Philosophy of Science (2026), **93**, 607–630
doi:10.1017/psa.2026.10197

PSA PHILOSOPHY OF SCIENCE ASSOCIATION
PHILOSOPHY OF SCIENCE

ARTICLE

Carnapian Inductive Logic for Exponential Smoothing

Simon M. Huttegger

Department of Logic and Philosophy of Science, University of California Irvine, Irvine, USA
Email: shuttegg@uci.edu

(Received 22 September 2025; revised 03 February 2026; accepted 04 February 2026; first published online 13 April 2026)

Abstract

This paper explores the inductive logic associated with exponential smoothing, the most widely used predictive rule that manifests the idea that more recent observations have a stronger influence on predictive probabilities than more remote ones. The main result shows that exponential smoothing can be derived from a set of plausible qualitative invariance assumptions about conditional probabilities. I discuss various aspects of the resulting inductive logic, including its connections to exchangeable processes, to Bayesian predictive inference and kernel methods in machine learning, as well as the philosophy of probabilistic invariance conditions and symmetries.

1. Introduction

Among the problems Carnap left to the future development of inductive logic is what Kuipers (1988) called "analogy by proximity." Kuipers distinguished analogy by proximity from analogy by similarity. The latter is about similarities between observational outcomes and how they influence predictive probabilities. Analogy by proximity refers to the temporal order of observations. According to Kuipers, there is analogy by proximity *in the future* and analogy by proximity *in the past*. The former gives rise to inductive logic for Markov chains, which is well understood (Skyrms 1991; Zabell 1995). Analogy by proximity in the past, on the other hand, has not received the same level of attention in the post-Carnapian literature on inductive logic.

The basic idea of analogy by proximity in the past is "that an outcome profits more from a recent self-occurrence than from a self-occurrence longer ago" (Kuipers 1988, 304). Kuipers, who traced this idea back to Carnap's *Logical Foundations of Probability* (Carnap 1950, 62–65), suggested an inductive rule that hewed closely to Carnapian inductive systems, but he did not offer principles from which such a rule can be derived. The main obstacle is, I believe, the fact that inductive rules that exhibit a recency effect are incompatible with the fundamental principle of Carnapian inductive logic: order invariance. That is, the order in which observations are made has no influence on inductive probabilities. If an inductive rule places more weight on

© The Author(s), 2026. Published by Cambridge University Press on behalf of Philosophy of Science Association.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

recent than on remote observations, order invariance clearly fails. In the absence of order invariance, it's not clear how to derive an inductive rule.

The main goal of this paper is to provide a foundation for analogy by proximity. Instead of modifying Carnapian inductive rules, I will consider *exponential smoothing*. Exponential smoothing is the most common predictive rule in scientific practice that exhibits the kind of recency effect described by Kuipers. It does so in a simple and elegant way by exponentially discounting past observations. Together with its variants it is widely used in forecasting, statistics, and machine learning. After reviewing Carnapian inductive logic in section 2, I will introduce exponential smoothing and its applications in section 3. In section 4, I will show how it can be derived from invariance principles in the spirit of Carnap's program. The relevant principles are, of course, not the same as the ones underlying Carnapian systems. Order invariance is replaced by the concept of *conditionally independent distributions* and by invariance assumptions about *relevance quotients*. Conditionally independent distributions were studied by Berti et al. (2004, 2021), who develop a general approach to statistical inference through predictive probabilities instead of statistical models. Relevance quotients were used by Costantini (1979) in the context of Carnapian inductive logic. They provide a flexible way to express invariance assumptions about observational outcomes. In section 5 I discuss some aspects of the inductive logic for exponential smoothing. Two further sections discuss additional aspects in more depth: its connections to kernel methods in machine learning (section 6) and to the concept of exchangeability in Bayesian statistics (section 7). I conclude in section 8 with some general remarks about the role of invariance principles in Carnapian and Bayesian inductive logic.

2. Carnapian inductive logic

The rules of inference studied by Carnap go back to Laplace. They are known as "rules of succession." A rule of succession takes as input a finite sequence of past observations. A sequence of observations is given by a sequence of random variables $\mathbf{X}_n = (X_1, \ldots, X_n)$. Each X_i takes a value in a finite set of observational outcomes $\{1, \ldots, t\}$. A rule of succession then outputs predictive probabilities for the next observation, X_{n+1}, given an assignment of values for $X_1, \ldots, X_n$:

$$\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_n) = \frac{n_i + \alpha_i}{n + \sum_j \alpha_j}. \tag{1}$$

The counts n_i represent the number of times outcome i appears among $X_1, \ldots, X_n$. The weight α_j determines the initial probability assigned to outcome j. In the Carnapian setting, the weights are positive if the sequence of random variables is infinite; otherwise, they can be negative. A Carnapian rule thus assigns probabilities to events about the immediate future given past observations.

The logical foundations of the predictive probabilities in (1) are fully understood. An especially lucid treatment is given in Zabell (1982), who derives (1) from three assumptions that are based on the ideas of Johnson (1932): *exchangeability*, *Johnson's sufficientness postulate*, and a *regularity* assumption. Regularity only requires that each finite sequence of outcomes has a positive probability. Exchangeability and Johnson's

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

sufficientness postulate are more substantial: they express *order invariance* for unconditional and conditional probabilities.

Exchangeability says that

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_{\tau(1)}, \ldots, X_n = x_{\tau(n)})$$

for each n, all possible sequences of outcomes $x_1, \ldots, x_n$, and for any permutation τ of $\{1, \ldots, n\}$. In words, a probability assignment is exchangeable if it is *invariant* under any reordering of the outcomes. A consequence of exchangeability is that the counts $n_1, \ldots, n_t$ are a *sufficient statistic*: knowing how often each outcome is observed among $X_1, \ldots, X_n$ completely determines the probability distribution $\mathbb{P}$ since each sequence with the same counts $n_1, \ldots, n_t$ must have the same probability.

Johnson's sufficientness postulate applies this idea to conditional probabilities. It says that observing an outcome i only depends on i, n_i, and the total sample size n:

$$\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_n) = f_i(n_i, n).$$

When combined with exchangeability and regularity, Johnson's sufficientness postulate implies that there exist constants α_j such that (1) holds.

The connections between Carnapian inductive systems and Bayesian inference are also well understood. The basic insight is the following. According to de Finetti's theorem, any exchangeable infinite sequence of random variables $X_1, X_2, \ldots$[1] is a mixture of multinomial distributions:

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \int_{\Delta_{t-1}} p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t} \, d\mu(p_1, \ldots, p_{t-1}). \tag{2}$$

Here, the numbers n_i are again the counts of observational outcomes. The integral ranges over the $(t-1)$-dimensional simplex $\Delta_{t-1} = \{(p_1, p_2, \ldots, p_t) : p_j \geq 0, \sum_j p_j = 1\}$. (The simplex has dimension $t-1$ since $p_t = 1 - \sum_{j=1}^{t-1} p_j$). The simplex Δ_{t-1} can be thought of as the set of all possible chances of outcomes: p_j is the chance that outcome j occurs at any stage of the observational process. The integrand in (2) articulates this by aggregating chances into an independent and identically distributed setup in which each of the n_j occurrences of j happens with chance p_j. However, the chances are unknown. That's why (2) integrates over all chances relative to a prior μ that represents the initial uncertainty about chances.

The prior μ can be used to forge a connection between de Finetti's theorem and Carnapian inductive logic. If μ takes on a specific functional form (when it is a so-called *Dirichlet prior*), then the predictive probabilities that result from the representation (2) *coincide* with the inductive system in (1) (with the parameters α_j determining the Dirichlet distribution).

There is no need here to go into the mathematical specifics, which are explained elsewhere (e.g., Skyrms 1996; Romeijn 2011). Two issues are more important for us. The first concerns the philosophical lessons to be drawn from the connection between Carnapian inductive logic and Bayesian statistics. I will say a bit more about this in the last section of this paper. For now, let me just make a brief remark about the bridge that has been built by putting inductive logic in contact with de Finetti's theorem. One way to read de Finetti's theorem is as a *pragmatic reduction* of chances to degrees of

[1] An infinite sequence is exchangeable if any finite initial segment of the sequence is exchangeable.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

belief (Skyrms 1984). According to this point of view, chances are nothing but stable components of an agent's exchangeable credences. What is missing in this picture is an explanation of the prior μ. The prior μ ranges over chances. But how can an agent elicit such a prior if chances are only something buried deep in the structure of her credences? The Carnapian approach, being purely predictive, circumvents this question. Exchangeability and Johnson's sufficientness postulate (as well as regularity) only speak about sequences of observations, and they do so in a way that fully determines a specific prior μ. These postulates can thus be viewed as an "observational" characterization of a specific class of priors (the Dirichlet priors) that specifies an unobservable theoretical structure.

