

# ON EXPLAINING THE SUCCESS OF INDUCTION

TOM F. STERKENBURG

ABSTRACT. Douven (2021) observes that Schurz’s meta-inductive justification of induction cannot explain the great empirical success of induction, and offers an explanation based on computer simulations of the social and evolutionary development of our inductive practices. In this paper, I argue that Douven’s account does not address the explanatory question that Schurz’s argument leaves open, and that the assumption of the environment’s induction-friendliness that is inherent to Douven’s simulations is not justified by Schurz’s argument.

## 1. INTRODUCTION

Douven (2021) offers an explanation of the success of induction to complement Schurz’s (2008; 2019) meta-inductive justification of induction. Schurz’s proposal is a refinement of Reichenbach’s “pragmatic” justification (see Salmon, 1991), and builds on results from the machine learning theory of *competitive online learning* or *prediction with expert advice* (see Cesa-Bianchi and Lugosi, 2006). Schurz argues that the provable optimality of *meta-induction*, a specific kind of success-based induction at the level of competing methods, also confers justification to *object-induction*, induction at the level of events, in virtue of the latter’s observed empirical success. Douven accepts Schurz’s argument, but points out that it leaves unexplained this empirical success, and sets out to offer such an explanation. Douven’s leading idea is that we must take account both of the social interaction and the evolutionary development of inductive reasoners, and this idea he backs up with computer simulations. Specifically, he embeds Carnap’s (1952) inductive methods in the Hegselmann-Krause (2002) model of opinion dynamics to model communities of interactive agents, and employs an evolutionary algorithm to simulate how selection on the basis of inductive performance quickly leads to communities of more inductively successful agents.

In this paper, I argue that Douven misdiagnoses the extent to which Schurz’s argument leaves the success of induction unexplained, and that as a consequence, his proposal does not suffice to remedy this. In particular, Douven’s simulations incorporate an assumption of induction-friendliness of the environment; and this assumption is not, as Douven suggests in passing, justified by Schurz’s argument.

In section 2, I summarize Schurz’s proposed justification (sect. 2.1) and exhibit two relevant limitations (sect. 2.2). In section 3, I discuss Douven’s proposed explanation, and argue that it falls short of complementing Schurz’s argument. I conclude in section 4. Finally, in an appendix I collect some formal performance guarantees that support my observations about the simulations’ inherent induction-friendliness.

---

*Date:* March 23, 2021.

This is a preliminary version. I welcome feedback.

## 2. SCHURZ’S META-INDUCTIVE JUSTIFICATION OF INDUCTION

**2.1. The argument.** I here give a very brief and non-technical summary of Schurz’s argument (for details, see [Schurz, 2019](#); [Sterkenburg, 2020](#)). The argument infers from, first, the *analytical* optimality of *meta-induction* (sect. 2.1.1) and, second, the *empirical* success of *object-induction* (sect. 2.1.2), a justification for the latter (sect. 2.1.3).

**2.1.1. Step one: the analytical optimality of meta-induction.** Schurz’s argument is formulated within the formal framework of *sequential prediction*. In a *prediction game*, we have a pool  $\Pi$  of *prediction strategies*, that each attempt to predict an (unending) succession of *outcomes* (say, values in the unit interval). In each round  $n$ , every prediction strategy  $P \in \Pi$  submits a *prediction*  $p_n$  for this round’s outcome (say, again, a value in  $[0, 1]$ ), after which this outcome  $e_n$  is revealed, and each  $P$ ’s prediction is scored with the help of some *error function* (say, the square error function, returning error  $(e_n - p_n)^2$ ). By each round  $n$ , all prediction strategies thus have a certain *cumulative error*, or sum of errors so far.

A *meta-inductive* prediction strategy is a strategy that makes its predictions for events by aggregating the predictions of all the members of  $\Pi$  (that we also call the *object-strategies*). It turns out that, by clever ways of aggregating, meta-inductive strategies can provably attain a certain *optimality*. That is, *no matter* the actual sequence of outcomes, such strategies will not do (much) worse than *any* of the strategies in  $\Pi$ . More specifically, they not only converge on the minimal (among all  $P \in \Pi$ ) average (over  $n$ ) error in the long run, but also satisfy certain tight bounds on the excess (relative to any  $P \in \Pi$ ) cumulative error in the short run.

