Erkenntnis
https://doi.org/10.1007/s10670-023-00761-9

ORIGINAL RESEARCH

Meta-inductive Justification of Inductive Generalizations

Gerhard Schurz[1]

Received: 12 February 2023 / Accepted: 19 November 2023

Abstract

The account of meta-induction (G. Schurz, Hume's problem solved: the optimality of meta-induction, MIT Press, Cambridge, 2019) proposes a two-step solution to the problem of induction. Step 1 consists in a mathematical a priori justification of the predictive optimality of meta-induction, upon which step 2 builds a meta-inductive a posteriori justification of object-induction based on its superior track record (Sect. 1). Sterkenburg (Br J Philos Sci, forthcoming. 10.1086/717068/) challenged this account by arguing that meta-induction can only provide a (non-circular) justification of inductive predictions for now and for the next future, but not a justification of inductive generalizations (Sect. 2). This paper develops a meta-inductive method that does provide an a posteriori justification of inductive generalizations, in the form of exchangeability conditions (Sect. 3). In Sect. 4, a limitation of the proposed method is worked out: while the method can justify weakly lawlike generalizations, the justification of strongly lawlike generalizations (claimed to hold for all eternity) requires epistemic principles going beyond meta-induction based on predictive success.

1 Introduction

The problem of induction was raised by David Hume 250 years ago. Hume argued that all standard methods of justification fail when applied to the task of justifying induction, broadly conceived as the projection of observed patterns from the past to the future. Most importantly, induction cannot be justified by induction from the observation of its past success as follows:

Inductive justification of induction: Induction has been successful in the past, thus by induction it will be successful in the future.

✉ Gerhard Schurz
schurz@hhu.de

[1] Department of Philosophy, Heinrich Heine University Duesseldorf, 40225 Duesseldorf, Germany

Published online: 30 March 2024

This argument is circular and without justificatory value. As Salmon (1957, p. 46) has pointed out, counter-induction (that predicts the opposite of induction) can be pseudo-justified in the same circular manner:

Counter-inductive justification of counter-induction: Counter-induction was unsuccessful in the past, thus by counter-induction it will be successful in the future.

That circular arguments may 'pseudo-justify' mutually inconsistent conclusions has been shown also for other kinds of circles (Douven, 2011, sec. 3.2).

A probabilistic version of the circularity problem besets the justification of a prior distribution for the predictions of event probabilities. While the predictive success of a chosen prior depends on the future course of events, this future course can only be assessed by probabilistic induction based on a given prior. Bayesians reply that the influence of the prior can be washed out by conditionalizing the posteriors on increasing amounts of evidence, but this reply has two limitations: (i) Not all prior distributions can be washed out in this way, not even in the long run. For example, Carnap's $m^{\dagger}$ measure that assigns a uniform probability (density) to all possible event-sequences cannot, because it makes inductive learning impossible (Carnap, 1950, pp. 564–566). (ii) In the short run the situation is worse: for every finite amount of evidence there exists a suitably biased prior that resists learning from this evidence (Schurz, 2019, sec. 4.5).

In conclusion, the crucial challenge of Hume's problem is to find a *non-circular* justification of induction. Such a justification has to be a priori in the sense that it does not assume anything about the future or the unobserved part of the world. A justification attempt of this sort was proposed in Reichenbach's "best-alternative" account to induction. Reichenbach (1949) argued that induction is the best one can do to achieve successful predictions. What Reichenbach attempted here is an *optimality justification*. Optimality justifications are epistemologically weaker than reliability justifications. They do not establish that a prediction method is reliable, in the sense that its predictive success is greater than a certain threshold that is greater than random success. An a priori demonstration of the reliability of induction is impossible because of the possibility of skeptical scenarios in which *no method* can be successful. Skeptical scenarios refute the possibility of reliability justifications, but are compatible with optimality justifications, because even in skeptical scenarios a method may be the 'best-of-a-bad-lot'.

Reichenbach's best-alternative account failed because Reichenbach applied it to induction at the level of events, called *object-induction*. What blocks Reichenbach's account is the possibility of methods that are superior to scientific object-induction, e.g., methods based on *clairvoyance*, or more generally, based on 'paranormal' information channels by means of which certain agents could 'see' the future, i.e., receive temporally backward-directed signals from it, independent from information about the past. This possibility cannot be excluded a priori (Skyrms, 1975, ch. III.4; Schurz, 2021).[1] Schurz (e.g., 2019) develops a new optimality approach by applying induction at the level of meta-methods, called *meta-induction*. Meta-induction can

[1] Reichenbach recognized the possibility of clairvoyance and argued that if a successful future-teller existed, then the inductivist could recognize this by applying induction to the success of prediction methods (1949, 476f). But Reichenbach didn't make anything out of this observation; in particular he did not

handle the clairvoyant objection because if there would really be an accessible prediction method that is superior to scientific induction, meta-inductivists would base their predictions on this method.[2]

More generally, meta-inductive methods try to find an optimal prediction method by basing their prediction on the predictions and the observed track records of the set of *accessible methods*. This set is called the *pool* of *candidate* methods. A method M is accessible to a subject S if M is at least *externally* accessible to S, which means that S has access to M's predictions but need not understand M. If S understands M and can perform M by itself, we say that M is *internally* accessible to S. A merely externally given method must be 'played' by a real agent. In contrast, internally accessible methods may be simulated by the meta-inductivist itself.

The candidate pool contains object-level methods of various sorts, including object-inductive methods as well as non-inductive or other 'alternative' methods. Object-inductive methods may be simple or complex; purely observation-based or based on scientific theories. A simple version of object-inductive methods are the Carnapian λ-rules employed in Douven's (2023) evolutionary simulations. They predict the probability of an unobserved binary event $e_{n+1} \in \{1,0\}$ conditional on a sequence of observed events $e_1,\ldots,e_n$ as

$$P(e_{n+1} = 1 \mid e_1, \ldots , e_n) = \frac{n_1 + \frac{\lambda}{2}}{n + \lambda},$$

where n_1 is the number of 1s in $e_1,\ldots,e_n$; $\lambda = 0$ yields the straight rule and $\lambda = 2$ Laplace's rule of succession. Generally considered, the number of possible inductive methods is countless, all of them projecting *some* sort of observed pattern to future or unobserved cases. Importantly, the meta-inductive account does not need a sharp general definition of an 'inductive method'. On the other hand, clear examples of a non-inductive method are counter-inductive methods, blind guessing methods, and agent-based methods relying on purported clairvoyance.

Besides object-level methods the candidate pool may also contain other meta-methods, which is important for two reasons: first, because only under this assumption meta-induction can be optimal also in regard to other accessible meta-methods, and second, because in this way an infinite regress of meta-levels is avoided. Abstractly speaking a prediction method is a device that (at any time point n) produces a prediction in application to a method-specific input. While the input of

Footnote 1 (continued)

try to show that by this observation the inductivist could have an equally high predictive success as the future-teller (cf. Skyrms 1975, ch. III.4).

[2] Pitts (2023) objects that that the approach of meta-induction cannot simultaneously admit the a priori possibility of clairvoyance and reject it on a posteriori grounds, because its standard rejection is based on its contradiction with the results of scientific object-induction. Schurz (2023, sec. 3) replies that the meta-inductive rejection of clairvoyance is not based on the a priori assumption of object-induction, but on the a posteriori ground of its poor empirical success record.

object-level methods consists of observed object-level events, that of meta-methods includes the success records and predictions of other methods.

By transforming the best-alternative account to the meta-level, the optimality of meta-induction becomes mathematically demonstrable. The demonstration is carried out in the framework of *prediction games*. A prediction game consist of an infinite sequence of (binary or real-valued) events e_1, e_2,…, whose (normalized) values lie in the real-valued interval [0,1], together with a given meta-inductive method MI and a candidate pool $\mathcal{C}$ of methods (or 'players') that are accessible to MI. In each round (or discrete time point) $n = 1, 2,\ldots$, each method $M \in \{MI\} \cup \mathcal{C}$ delivers a prediction $pred_{n+1}(M)$ of the next event e_{n+1}. The predictions, too, take their values in [0,1]. Importantly, it is permitted to predict weighted averages or probabilities of event values. The predictions are scored by a convex loss function, $loss(pred_n,e_n)$, normalized within the interval [0,1]. Convexity means that for any two predictions and weight $w \in [0,1]$ the loss of their w-weighted average is not greater than the w-weighted average of their losses. Typical convex loss functions are the absolute distance $|pred_n - e_n|$ between e_n and $pred_n$, or the squared distance $(pred_n - e_n)^2$ that is especially suited for the *prediction of probabilities* (see Sect. 2). The score obtained by a method M in round n is defined as $1 - loss(pred_n(M),e_n)$, M's *cumulative* score or absolute success achieved in round n, $S_n(M)$, is the sum of M's scores until round n, and M's *average* score or success *rate* at round n, $s_n(M)$, is defined as $S_n(M)/n$. For binary predictions with an absolute loss function, their average score is identical to their truth frequency.