The second issue has to do with the limitations of exchangeability. Exchangeability, which expresses order invariance, is clearly not always a good assumption. Carnapian inductive systems and de Finetti's theorem have been extended in many different ways.[2] What most of these generalizations have in common is a commitment to order invariance at some level. For instance, Markov chains involve chances of transitioning from one observational state to another (instead of chances of outcomes *simpliciter*). Carnapian inductive logic for Markov chains requires order invariance within certain subsequences of observations, namely those that follow a particular outcome. The relevant concept, *Markov exchangeability*, is weaker than exchangeability, but it does assume a certain kind of order invariance.

However, nonstationary environments don't exhibit the kind of stability that underlies exchangeability and its generalizations. Instead of stable patterns that persist throughout the observational process, such environments are characterized by transient short-term trends. If we find ourselves in an epistemic situation of this kind, order invariance may not be a plausible assumption.

3. Exponential smoothing

In the forecasting literature, exponential smoothing is one of the most prominent rules for nonstationary epistemic settings. Exponential smoothing goes back to the work of Brown (1959) and Holt (1957) in operations research. Using the notation of the last section, the rule is given by the following update procedure:

$$\mathbb{P}(X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n) = \beta^n \mathbb{P}(\boldsymbol{X}_1 = i) + (1 - \beta) \sum_{m=1}^{n} \beta^{n-m} \chi_i(x_m). \quad (3)$$

Here, the parameter β is in [0,1], and χ_i is the indicator of outcome type i (i.e., $\chi_i(x_m) = 1$ if $x_m = i$ and $\chi_i(x_m) = 0$ otherwise). According to (3), the initial probability of i, $\mathbb{P}(X_1 = i)$, influences its predictive probability but is discounted, unless $\beta = 1$, in which case there is no learning. The remaining predictive weight, $1 - \beta$, comes from observations of i. More recent observations of i are weighed more heavily, the weights being regulated exponentially by the parameter β.

In contrast to Carnapian inductive logic and most of its variants, exponential smoothing ages out older observations. As a result, it clearly fails to be order invariant. The time step at which an observation of type i was made matters a lot in

[2] See, inter alia, Diaconis and Freedman (1980), Skyrms (1991), Zabell (1992), Paris and Vencovská (2015), or Huttegger (2019).

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

determining the predictive probability of observing another i. Thus, exponential smoothing is a qualitatively plausible predictive system for settings in which trends and patterns do not need to persist for long periods of time.

The question I wish to answer goes beyond the broad features of exponential smoothing and concerns its specific mathematical form. Aging out of past observation can be done in many different ways. Is there something special or principled about the exponential discounting of the past as expressed in (3)? In the next section, we will see that inductive logic can shed some light on this question. It does so by identifying the types of *invariance* that underlie exponential smoothing. My approach turns out to be different from the usual approach, which is based on ideas about order invariance. While my analysis does not rely on order invariance, it is inspired by an alternative axiomatic derivation of Carnapian inductive logic that was put forward by Costantini (1979). Costantini's paper addresses a technical wrinkle in the standard derivation of Carnapian inductive logic, which relies on Johnson's sufficientness postulate. It is well known that the postulate is empty when there are only two types of observations, for then the number of times one type was observed and the total number of observations determine the number of times the other type was observed. In his paper, Costantini replaced Johnson's sufficientness postulate with an invariance requirement for *relevance quotients*. In the context of Carnapian inductive logic, relevance quotients are ratios of the form

$$\frac{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)}{\mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})}.$$

This ratio quantifies the relevance an observation j has on the predictive probability of i, given the history of observations $\mathbf{X}_{n-1}$. Costantini's invariance condition on relevance quotients states that for any $i \neq j$ and $k \neq l$,

$$\frac{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)}{\mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})} = \frac{\mathbb{P}(X_{n+1} = k \mid \mathbf{X}_{n-1}, X_n = l)}{\mathbb{P}(X_n = k \mid \mathbf{X}_{n-1})}. \tag{4}$$

The relevance quotient on the left side of this equation has the same number of times i has been observed, n_i, in the numerator and the denominator. The right side likewise fixes n_k. The invariance condition says that the effect an observation of type $j \neq i$ has on the predictive probability of i is the same as the effect an observation of type $l \neq k$ has on the predictive probability for k, regardless of i, j and k, l (as long as each pair has distinct outcomes). Costantini's invariance of relevance quotients entails (1) in the presence of exchangeability and regularity, and it does so whenever there are at least two types of observational outcomes.

In the next section, we will see that invariance conditions for relevance quotients are also key to understanding the principles that underlie exponential smoothing. The other key is a generalization of exchangeability that applies to settings in which order invariance fails.

Before spelling this out in more detail, let me comment briefly on the significance of exponential smoothing. In data science and machine learning, rules of succession (i.e., Carnapian inductive logic) fall under the rubric of additive smoothing. Smoothing methods, such as additive or exponential smoothing, are often used to filter out noise from data in order to discover underlying trends or patterns. This is related to curve fitting, but is mathematically the same as predicting unobserved data points given a set of observed data, which is the principal domain of inductive logic. Exponential smoothing has

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

applications in a wide array of fields.[3] Especially noteworthy are recent applications of exponential smoothing in machine learning, such as hybrids of exponential smoothing and recurrent neural networks (Smyl 2020; Smyl et al. 2023). In these settings, exponential smoothing is used to identify the main components of highly complex data sets. Thus, aside from the intrinsic value of understanding the inductive logic of exponential smoothing, the principles studied in the next section may also enhance our understanding of certain machine learning algorithms that utilize exponential smoothing.

4. Inductive logic for exponential smoothing

It is well known that exponential smoothing (3) can be equivalently expressed as a convex combination of previous predictive probabilities and current observations. We prove this result for the sake of completeness.

Lemma 1. Exponential smoothing (3) is equivalent to

$$\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = x_n) = \beta \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1}) + (1-\beta)\chi_i(x_n). \tag{5}$$

Proof. Suppose (3) holds for all n. Then

$$\beta \mathbb{P}(X_n = i \mid X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) + (1-\beta)\chi_i(x_n)$$

is equal to

$$\beta\Big[\beta^{n-1}\mathbb{P}(X_1 = i) + (1-\beta)\sum_{m=1}^{n-1}\beta^{n-1-m}\chi_i(x_m)\Big] + (1-\beta)\chi_i(x_n),$$

which simplifies to

$$\beta^n \mathbb{P}(X_1 = i) + (1-\beta)\sum_{m=1}^{n}\beta^{n-m}\chi_i(x_m).$$

This yields one direction. Now suppose that (5) holds. Then (3) follows by simply applying the recursion to each n. QED

The convex combination (5) says that the old predictive probability of i receives weight β, with the remaining weight being attached to a new observation of i if it occurs; otherwise, the new weight goes to a different category.

I am now going to introduce a set of postulates that together suffice to derive exponential smoothing as a predictive system. In what follows, we consider an infinite sequence of random variables $X_1, X_2, \ldots$ taking on values in a finite set that has at least two elements. The finite case is a bit more complex and will be discussed in the next section.

As mentioned above, *regularity* is a basic assumption in Carnapian inductive logic. We also assume it for exponential smoothing, since it guarantees that all conditional probabilities are elementary. Formally, regularity says that

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) > 0 \tag{6}$$

for all n and all possible combinations $x_1, \ldots, x_n$ of observational outcomes.

[3] For a detailed survey, see Gardner (2006).

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

The second assumption is new and has more substance. It was introduced in Berti et al. (2004), who refer to it as distributions being *conditionally independent*. The sequence $X_1, X_2, \ldots$ is conditionally independently distributed if

$$\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_n) = \mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n) \tag{7}$$

for all n and i, and for all $k \geq 2$. Conditional independence in distribution is an invariance condition for predictive probabilities. Conditional on the evidence at time n, any future observation of an outcome has the same probability regardless of when it occurs.

Exchangeable sequences of random variables are conditionally independently distributed, but a conditionally independently distributed sequence of random variables can fail to be exchangeable (Berti et al. 2004, 2021). As it turns out, conditional independence in distribution is the probabilistic invariance required for an inductive logic of exponential smoothing, where it plays the same role as exchangeability does in the standard setting. However, its significance goes beyond this particular context. It is a cornerstone of predictive approaches in statistics (Fong et al. 2023; Falck et al. 2024).

The remaining postulates make use of relevance quotients. They are slightly different from Costantini's relevance quotients (4) in that they refer to X_{n+1} in the numerator and the denominator. In the presence of conditional independence, there is no difference between the two ways of expressing relevance quotients, because (7) implies that $\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}) = \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})$ and $\mathbb{P}(X_{n+1} = i) = \mathbb{P}(X_n = i)$. Since exchangeability implies conditional independence (7), X_{n+1} could be used for Costantini's relevance quotients (4) in the same way. I prefer using just X_{n+1} because it fixes a single hypothesis about future observations that is evaluated in the context of inductive relevance.