The details depend on the kind of prediction game (types of outcomes, predictions, and error function; see [Cesa-Bianchi and Lugosi, 2006](#)). But we can say in general that such optimal meta-inductive strategies predict by weighted averages over the predictions of the object-strategies, where the weights in each round depend on the latters’ cumulative error by that round. For instance, the central *exponentially-weighted-average* meta-inductivist weighs strategy  $P$  by  $\exp(-\eta \cdot \text{error}_n(P))$  for some optimizable parameter  $\eta > 0$ , thus assigning strategies with lower error (much) more weight.

What is, in sum, important for Schurz’s argument is that for a wide class of prediction games there exist meta-inductive strategies that are optimal in the previous sense, and that these optimal strategies predict by combining performance-weighted predictions of the object-strategies. In particular, the object-strategy (or strategies) that so far has (have) been performing best receive(s) the largest weight. It is in that sense that we can say that a meta-inductive strategy *favors* strategies as to their relative success so far, and favors most the strategy (strategies) that has (have) been most successful. Abstracting away from the precise prediction game, and hence which precisely is the strategy that I will now simply call “the meta-inductive strategy MI,” we can summarize the first step of Schurz’s argument as the analytical statement that

- (A) The meta-inductive strategy MI, that at each point in time favors strategies to the extent of their relative success so far, is an optimal strategy.

**2.1.2. Step two: the empirical success of object-induction.** The second step is the empirical observation that *object-induction* has turned out to be the most successful prediction strategy so far. Object-induction does not need to be precisely defined

(Sterkenburg, 2020, 533ff): we assume for the argument that it is agreed that there is something called inductive method that we use “in the real world” (Schurz, 2019, 85), and that we can distinguish from alternative, and (up to now!) predictively much less successful, non-inductive methods. No matter how we map this circumstance from the real world to a formal prediction game, we will end up with one or more object-inductive methods (I will simply talk about “the object-inductive strategy OI”) and a collection of alternative methods, where in the round  $n$  that corresponds to this point in history the first have a (much) lower cumulative error than the latter. Thus we can summarize the second step of Schurz’s argument as the statement that

- (E) As a matter of empirical fact, the object-inductive strategy OI has been, at this point in time, the most successful strategy (among all its non-inductive competitors).

2.1.3. *Conclusion: meta-induction favors object-induction.* From (A) and (E) it follows that

- (C) The meta-inductive strategy MI for the pool  $\Pi$  of OI and its non-inductive competitors, an optimal strategy for  $\Pi$ , favors most, at this point in time, the object-inductive strategy OI.

Note, however, that (C) does not yet express the desired justification of OI. For this, we need to take the further step from the fact that an optimal strategy at this time favors OI most to the claim that the latter is justified. A closer look at this step brings out some limitations in the kind of justification that follows.

2.2. **Two limitations.** As noted before (Sterkenburg, 2020, 538), one limitation is that the argument can at best provide a justification for sticking with the object-inductive strategy *for now*.

The empirical premise (E) only asserts the success of OI in the past. It is perfectly consistent with (E) that in the future, other strategies become more successful than OI. In that case, the analytically optimal strategy MI will cease to give the largest weight to OI: it will cease to favor OI most. If we accept that a method being favored most by the optimal strategy MI constitutes a justification for following this method’s predictions, then the most that Schurz’s argument can therefore establish is that we are justified in following OI’s predictions *for now*. We do not have a justification for always following the inductive method: the justification we have attaches to the next couple of rounds of inductive predictions. One could say that we do not have a justification for thinking that inductive method is a *good method*; and this is perhaps not too surprising, given that Schurz’s argument does not use any particular properties of inductive method, other than its past success.<sup>1</sup>