Based on theorems in machine learning (Cesa-Bianchi & Lugosi, 2006), Schurz shows that a certain form of meta-induction, called attractivity-based meta-induction, is universally optimal among all accessible methods. MI predicts a weighted average of the predictions of the candidate methods[3]:

(1) $pred_{n+1}(MI) = \Sigma_{1 \le i \le m} w_n(M_i) \cdot pred_{n+1}(M_i)$,
using the following normalized (w) and unnormalized (w′) weights: $w'_n(M_i) = \exp(\eta \cdot S_n(M_i))$, with $\eta = \sqrt{8 \cdot \ln(m)/(n+1)}$ (Schurz, 2019, (6.8)),
and $w_n(M_i) = w'_n(M_i) / \Sigma_{1 \le i \le m} w'_n(M_i)$.

(2) *Optimality result for MI:* For every possible event sequence e_1, e_2,… and candidate pool $\mathcal{C} = \{M_1,\ldots,M_m\}$:

(2.1) MI's 'long run' average score ($\lim_{n\to\infty} s_n(MI)$) is never worse and sometimes better than that of the best candidate methods.

(2.2) In the 'short run', small losses of MI compared to the actually best method, so-called 'regrets', are unavoidable, because MI bases its prediction of the

[3] The weights in (1) are equivalently definable as negative exponentials $w'_n(M) = \exp(-\eta \cdot Loss_n(M))$ of the cumulative losses $Loss_n(M) =_{def} \sum_{1 \le i \le n} loss(pred_i(M),e_i)$, since $Loss_n(M) = n - S_n(M)$ and the constant factor $\exp(-\eta \cdot n)$ cancels by normalization.

> next event on the *past* track records of the candidate methods; however, these regrets have the following tight worst-case bound:[4]
> $\max(\{\mathrm{suc}_n(\mathrm{M}_i)\colon 1 \le i \le m\}) - \mathrm{suc}_n(\mathrm{MI}) \le 1.77 \cdot \sqrt{\ln(m)\,/\,n}$,
> so they become quickly negligible when the number of rounds (n) grows larger than the number of competing methods (m) (ibid., th. 6.9).

This optimality result holds for all prediction games, even in 'paranormal' environments that host clairvoyants or adversarial methods that try to deceive MI, as well as in chaotic environments in which the method's average scores do not converge to a stable performance ordering but oscillate around each other forever. Note that not all meta-inductive methods are universally optimal in the sense of result (2). For example, the method Imitate-the-best (ITB) that always predicts what the hitherto most successful candidate method predicts is provably non-optimal (Schurz, 2019, sec. 6.3), as well as success-weighted meta-induction that directly uses the method's success rates as their weights (ibid., sec. 6.8.1). MI is optimal because the weight its assigns to a candidate method reflects its 'attractivity' (or 'regret'), which means that $w_n(M_i)$ increases with M_i's success but converges to zero if M_i continues to perform worse than MI. So if the candidate pool contains a sustainably superior method M*, MI will soon assign almost all weight to M* and converge predictively towards ITB.

Prediction games and corresponding optimality results have been generalized in three respects: (1.) to discrete prediction games with non-convex loss functions, (2.) to prediction games with an increasing pool of candidate methods, and (3.) to action games in which choices of actions instead of predictions are optimized (Schurz, 2019, sec. 6.7, 7.3, 7.5).

By itself, the a priori optimality of meta-induction does not entail anything about the rationality of object-induction. Which prediction method, or combination of methods, is meta-inductively evaluated as optimal is an a posteriori matter that depends on the empirically given track record of the accessible methods. The possibility of superior non-inductive methods cannot be excluded a priori. However, the a priori justification of meta-induction bestows us the following a posteriori *justification* of object-induction: to the extent that object-inductive prediction methods were observed as more successful than all accessible non-inductive methods, we are meta-inductively justified in continuing to favor object-inductive prediction methods in the future. This argument is no longer circular, because a non-circular justification of meta-induction has been independently established.

Summarizing, Schurz' account of justifying induction consists of two parts: (i) the a priori (mathematical) justification of meta-induction, and (ii) the a posteriori (empirical) justification of object-induction based on (i). The next section discusses

[4] The loss term in (ii) simplifies the loss term of th. 6.9(ii) of Schurz (2019) and follows from it, since $\sqrt{2 \cdot \ln(m)/n} + \sqrt{\ln(m)/8 \cdot n^2} \le \left(\sqrt{2} + \sqrt{1/8}\right) \cdot \sqrt{\ln(m)/n}$ and $\left(\sqrt{2} + \sqrt{1/8}\right) \le 1.77$.

a recent challenge to the meta-inductive account by Sterkenburg (forthcoming), who (in reply to a paper of Douven, 2023) argues that meta-induction can offer a justification of object-inductive predictions only for now and for the next time point, but not for the infinite future.

2 Sterkenburg's Challenge: What Can Meta-induction Justify?

Douven (2023) supports the meta-inductive approach, arguing that "this justification has accomplished something quite remarkable" (ibid., p. 384), but he continues that this "account leaves an important aspect of our object-inductive practices unexplained, to wit, that these practices have been *highly* successful" (ibid., p. 382). Douven is asking here for an a posteriori explanation of the high success of object-induction based on contingent properties of our actual world. Clearly, the a priori results about meta-induction cannot answer this question and one should not expect that they can. The obvious *minimal* explanation of the observed success of (object-) induction must be that the part of the world to which induction has been applied in the past was to some degree *uniform*, i.e., it exhibited certain (strict or statistical) regularities that have been inductively projected. This explanation follows even *analytically*, since the success of any particular method of induction rests on the projectability of a certain pattern from observed to unobserved events, and this projectability constitutes a regularity that explains the method's success. For Douven this minimal explanation, although correct, is too weak: it does not explain why methods of induction developed as rapidly and successfully as they actually did. Based on an evolutionary simulation of different learning methods Douven (2023) provides a deeper explanation of the success of induction that, roughly speaking, consists in social learning combined with an evolutionary optimization of learning parameters. This is not the place to review the results of Douven's fascinating work but to explore its epistemological assumptions and their justification by meta-induction. Douven understands the notion of "success" not merely as induction's success as observed in the past, but as induction's *general* success, including its future success. As a consequence, Douven's question faces the problem of induction: before one can ask for an *explanation* of the general success of induction, one has to be *justified* in believing in this success. Douven is aware of the induction problem involved in his question and writes: "There is strong inductive evidence that induction is highly successful, which requires that our world satisfy certain uniformity conditions. The reliance on induction here is entirely justified in light of Schurz's results" (2023, p. 402).

In conclusion, Douven (2023) argues that the application of meta-induction to the superior track record of scientific induction provides an a posteriori justification of the general success or rationality of object-induction, and he offers an explanation of this general success. This is the point where Sterkenburg's challenge hooks in. According to Sterkenburg, the account of meta-induction cannot provide what Douven's evolutionary explanation requires, namely a justification of the general rationality of object-induction (with "induction" Sterkenburg means "object-induction"). Sterkenburg (forthcoming, sec. 6, §§ 2–4) argues that meta-induction can only provide

i. a justification of the rationality of induction that holds *for now*, but not necessarily for all times, and
ii. a justification that holds only in application to the *next* time point, but *not* for the indefinite future.

Sterkenburg does not clearly distinguish between these two versions of his objection. For example, when he writes that meta-induction cannot show that object-induction "is a good procedure to follow in general", it remains unclear whether he means this in the sense of (i) or (ii). Yet the distinction between the two versions of the objection is highly important.