I will refer to the first postulate about relevance quotients as *invariance of inductive relevance*. It comes in two parts. The first part states that

$$\frac{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)}{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})} = \frac{\mathbb{P}(X_{n+1} = i \mid X_1 = j)}{\mathbb{P}(X_{n+1} = i)} \tag{8}$$

for all n, i, and $j \neq i$: the relevance j has for an outcome i at time $n + 1$ is the same as the initial relevance j has on that outcome regardless of the intermediate history. The same is assumed for the relevance an outcome has on all other categories taken together. That is, for all n and for all outcomes i,

$$\frac{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1}, X_n = i)}{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1})} = \frac{\mathbb{P}(X_{n+1} \neq i \mid X_1 = i)}{\mathbb{P}(X_{n+1} \neq i)}. \tag{9}$$

Clearly, (9) does not just follow from (8), or vice versa. The denominators and numerators in (9) are sums, but the ratio of these sums is generally not equal to the sum of ratios like (8). Equation (9) is a somewhat roundabout way of expressing the inductive relevance that a category has on itself.

I call the final postulate *invariance of initial relevance*. It also has two parts. First,

$$\frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_2 = i)} = \frac{\mathbb{P}(X_2 = k \mid X_1 = l)}{\mathbb{P}(X_2 = k)} \tag{10}$$

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

for all i, j, k, l such that $i \neq j$ and $k \neq l$. Second,

$$\frac{\mathbb{P}(X_2 \neq i \mid X_1 = i)}{\mathbb{P}(X_2 \neq i)} = \frac{\mathbb{P}(X_2 = j \mid X_1 = i)}{\mathbb{P}(X_2 = j)} \tag{11}$$

for all i and all $j \neq i$. According to (10), the initial relevance an outcome has on another is the same across all pairs of distinct outcomes, and (11) says that the relevance an outcome has on all the others, taken together, is the same as the relevance it has on each distinct outcome, taken individually. Thus, there is no privileged outcome that is affected by i more or less positively than the other categories.

It is now not difficult to prove that exponential smoothing can be derived from the aforementioned axioms.

Theorem 1. Suppose $X_1, X_2, \ldots$ is an infinite sequence of random variables taking values in a finite set $\{1, \ldots, t\}$, $t \geq 2$, that is regular (6), conditionally independent (7), satisfies invariance of inductive relevance, (8) and (9), as well as invariance of initial relevance, (10) and (11). Then there exists a $\beta \in (0, 1]$ such that

$$\mathbb{P}(X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n) = \beta^n \mathbb{P}(X_1 = i) + (1 - \beta) \sum_{m=1}^{n} \beta^{n-m} \chi_i(x_m)$$

for all n. The parameter β is given by $\beta = \mathbb{P}(X_2 = i \mid X_1 = j)/\mathbb{P}(X_1 = i)$, where i and j can be any two distinct outcomes.

Proof. It suffices to show that there is a $\beta \in (0, 1]$ such that the convex recursion (5) holds for all n. Regularity is used throughout the proof without being mentioned explicitly.

We consider two cases. First, suppose $x_n = j \neq i$. Then (5) is given by

$$\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = x_n) = \beta \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1}). \tag{12}$$

Let

$$\beta_{ij} = \frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_1 = i)}.$$

Conditional independence in distribution (7) implies

$$\frac{\mathbb{P}(X_{n+1} = i \mid X_1 = j)}{\mathbb{P}(X_{n+1} = i)} = \frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_1 = i)} = \beta_{ij}.$$

Invariance of inductive relevance (8) and (7) imply that

$$\beta_{ij} = \frac{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)}{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})} = \frac{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)}{\mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})}.$$

Hence,

$$\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j) = \beta_{ij} \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1}).$$

Now, invariance of initial relevance (10) implies that $\beta_{ij} = \beta_{kl}$ for all $i \neq j$ and $k \neq l$. So we can set $\beta = \beta_{ij}$ for any two distinct outcomes i,j. As a consequence, (12) holds for this β.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

For the second case, consider again (5) and let $x_n = i$. We need to show that for the β determined in the previous case the recursion (5) is satisfied:

$$\begin{aligned}\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = i) &= \beta\mathbb{P}(X_n = i \mid \mathbf{X}_{n-1}) + 1 - \beta \\ &= 1 - \beta(1 - \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})). \end{aligned} \tag{13}$$

To this end, note first that (7) and invariance of inductive relevance (9) imply that

$$\frac{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1}, X_n = i)}{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1})} = \frac{\mathbb{P}(X_2 \neq i \mid X_1 = i)}{\mathbb{P}(X_2 \neq i)},$$

while invariance of initial relevance, (10) and (11), together with (7) yield

$$\begin{aligned}\frac{\mathbb{P}(X_2 \neq i \mid X_1 = i)}{\mathbb{P}(X_2 \neq i)} = \frac{\mathbb{P}(X_2 = j \mid X_1 = i)}{\mathbb{P}(X_2 = j)} = \frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_2 = i)} \\ = \frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_1 = i)} = \beta,\end{aligned}$$

where β is the same number as in the previous paragraph. Putting these equations together, we get

$$\frac{1 - \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = i)}{1 - \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})} = \beta.$$

This is equivalent to

$$1 - \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = i) = \beta(1 - \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})).$$

Equation (13) follows from this by applying (7) to $\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})$. Altogether, this shows that there exists a β such that (3) holds for all n and i.

It remains to show that $0 < \beta \leq 1$. If $\beta < 0$, then

$$\frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_1 = i)} < 0,$$

which is impossible. If $\beta = 0$, then

$$\frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_1 = i)} = 0.$$

Hence $\mathbb{P}(X_2 = i \mid X_1 = j) = 0$, and so

$$\mathbb{P}(X_1 = i, X_2 = j) = \mathbb{P}(X_2 = i \mid X_1 = j)\mathbb{P}(X_1 = j) = 0,$$

which contradicts regularity.

Now suppose $\beta > 1$. We know that, for all n,

$$\mathbb{P}(X_{n+1} = i \mid X_1 = x_1, \ldots X_n = x_n) = \beta^n\mathbb{P}(X_1 = i) + (1 - \beta)\sum_{j=m}^{n} \beta^{n-m}\chi_i(x_m).$$

Suppose that $x_m \neq i$ for $m = 1, \ldots, n$. By regularity, this event has positive probability. Then

$$\mathbb{P}(X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n) = \beta^n\mathbb{P}(X_1 = i).$$

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

This can be done for any n. Since $\beta > 1$, $\beta^n \to \infty$. Hence, because $\mathbb{P}(X_1 = i) > 0$ there is an N such that

$$\mathbb{P}(X_{n+1} = i \mid X_1 = x_1, \ldots, X_N = x_N) > 1,$$

which is impossible. QED

5. Discussion

5.1. Finite forms of exponential smoothing

The last part of the proof of Theorem 1 indicates that for finite forms of exponential smoothing the parameter β may be greater than one. If $\beta > 1$, then the predictive probability of observing an outcome of type i decreases with the number of times i is observed. Hence, the predictive probability of i is greatest if it has never been observed. In that case,

$$\mathbb{P}(X_{n+1} = i \mid X_1 = x_1, \ldots, X_n = x_n) = \beta^n \mathbb{P}(X_1 = i).$$

This needs to hold for any outcome i. Let $p = \max_i \mathbb{P}(X_1 = i)$. Then exponential smoothing is well defined only if

$$\beta^n p \leq 1,$$

or equivalently,

$$n \leq -\frac{\log p}{\log \beta}.$$

By contrast, there are no such restrictions if $0 < \beta \leq 1$. If $\beta = 1$, then the predictive probabilities for each outcome i is constant and equal to $\mathbb{P}(X_1 = i)$. If $0 < \beta < 1$, exponential smoothing satisfies a well-known principle of inductive logic called *instantial relevance*. In the present context, the principle of instantial relevance says that

$$\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = i) > \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)$$

for any $j \neq i$. In other words, observing an outcome has a positive influence on its predictive probability. Now, if $\beta > 0$, then

$$\beta \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1} = x_{n-1}) + (1 - \beta) > \beta \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1} = x_{n-1})$$

if and only if $\beta < 1$. Thus, the more of i we see, the more the predictive probability of i will increase, and the predictive probabilities of other types of outcomes will decrease accordingly. As a result, for exponential smoothing with $0 < \beta \leq 1$, any finite sequence of random variables can be extended to an infinite number of observations.

This is similar to standard Carnapian inductive logic. Certain settings of the parameters restrict the observational process to be finite, while others allow it to be infinite (Zabell 1982).

5.2. Urn models

The remarks about finite forms of exponential smoothing are closely related to urn models. The standard urn model for Carnapian inductive logic is the Polya urn. In the simplest case, each type of outcome is associated with a token. At each round, a token

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

is chosen at random from the urn. It is put back together with another token of the same type. The Polya urn model generates the predictive probabilities that correspond to an infinite observational process. The initial composition of the urn can be adjusted to weigh the tokens in an unequal way to take into account the parameters α_i of the inductive system (1). A finite observational process can have negative parameters α_i (Zabell 1982). This corresponds to choosing from an urn without replacement: α_i is the weight of the i tokens in the urn, from which 1 is subtracted each time a token of type i is chosen.