Note that this is not the same point as the basic observation that Schurz’s argument does not show the *reliability* of object-induction. We also do not have a proof of the reliability of *meta-induction*, but we do have an analytic argument for the *optimality* of MI, and in that sense we have a justification for believing

<sup>1</sup>In fact, even this justification for OI’s immediate prediction needs a qualification. It would go through if OI’s and MI’s immediate predictions are more or less identical, but this is not yet guaranteed if MI “favors most” (gives the largest weight to) OI. If the remaining weight is given to other strategies that predict wildly different from OI, then, if the size of this weight is still significant, there can be a substantial difference (cf. Shogenji, 2020, 230f). To exclude this possibility we would have to strengthen the empirical statement (E) and proclaim the *much greater* success of object-induction (Sterkenburg, 2020, 538).

MI is a good method (if only in a “best-alternative” sense). The point is that this optimality justification for MI only partially transfers to the object-inductive strategy: we do not have a justification for thinking OI is a good method (be it in the sense of reliability or of optimality), but only for sticking with it for now.

There is actually a further limitation in the scope of the above justification. Moreover, this is a limitation that also pertains to the optimality justification for the meta-inductivist. Namely, this justification only concerns predictions *about the next round*. The optimality results as stated, in the framework of sequential prediction, only pertain to the scores over the immediate, one-step predictions. As stated, they do not pertain to predictions, in any round, about rounds further in the future (say, a prediction in round  $n$  about the event in round  $n + 10$ , or indeed of the events in rounds  $n + 1$  to  $n + 10$ ). But this only gives us a justification for following MI’s one-step predictions. Similarly, since OI’s predictions are justified in virtue of being in line with the predictions of the optimal MI, we only have a justification for following, for now, OI’s one-step predictions.

Note that the limitation here is not that we do not have a justification for  $k$ -step predictions *for fixed*  $k$ . This might indeed be overcome by slightly tweaking the optimality result.<sup>2</sup> The limitation is that we do not have a justification for  $k$ -step predictions *for any*  $k$  *simultaneously*. The limitation is that the argument does not allow us to say that we are justified (for now) in adopting OI’s predictions, not only for the next outcome, but *also* for the outcome two steps ahead, *and* the one after that, and so forth.<sup>3</sup> I am not claiming that this limitation cannot be overcome by tweaking the argument and/or the optimality results. But this does require more work; as it stands, Schurz’s argument does not give us this. And a much more formidable challenge still would be to extend the argument to the *infinite* future, to what we may call *universal* inductions. These are predictions of the kind, “all next outcomes will be 1’s,” or “the limit relative frequency of 1’s will be  $1/2$ .” Schurz’s argument does not give us a justification for following OI in such predictions.

I think it would still be an important result to have a justification for following, *for now*, object-induction’s *immediate* predictions. Nevertheless, this is a more limited result than what is normally understood by a justification of induction, and this limitation will be relevant to Douven’s proposal.

### 3. DOUVEN’S EXPLANATION OF THE SUCCESS OF INDUCTION

In this section, I will evaluate Douven’s proposed explanation. I will first discuss the explanatory challenge that Schurz’s argument leaves open (sect. 3.1). Next I will give a summary of Douven’s simulations (sect. 3.2), and point out that their inherent induction-friendliness means they cannot provide the required explanation (sect. 3.3). Finally, I will argue that this induction-friendliness is not justified by an appeal to Schurz’s argument (sect. 3.4).

<sup>2</sup>It is trivial to reformulate the optimality result to the case of making  $k$  predictions ahead in each round  $nk$  for  $n \in \mathbb{N}$  (we simply reformulate the game to make every consecutive sequence of  $k$  rounds one round). More relevant is the case of making  $k$  predictions ahead in each round  $n$ : this is slightly more involved but I suspect still allows for a similar optimality result—for fixed  $k$ .

<sup>3</sup>A problem I see for extending the argument to this case is that the different predictions might be *inconsistent*: by which I mean that, for instance, the optimal 1-step meta-inductivist predicts the outcome 1, while the optimal 3-step meta-inductivist predicts the outcome 000.