We think that objection (i) can be easily remedied. With "for now" it is meant that the justification is relative to the *present* evidence. What has to be done to deal with this objection is to recall the central distinction between an a priori and an a posteriori justification. What Sterkenburg's point (i) accentuates is that the meta-inductive justification of an object-level method is always a posteriori, relative to the given track record of the competing methods for given prediction targets, and could be overruled by future evidence. Sterkenburg is right that an a posteriori justification of object-induction is weaker and more restricted than an a priori justification. But given the insight that an a priori justification of object-induction is impossible, the replacement of this unaccomplishable goal by the attainable goal of a non-circular a posteriori justification should not be seen as a disadvantage, but as epistemological progress. The a posteriori nature of the given justification of object-induction is not different from the confirmation of *scientific theories*, which, too, is always relative to the actual evidence and can be overthrown by future evidence. In footnote 2 of his article, Sterkenburg suggests that this analogy is problematic, but I cannot see why, because theories can of course figure as prediction methods and be meta-inductively evaluated by their predictive success.

In conclusion, the "for now" argument seems to have an easy treatment. In contrast, objection (ii) constitutes a genuine challenge for the meta-induction project. This objection says that even conditional on the actual success records that favor object-induction, all that meta-induction can justify is our belief in the rationality of object-induction in application to the *next* time point, but not for the infinite future. For prediction games as standardly presented, this objection seems to apply, because the prediction or hypothesis recommended by meta-induction concerns only the next time point.

To defeat this objection we first have to ask: what would the recommendation of a prediction method M, conditional on the *present* evidence, for all future time points precisely mean? Clearly not that M is the recommended method at all future time points, as this would amount to an a priori justification. It can only mean that M is the *presently* recommended method for *predictions* of events *at all* *future* time points. Showing this requires the extension of the notion of a prediction game from predictions of next events to predictions of events in the distant future. We will see soon that this is indeed possible. So what we have to show, to rebut Sterkenburg's second objection, is that in sufficiently induction-friendly environments meta-induction is likely to favor an object-inductive prediction method that is *intrinsically general* in that it applies not only to the next event but equally to events in the distant future. For example, the Carnapian λ-rules are intrinsically general in this sense; our formal explication of this generality will be the principle of exchangeability (see Sect. 3).

The next section is devoted to the development of a method by which a meta-inductive a posteriori justification of inductive generalizations is indeed possible. Before we get to this we have to clarify a more fundamental epistemological question, concerning the relation between the *justification of an epistemic method* and the *justification of the beliefs* recommended by this method. In each round, meta-induction gives us an a posteriori justification of a particular method (if it accumulates nearly all weight) or of a combination of methods, because by following the meta-inductively recommended method we optimize predictive success. This does not automatically imply, however, that we are justified in believing the individual *predictions* recommended by that method, for reasons that will be immediately explained. Sterkenburg does not consider this problem, but it comes up when he argues that meta-induction should give us reason to *believe* that induction is successful. If this is interpreted as having reason for believing that inductive predictions are probably true, then this is not granted but requires more argumentation. There are two major reasons for why the justification of a method may not be sufficient to justify the belief in its recommended predictions:

(1.) The first reason pertains to *qualitative* yes-or-no predictions, e.g., whether it will rain tomorrow or not. In some environments a meta-inductively optimal method may be only slightly better than random guessing, and its success probability may be too small to adopt its prediction as a qualitative belief and base one's actions on it. In this case a rational utility maximizer will still apply the optimal method but *suspend judgement* in regard to the predictions's truth value. What can at most be reasonably inferred from a meta-inductive event-prediction is that it is more probable than all competing event-predictions. This is a rather weak belief. What is really important is a good estimation of the *numerical* probabilities of the possible events. Fortunately, this can be provided by meta-induction, namely by applying meta-induction to the predictions of probabilities. This is called a *probabilistic prediction game* (Schurz, 2019, sec. 7.1; Sterkenburg, 2020). Here, the candidate methods predict probability distributions over the possible values of the next event, conditional on the past events. The deviation of the predicted probabilities from the event value that has been actually realized is scored by a *proper* scoring function such as the quadratic loss (which is assumed by us). It is well-known that under proper scoring, forecasters will maximize their average score if they predict the objectively correct probabilities. The meta-inductivist predicts a weighted average of the distributions in the candidate pool $\mathcal{C}$, with weights defined as in (1) of sec. 1. Also Douven (2023) and Sterkenburg (forthcoming) develop their accounts in the framework of probabilistic games and the remainder of this paper will focus on probabilistic games.

One advantage of meta-inductive probability aggregation over Bayesian learning by conditionalization lies in the fact that is not restricted to a particular class of prior distributions, but permits the inclusion of any prior distribution in one's candidate pool. What is more important: the optimality theorem of Sect. 1 applies equally to probabilistic prediction games. For these games it yields the result that meta-induction provides an optimal probability distribution over the next event(s) conditional on the evidence about past events (Schurz, 2019, memo (7.2)). Since in probabilistic games, the methods are given as probability distributions, it seems that the a posteriori justification of a method and that of a particular degree of belief coincide in this case, and therefore the problem of the transition from the a posteriori justification of a method to the justification of a degree of belief is solved.

(2.) Even the latter conclusion is not generally warranted, but only if the optimal probabilities are about observable events that are relevant to our success in actions, so that rational utility maximizers must implicitly adopt these optimal probabilities in order to maximize their expected utility. This becomes different if the prediction method is theory-based and involves *theoretical* (i.e., non-observable) concepts or variables. The meta-inductive justification of a theory-based prediction method justifies our belief in the empirical predictions of this method, but not necessarily our belief in its theoretical superstructure, as this superstructure is not directly practically relevant. For example, if I use the empirical predictions of Newton's laws to predict the trajectory of a planet, then (by the arguments below) I implicitly adopt the belief in these empirical predictions, but I need not necessarily believe in the reality of gravitational forces. Thus, meta-induction leaves room for the *suspension of judgement* in regard to theoretical claims. This corresponds to the instrumentalistic position in philosophy of science, exemplified for example by van Fraassen (1989), according to which we are warranted to believe in the *empirical adequacy* of scientific theories (their past and future predictive success), but not in their realistic truth. This does not mean that scientific theory realism is unjustifiable; it only means that meta-induction by itself does not give us this justification. The transition from the empirical adequacy of a theory to its theoretical truth does not correspond to a (meta-) inductive, but to an *abductive* inference (or inference to the best explanation). Its justification requires stronger epistemological assumptions that cannot be discussed here (Schurz, 2022, sec. 5.2); only in Sect. 4, where we discuss the limitations of our approach, the abductive justification of theoretical assumptions will become relevant.

We argue that at least for beliefs whose truth-probabilities are immediately relevant to our success in actions, the suspension of judgement is *not* an alternative because we are *forced* to act in some way, and as rational utility maximizers we will act *as if* we have certain degrees of belief, namely those degrees of beliefs from which we think they maximize predictive success and expected utility. Therefore we propose the following optimality principle that justifies the transition of the justification of a method to the justification of the beliefs recommended by it:

(3) *Optimality principle:* If probabilistic meta-induction recommends a probability distribution P_{MI} as optimal in the class $\mathcal{C}$ of accessible probabilistic methods and a rational utility maximizer X is practically forced to act *as if* she prefers one of the distributions in $\mathcal{C}$, then it is rational for X to adopt P_{MI} as her degrees of belief.

The reason behind (3) is that we regard it as an a priori requirement of *cognitive coherence* that our explicit (degrees of) beliefs should agree with the implicit 'as-if' beliefs that are embodied in our actions. The program of retrieving a person's implicit degrees of belief from her actions and utilities is prominently realized in the idea of defining rational degrees of beliefs in terms of fair betting quotients (going back to Ramsey and de Finetti). Standard Bayesians tend to *identify* a person's degrees of beliefs with these implicit degrees of beliefs. Under this assumption the above-mentioned principle of cognitive coherence becomes analytically true. For rational justifications, however, it is important that a person's degrees of beliefs are explicit in the sense of being *consciously accessible* by introspection, which need not be true for implicit degrees of belief. In fact, findings in cognitive psychology have discovered biases of people's self-assessment of their degrees of beliefs that point in the direction of overconfidence (Hoffrage, 2004). Therefore the coherence principle expresses an epistemologically important rationality condition.

The optimality principle shows us the principled way how a meta-inductive justification of inductive generalizations can work. We have to show two things: (1.) In suitably induction-friendly environment the chances are high that the meta-inductively recommended probability distribution P_{MI} is intrinsically general, in the sense of being (approximately) exchangeable, which means roughly speaking that P_{MI} is applicable to arbitrary future time points conditional on arbitrary sequences of past events. (2.) If P_{MI} turns out to be (approximately) exchangeable, then this does not rest on accidental features of the actual sequence of events or the chosen candidate pool of methods, so that it is practically not possible to reach the same success with a non-exchangeable method. This means that the utility maximizer must act as if his degrees of belief are exchangeable, which by principle (3) justifies him to believe in the principle of exchangeability. Points (1.) and (2.) will be worked out in the next section.