It is not difficult to design an urn model for exponential smoothing. Initially, each token (outcome) i is assigned a weight $\tau_i = \mathbb{P}(X_1 = i)$. Tokens are chosen according to weight. Suppose $0 < \beta \leq 1$. If, on the first trial, i is chosen, its initial weight τ_i decreases by a factor of β, just like the remaining weights $\tau_j, j \neq i$, and a token of weight of $1 - \beta$ is added for type i. Then the process repeats. At each stage, the current weight of a token decreases by β and a token of weight $1 - \beta$ is added to the chosen type. If $\beta > 1$, then this amounts to increasing the initial weights of each token by a factor of β and subtracting the weight of the observed token by $1 - \beta$.

5.3. Are the postulates necessary?

Theorem 1 says that the axioms laid out in section 4 are sufficient for exponential smoothing. Are they also necessary?

The regularity postulate (6) is clearly not necessary. If $\beta = 0$, then the predictive probability of an outcome i is equal to one if it was observed; otherwise, it is zero. Consequently, the prior probability of any sequence of mixed outcomes (sequences showing different types of outcomes) is equal to zero. If $\beta > 0$, one might have $\mathbb{P}(X_1 = i) = 0$ for some outcome i, contradicting regularity.

Assuming regularity, the necessity of the other postulates can be demonstrated. I start with conditional independence in distribution. That (7) is a consequence of exponential smoothing follows from a more general result Berti et al. (2021), Example 7. I provide an elementary proof for completeness.

Proposition 1. Suppose $X_1, X_2, \ldots$ is a sequence of random variables taking values in a finite set $\{1, \ldots, t\}$, $t \geq 2$, and suppose that the regularity postulate (6) holds for all n. If predictive probabilities are determined by exponential smoothing (3) for some $\beta \in (0, 1]$, then the sequence $X_1, X_2, \ldots$ is conditionally independently distributed.

Proof. We start with proving the following auxiliary result: for all $k \geq 1$ and for all $i, x_1, \ldots, x_{k-1}$ in $\{1, \ldots, t\}$, $\mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1})$ is equal to

$$\sum_{x_k} \big[\mathbb{P}(X_{n+k} = x_k \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) \times \mathbb{P}(X_{n+k+1} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k} = x_k)\big]. \tag{14}$$

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

By (5), the sum in (14) is equal to

$$\sum_{x_k}[\mathbb{P}(X_{n+k} = x_k \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) \\ \times (\beta\mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) + (1-\beta)\chi_i(x_k))],$$

which is equal to

$$\mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) \\ \times (\beta\mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) + 1 - \beta) \\ + \Big[\beta\mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) \\ \times \sum_{x_k \neq i} \mathbb{P}(X_{n+k} = x_k \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1})\Big].$$

This, in turn, is the same as

$$\mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) \\ \times \left[\beta \sum_{x_k} \mathbb{P}(X_{n+k} = x_k \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) + 1 - \beta\right].$$

Then (14) follows since $\beta \sum_{x_k} \mathbb{P}(X_{n+k} = x_k \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) = \beta$.

Consider now

$$\mathbb{P}(X_{n+k+1} = i \mid \mathbf{X}_n) = \sum_{x_1} \cdots \sum_{x_k} \mathbb{P}(X_{n+k+1} = i, X_{n+1} = x_1, \ldots, X_{n+k} = x_k \mid \mathbf{X}_n).$$

The sums on the right side can be written as follows:

$$\sum_{x_1} \mathbb{P}(X_{n+1} = x_1 \mid \mathbf{X}_n) \\ \times \sum_{x_2} \mathbb{P}(X_{n+2} = x_2 \mid \mathbf{X}_n, X_{n+1} = x_1) \\ \times \cdots \times \sum_{x_k} \mathbb{P}(X_{n+k} = x_k \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1}) \\ \times \mathbb{P}(X_{n+k+1} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k} = x_k).$$

By the auxiliary result (14), the last sum is equal to $\mathbb{P}(X_{n+k} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-1} = x_{k-1})$, the preceding sum is equal to $\mathbb{P}(X_{n+k-1} = i \mid \mathbf{X}_n, X_{n+1} = x_1, \ldots, X_{n+k-2} = x_{k-2})$, and so on. The same procedure applies all the way to the first sum, which again by (14) is equal to $\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_n)$. Hence, $\mathbb{P}(X_{n+k+1} = i \mid \mathbf{X}_n) = \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_n)$ for each i and all $k \geq 1$. QED

Taking regularity for granted, invariance of inductive relevance and invariance of initial relevance are also necessary conditions for exponential smoothing.

Proposition 2. Suppose $X_1, X_2, \ldots$ is a sequence of random variables taking values in a finite set $\{1, \ldots, t\}$, $t \geq 2$, and suppose that the regularity postulate (6) holds for all n. If predictive probabilities are determined by exponential smoothing (3) for some $\beta \in (0, 1]$, then invariance of inductive relevance, (8) and (9), as well as invariance of initial relevance, (10) and (11), hold for all n.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

Proof. Assume (3) and regularity. If $j \neq i$ we have

$$\frac{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)}{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})} = \frac{\beta \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1}) + (1-\beta)\chi_i(j)}{\mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})} = \frac{\beta \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})}{\mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})} = \beta.$$

The first equality follows from (5) and from Proposition 1. For the same reasons we have

$$\frac{\mathbb{P}(X_{n+1} = i \mid X_1 = j)}{\mathbb{P}(X_{n+1} = i)} = \frac{\mathbb{P}(X_2 = i \mid X_1 = j)}{\mathbb{P}(X_2 = i)} = \frac{\beta \mathbb{P}(X_1 = i) + (1-\beta)\chi_i(j)}{\mathbb{P}(X_1 = i)} = \beta. \quad (15)$$

It follows that (8) is true:

$$\frac{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = j)}{\mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})} = \frac{\mathbb{P}(X_{n+1} = i \mid X_1 = j)}{\mathbb{P}(X_{n+1} = i)}.$$

Furthermore, (15) also shows that $\beta = \mathbb{P}(X_2 = i \mid X_1 = j)/\mathbb{P}(X_2 = i)$ for all pairs of distinct indices i, j, which shows that (10) is true.

Now consider any outcome i:

$$\frac{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1}, X_n = i)}{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1})} = \frac{1 - \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1}, X_n = i)}{1 - \mathbb{P}(X_{n+1} = i \mid \mathbf{X}_{n-1})} = \frac{1 - \beta \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1}) - (1-\beta)\chi_i(i)}{1 - \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})} = \frac{\beta(1 - \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1}))}{1 - \mathbb{P}(X_n = i \mid \mathbf{X}_{n-1})} = \beta,$$

where we again used (5) and Proposition 1. We use the same facts to obtain

$$\frac{\mathbb{P}(X_{n+1} \neq i \mid X_1 = i)}{\mathbb{P}(X_{n+1} \neq i)} = \frac{\mathbb{P}(X_2 \neq i \mid X_1 = i)}{\mathbb{P}(X_2 \neq i)} = \frac{1 - \beta \mathbb{P}(X_1 = i) - (1-\beta)}{1 - \mathbb{P}(X_1 = i)} = \beta. \quad (16)$$

It follows that

$$\frac{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1}, X_n = i)}{\mathbb{P}(X_{n+1} \neq i \mid \mathbf{X}_{n-1})} = \frac{\mathbb{P}(X_{n+1} \neq i \mid X_1 = i)}{\mathbb{P}(X_{n+1} \neq i)},$$

and so (9) is true. In addition, (16) entails

$$\frac{\mathbb{P}(X_2 \neq i \mid X_1 = i)}{\mathbb{P}(X_2 \neq i)} = \beta = \frac{\mathbb{P}(X_2 = j \mid X_1 = i)}{\mathbb{P}(X_2 = j)}.$$

Thus, (11) is true as well. QED

5.4. Variants of exponential smoothing

We have focused on the simplest type of exponential smoothing. There are many variants that modify various aspects of the simple model (Gardner 2006). For example, exponential smoothing (3) does not take into account seasonal patterns, which are important in many forecasting problems, such as weather forecasting or inventory supply forecasting. It also fails to account for trends in the data (e.g., the current direction of change). Variants of exponential smoothing that incorporate trends and

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

seasonal effects are considerably more complex, and at the moment I do not know whether there are simple invariance principles from which they can be derived.

On the other hand, exponential smoothing can be easily combined with insights from the literature on generalizations of de Finetti's theorem, for instance those known as *partial exchangeability* (De Finetti 1980; Diaconis and Freedman 1980). One example of partial exchangeability, Markov exchangeability, was briefly mentioned above. Along the same lines, we can consider outcomes that are of different *types* and satisfy the postulates of exponential smoothing *only relative to a type*. As with Markov exchangeability, these types can be defined by the predecessor state of an outcome. It is not difficult to see how to extend the axiomatic approach to exponential smoothing given here to such settings.

