then $P(E_2)$ will be lower or equal to $P(E_1)$ in any model that represents H[3]. On that basis $E_1$ would be preferable as potentially more probable. However, this criterion seems to be inappropriate, since one can think of another $E_0$ taken from the raw data that will be more probable than $E_{1,}$ but intuitively less adequate than $E_1$. $E_0$ could be for example "there was at least one male in the population after twelfth trial". This approach is incorrect, but elucidates the two ways in which E can be wrongly constructed from the raw data in the case of testing koalas' sex ratio: E can take the form that expresses too little information, like in the case of $E_0$, or such that expresses too much information, like in the case of $E_2$.

For the above reasons, I state to render $E_1$ the best 'fitting' for the given hypothesis, based on the criterion that I will refer to as "best accuracy of evidence": evidence should be saying about as many features of the data that are stated by the hypothesis about a hypothesized state of affairs, as possible; and about as little features of the data that do not occur in the hypothesis statement, as possible. The hypothesis says about the actual fraction or long run frequency of pouch young males in a koala population. In accordance to this, $E_0$ is not the best choice: it does say something about the ratio, namely that the male fraction after the 12th record is greater than 1/12, but the precision is too low comparing the hypothesis' precision level. The $E_2$ on the other hand is too precise by providing some additional information about the order of outcomes in a sample after the 12'th record. Evidence $E_1$ is the most accurate for the hypothesis in question.

Note that it may resemble the calibration process, in which a researcher having the evidence calibrates the model by choosing, from a bundle of models accessible within a more general theory, the one that best fits the data at hand. The difference is that the presented criterion would be a sort of reversed calibration: a researcher starts from the ontological level by analyzing the content of the ecological hypothesis, to construct the best-fitting formulation of evidence from the possibilities embedded in the raw set of data.

---

[3] In fact, when both models are compared, straight inequality holds: substituting the same number of successes and failures in both models P(E) will be bigger in the case of binomial; $P(E_1)$ and $P(E_2)$ in two models could be equal only in the case where number of failures is infinite, which is prohibited in the binomial model.

At this point it might become more urgent to reflect more on what the ecological hypothesis really states and how it is represented by two statistical models. I will answer this question in the following paragraph to provide a second alternative criterion based on a different aspect – choosing the model that corresponds better to the hypothesized state of affairs.

## 4. Choice based on model comparison

It is obvious that models for $E_1$ and $E_2$ are two different mathematical functions so in this sense statistical hypotheses constructed with the use of them are different. Nevertheless both models might be seen as equivalent in a sense of representing certain aspects of phenomena that stands for hypothesized state of affairs. Especially when they both have the same *p*.

However, the models are not equivalent. The essential differences (between the models) can first be noticed from evidential level, without referring to the models' mathematical structure. Imagine that a researcher has not started collecting the data and states a certain ecological point hypothesis about a ratio. In such a setup, one possible situation is that H is true (the ratio in a population is *p*), and the evidence is complete. In such a situation E should be a consequence of H. Assume, in the case of koalas, a possibility in which: the hypothesis was that the ratio was 0.5 and it was true, the population was 12 and a researcher scrutinized all pouch youngsters (unaware of her evidence's completeness) ending up with data collection comprised of 6 males (successes) and 6 females. In a condition of being complete this evidence should be a consequence of the true hypothesis. Accordingly, the evidence representing $S_1$, in its form which quite explicitly expresses a stopping rule, would be that 6 males were recorded until the twelfth trial. Now, if $ratio_H = ratio_E = 0.5$ and $n_s = N_P = 12$ then it is always true that until the 12-th trial a researcher has six males recorded. But it is not always true – given the same premises – that, in case of $S_2$, until the 6-th male a researcher will have 6 females in the record. As a result, stopping rule $S_2$ could be seen as yielding logical inconsistency in this quite special case of possible situation, yet the problem is not inconsistency, but the fact, that within the design of $S_2$ a presented situation is impossible. $S_2$ is associated with acceptance of a concept of E that can be comprised of infinite number of trials. Condition for E being possibly infinite is to assure population to be infinite – concept of always potentially infinite E cannot be introduced when the population is possibly finite. Now, if the population is to be infinite, the evidence's completeness condition for the presented situation seems to be unclear or impossible.

Now, it is explicit that the essential difference lies in the different assumptions about the size of the population: in case of $S_1$ it can be finite, in case of $S_2$ it cannot. Looking straightaway at the models this fundamental difference is seen directly:

sample space in the binomial distribution is finite, whilst in the negative binomial distribution it is infinite. A number of elementary events in sample space cannot be infinite if the space is to represent outcomes of drawing without replacement from finite population. Thus, the model for $E_2$ with infinite reference space assumes infinite population. On the other hand, finite $E_1$ can be drawn from finite population as well as from infinite population, so the binomial model does not assume finiteness nor infinity of the population.

As a consequence, a choice of different population concept determines a choice of a model. Whether a researcher takes the population to be finite or infinite is not arbitrary. Assumption of infinity is less plausible. If the population is considered as a current set – it is obviously finite. If one considers it temporarily – infinity might be conceivable, but in this case finiteness is still more realistic assumption. Not least because of evolutionary impermanence of the species, or cosmological prediction of the end of the universe in its present form. For this reason the binomial distribution of evidences of the form of $E_1$ is epistemically better, because it does not grant an assumption that is unrealistic in the light of our scientific knowledge.