3 Meta-inductive Justification of Exchangeability

The most basic kind of an inductive-probabilistic generalization is de Finetti's (1937/64) principle of *exchangeability* (or symmetry, as Carnap, 1950, 434ff., called it). Exchangeability asserts that the probability distribution is invariant w.r.t. arbitrary permutations of the individual constants a_i or time points $i \in \mathbb{N}$ enumerating the individual events of the given sequence. We first introduce some technical notions needed for expressing this principle:

- P(−) ranges over epistemic probability functions (interpreted as rational degrees of beliefs), and $P_{i,n}$ is the probability distribution predicted by method no. i at time n. We write P_i (instead of M_i) for a probabilistic prediction method; thus $P_i \in \mathcal{C}$. Probabilities are expressed in the terminology of mathematical variables; we include the special case of a binary event variable expressed in logical notation (which for some is more easily understandable).

- The *prediction target* is represented by an *event variable* E that is formally a function from the *domain* of natural numbers $\mathbb{N} = \{1,2,\ldots\}$ representing individuals or time points into a value space Val, $E{:}\mathbb{N}\rightarrow \mathrm{Val}$, where $\mathrm{Val} = \{w_1,\ldots,w_q\}$ is a finite space of possible values of E. For example, E may be the event variable "weather conditions" with value space {sunny, cloudy, rainy, snowy}.
- $E_i =_{def} E(i)$ denotes the result of the realization of E at time point (or individual) i; thus $E_i \in \mathrm{Val}$. A binary event variable is denoted as $\pm E$ with value space $\mathrm{Val} = \{1,0\}$, where 1 designates E and 0 $\neg E$; so the realizations of $\pm E$ expressed in a logical language are $\pm E_i \in \{Ea_i, \neg Ea_i\}$.
- e_i is the *actual* event value at time i and $P(e_i)$ abbreviates $P(E_i{=}e_i)$, i.e. the predicted probability that the outcome of E's realization at time i is the particular value that actually occurred. The comma stands for conjunction; so $P(e_1,\ldots,e_n)$ stands for $P(E_1{=}e_1 \wedge \ldots \wedge E_n{=}e_n)$.
- Similarly, $P(E_i = v)$ is the probability that the outcome of E's realization at time i is the particular value v, where v and v_i are *variables* ranging over arbitrary event values in Val. $P(E_i)$ denotes a distribution $P(E_i{=}v)$ over all possible values $v \in \mathrm{Val}$.

Using this terminology, the prediction $pred_{n+1}(P_i)$ of each method $P_i \in \mathcal{C}$ delivered in round n is a probability distribution

$$P_{i,n}(E_{n+1} | e_1, \ldots, e_n),$$

conditional on the actual event values $e_1,\ldots,e_n$ and possibly on further method-specific evidence that is left implicit. Attractivity-based meta-induction predicts the weighted average of these probability distributions.

$$P_{MI,n}(E_{n+1} | e_1, \ldots, e_n) = \Sigma_{1\le i\le m} w_n\left(P_i\right) \cdot P_{i,n}(E_{n+1} | e_1, \ldots, e_n)$$

The principle of exchangeability can now be expressed as follows:

(4) *Exchangeability* of a distribution P for a given event variable E (conditionalized version):

For every $n \in \mathbb{N}$, $v_1,\ldots,v_k \in \mathrm{Val}(E)^{n+1}$and permutation $\pi{:}\mathbb{N}\rightarrow\mathbb{N}$ permuting finitely many individual indices:

$P(E_{n+1}{=}v_{n+1} \mid E_1{=}v_1,\ldots,E_n{=}v_n) = P(E_{\pi(n+1)}{=}v_{n+1} \mid E_{\pi(1)}{=}v_1,\ldots,E_{\pi(n)}{=}v_n)$.

In particular, for a binary event variable $\pm E$ in logical notation:

$P(Ea_{n+1} \mid \pm Ea_1 \wedge \ldots \wedge \pm Ea_n) = P(Ea_{\pi(n+1)} \mid \pm Ea_{\pi(1)} \wedge \ldots \wedge \pm Ea_{\pi(n)})$,

for all $\pm Ea_i \in \{Ea_i, \neg Ea_i\}$.

Exchangeable epistemic probabilities of events depend only on their observed frequencies, but not on the order of the observed events. For example, $P(Ea_4 \mid Ea_1 \wedge \neg Ea_2 \wedge Ea_3) = P(Ea_7 \mid Ea_2 \wedge \neg Ea_1 \wedge Ea_4) =$ P(an unobserved individual is E | two out of three observed individuals were E). Informally speaking this entails the *inductive* assumption that the probabilistic tendencies of individuals do not depend on their mere location in space and time. Exchangeable epistemic probabilities are adequate for event sequences resulting from independent repetitions of some physical process, called 'random experiment', whose unknown *statistical* probabilities (expressed by lower-case p) are IID (independently identically distributed) and, therefore, satisfy the product rule, $p(v_1,v_2) = p(v_1) \cdot p(v_2)$ (Gillies, 2000, 71, 77). By de Finetti's famous representation theorem, an exchangeable (conditional) epistemic probability function can be identified with an *expectation value* of *statistical* probabilities (p's), whose weights are the posteriors of the possible p's (cf. Spielman, 1976; Schurz, 2019, prop. 4.2).

We argue that belief in the exchangeability of P_{MI} is meta-inductively justified if three conditions are satisfied that are significantly stronger than the meta-inductive justification conditions for single event predictions. These conditions are:

1. *Approximate P_{MI}-exchangeability:* Conditional on the actual track record, the meta-inductively optimal a posteriori distribution $P_{MI,n}$ satisfies exchangeability for the given event variable E, at least approximately, for all prediction games that have been performed.
2. *Sufficient track record:* Sufficiently many probabilistic prediction games with prediction target E, applied to varying sequences of E-events, have been performed, to exclude the practical possibility of designing exchangeability or non-exchangeability by artificial means.
3. *Minimal richness:* The candidate pool contains at least (i) a basic object-inductive rule (straight rule or a Carnapian λ rule with small λ), which (given a sufficient track record) will detect an objective IID distribution if it is there, and (ii) for every event value w ∈ Val(E) a method P^w predicting constantly a high probability of w, which prevents that P_{MI} can be exchangeable if the event sequence is not governed by statistical IID probabilities.

The a posteriori justification of our belief in the exchangeability of $P_{MI,n}$ given these three conditions proceeds as indicated above (where with "exchangeability" I mean "approximate exchangeability"): By 1. we are entitled to employ the exchangeable distribution P_{MI} and because of conditions 2. and 3. it is not practically possible for the utility maximizer to reach the same success with a non-exchangeable method; so by the optimality principle (3) her belief in the exchangeability of P_{MI} is justified. In what follows the three conditions are motivated in more detail.

3.1 Condition 1: Approximate P_{MI}-Exchangeability

Of course it cannot be expected that after a finite amount of evidence P_{MI} is *precisely* exchangeable. Therefore we merely require *approximate* exchangeability, meaning that the average difference between two permuted probabilities is smaller than a small pragmatically fixed approximation threshold ε:

(5) *Approximate exchangeability* of P for event variable E:
For all permutations $\pi:\mathbb{N}\rightarrow\mathbb{N}$, $n \in \mathbb{N}$ and $v_1,\ldots,v_{n+1} \in Val(E)^{n+1}$:$|P(E_{n+1}=v_{n+1} \mid E_1=v_1,\ldots,E_n=v_n) - P(E_{\pi(n+1)}=v_{n+1} \mid E_{\pi(1)}=v_1,\ldots,E_{\pi(n)}=v_n)| \leq \varepsilon$.

Since P_{MI} is a weighted average of candidate distributions, P_{MI} will be exchangeable in a robust weight-invariant sense if and only if all candidate distributions in $\mathcal{C}$ of significant weight are themselves exchangeable. The if-direction of this claim is proved in theorem (11) below. Concerning the only-if direction, it may of course happen that the weights of two non-exchangeable distributions P_1 and P_2 are just so adjusted that P_{MI} becomes exchangeable, but this fine-tuning breaks down for minimal weight-variations. For example, assume a non-exchangeable P_1 with $P_{1,n}(Ea_{n+1}|S)=0.1$ conditional on a sequence S of n ±E-events and $P_{1,n}(Ea_{\pi(n+1)}|\pi S)=0.9$ conditional on the permuted sequence πS, and a second non-exchangeable distribution P_2 with $P_{2,n}(Ea_{n+1}|S)=0.9$ but $P_{2,n}(Ea_{\pi(n+1)}|\pi S)=0.1$. Then with weights $w_n(P_1)=w_n(P_2)=0.5$ $P_{MI,n}$ will come out as exchangeable, but small weight deviations will make $P_{MI,n}$ non-exchangeable.