6. Kernels and strategies

Exponential smoothing is an example of a class of updating strategies explored in Berti et al. (2021, 2023) and Fong et al. (2023). To study their common structure more generally, we will drop the assumption of having only a finite set of outcomes $\{1, \ldots, t\}$. Everything I say below applies to this case as well, of course, but the greater generality will allow us to gain insight into connections between different predictive systems and their points of contact with Carnapian inductive logic.

Let Ω be a Polish space and $\mathfrak{B}$ its Borel sigma-algebra, and let $(\Omega^\infty, \mathfrak{B}^\infty)$ be the corresponding infinite product space (Ω^∞ is the the set of all infinite sequences of elements in Ω and $\mathfrak{B}^\infty$ the corresponding product sigma-algebra). The space $(\Omega, \mathfrak{B})$ is the set of observable outcomes, and $(\Omega^\infty, \mathfrak{B}^\infty)$ can be thought of as representing the process of repeatedly observing these outcomes.

A *kernel*, κ, on $(\Omega, \mathfrak{B})$ is a measurable map from Ω to the set of all probability distributions on $(\Omega, \mathfrak{B})$, denoted by $\mathscr{P}$. More precisely,

- $\kappa(\boldsymbol{\omega})$ is in $\mathscr{P}$ for all $\boldsymbol{\omega}$ in Ω, and
- $\kappa(\cdot)(B)$ is $\mathfrak{B}$-measurable for every B in $\mathfrak{B}$.

In the theory of Markov chains, kernels give the probabilities of transitioning from a state to any measurable set of states. In the present context, κ can be understood as expressing the effect of an observation x on the probabilities assigned to the next observation. This role will become clearer after we have introduced the notion of a *strategy*, which goes back to Dubins and Savage (1965). Let $\mathfrak{B}^n$ be the sigma algebra of subsets of Ω^∞ generated by the n-fold product of $\mathfrak{B}$. Heuristically, $\mathfrak{B}^n$ captures all events about the first n observations. A strategy σ is a sequence of measurable mappings σ_n from Ω^∞ to $\mathscr{P}$. More precisely,

- σ_0 is an element of $\mathscr{P}$ and $\sigma_n(\boldsymbol{\omega}^\infty)$ is in $\mathscr{P}$ for each $\boldsymbol{\omega}^\infty \in \Omega^\infty$, and
- $\sigma_n(\cdot)(B)$ is $\mathfrak{B}^n$-measurable for each $B \in \mathfrak{B}$ for $n = 1, 2, \ldots$

Taking $(\Omega^\infty, \mathfrak{B}^\infty)$ as a model of an observational process, a strategy σ defines a unique predictive distribution σ_n that depends only on the first n observations. Thus, a strategy σ can be thought of as an inductive logic.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

Kernels can be used to define strategies (Berti et al. 2021, 2023). For example, for any n let $\beta_n : \Omega^\infty \to [0,1]$ be $\mathfrak{B}^n$-measurable and let κ be a kernel on $(\Omega, \mathfrak{B})$. Then

$$\sigma_{n+1}(\boldsymbol{\omega}^\infty)(B) = \beta_n(\boldsymbol{\omega}^\infty)\sigma_n(\boldsymbol{\omega}^\infty)(B) + (1 - \beta_n(\boldsymbol{\omega}^\infty))\kappa(\boldsymbol{\omega}_{n+1})(B) \tag{17}$$

defines a strategy, where $\boldsymbol{\omega}_{n+1}$ is the $(n+1)$st element of $\boldsymbol{\omega}^\infty$. The kernel κ expresses the influence of the most recent observation on predictive probabilities and is weighed by $1 - \beta_n$ against the immediately preceding predictive probability σ_n, which has weight β_n. Thus, predictive probabilities are convex combinations in which the influence of the kernel is tempered by past predictive assessments with weights that are allowed to depend on the full history of the process.[4]

As noted in Berti et al. (2021, 2023), it follows from the Ionescu–Tulcea theorem that for any strategy σ there exists a probability measure $\mathbb{P}_\sigma$ on $(\Omega^\infty, \mathfrak{B}^\infty)$ that almost surely agrees with the predictive probabilities σ_n for every n. That is, let $X_n(\boldsymbol{\omega}^\infty) = \boldsymbol{\omega}_n$, where $\boldsymbol{\omega}_n$ is the nth coordinate of $\boldsymbol{\omega}^\infty$. Then

$$\mathbb{P}_\sigma(X_{n+1} \in B \mid \mathfrak{B}_n)(\boldsymbol{\omega}^\infty) = \sigma_n(\boldsymbol{\omega}^\infty)(B) \tag{18}$$

for all $\boldsymbol{\omega}^\infty$ in a set of $\mathbb{P}_\sigma$-probability one and all B in $\mathfrak{B}$. This allows us to seamlessly switch between two points of view: one where we have a probability over the infinite model of an observational process, and another where we have sequences of conditional probabilities that assign predictive probabilities to the next stage.

If the kernel κ satisfies two conditions (see (19) and (20) below), then strategies (17) give rise to a conditionally independent distribution: the random variables $X_1, X_2, \ldots$ in (18) are conditionally independently distributed according to P_σ (Berti et al. 2021, Theorem 5).

Exponential smoothing is clearly an instance of this scheme. To see this, take $\beta_n(\boldsymbol{\omega}^\infty) = \beta$ for all n and all $\boldsymbol{\omega}^\infty$, and let the kernel κ be given by the Dirac delta function:

$$\delta_{\boldsymbol{\omega}}(B) = \begin{pmatrix} 1 & \text{if} \boldsymbol{\omega} \in \mathrm{B}, \\ 0 & \text{otherwise}. \end{pmatrix}$$

By Lemma 1 this characterizes exponential smoothing in the finite case (just substitute δ for the indicator function with $B = \{i\}$). The kernel δ captures the way in which the most recent observation influences the predictive probabilities of future outcomes in perhaps the simplest way possible. It says that only the observation that is actually made matters for prediction. Its influence is hedged against previous predictive probabilities by the weights β and $1 - \beta$. Carnapian inductive logic has the same kernel δ. If we set

$$\beta_n = \frac{n + \lambda}{n + 1 + \lambda},$$

then (17) is equivalent to the Carnapian rule (1) when Ω is finite, with $\alpha_i = \lambda\sigma_0(\{i\})$ for each i in Ω. The kernel δ, together with the weights, expresses something similar

[4] The case considered here is actually a special case of the class of strategies studied in Berti et al. (2021, 2023). I focus on the simpler case because the extra layer of sophistication doesn't affect the discussion.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

to Johnson's sufficientness postulate: Only the observations actually made matter while their positions in the data stream are irrelevant.

The kernel δ thus is the common core of exponential smoothing and Carnapian inductive logic. Only the weights β_n are different. But there are many epistemic situations where δ is not a plausible choice of kernel. I already mentioned symmetries like partial exchangeability. There also has been a long debate in inductive logic about the importance of analogy in predictive inference.[5] One type of analogy is the influence an observation may have on the predictive probability of similar outcomes. With a kernel like δ, there is no such analogy influence. But the kernel framework is very flexible and allows the investigator to express all kinds of relationships between different observations. For several examples of practical significance see Berti et al. (2021, 2023) and Fong et al. (2023). In particular, a kernel κ can assign a distribution over outcomes that is guided by a similarity relation between outcomes. The only constraints that kernels are usually asked to satisfy are the following (Berti et al. 2021, 705). First, for any B in $\mathfrak{B}$,

$$\sigma_0(B) = \int_\Omega \kappa(\boldsymbol{\omega})(B)\sigma_0(d\boldsymbol{\omega}). \tag{19}$$

Second, there is a set A in $\mathfrak{B}$ such that $\sigma_0(A) = 1$ and for any $\boldsymbol{\omega}$ in A and B in $\mathfrak{B}$,

$$\kappa(\boldsymbol{\omega})(B) = \int_\Omega \kappa(\boldsymbol{\omega}')(B)\kappa(\boldsymbol{\omega})(d\boldsymbol{\omega}'). \tag{20}$$

The first constraint (19) is known in Bayesian epistemology as the "weak reflection principle." If κ is considered the future probability assignments of an agent, (19) says that the prior expectation of the future probability of any event is equal to its current probability. The weak reflection principle and its variants are often understood as touchstone principles of rational learning.[6]

The second constraint (20) is not as well known as the weak reflection principle. It goes back to at least Blackwell (1942), who referred to it as "idempotence."[7] Idempotence can be thought of as a weak reflection principle for the kernel κ itself. From the point of view of each $\kappa(\boldsymbol{\omega})$, all possible updates agree with the probabilities assigned by $\kappa(\boldsymbol{\omega})$ on average. This means that different updates κ endorse each other as legitimate. The kernel, as a whole, is internally consistent.

The two principles—weak reflection and idempotence—are very mild. They put almost no limits on what can serve as a kernel. In particular, a large number of analogical influences that observational outcomes may have on one another become expressible in the language of kernels.