**3.1. The need for explanation.** Douven’s starting point is that Schurz’s proposal leaves unexplained the great empirical success of our inductive methods.<sup>4</sup> This is surely right.

One concern with Reichenbach’s original argument for the justification of induction was that it seems too good to be true that such an informative conclusion could be derived, be it pragmatically, by purely a priori or *deductive* reasoning (e.g., Black, 1954, 181ff). What makes Schurz’s argument so innovative is that it combines Reichenbach’s argument with an informative *empirical* component: induction’s past success. However, this does mean that induction’s past success is simply taken as an input component to the argument: it is not itself derived from or explained by the argument.

In earlier proposals, the justification of induction is much more intertwined with an explanation of its success. A justification of inductive method in the form of reason to believe induction is *reliable* would immediately give us an explanation for the success of inductive method: it is successful because it is bound to work well. For instance, in proposals that stipulate some general principle (or more specific postulates) of induction-friendliness of the world (e.g., Russell, 1948; Burks, 1953), this stipulation can serve at the same time to justify induction and to explain its success (induction is successful *because* the world has the right structure).

Proposals of the latter kind are, of course, hopelessly question-begging (Salmon, 1953), and another merit of Schurz’s argument, like Reichenbach’s, is that it has no need for such principles. But Schurz’s argument does not give us a reason to believe that induction is reliable, nor that induction is optimal. As noted in sect. 2.2 above, Schurz’s argument does not give us a justification for believing induction is a good method. This also means that it does not in the above way automatically give us an explanation of the success of inductive method. Again, the argument needs to be explicitly supplied with the empirical fact of the past success of induction.

In short, Douven is right to pose the explanation for the success of induction as an important question that Schurz’s account leaves open. However, Douven’s own phrasing of what is missing—namely, an explanation why “these [inductive] practices have so far been *highly* successful” (2021, 2, emphasis original)—actually suggests that he believes Schurz’s argument *does* give an explanation for the success of induction, but one that falls short of an explanation of its *great* success.<sup>5</sup> Here lies the root of the shortcoming of Douven’s simulations that I will discuss in sect. 3.3 below. But let me first provide a brief overview of the simulations.

**3.2. The simulations.** Douven (2021, 2) writes,

To explain the substantial success of induction, we turn to insights from social epistemology, . . . and make these formally explicit in a framework that combines the Hegselmann-Krause model for opinion dynamics and Carnap’s system of  $\lambda$ -rules. Within this framework, we use a standard evolutionary algorithm to show how communities of interacting agents may, through variation and selection, come to consist of highly successful inductive reasoners.

<sup>4</sup>This is also noted by Shogenji (2020, 231).

<sup>5</sup>In this context Douven also writes that “to satisfy the optimality requirement, induction only needs to be as successful as random guessing” (ibid., 2, also see 4), but this is a puzzling claim that I do not see supported by Schurz’s argument.

The inductive problem in Douven’s simulations is to estimate the unknown bias  $\tau$  of a Bernoulli process (a sequence of experiments where on each repetition a successful outcome occurs with probability  $\tau$ , the prime example being repeated coin tossing with constant probability  $\tau$  of heads).

The inductive methods in the simulations are based on Carnap’s  $\lambda$ -rules (1952). Following such a rule, agent  $i$  would return on evidence  $e_n$  (a sequence of  $n$  outcomes) the estimate

$$(1) \quad c_i(e_n) = \frac{n_1 + \lambda_i/2}{n + \lambda_i},$$

where  $n_1$  denotes the number of successes in  $e_n$ , and  $\lambda_i \in \mathbb{R}^{\geq 0}$  is an agent parameter of choice.