As explained, a successful probability distribution can only be expected to be robustly exchangeable for event-sequences whose finite frequencies converge to statistical IID probabilities. The question is how this property of event sequences can be recognized by meta-induction. This is the point where the two further conditions come into play.

3.2 Condition 2: Sufficient Track Record

In a standard probabilistic prediction game, all that is meta-inductively determined are MI's probabilities of the possible values (v) of the *next* event conditional on the sequence of *actual* past events, i.e.

(6) $P_{MI,n}(E_{n+1}|e_1,\ldots,e_n)=\Sigma_{1\leq i\leq m}w_n(P_i)\cdot P_{i,n}(E_{n+1}|e_1,\ldots,e_n)$, for n = the present time.

However, the exchangeability condition applies to probabilistic predictions of arbitrary future events E_{n+m}, $P_{MI,n}(E_{n+m}|e_1,\ldots,e_n)$; so we have to devise games for predictions in the more distant future. Moreover, the exchangeability condition refers to predictions conditional on arbitrary *possible* event sequences of the form $E_1=v_1,\ldots,E_n=v_n$ and it requires the invariance of P_{MI} for all event-reorderings. If P_{MI} would be determined only conditional on the actual event sequence, P_{MI} could be artificially *made* exchangeable, just by extending P_{MI} to all *non-tested* event-sequences in a way so that exchangeability holds. More importantly, by a similarly trivial definitional extension P_{MI} could easily be made non-exchangeable. For example, assume $\mathcal{C}$ contains a successful exchangeable probabilistic method P (e.g., straight rule) and $v_1,\ldots,v_n$ is a sequence of event values that permutes the sequence of actual values $e_1,\ldots,e_n$, i.e., preserves their frequencies. Then we could add a non-exchangeable distribution $P_{non\text{-}ex}$ to $\mathcal{C}$ that predicts deviantly just for this sequence and predicts otherwise like P. Assuming that P and $P_{non\text{-}ex}$ are the only methods of significant weight, the meta-inductive aggregation of the two distributions will be non-exchangeable, too. But as soon as we perform another prediction game with a realization of the permuted event sequence $v_1,\ldots,v_n$, then assuming IID conditions the success of $P_{non\text{-}ex}$ will disappear and P_{MI}'s exchangeability is restored. Surely one could go on in artificially designing non-exchangeable variants of P for further untested event sequences, but the more event sequences have been tested, the more complicated and arbitrary these piecewise concatenated distributions will become, so that for reasons of cognitive economy their inclusion in the candidate pool will be unreasonable and sooner or later even unfeasible.

In conclusion, to block the practical possibility of making distributions exchangeable or non-exchangeable by artificial design, we have to perform not only one, but *many* prediction games, generating meta-inductive probabilities $P_{MI,n}(E_{n+m}|E_1=v_1, \ldots E_n=v_n)$ not only for the actual event sequence $e_1,\ldots,e_n$, but for various further sequences $v_1,\ldots v_n$, and not only up to the *next* time n+1 but up to arbitrary time points n+m in the distant future. At this point an important *generalization* of prediction games has to be introduced. The optimality theorem for MI relies merely on the fact that all methods in $\mathcal{C}$ are applied to the same sequence of prediction tasks and are scored by a proper loss function that determines MI's weights and predictions; the particular format of these tasks is inessential. For example, instead of predictions of the next event, the prediction task may consist in predictions of events lying in the distant future, or in predictions of finite sequences of future events, or of event-averages in samples of future events, which corresponds to the standard method of training and test samples (Schurz, 2019, sec. 7.4). To test for exchangeability, the weights of the candidate methods are determined separately in each of these varied prediction games, so they need not be the same in these games, although as will be argued below, under IID conditions the weight of all successful methods will be approximately the same.

Note that in extended prediction games, the time points of the events and of the rounds of the game need not coincide: in round n probabilities of the form $P_{i,n}(e_{j_{n+1}}|e_{j_1}, \ldots, e_{j_n})$ are predicted. Even prediction tasks of the form $P(e_3|e_4, e_5, e_2, e_1)$ are admitted, in which an event is post-facto predicted by events lying in its past and in its future; these tasks are called *interpolations*.

We argue that belief in the approximate exchangeability of P_{MI} is meta-inductively justified only if MI's optimal probabilities in prediction games with permuted but frequency-invariant event-sequences are ε-approximately equal. One may object that prediction games for distant futures are practically impossible, because their iteration would take too long. This objection can be defeated by a further generalization: the 'predicted' events need not actually lie in the future, they may also lie in the past. All that is required is that the predicted events are *use-novel* in the sense of Worrall (2006), which means that their knowledge has not been used in the design of the methods. Thus, "predictions" may be applied *post-facto* to known events in the past (e_{n-m}, …,e_{n-k} for $k < m < n$), by 'predicting' past events based on the values of their predecessor events, provided the methods are defined independently from knowledge of the 'predicted' events (cf. Thorn & Schurz, 2020 for post-facto prediction games based on real-world data). The difference between the time of the predicted event and that of the latest conditionalizing event is called the *prediction interval*. By post-facto predictions, prediction intervals may be extended to time spans as long as the entire past for which we possess information about the prediction target. Examples of long-term scientific post-facto predictions are the predictions of ice ages (occurring in intervals of about 100,000 years), volcano eruptions (in intervals of around 50 years) or climatic changes.

A possible objection against post-facto predictions could point out we can only acquire knowledge about past events by employing induction.[5] Is this circular? To infer the temperature on our earth some hundred thousands of years ago, scientists use empirical laws correlating present observations with the temperature at that time. One example is the empirical correlation between the temperature at that time and the oxygen isotope composition of Greenland ice layers from that historical age (cf. Petit et al., 1999). The latter laws have been justified by (object-) inductive methods applied to samples of actual observations, where these (object-) inductive methods are in turn justified by meta-induction applied to their success records. This is not circular, because the a posteriori justification of an inductive-probabilistic method is relative to the prediction target(s) for which its tracks record has been determined. The inductive generalizations that have been justified in experiments about the predictability of temperatures from the oxygen isotope composition of preserved layers of ice are used, in a second step, to justify the predictive success of long-term post-facto predictions of climatic changes.

Our opponent could push his objection further and argue that even with post-facto predictions the exchangeability of P_{MI} can only be tested for events within prediction intervals that are within one's epistemic reach. We call these intervals predictively *assessable*. They may cover long time spans, for some prediction targets even as long as the life time of our universe (15 billion years), but *not* infinitely long, because the infinite future is not predictively assessable. In what follows we call generalizations over predictively assessable time span *weak* generalizations, as opposed to *strong* generalizations that are claimed to hold for the whole infinite future. Since weak and strong generalizations are predictively equally successful, we have to admit that the justification of the transition from a weak to a strong generalization goes beyond what meta-induction over success rates can give us. One obvious possibility of justifying this transition is by reasons of *simplicity*; this question will be taken up in Sect. 4.

Conditions 1 and 2 specify the track record of prediction methods required for justifying the belief that P_{MI} is approximately exchangeable and, thus, that the event

[5] This was pointed out by a referee.

sequence is governed by an IID. Critics might object that even under IID conditions these conditions need not obtain for certain 'degenerated' candidate pools. At this point our third condition steps in.