In mathematics, the term "kernel" is used to refer to different objects. One is that of a transition probability, which we have been using so far. Another one is being actively investigated in machine learning, especially in the context of the apparent ability of deep learning methods to both interpolate observed data and generalize to

[5] Carnap himself was concerned with this issue, especially in his mature inductive logic (Carnap 1971, 1980). Of special relevance are his "attribute spaces" (Sznajder 2016; Poth 2023). See Skyrms (1993), Romeijn (2006), and Huttegger (2019) for accounts of analogical inference and discussions of the literature.

[6] For a discussion of reflection principles and an overview of the literature, see Huttegger (2017).

[7] See also Gaifman (1988). Huttegger and Nielsen (2020) explain it in the context of Bayesian updating.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

new observations by massive overparametrization; this is in contrast to received statistical wisdom, which holds that overparametrization allows interpolation of existing data but generalizes poorly to new situations. Belkin (2021) provides an in-depth overview and discussion of the salient issues. He focuses on standard deep learning: optimizing parameters of deep neural networks to obtain a predictive function, for example in classification and regression problems. In Bayesian deep learning, Bayesian inference is applied to updating a prior over parameters to posteriors in order to account for uncertainty. In both cases, kernel machines can be used to understand the success of certain deep neural networks. Belkin (2021) discusses kernels in the setting of standard deep learning. Similar kernels play an important role in Bayesian deep learning (see Wilson (2014) and Murphy (2023), Chapters 17 and 18). What they have in common is the idea that a kernel fixes the degree of similarity between data points. This allows inference from the outputs associated with previously observed input data x_i to the outputs associated with a newly observed input datum x.

Kernels, in this strand of the literature, are defined very generally as functions from pairs of data into the real numbers that usually satisfy certain smoothness properties, and not as transition probabilities as in Bayesian predictive inference. But there is a family resemblance between these two types of kernels. Both establish similarity connections between observational outcomes and are versatile tools for making predictive inferences within a broadly probabilistic framework. From the point of view of inductive logic, they can be thought of as expressing analogy influences for predictive inferences, where analogy influences are expressed by fixing a kernel.

Bayesian predictive inference goes significantly beyond the kernels introduced above, in a way that may prove useful for inductive logic and for machine learning. A kernel κ determines how the most recent observational outcome influences predictive probabilities. This implicitly assumes that the predictive influence of that outcome is independent of the history of the observational process. This might be a reasonable assumption in many, but certainly not all, situations. However, the approach to predictive inference through strategies already incorporates the conceptual tools to generalize prediction beyond kernels. Instead of a kernel, one can fix a strategy α that captures predictive influences, as in Berti et al. (2023). More precisely, let α be a strategy and define the strategy σ by

$$\sigma_{n+1}(\boldsymbol{\omega}^\infty)(B) = \beta_n(\boldsymbol{\omega}^\infty)\sigma_n(\boldsymbol{\omega}^\infty)(B) + (1 - \beta_n(\boldsymbol{\omega}^\infty))\alpha_{n+1}(\boldsymbol{\omega}^\infty)(B). \tag{21}$$

In (21), the influence of the most recent observation $\boldsymbol{\omega}_{n+1}$ is not just hedged against the last predictive probability, σ_n, as in (17), but also potentially mediated through the observations $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n$.[8] This allows for very fine-grained evaluations of future outcomes. The kernel methodology can be recovered by considering *Markovian* strategies, where the strategy α depends only on the most recent observation. Many intermediate cases are possible. For example, α may depend on two or three previous observations. This makes room for analogy influences that are considerably more

[8] Strategies of the form (21) give rise to a conditionally independent distribution under a martingale condition; see Berti et al. (2023), Theorem 2.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

general than the similarity relations between individual data that can be expressed within a kernel framework.

How does the inductive logic approach to exponential smoothing and, more generally, to predictive updating compare to the approach through kernels and strategies? According to the latter approach, the impact of observations on predictive probabilities is expressed by kernels or, more generally, strategies, whereas within inductive logic they are expressed by qualitative postulates for conditional probabilities. The two approaches are complementary. They both improve our understanding of certain predictive regimes. One salient difference between the two lies in the qualitative nature of Carnapian inductive logic. The methodology surveyed in this section requires the choice of a particular kernel, or strategy, that is, a function that takes on real values expressing similarities between data. Carnapian inductive logic, by contrast, considers classes of probability assignments and aims to isolate qualitative principles that give rise to particular families of predictive systems. By qualitative principles I mean relations between unconditional probabilities and conditional probabilities, such as exchangeability, Johnson's sufficientness postulate, conditional independence, or the invariance constraints for relevance quotients discussed above. As such, Carnapian inductive logic provides a distinct contribution to our understanding of predictive systems.

7. Exponential smoothing and exchangeability

Exchangeability implicitly assumes a kind of stationarity that is absent in exponential smoothing. By de Finetti's theorem, exchangeable distributions can be represented as mixtures of independent and identically distributed data streams. There appears to be no such inherent underlying structure in exponential smoothing. In fact, exponential smoothing seems to be designed for time-varying learning environments. There are, nevertheless, close connections between exponential smoothing and general conditionally independent distributions, on the one hand, and exchangeable distributions, on the other.

The key insight is that, even if they fail to be exchangeable, conditionally independent distributions are asymptotically similar to exchangeable distributions.[9] There are two main results supporting this insight. Let $(\Omega^\infty, \mathfrak{B}^\infty)$ and X_n be as in the preceding section. Suppose that $X_1, X_2, \ldots$ is conditionally independently distributed according to some measure $\mathbb{P}$. Consider the empirical measures

$$\mu_n(\boldsymbol{\omega}^\infty)(B) = \frac{1}{n}\sum_{m=1}^{n} \delta_{X_m(\boldsymbol{\omega}^\infty)}(B),$$

where B is in $\mathfrak{B}$. The measures μ_n quantify the average number of times the observational process visits B. Each μ_n is a random probability measure on $(\Omega, \mathfrak{B})$. It can be shown that the empirical measures almost surely converge weakly to a random probability measure μ on $(\Omega, \mathfrak{B})$. More precisely,

$$\mu_n(\boldsymbol{\omega}^\infty)(B) \to \mu(\boldsymbol{\omega}^\infty)(B)$$

[9] For what follows, cf. Berti et al. (2004) and Berti et al. (2021, 2023).

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

for every B is in $\mathfrak{B}$ and for every $\boldsymbol{\omega}^{\infty}$ in a set of $\mathbb{P}$-probability one. Since the sequence μ_n is bounded, we have

$$\mathbb{E}(\mu(B) \mid X_1, \ldots, X_n) = \lim_k \mathbb{E}(\mu_k(B) \mid X_1, \ldots, X_n) = \lim_k \frac{1}{k} \sum_{i=n+1}^{k} \mathbb{P}(X_i \in B \mid X_1, \ldots, X_n)$$
$$= \mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n).$$

The last equality follows from the assumption that the sequence of random variables is conditionally independently distributed. The second equality follows since the first n terms of the sum, which are either zero or one, vanish in the limit. Since the sequence of conditional probabilities $\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n)$ is a martingale, the martingale convergence theorem implies that

$$\mathbb{P}(X_{n+1} \in B \mid X_1, \ldots, X_n) \to \mu(B)$$

almost surely.

Everything that has been said thus far also holds if the observational process is exchangeable. But in the exchangeable case, something stronger is true. The random measure μ completely determines the full probability distribution of an exchangeable sequence of random variables. By de Finetti's theorem, the sequence is distributed independently with common measure μ, and the uncertainty about μ is expressed by a prior over the set of all probability measures on $(\Omega, \mathfrak{B})$.

In the exchangeable case, the measure μ can be understood as representing the chance setup that generates the observational process. This is the stationarity underlying an exchangeable process. For exponential smoothing and other predictive systems that are conditionally independently distributed but not exchangeable, μ cannot be interpreted in this way. Rather, from what has been said above it follows that μ represents the best estimate of the predictive probability measure and that the predictive probability of an observable outcome is, in turn, the expectation of its best estimate.

A second result shows that the connection between exponential smoothing and exchangeability is even closer. Suppose $X_1, X_2, \ldots$ is not exchangeable but is conditionally independently distributed according to a probability $\mathbb{P}$. Then there exists a sequence of random variables $Y_1, Y_2, \ldots$ that is exchangeable with probability distribution $\mathbb{Q}$, and

$$(X_n, X_{n+1}, \ldots) \to (Y_1, Y_2, \ldots)$$

in distribution as $n \to \infty$. Thus, the probability of $(X_n, X_{n+1}, \ldots)$ being in a certain range of values is approximately equal to the probability of $(Y_1, Y_2, \ldots)$ being in that range of values. This does not guarantee that all probabilistic inferences drawn based on the first sequence are the same as those drawn about the exchangeable sequence, but it does guarantee that bounded random variables that depend on the sequences are evaluated the same asymptotically. As a result, there is a sense in which any sequence of random variables that is conditionally independently distributed is asymptotically exchangeable (even if it is not outright exchangeable).