For agent  $i$ ’s actual estimate  $p_i(e_n)$ , however, Douven conjoins these rules with a “social” component, that is based on the Hegselmann-Krause model of bounded confidence updating (Hegselmann and Krause, 2002). Write

$$(2) \quad s_i(e_n) = \frac{\sum_{j \in X_i(u)} p_j(e_{n-1})}{|X_i(u)|},$$

where  $X_i(u) = \{j : |p_i(e_n) - p_j(e_n)| \leq \epsilon_i\}$  are the agents  $j$  whose previous estimates  $p_j(e_{n-1})$  were  $\epsilon_i$ -close to agent  $i$ ’s estimate  $p_i(e_{n-1})$ , with  $\epsilon_i \in [0, 1]$  an agent parameter of choice. Then the estimate of agent  $i$  based on  $e_n$  is given by the mixture

$$(3) \quad p_i(e_n) = \alpha_i \cdot s_i(e_n) + (1 - \alpha_i) \cdot c_i(e_n),$$

with  $\alpha_i \in [0, 1]$  an agent parameter of choice.

Douven runs two sets of simulations. In the first set (2021, sec. 4.1), the previous type of agents (“Carnapians”) are set against two other types: “dogmatists” who uniformly randomly pick an estimate and forever stick with it, and “randomizers” who after each outcome uniformly randomly pick a new estimate. For each simulation, a “parent population” of twenty agents of each type is generated (with, for the Carnapians, uniformly randomly generated parameters  $\lambda \in (0, 50)$  and  $\epsilon, \alpha \in (0, 1)$ ). The second set of simulations (2021, sec. 4.2) features Carnapians only. For each simulation, an initial number of fifty such agents is randomly generated.

In both cases, each individual simulation tracks a fixed number of *generations* of agents. Each generation starts with a “parent population” (initially generated as above, subsequently the survivors of the previous generation) that is extended to a twice as large “full population” using “cross-over and mutation operations.” These are the first steps of the “standard evolutionary algorithm” (namely, the Nondominated Sorting Genetic Algorithm, or more specifically still, the version NSGA-II, Deb et al., 2002) that Douven employs to model the evolutionary process. This full population is then submitted to 25 *tests*, each of which proceeds as follows.

A bias  $\tau \in (0, 1)$  is generated uniformly randomly, and for each agent in the population one hundred outcomes are generated from the Bernoulli- $\tau$  process, in response to each of which the agent updates its estimates. After this, all agents are *scored*, along two dimensions: the sum of squared errors  $(\tau - p_i(e_n))^2$  of all estimates (“accuracy”) and the number of outcomes it took for the agent’s estimates to arrive and stay within a distance of 0.1 from  $\tau$  (“speed”).

Finally, all agents' average scores over the 25 tests are used as input for the NSGA-II algorithm to select in a specific way ("nondominated sorting" and breaking ties by "crowding distance," see *ibid.*, 10 for the details) the fifty percent most successful agents. These are the survivors that form the parent population for the next generation.

That concludes my summary of the set-up of Douven's simulations. The main results are the following.<sup>6</sup>

In the first set of simulations (Carnapians together with dogmatists and randomizers), we find that randomizers and then dogmatists quickly die out. Randomizers make it (on average over all simulations) barely to the second generation; dogmatists at most to the fourth (*ibid.*, fig. 4). This is despite the fact that randomizers and dogmatists were "dominant" in the cross-over operation used to expand the parent populations, "loading the dice *against* the Carnapians" (*ibid.*, 13).

In the second set of simulations (Carnapians only), we see that the population's mean score (i.e. mean accuracy and speed) over the generations quickly improves. Within the first 25-30 generations (on average over all simulations) there is especially rapid progress, and in the remainder (of a total of one hundred generations) there is slight improvement still (*ibid.*, fig. 6). The mean values of the  $\alpha$  and  $\epsilon$  parameters in the population increase by about 0.2 and 0.1, respectively, while the mean  $\lambda$  values show a steep drop from 50 to slightly below 2 (*ibid.*, fig. 7).

This is beautiful work, and a great example of computational philosophy in action (Mayo-Wilson and Zollman, 2021). But in what sense does this give us an explanation of the success of induction?