3.3 Condition 3: minimal richness of $\mathcal{C}$

If the candidate pool $\mathcal{C}$ is not minimally rich, then even if the event sequence is IID, a non-exchangeable candidate method may predict better than an exchangeable one, with the result that P_{MI} will also come out as non-exchangeable (we say "exchangeable" short for "approximately-exchangeable"). For example, assume the objective probability of a binary event, p(1), is 0.6, and $\mathcal{C}$ contains two methods, P_1 being exchangeable and predicting constantly $P(1)=0.8$, and P_2 being non-exchangeable and predicting $P(1)=0.7$ on even times and $P(1)=0.5$ on odd times. Then P_2 will perform better and attain a higher weight than P_1, whence also P_{MI} will be non-exchangeable, although the event sequence is in fact IID. Therefore certain methods that figure as indicators of IID- respectively non-IID conditions have to be included in the pool, as follows:

(1.) Some basic object-inductive probabilistic method has to be included in $\mathcal{C}$ that detects that the event sequence is IID (if it is IID) by maximally exploiting it. For simplicity we assume that this is the straight rule, P_{st}, that transfers the observed frequency to the predicted instance. The long-run optimality of P_{st} for IID sequences drives the weights of all inferior and in particular of all non-exchangeable methods to low values. Informally this is seen as follows: Assume a binary event sequence governed by an objective IID probability $p(1)=p$. It is well-known that the method P_p that constantly predicts p minimizes the expected cumulative (squared) loss among all (non-clairvoyant) methods. The standard error of the frequency after n rounds compared to p is $\sqrt{p \cdot (1-p)/n}$. So P_{st}'s expected (squared) per-round regret compared to P_p is $p \cdot (1-p)/n$, which quickly converges to zero. Every prediction method P* that differs from P_{st} for a non-negligible share of times experiences an additional loss compared to P_{st} that accumulates with n; in particular this must hold for a non-exchangeable P*. Since weights are negative exponentials of the cumulative losses (cf. fn. 3), P*'s expected weight becomes small for increasing n. This is cashed out in the following theorem, in which $\mathbf{E}(w_n(P))$ denotes the *expected weight* of method P as calculated from its expected cumulative loss:

(10) *Theorem:* Assume a binary event sequence $e_1, e_2, \ldots$ governed by an objective IID probability $p(E=1)=p$. Let P* be a suboptimal prediction method that predicts $p+\delta$ in q% of times and predicts like P_{st} otherwise. Then the ratio between the expected weights of P_{st} and P* after n rounds is
$\mathbf{E}(w_n(P_{st}))/\mathbf{E}(w_n(P^*)) \geq \exp(\eta \cdot q \cdot [n \cdot (\delta^2 + 2 \cdot \delta \cdot p \cdot (1-p)) - \log_2(n+1) \cdot p \cdot (1-p)])$,
with $\eta = \sqrt{8 \cdot \ln(m)/(n+1)}$ the constant in (1).

Example: If $\delta=0.2$, $q=40\%$, $p=0.7$, $m=100$, and $n=1000$,
then $\mathbf{E}(w_n(P_{st})) / \mathbf{E}(w_n(P^*)) \geq 11{,}638$.

Proof See appendix.

In conclusion, under IID conditions only methods that predict values that are close to the true statistical probability will attain significant weight for increasing n. All these methods are approximately exchangeable and their weights are approximately equal over all tested event sequences. This implies that also $\mathrm{P}_{\mathrm{MI,n}}$ will be approximately exchangeable, being exposed to a small additional approximation loss as computed in theorem (11) below. In this theorem we abbreviate, for any given sequence of event values $\mathrm{v}_1,\ldots,\mathrm{v}_{\mathrm{n}+1}\in \mathrm{Val}^{\mathrm{n}+1}$ and permutation function π, the absolute difference between the probabilities for the sequence and the permuted sequence predicted by method P as:

$$\Delta_{\mathrm{P,n},\pi} =_{def} |\mathrm{P}_{\mathrm{n}}(\mathrm{E}_{\mathrm{n}+1}=\mathrm{v}_{\mathrm{n}+1}|E_1=\mathrm{v}_1,\ldots,\mathrm{E}_{\mathrm{n}}=\mathrm{v}_{\mathrm{n}}) - \mathrm{P}_{\mathrm{n}}(E_{\pi(\mathrm{n}+1)}=\mathrm{v}_{\pi(\mathrm{n}+1)}|\mathrm{E}_{\pi(1)}=v_1,\ldots,\mathrm{E}_{\pi(n)}=\mathrm{v}_{\mathrm{n}})|.$$

Similarly for $\Delta_{\mathrm{MI,n},\pi}$. We abbreviate $\mathrm{P}_{\mathrm{n}}(\mathrm{E}_{\mathrm{n}+1}=\mathrm{v}_{\mathrm{n}+1}|\mathrm{E}_1=\mathrm{v}_1,\ldots,\mathrm{E}_{\mathrm{n}}=\mathrm{v}_{\mathrm{n}})$ as $\mathrm{P}(\boldsymbol{v})$ and likewise the permuted prediction as $\mathrm{P}(\pi(\boldsymbol{v}))$.

(11) *Theorem:* Let $\boldsymbol{S}(\mathrm{n})$ be the set of candidate methods P that are ε-approximately exchangeable at time n, $1-\varepsilon_1$ their weight sum for their prediction of $\mathrm{P}(\boldsymbol{v})$, and ε_2 an upper bound of the sum of their (unsigned) weight-differences, $\Sigma_{\mathrm{P}\in\boldsymbol{S}(\mathrm{n})}$ $|\mathrm{w}_{\mathrm{n}}(\mathrm{P})-\mathrm{w}_{\mathrm{n},\pi}(\mathrm{P})|$, between the weights for the prediction $\mathrm{P}(\boldsymbol{v})$ and the permuted prediction $\mathrm{P}(\pi(\boldsymbol{v}))$. Then:

$$\Delta_{\mathrm{MI,n},\pi}\leq \varepsilon_1+(1-\varepsilon_1)\cdot\varepsilon+ \varepsilon_2.$$

Proof See appendix.

(2.) On the other hand, if the event sequence is not IID, then there must be methods in $\mathcal{C}$ that can detect this non-exchangeability. This is the purpose of the methods P^{w} that constantly predict a high probability of the value w (for all w ∈ Val). To see how this works, assume again that the objective probability p(1) of a binary event $\mathrm{E}=1$ changes at time $\mathrm{n}+50$ from 0.1 to 0.9 (for similar arguments cf. Gillies, 2000, pp. 69–83). A realistic example of this sort is the increased probability of a tornado occurrence (per season) in Western Europe. $\mathrm{P}_{\mathrm{st,n}}$ would transfer the observed frequency indiscriminately to all future events $\mathrm{e}_{\mathrm{n}+\mathrm{m}}$ and would be moderately successful in games predicting the near future (small m) but unsuccessful in games predicting the distant future ($\mathrm{m}\geq 50$), but if P_{st} is the only method in $\mathcal{C}$, then $\mathrm{P}_{\mathrm{MI,n}}$ would nevertheless imitate $\mathrm{P}_{\mathrm{st,n}}$ and stay exchangeable. At this point the constant methods P^{w} come into play; in the binary case P^1 and P^0. The distribution $\mathrm{P}^1_{\mathrm{n}}$ would receive a low weight for predicting events in the near and a high weight for predicting events in the distant future, and vice versa for $\mathrm{P}^0_{\mathrm{n}}$. Based on the success rates of $\mathrm{P}_{\mathrm{st,n}}$, $\mathrm{P}^1_{\mathrm{n}}$ and $\mathrm{P}^0_{\mathrm{n}}$ for different prediction intervals, $\mathrm{P}_{\mathrm{MI,n}}$ would predict a value close to 0.1 for the near future and a value close to 0.9 for the distant future; so $\mathrm{P}_{\mathrm{MI,n}}$ would not be approximately exchangeable.

We emphasize that the minimal richness condition does not impose a dogmatic restriction, since arbitrary other methods may be included in the pool. The condition merely exploits mathematical results that grant to certain object-level methods certain success rates in particular types of environments. This guarantees that IID-probabilities will be detected if they are there; and similarly, deviations from them will be recognized if they are present.

In conclusion, we argue that if we perform many prediction games over varying event sequences and prediction intervals (condition 2) with minimally rich candidate pools (condition 3), and if in all these games the meta-inductivist's probability function P_{MI} comes out as approximately exchangeable (condition 1), then no further practically feasible way is open to achieve the same success with a non-exchangeable probabilistic method. So by the optimality principle (3) the epistemic agent is a posteriori justified to believe in the exchangeability of P_{MI}, relative to the presently available evidence.

Still more is possible. Recall that an exchangeable P is identical with an expectation value of statistical probabilities. So, if we are justified in believing that P_{MI} is exchangeable, we are also justified in believing that the predicted events are governed by an objective-statistical probability distribution, so that we can interpret P_{MI} as an optimal estimation of the *expected statistical probability* p_{MI}, i.e., $P_{MI}(E_{n+1}=v|e_1,\ldots,e_n)=p_{MI}(E_x=v)$, with "$E_x$" denoting the event at variable time points.