The connection between exponential smoothing and exchangeability speaks to a point raised by Kuipers (1988) about analogy by proximity in the past. According to Kuipers, it seems doubtful that there are "objective processes" that would give rise to an inductive rule like exponential smoothing. For Carnapian inductive logic there is

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

such a process. As we have seen in section 2, from a parametric Bayesian perspective Carnapian inductive rules derive from independent multinomial distributions, which can be thought of as an underlying objective stochastic process whose parameters are estimated by the inductive rule. By the result described above, exponential smoothing approximates such a process in the long run, even though it is not fully determined by it. This process can be thought of as an objective process that is weakly connected to exponential smoothing.

8. Invariance axioms as an expression of the meaning of data

For a subjective Bayesian, invariance assumptions and probabilistic symmetries are simply the probability judgments of a rational agent (de Finetti 1974; Skyrms 1984; Jeffrey 1992; Huttegger 2017). If an agent's subjective probabilities are exchangeable, they can be represented as a mixture of independent and identically distributed probabilities by de Finetti's theorem. If they are, in addition, regular and satisfy Johnson's sufficientness postulate, they give rise to a Carnapian predictive system. The same is true for the invariance postulates that underlie exponential smoothing. Yet, whether subjective probabilities satisfy any of these conditions is a contingent fact about an agent. In particular, there is nothing irrational about probabilities that fail to be exchangeable or to possess some other kind of invariance.

Carnap, at least in his early writings on the subject, had a different view of invariance conditions. In Carnap (1950) he thought of inductive logic as a branch of formal semantics. On this view, invariance conditions reflect certain features of the underlying language. The language is first-order with relations and individual constants. Exchangeability is a consequence of considering probability as a generalization of logical consequence.[10] The probability of a statement given another is thought of as expressing the degree of support the second has on the first. Since logical consequence is invariant under replacing the individual constants of the language, probability should, according to Carnap, have the same invariance.

Carnapian inductive logic takes probabilistic invariance conditions to be *determined by* semantic features of the underlying logical language. Bayesian inductive logic also has an underlying conceptual framework (usually a set-theoretic and not a language-based one), but probabilistic invariance conditions are additional *independent constituents* of the inductive system. The two approaches obviously have a lot in common. Some commentators have claimed that Carnap's views evolved toward an effectively Bayesian position.[11] There is some evidence for this perspective, but it can also be criticized for good reasons, especially the fact that Carnap himself never seems to have endorsed an interpretation of probability as subjective degree of belief, emphasizing instead the analyticity of probability assignments within inductive logic (Sznajder 2016). Needless to say, there is a large number of open questions, both historical and systematic, around Carnapian and Bayesian inductive logic, what their points of contact are, and how they potentially differ. This is not the place to embark on the extensive enterprise of discussing these issues. What I can do here is point out a role of probabilistic invariance and symmetry assumptions that I think both Carnapians and Bayesians can endorse.

[10] In Carnap's terminology, probability is degree of confirmation and exchangeability is symmetry.
[11] See Zabell (1988), Earman (1992), Skyrms (1996), and Galavotti (2007).

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

What I'm going to suggest is not an original point of view. It is implicit in some of the writings of Carnap and de Finetti, and it has been expressed in one form or another by Bayesians who try to engage with the role of priors in inductive inference instead of reflexively dismissing them as arbitrary. Within a sufficiently rich setting, certain features of a prior probability assignment can be understood as "inductive assumptions" in the sense of Howson (2000) and Romeijn (2005). In order to explain inductive assumptions, it will be helpful to refer to a distinction between "data" and "evidence" in philosophy of statistics. As explained in Morey et al. (2016), data are raw entries in statistical databases, such as sequences of zeros and ones, survey responses, or measurement entries. They acquire the status of evidence—in the sense of statistical evidence—by putting them in contact with statistical hypotheses, h_i, in terms of likelihoods, $\mathbb{P}(d \mid h_i)$, where d represents a stream of raw data. Together with a prior over hypotheses, this allows inferences about the posterior probability of hypotheses and unobserved data (by Bayes' theorem). One way to think about this is that d, by itself, is a purely syntactical object. It acquires meaning by specifying a theoretical context regarding how it interacts with statistical hypotheses and how it affects the probability of unobserved data. Evidence is therefore context-dependent:

> To our mind, the context-sensitivity of evidence is an apt expression of the widely held view that empirical facts do not come wrapped in their appropriate interpretation. The same empirical facts will have different interpretations and different evidential value in different situations. We ourselves play a crucial part in this interpretation, by framing the empirical facts in a theoretical context or more concretely, in a statistical model. (Morey et al. 2016, 10)

Here, inductive assumptions about an observational process are expressed by adopting a certain set of hypotheses, likelihoods, and priors. What I would like to point out is that often inductive assumptions can alternatively be expressed by invariance conditions on the probabilistic structure of data streams, directly and without recourse to statistical hypotheses. Consider exchangeability as an example. If observational outcomes are assumed to be exchangeable, then all one can see in a data sequence are frequencies. The same is true, by de Finetti's theorem, in case one assumes an underlying independent multinomial statistical model, as well as for other types of invariance, like Markov exchangeability, which allows one to see nothing but transition frequencies. In all these cases, invariance assumptions and probabilistic symmetries express the meaning of data. One can look at one and the same sequence of observations and pay attention to different aspects of it, depending on the assumed invariance conditions.

The invariance conditions of exponential smoothing serve the same purpose. Whenever a particular observation is made, it has the same influence on inductive probabilities. The influence is fixed from the start and guides the whole process. This is a consequence of conditional independence in distribution and of invariance of inductive relevance. Thus, an agent who adopts these invariance conditions will also count the number of times an observation was made, but will weigh them according to how recently it was made. Such an agent observes different aspects of a data sequence than an agent who adopts an exchangeable probability. They frame the empirical facts under different inductive assumptions.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

The doctrine that inductive assumptions (invariance conditions, statistical models) shape the meaning of data is part of the common ground between Bayesian and Carnapian inductive logic. This leaves many questions unanswered. Most importantly, there is an unresolved gap between the semantic role of inductive assumptions—they express the meaning of data—and their epistemological role in inductive inference. Whether we assume exchangeability, invariance of inductive relevance, or any other invariance condition does not just say something about what we see in the data; it has real consequences about how our probabilities respond to evidence. Carnap tried to resolve these issues within his voluntarist metaphilosophy (the tolerance principle), in which inductive assumptions arise from a merely pragmatic choice of a conceptual framework. De Finetti, as mentioned above, confronted them by regarding inductive assumptions as nothing more than expressions of personal belief. Is there a connection between these two approaches? Do they have the same kind of pragmatism in common? Do these approaches account for the rationality of scientific reasoning? I set these questions aside for now and hope to revisit them in the future.

9. Conclusion

Exponential smoothing is the canonical model in science for analogy by proximity in the past. We have shown in this paper that it can be given a full treatment within Carnapian inductive logic. There is a set of plausible axioms that are jointly sufficient for exponential smoothing. Several axioms are also necessary (with certain qualifications). The resulting inductive logic connects to topics in data science and machine learning, in predictive statistics, and to more general issues in philosophy of science. On the technical side, conditionally independent distributions and relevance quotients might prove useful in studying other predictive systems.

Acknowledgments. For discussions and comments I would like to thank Hannes Leitgeb, Krzysztof Mierzewski, Aydin Mohseni, Ignacio Ojea Quintana, Jan-Willem Romeijn, Hannah Rubin, Teddy Seidenfeld, Brian Skyrms, Tom Sterkenburg, Marta Sznajder, Sandy Zabell, Francesca Zaffora Blando, and Kevin Zollman. I'm also grateful to audiences at Carnegie Mellon University, the University of Rochester, the 2025 *Formal Epistemology Meets Philosophy of AI* workshop at the MCMP, and the EPSA 2025 in Groningen. Finally, I thank the John Simon Guggenheim Foundation for their generous support of my research.

Funding statement. None to declare.

Declarations. None to declare.

References

Belkin, Mikhail. 2021. "Fit Without Fear: Remarkable Mathematical Phenomena of Deep Learning Through the Prism of Interpolation". *Acta Numerica* 30:203–48. doi: https://doi.org/10.1017/S0962492921000039.

Berti, Patrizia, Emanuela Dreassi, Fabrizio Leisen, Luca Pratelli, and Pietro Rigo. 2023. "Bayesian Predictive Inference Without a Prior". *Statistica Sinica* 33:2405–29. doi: https://doi.org/10.5705/ss.202021.0238.

Berti, Patrizia, Emanuela Dreassi, Luca Pratelli, and Pietro Rigo. 2021. "A Class of Models for Bayesian Predictive Inference". *Bernoulli* 27:702–26. doi: https://doi.org/10.3150/20-BEJ1255.