**3.3. The explanation.** As we saw, the simulations are to "show how communities of interacting agents may ... come to consist of highly successful inductive reasoners" (2021, 2). Further on, Douven writes in similar vein that the aim is "to complement Schurz's justification of induction by showing how inductive reasoning may have come to be the powerful cognitive tool that it is" (*ibid.*, 4). The explanation envisioned is thus essentially a plausible narrative or reconstruction of how evolutionary and social factors can lead to very successful inductive reasoners.

More precisely, this would be a reconstruction of how evolutionary and social factors can lead to the selection of (very) successful inductive reasoners, *given* that (certain) inductive reasoners are, in fact, reliable, and can be expected to be (very) successful. The simulations start out with agents following methods that include (in the first simulation) or are restricted to (the second simulation) variations of the  $\lambda$ -rules, and, crucially, presuppose an environment for which  $\lambda$ -rules are well-equipped. If the inductive problem is to estimate the unknown bias of an i.i.d. or Bernoulli process, and this is the inductive problem in the tests that are run in the simulations, then  $\lambda$ -rules are good choices of method: they can be seen to especially

---

<sup>6</sup>Douven has made the Jupyter notebooks with the code available at <https://osf.io/6njgd/>.



fit this problem,<sup>7</sup> and come with certain probabilistic guarantees of success.<sup>8</sup> To put this point differently, if we identify inductive methods with the  $\lambda$ -rules, then the simulations presuppose an *induction-friendly* environment.

It is even the case here that some of the  $\lambda$ -rules are more reliable (i.e., can be expected to be more successful) than others. The inductive problem that the agents are in each test subjected to is, more precisely, to estimate an unknown bias of a Bernoulli process that is itself uniformly randomly generated. If one also takes into account the way the bias is sampled, then one can show that some choices of  $\lambda$  are better than others; in particular, for squared error, the choice  $\lambda = 2$  is optimal.<sup>9</sup>

Furthermore, from the simple fact that in the simulations the *same* test is run every time, and that some agents are better in this test than others, it follows that greater success in earlier tests is indicative of greater success in further tests. This means that the presumed set-up is also explicitly *meta-induction-friendly*: a *meta-procedure*, that picks out the best performing agents in earlier collections of tests (like an evolutionary algorithm), can be expected to narrow down the class of agents to those that are *best*, and that can hence be expected to continue to be very successful.<sup>10</sup> And indeed we see in the first set of simulations that the Carnapians survive at the expense of the randomizers and dogmatists, and in the second set of simulations that the mean value of  $\lambda$  for the later generations “is not significantly different from 2” (Douven, 2021, 15f).<sup>11</sup>

<sup>7</sup>The  $\lambda$ -rules for binomial outcomes correspond to Bayesian prediction methods with a prior that is a mixture over all possible Bernoulli processes (Good, 1965, 17; Rosenkrantz, 1977, 71f). (More precisely, they correspond to a subset of the mixtures weighted by a beta distribution over the possible biases; the more general  $\lambda$ - $\gamma$  continuum of inductive rules (Carnap and Stegmüller, 1959) for binomial outcomes corresponds exactly to the class of these mixtures (see Festa, 1993, 66).) One can say that the  $\lambda$ -rules are methods that operate under the inductive assumption that the data is generated by some Bernoulli process, exactly the situation at hand.

<sup>8</sup>In particular, one can formally derive their expected speed and accuracy and show this to be much better than the “non-inductive” dogmatist or randomizer. See the [appendix](#).

<sup>9</sup>This may not be so surprising, when one realizes that the probabilistic model—a uniform mixture over Bernoulli processes—is exactly the full Bayesian distribution corresponding to the  $\lambda$ -rule with  $\lambda = 2$  (the corresponding beta distribution is the uniform distribution). Douven writes that “there has been some discussion about whether there might be an optimal value for  $\lambda$ , but this has not led to any definitive answer” (ibid., 8). There is no optimal value *absent* any assumptions on  $\tau$ ; but in Douven’s set-up, there is. Or rather, there are optimal *values*, since they are relative to the way we score agents. For convergence speed as defined by Douven, the optimal value (perhaps more surprisingly) does not equal 2; it lies a little below it. See the [appendix](#) for details.