There will of course be domains and prediction targets whose frequencies do not converge to IID probabilities. In these domains a meta-inductive justification of exchangeability and expected statistical probabilities is not possible, although the meta-inductive optimality theorem for predictions still applies. However, the proposed account can be generalized from IID distributions to more complex forms of inductive regularities, such as Markov chains. In Markov chains of kth order, an event's probability depends on the k previous events. Markov chains are not exchangeable in the basic sense, but in a generalized sense, called *partial exchangeability*. The latter condition asserts that all blocks of $n \geq k$ consecutive events with the same starting event and the same transition frequencies between two events have the same probability (Diaconis & Friedman, 1980). An elaboration of our argument for these and other more complex forms of inductive generalizations is left to future work.

We now come to the final step: the same procedure can be applied to the success rates of prediction methods. Here we assume a prediction game in which the success rates of various prediction methods are predicted based on their track record. Provided the success frequencies are governed by statistical success probabilities, then by applying meta-induction to that game we obtain optimal estimations of the statistical reliabilities of methods that are turned into a posteriori justified beliefs about their statistical reliabilities by the optimality principle.

Does the last result contradict the distinguishing property of the meta-inductive approach that it can offer only an optimality justification, but not a reliability justification (as argued by Sterkenburg, forthcoming, in several places, e.g. in sec. 3.1, 6th §)? No, the contradiction is only apparent, because this distinguishing property concerns only the a priori part of the meta-inductive account—which is, of course, crucially relevant for the possibility of a non-circular justification. It is impossible to give an a priori justification of the reliability of meta-induction; the only possible a priori justification is an optimality justification. But based on this a priori optimality justification it is indeed possible, under certain observable conditions, to give an a posteriori justification of estimated statistical probabilities, and therefore also of expected statistical reliabilities.

4 Weak versus Strong Generalizations: From Meta-induction to Abduction

In the previous section we described the meta-inductive justification of weak inductive generalizations, in the form of exchangeability principles over predictively assessable time spans. In this section we discuss the transition from weak to *strong* generalizations, i.e., universal generalizations over the whole infinite future. Both weak and strong generalizations may be strict or probabilistic (see below). The justification of the transition from weak to strong generalizations requires epistemic principles that go beyond optimal predictive success, since their assessable predictive success is precisely the same. In this section we discuss epistemic principles that may warrant this transition, emphasizing that the justification of these principles goes beyond the scope of this paper.

Consider special science generalizations like the following:

(12) *Weak generalizations:*
All ravens are black.
95% of all birds can fly.

That all ravens are black is a frequently used example of a general law. But biologically informed persons know that this generalization does not express a law of nature (and likewise for the birds-can-fly example). Color mutations of ravens are possible. However, until now no color-mutated ravens are known (except for ravens suffering from leucism). So until now the generalization "All (non-leucistic) ravens are black" is unfalsified. Should we *interpret* this generalization as a weak or strong one? We think that most biologists would be careful and assess this generalization merely for the "practically foreseeable future", e.g. for the next 1000 or even 10,000 years. Different prediction intervals, taking different risks, are possible, but our main point is that when scientists assert a generalization like this, they usually do not mean that it holds for the infinite future, but merely for the predictively assessable future.

Prima facie, there are two epistemic criteria regulating the acceptability of generalizations: *safety* versus *simplicity*. The two criteria pull in opposite directions. Weak generalizations are probabilistically safer than strong ones, which speaks in favor of weak generalizations. But strong generalizations are simpler, because the precise formulation of the restriction of "assessable" future time spans is complicated and depends on various contextual factors, which speaks in favor of strong generalizations.

The simplicity of a generalization G may be measured by the shortest length of a formula expressing G in a given system of primitive symbols (cf. the criterion of "minimal description length" in machine learning; Grünwald, 2000). Simplicity is an *instrumentalistic* evaluation criterion that is prima facie unrelated to the truth chances of generalizations. It is an inverse measure of the *cognitive cost* of a prediction method that is not covered by standard meta-induction based on predictive success. One could attempt to include simplicity in the meta-inductive evaluation by adding a simplicity component to the scoring function. There are different possibilities of aggregating the predictive success score with a simplicity score. A simple way (proposed in Schurz, 2019, sec. 7.4) is to add an additional complexity term to the predictive loss of P_i, with a subsequent renormalization. Provided that the aggregated score is convex, the meta-inductive optimality theorem will still apply for this aggregated score. Another possibility would be to use simplicity as a *ceteris paribus* criterion, establishing a preference among predictively equally successful methods.

The justification of the preference for strong over weak generalizations by their greater simplicity is *instrumentalistic* in nature: it does not entail a justification of the truth-closeness of strong generalizations. The simplicity advantage is certainly not all what scientists mean when they asserts a strong generalization. Consider the following physical science generalizations:

(13) *Strong generalizations:* All salt dissolves in water (at normal temperatures).

All Cs^{137}-atoms decay with probability 0.5 within 30.12 years.

Why are physicists or chemists so sure that these generalizations are strongly universal? Because they are regarded as genuine *laws of nature*, or *physical necessities*. This justification is more than a mere reason of *simplicity*. It attempts to offer an *explanation* for the general truth, compressed in the physical necessity claim. In other words, this justification is based on an inference of *abduction*, or *inference to the best explanation* (IBE).

The justification of abductions or IBEs, if possible at all, is more difficult than the justification of inductive inferences and we cannot give here a comprehensive account of the involved problems. Strong generalizations that are abductively justified by claims of physical necessities are called *strongly lawlike*. The strong lawlikeness of generalizations figures as the explanation of their meta-inductively justified generality—which without this abductive step would merely have a weak interpretation.

Strongly lawlike generalizations are considered as *unbreakable* by (human or other) interventions. In contrast, weak generalizations are breakable by interventions, although they are clearly not just accidental generalizations such as "All apples in this basket are red". Rather, they are the effect of a combination of contingent conditions and lawlike mechanisms, whence we call them *weakly lawlike*. For example, that all ravens are black is caused by the fact that their color is the phenotypic expression of certain raven *genes* that cannot be easily altered. But they could be altered, and this is the difference to the solubility of salt in water, which can be derived from the application of fundamental physical equations to water and salt molecules.

Neglect of the difference between weak and strong lawlikeness has caused misunderstandings concerning the term “law of nature”. Those philosophers who argue that no genuine laws can be found in the special (non-physical) sciences (e.g. Earman & Roberts, 1999) understand laws in the strong sense, while those who argue that each special science has its own laws (e.g. Lange, 2009) understand laws in the weak sense.

The epistemic justifiability of abductive inferences to observation-transcending explanations (such as physical necessities) is controversial. Some philosophers reject these inferences throughout (e.g. van Fraassen, 1989), others accept abductions only in the natural sciences but not in metaphysics (e.g. Ladyman, 2012), while others advocate abductive inferences also in philosophy (e.g. Armstrong, 1983; Williamson, 2016). Schurz (2022, sec. 5.2) argues that theory-generating abductions are justified under restricted conditions: the abducted hypothesis must be instrumentally optimal and must appear as an isomorphic submodel in all of its equally successful competitors.

5 Conclusion

Let us summarize. What is possible is a meta-inductive justification of weak (inductive) generalizations based on the assessable predictive success records of the accessible methods. What is also possible is a meta-inductive justification of strong generalizations based on their instrumentalistic success records, aggregating predictive success and simplicity into one score. But what is not possible is a meta-inductive justification of strong generalizations in terms of their probable truth. This justification must be based on an abduction, explaining meta-inductively justified generalizations by their nature as strongly lawlike generalizations, or physical necessities.

In conclusion, the justification of strongly lawlike generalization requires an interplay of a (meta-) inductive and an abductive inference step. In this interplay, meta-induction has the task of providing an a posteriori justification of the explanandum of the abduction, which is a weakly lawlike inductive generalization over assessable prediction intervals. The inference to a strongly lawlike generalization proceeds by an abduction, whose epistemic justifiability is controversial, is more difficult than the justification of induction and lies beyond the scope of this paper.

Appendix

Proof of theorem (10) Assume the conditions of theorem (10).

(a) Under IID conditions, the predictor with minimal expected cumulative (squared) loss is the constant p-predictor P_p, whose expected (squared) loss in each round is $p\cdot(1-p)^2+(1-p)\cdot p^2$. So P_p's cumulative (squared) loss after n rounds is $\text{Loss}_n(P_p)=n\cdot[p\cdot(1-p)^2+(1-p)\cdot p^2]=n\cdot p\cdot(1-p)$.

(b) The expected deviation of P_{st}'s prediction from the true value p at time n (due to the deviation of the finite frequencies from the limit) is $\sqrt{p\cdot(1-p)/n}$, as explained in the text. So the expected squared surplus loss of P_{st} after n rounds (compared to $\text{Loss}_n(P_p)$) is given as $\Sigma_{1\le i\le n}p\cdot(1-p)\cdot(1/i)$, which is abbreviated as S.