## 5. Conclusions

Recollecting the conclusions from the above I state that:

> the data, the model, and H that correspond with sampling strategy $S_1$ is different to, and not equivalent to those that correspond with $S_2$.

There are two frequentist criteria for discerning the better approach from the two presented:

(1) evidence-perspective criterion: choose a formulation of evidence which precision level better corresponds to the hypothesis' precision level

(2) model-perspective criterion: choose a model that bears less unrealistic assumptions.

According to both criteria the model and the evidence associated with stopping rule $S_1$ are clearly more adequate for testing the hypothesis in question than those related to $S_2$.

What model shall a researcher choose when she knows that the data were collected with sampling strategy $S_2$? I state that she is allowed, or even encouraged, to suspend the additional body of evidence related to $S_2$ (about the order), and select from the data the evidence $E_1$ representing the more relevant binomial model.

Back to the question, whether the outcome depends on the stopping rule, two answers comes to mind. The sampling procedure affects the properties of collected data (like for example a number of trials). In this sense it is plausible to say that sampling results depend on the stopping rule. One might state that a researcher cannot ignore the sampling strategy even if she explained it in the evidential form, and that she is not allowed to choose the body of evidence $E_1$ from the data that were obtained by means of $S_2$. In the case of such an approach the outcome depends on sampling strategy. But this is unproblematic for a frequentist ecologist, who will avoid arbitrariness in choosing the sampling procedure. With reference to the presented criteria, $S_2$ is clearly an inadequate method for the discussed H[4].

On the other hand, if one agrees that after explaining evidential forms that represent $S_1$ and $S_2$ a researcher can suspend the information about the stopping rule, then McCarthy's objection is defeated by the following possibility. Two researchers use two different sampling rules and come up with exactly the same set of raw data. Then they both, driven by presented epistemic criteria, take from the data the same body of evidence $E_1$ and use the same model. In consequence, they end up with the same outcome of testing, regardless of the difference in a rule that guided them to collect their data. From such a perspective it can be said, that the outcome does not depend on the stopping rule.

For the reason above, I reject the McCarthy's objection: the data in both hypothetical examples are not equivalent; an outcome does not dependent on sampling procedure, but on what H is being tested; there is no problematic arbitrariness – for testing the ecological hypothesis in question with the use of classical framework, data $E_2$ together with $H_2$ are clearly inadequate.

In scientific practice epistemic virtues are not necessarily a decisive element. Choosing the stopping rule $S_2$ could turn out to be more pragmatically pertinent. As I tried to argue, despite such a way of sampling, a researcher can be justified in making use of a formulation of evidence different from the one imposed by the way the data was collected. As the analyzed example was a simple thought experiment, provided solutions do not aspire to have much value of practical application. The discourse I presented is rather intended to be a theoretical argument in defense of the frequentist NHST as subjected to problem defined by Lindley & Phillips. Yet, I did not mean to provide the argument for superiority of this methodology over other alternatives, in particular Bayesian. I aimed to show that a frequentist researcher can have a solution for Lindley & Phillips' stopping rule problem within her methodological scaffolding. For classical hypothesis testing the crucial task is to design experiments and sampling frameworks that will be appropriate with respect to research goals. It can be contested whether it is an advantage or a draw-

[4] If the number of trials in the discussed data set was to be large, the difference would be of less importance.

back, but no doubt there is an analogy between frequentists and Bayesians: "To a Bayesian, there are no bad models, just bad beliefs" (Dennis 1996).

## References


Bovens L., Ferreira J.L., 2010, *Monty Hall drives a wedge between Judy Benjamin and the Sleeping Beauty: a reply to Bovens*, "Analysis" 70: 473-481.

Carnap R., 1947, *On the application of inductive logic*, "Philosophy and Phenomenological Research" 8: 133-148.

Carnap R., 1950, *Logical Foundations of Probability*, Chicago: University of Chicago Press.

Dennis B., 1996, *Discussion: Should ecologists become Bayesians?*, "Ecological Applications" 6: 1095-1103.

Efron B., 2005, *Bayesians, Frequentists, and Scientists*, "Journal of the American Statistical Association" 469: 1-5.

Gorroochurn P., 2012, *Classic Problems of Probability*, Hoboken: Wiley.

Kéry M., Schaub M., 2012, *Bayesian Population Analysis using WinBUGS: A hierarchical perspective*, London: Academic Press.

Laplace P. S., 1812, *Théorie analytique des probabilities*, Paris: Courcier.

Lindley D. V., 1957, *A statistical Paradox*, "Biometrika": 44: 187-192.

Lindley D. V., Phillips L. D., 1976, *Inference for a Bernoulli Process*, "The American Statistician" 30: 112-119.

Link W.A., Barker R.J., 2010, *Bayesian Inference with ecological applications*. London: Academic Press.

Mayo D., 1996, *Error and the Growth of Experimental Knowledge*, Chicago: Chicago University Press.

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

Stephens P.A., Buskirk S.W., del Rio C.M. 2006, *Inference in ecology and evolution*, "Trends in Ecology and Evolution" 22: 192-197.

Strevens M., 2013, *Tychomancy: Inferring Probability from Causal Structure*, Cambridge, MA: Harvard University Press.

von Mises R., 1919, *Grundlagen der Wahrscheinlichkeitsrechnung*, "Mathematische Zeitschrift" 5: 52-99.

Williamson J., 2013, *Why Frequentists and Bayesians Need Each Other*, Erkenntnis 78: 293-318.