Berti, Patrizia, Luca Pratelli, and Pietro Rigo. 2004. "Limit Theorems for a Class of Identically Distributed Random Variables". *Annals of Probability* 32:2029–252. doi: https://doi.org/10.1214/009117904000000676.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

Blackwell, David. 1942. "Idempotent Markoff Chains". *Annals of Mathematics* 43:560–67. doi: https://doi.org/10.2307/1968811.
Brown, Robert G. 1959. *Statistical Forecasting for Inventory Control*. New York: McGraw-Hill.
Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago, IL: University of Chicago Press.
Carnap, Rudolf. 1971. "A Basic System of Inductive Logic, Part 1". In *Studies in Inductive Logic and Probability*, edited by Rudolf Carnap and Richard C. Jeffrey, vol. 1, 33–165. Berkeley, CA: University of California Press. doi: https://doi.org/10.1525/9780520334250-003.
Carnap, Rudolf. 1980. "A Basic System of Inductive Logic, Part 2". In *Studies in Inductive Logic and Probability*, edited by Richard C. Jeffrey, vol. 2, 7–155. Berkeley, CA: University of California Press. doi: https://doi.org/10.1525/9780520318328-002.
Costantini, Domenico. 1979. "The Relevance Quotient". *Erkenntnis* 14:149–57. doi: https://doi.org/10.1007/bf00196730.
de Finetti, Bruno. 1974. *Theory of Probability*, vol. 1. New York: John Wiley & Sons.
De Finetti, *Bruno*. 1980. "On the Condition of Partial Exchangeability". In *Studies in Inductive Logic and Probability*, edited by Richard C. Jeffrey, vol. 2, 193–206. Berkeley, CA: University of California Press. doi: https://doi.org/10.1525/9780520318328-005.
Diaconis, Persi and David Freedman. 1980. "De Finetti's Generalizations of Exchangeability". In *Studies in Inductive Logic and Probability*, edited by Richard C. Jeffrey, vol. 2, 233–49. Berkeley, CA: University of California Press. doi: https://doi.org/10.1525/9780520318328-007.
Dubins, Lester E. and Leonard J. Savage. 1965. *How to Gamble If You Must: Inequalities for Stochastic Processes*. New York: McGraw Hill.
Earman, John. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.
Falck, Fabian, Ziyu Wang, and Christopher C. Holmes. 2024. "Are Large Language Models Bayesian? A Martingale Perspective on In-Context Learning". ICLR 2024 Workshop on Secure and Trustworthy Large Language Models.
Fong, Edwin, Chris Holmes, and Stephen G Walker. 2023. "Martingale Posterior Distributions". *Journal of the Royal Statistical Society Series B: Statistical Methodology* 85:1357–91. doi: https://doi.org/10.1093/jrsssb/qkad005.
Gaifman, Haim. 1988. "A Theory of Higher Order Probabilities". In *Causation, Chance and Credence*, edited by Brian Skyrms and William L. Harper, 191–219. Dordrecht: Kluwer. doi: https://doi.org/10.1007/978.
Galavotti, Maria C. 2007. "Confirmation, Probability, and Logical Empiricism". In *The Cambridge Companion to Logical Empiricism*, edited by Alan W. Richardson and Thomas Uebel, 117–35. Cambridge: Cambridge University Press. doi: https://doi.org/10.1017/CCOL0521791782.006.
Gardner, Everette S. 2006. "Exponential Smoothing: The State of the Art – Part II". *International Journal of Forecasting* 22:637–66. doi: https://doi.org/10.1016/j.ijforecast.2006.03.005.
Holt, C. C. 1957. "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages". ONR Memorandum, vol. 52. Pittsburgh, PA: Carnegie Institute of Technology.
Howson, Colin. 2000. *Hume's Problem: Induction and the Justification of Belief*. Oxford: Oxford University Press.
Huttegger, Simon M. 2017. *The Probabilistic Foundations of Rational Learning*. Cambridge: Cambridge University Press.
Huttegger, Simon M. 2019. "Analogical Predictive Probabilities". *Mind* 128:1–37. doi: https://doi.org/10.1093/mind/fzw049.
Huttegger, Simon M. and Michael Nielsen. 2020. "Generalized Learning and Conditional Expectation". *Philosophy of Science* 87:868–83. doi: https://doi.org/10.1086/710543.
Jeffrey, Richard C. 1992. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.
Johnson, William E. 1932. "Probability: The Deductive and Inductive Problems". *Mind* 41:409–23. doi: https://doi.org/10.1093/mind/XLI.163.281.
Kuipers, Theo A. F. 1988. "Inductive Analogy by Similarity and Proximity". In *Analogical Reasoning*, edited by David H. Helman, 299–313. Dordrecht: Springer. doi: https://doi.org/10.1007/978-94-015-7811-0_14.
Morey, Richard D., Jan-Willem Romeijn, and Jeffrey N. Rouder. 2016. "The Philosophy of Bayes Factors and the Quantification of Statistical Evidence". *Journal of Mathematical Psychology* 72:36. doi: https://doi.org/10.1016/j.jmp.2015.11.001.
Murphy, Kevin P. 2023. *Probabilistic Machine Learning: Advanced Topics*. Cambridge, MA: MIT Press.
Paris, Jeffrey and Alena Vencovská. 2015. *Pure Inductive Logic*. Cambridge: Cambridge University Press.

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press

Poth, Nina. 2023. "Probabilistic Learning and Psychological Similarity". *Entropy* 25:1407. doi: https://doi.org/10.3390/e25101407.
Romeijn, Jan-Willem. 2005. "Bayesian Inductive Logic". PhD diss., Rijksuniversiteit Groningen.
Romeijn, Jan-Willem. 2006. "Analogical Predictions for Explicit Similarity". *Erkenntnis* 64:253–80. doi: https://doi.org/10.1007/s10670-005-0232-8.
Romeijn, Jan-Willem. 2011. "Inductive Logic and Statistics". In *Handbook of the History of Logic*, vol. 10, edited by Dov M. Gabbay, Stephan Hartmann, and John Woods, 625–50. Amsterdam: Elsevier. doi: https://doi.org/10.1016/B978-0-444-52936-7.50015-X.
Skyrms, Brian. 1984. *Pragmatics and Empiricism*. New Haven, CT: Yale University Press.
Skyrms, Brian. 1991. "Carnapian Inductive Logic for Markov Chains". *Erkenntnis* 35:439–60. doi: https://doi.org/10.1007/BF00388296.
Skyrms, Brian. 1993. "Carnapian Inductive Logic for a Value Continuum". *Midwest Studies in Philosophy* 18 (1):78–89.
Skyrms, Brian. 1996. "Inductive Logic and Bayesian Statistics". In *Statistics, Probability, and Game Theory: Papers in Honor of David Blackwell*, edited by Thomas S. Ferguson, Lloyd S. Shapley, and James B. MacQueen, 321–336. Cleveland, OH: Institute of Mathematical Statistics. doi: https://doi.org/10.1214/lnms/1215453580.
Smyl, Slawek. 2020. "A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting". *International Journal of Forecasting* 36:75–85. doi: https://doi.org/10.1016/j.ijforecast.2019.03.017.
Smyl, Slawek, Grzegorz Dudek, and Paweł Pełłka. 2023. "ES-dRNN: A Hybrid Exponential Smoothing and Dilated Recurrent Neural Network Model for Short-Term Load Forecasting". *IEEE Transactions on Neural Networks and Learning Systems* 35:11346–58. doi: https://doi.org/10.48550/arXiv.2112.02663.
Sznajder, Marta. 2016. "What Conceptual Spaces can do for Carnap's Late Inductive Logic". *Studies in History and Philosophy of Science Part A* 56:62–71. doi: https://doi.org/10.1016/j.shpsa.2015.12.001.
Wilson, Andrew G. 2014. "Covariance Kernels for Fast Automatic Pattern Discovery and Extrapolation with Gaussian Processes". PhD diss., University of Cambridge.
Zabell, Sandy L. 1982. "W. E. Johnson's 'Sufficientness' Postulate". *The Annals of Statistics* 10:1090–99. doi: https://doi.org/10.1214/aos/1176345975.
Zabell, Sandy L. 1988. "Symmetry and its Discontents". In *Causation, Chance and Credence*, edited by Brian Skyrms and William L. Harper, 155–90. Dordrecht: Kluwer.
Zabell, Sandy L. 1992. "Predicting the Unpredictable". *Synthese* 90:205–32. doi: https://doi.org/10.1007/BF00485351.
Zabell, Sandy L. 1995. "Characterizing Markov Exchangeable Sequences". *Journal of Theoretical Probability* 8:175–78. doi: https://doi.org/10.1007/bf02213460.

Cite this article: Huttegger, Simon M. 2026. "Carnapian Inductive Logic for Exponential Smoothing." *Philosophy of Science* 93 (3):607–630. https://doi.org/10.1017/psa.2026.10197

https://doi.org/10.1017/psa.2026.10197 Published online by Cambridge University Press