(c) By (a) and (b), the expected cumulative loss of P_{st} is
$\mathbf{E}(\text{Loss}_n(P_{st}))=n\cdot p\cdot(1-p)+S$.

(d) The expected per-round loss of the constant $p-\delta$ predictor, $P_{p-\delta}$, compared to P_p, is $p\cdot((1-p+\delta)^2-(1-p)^2)+(1-p)\,((p-\delta)^2-p^2)=\delta^2+2\cdot\delta\cdot p\cdot(1-p)$. Since P* predicts like $P_{p-\delta}$ in q·100% of all times and otherwise like P_{st}, the expected cumulative loss of P* is given as
$\mathbf{E}(\text{Loss}_n(P^*))=(1-q)\cdot[n\cdot p\cdot(1-p)+S]+q\cdot[n\cdot(p\cdot(1-p)+\delta^2+2\cdot\delta\cdot p\cdot(1-p))]$.

(e) The difference between the expected losses of P* and P_{st} is computed as
$\mathbf{E}(\text{Loss}_n(P^*))-\mathbf{E}(\text{Loss}_n(P_{st})=$
$=-q\cdot[n\cdot p\cdot(1-p)+S]+q\cdot[n\cdot(p\cdot(1-p)+\delta^2+2\cdot\delta\cdot p\cdot(1-p))]$
$=q\cdot[n\cdot(\delta^2+2\cdot\delta\cdot p\cdot(1-p))-S]\ge q\cdot[n\cdot(\delta^2+2\cdot\delta\cdot p\cdot(1-p))-\log_2(n+1)\cdot p\cdot(1-p)]$,
because $S=_{def}p\cdot(1-p)\cdot\Sigma_{1\le i\le n}(1/i)$ and $\log_2(n+1)$ is an upper bound of $\Sigma_{1\le i\le n}(1/i)$.

(f) The weight ratio is given (by (2) in Sect. 1) as
$\mathbf{E}(w_n(P_{st}))/\mathbf{E}(w_n(P^*))=\exp(\eta\cdot(\mathbf{E}(\text{Loss}_n(P^*)-\mathbf{E}(\text{Loss}(P_{st}))$ (by fn. 3) which by inserting the inequality in (e) gives the result. Q.E.D.

Proof of (11) We assume (without restricting the assumptions) that $P_{MI,n}(\nu)\le P_{MI,n}(\pi(\nu))$. Then:

$\Delta_{MI,n,\pi}=_{def}P_{MI,n}(\nu)-P_{MI,n}(\pi(\nu)) \le \varepsilon_1+\Sigma_{P\in\mathcal{S}(n)}(w_n(P)\cdot P(\nu)-w_{n,\pi}(P)\cdot P(\pi(\nu))$,
since $\Sigma_{P\notin\mathcal{S}(n)}(w_n(P)\cdot P(\nu)-w_{n,\pi}(P)\cdot P(\pi(\nu)) < \Sigma_{P\notin\mathcal{S}(n)}w_n(P)=\varepsilon_1$,
$\le \varepsilon_1+\Sigma_{P\in\mathcal{S}(n)}w_n(P)\cdot(P(\nu)-P(\pi(\nu))+\Sigma_{P\in\mathcal{S}(n)}(w_n(P)-w_{n,\pi}(P))\cdot P(\nu)$,
since $\Sigma_i(a_i\cdot b_i-c_i\cdot d_i)=\Sigma_i a_i\cdot(b_i-d_i)+\Sigma_i(a_i-c_i)\cdot d_i$,
$\le \varepsilon_1+(1-\varepsilon_1)\cdot\varepsilon+\varepsilon_2$,
since $\Sigma_{P\in\mathcal{S}(n)}w_n(P)\cdot(P(\nu)-P(\pi(\nu)) \le \Sigma_{P\in\mathcal{S}(n)}w_n(P)\cdot|P(\nu)-P(\pi(\nu)|$
$\le \Sigma_{P\in\mathcal{S}(n)}w_n(P)\cdot\varepsilon=(1-\varepsilon_1)\cdot\varepsilon$,
and $\Sigma_{P\in\mathcal{S}(n)}(w_n(P)-w_{n,\pi}(P))\cdot P(\pi(\nu)) \le \Sigma_{P\in\mathcal{S}(n)}|w_n(P)-w_{n,\pi}(P)|\cdot P(\pi(\nu))$
$\le \Sigma_{P\in\mathcal{S}(n)}|w_n(P)-w_{n,\pi}(P)| = \varepsilon_2$. Q.E.D.

Acknowledgements For valuable help I am indebted to Igor Douven, Tom Sterkenburg, Tomoji Shogenji, Leah Henderson, Peter Brössel and Brian J. Pitts.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the DFG (Deutsche Forschungsgemeinschaft), research unit FOR 2495.

Declarations

Conflict of Interest No conflict of interest.

References

Armstrong, D. M. (1983). *What is a law of nature?* Cambridge University Press.

Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge University Press.

De Finetti, B. (1937/64). Foresight, its logical laws, its subjective sources. In H. Kyburg & H. Smokler (eds.), *Studies in subjective probability* (pp. 93–159). John Wiley 1964.

Diaconis, P., & Friedman, D. (1980). De Finetti's theorem for Markov chains. *The Annals of Probability, 8*, 115–130.

Douven, I. (2011). Abduction. In *Stanford encyclopedia of philosophy*. http://plato.stanford.edu/entries/abduction/

Douven, I. (2023). Explaining the success of induction. *The British Journal for the Philosophy of Science*, 74, 381 – 404 https://doi.org/10.1086/714796

Earman, J., & Roberts, J. (1999). Ceteris paribus, there Is no problem of provisos. *Synthese, 118*, 439–478.

Gillies, D. (2000). *Philosophical theories of probability*. Routledge.

Grünwald, P. (2000). Model selection based on minimal description length. *Journal of Mathematical Psychology, 44*, 133–152.

Harman, G. (1965). The inference to the best explanation. *Philosophical Review, 74*, 88–95.

Hoffrage, U. (2004). Overconfidence. In R. Pohl (Ed.), *Cognitive illusions* (pp. 235–254). Psychology Press.

Ladyman, J. (2012). Science, metaphysics and method. *Philosophical Studies, 160*(1), 31–51.

Lange, M. (2009). *Laws and lawmakers*. Oxford University Press.

Petit, J.-R., et al. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature, 399*, 429–436.

Pitts, J. B. (2023). Does meta-induction justify induction: Or maybe something else? *Journal for the General Philosophy of Science*, 54(3), 393-419. https://doi.org/10.1007/s10838-022-09620-7/

Reichenbach, H. (1949). *The theory of probability*. University of California Press.

Salmon, W. C. (1957). Should we attempt to justify induction? *Philosophical Studies, 8*(3), 45–47.

Schurz, G. (2019). *Hume's problem solved: The optimality of meta-induction*. MIT Press.

Schurz, G. (2021). Reichenbach's best alternative account to the problem of induction. *Synthese, 99*, 10827–10838.

Schurz, G. (2022). Optimality justifications and the optimality principle: New tools for foundation-theoretic epistemology. *Noûs, 56*, 972–999.

Schurz, G. (2023). In search for optimal methods: New insights about meta-induction. *Journal for the General Philosophy of Science*, 54(3), 491–521. https://doi.org/10.1007/s10838-023-09649-2

Skyrms, B. (1975). *Choice and chance* (4th ed.). Dickenson.

Spielman, S. (1976). Exchangeability and the certainty of objective randomness. *Journal of Philosophical Logic, 5*, 399–406.

Sterkenburg, T. (2020). The meta-inductive justification of induction. *Episteme, 17*, 519–541.

Sterkenburg, T. (forthcoming). On explaining the success of induction. *The British Journal for the Philosophy of Science*. https://doi.org/10.1086/717068/

Thorn, P., & Schurz, G. (2020). Meta-inductive prediction based on attractivity weighting: An empirical performance evaluation. *Journal of Mathematical Psychology, 89*, 13–30.

Van Fraassen, B. (1989). *Laws and symmetry*. Clarendon Press.

Williamson, T. (2016). Abductive philosophy. *The Philosophical Forum, 47*, 263–280.

Worrall, J. (2006). Theory-confirmation and history. In C. Cheyne & J. Worrall (Eds.), *Rationality and reality* (pp. 31–61). Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.