<sup>10</sup>This meta-induction-friendliness is to be distinguished from the optimality of meta-induction that Schurz invokes. There are, first, some differences between optimal performance-weighting procedures and the evolutionary meta-procedure. The latter could be seen as irrevocably throwing out badly performing object-strategies, which is risky as circumstances could radically change and discarded strategies might have started doing much better. What I here call the simulations’ meta-induction-friendliness rules this out: success in the past is “projectible” to success in the future. This is more demanding than Schurz’s optimality, that holds whenever: in a world where success is no indication of further success, *all* agents, the optimal meta-inductive procedure included, will not perform very well in an absolute sense.

<sup>11</sup>I ignore here the element of social interaction between the agents. Douven’s “Carnapians” are not “pure”  $\lambda$ -rules but mixed with a social component. But there is no obvious reason why this would change the fact that in the assumed model “inductive methods” are good and some among them are best, and this is borne out by the simulations. Of course, it is still of much interest if the social component makes a community of agents on average more successful or even speeds up the evolutionary selection of the best agents (also see Douven and Wenmackers, 2017; Douven, 2019). This provides support for Douven’s *intended* explanation—discussed in the main text shortly.



One might at this point be tempted to conclude that all the simulations have shown us is that “in an environment that is both induction- and meta-induction-friendly, a meta-inductive procedure will pick out (very) successful inductive reasoners.” But simply pointing out that this is indeed not a very interesting statement is itself an uninteresting criticism. First, one could not reasonably hope for more: in simulations that robustly show that a meta-procedure converges on successful inductive reasoners, we are always going to find structural assumptions that make this possible, and that we can subsequently interpret as assumptions of (meta) induction-friendliness. Second and more important, Douven is clearly aware of the (meta) induction-friendliness of the simulations’ environment, and equally clearly does not seek to give an explanation of how there can exist (very) successful inductive reasoners in the first place. If he did seek the latter, then to explicitly assume that the environment is such that there can exist (very) successful inductive reasoners is indeed to beg the question. Rather, the goal is an explanation of how we could have quickly converged on adopting (very) successful inductive procedures among less successful procedures, where it is taken as a *given* that (the world is such that) there *exist* such (very) successful inductive procedures. Precisely, the proposed explanation is a plausible reconstruction of how the specific meta-inductive process of evolution, in combination with the specific dynamics of social interactions, could have quickly lead us to pick out increasingly good inductive procedures. The question of the explanatory value of the simulations then becomes the question whether they are indeed “realistic enough to provide insight into how we may have become such skilled users of induction” (Douven, 2021, 11).

Clearly, however, this more limited aim falls short of addressing the question identified in sect. 3.1 above. Schurz’s argument does not just leave open how we could have converged on highly successful inductive methods, given that (the world is such that) there exist successful inductive methods; it leaves fully open why induction is successful (why inductive methods are successful) in the first place. We can already conclude that Douven’s *intended* explanation, even if it were successful, would not suffice to complement Schurz’s meta-inductive justification of induction in the sense of settling the explanatory question that Schurz’s argument leaves open.

Furthermore, even within the scope of Douven’s more limited explanation one could wonder about the assumption of induction-friendliness. As noted above, the question at stake here is how realistic the reconstruction is. In particular, how accurately do the simulations model the actual historical process of the development of our inductive procedures? I will not here go into this question in any depth, apart from noting that it still leads us back to the assumption of induction-friendliness. Namely, this assumption takes here the role of a modeling choice. In the simulations’ model of the real-world process of the development of our inductive reasoning, the real world itself is modeled as being induction-friendly. This raises the question of the reason for this modeling choice. Why does it make sense to model our world as being induction-friendly?

Douven briefly addresses this question at the very end of his article. His answer is that this assumption is justified by Schurz’s argument.<sup>12</sup>

<sup>12</sup>Douven only makes this claim, as said, at the conclusion of his article, and in response to a referee comment. It is thus only fair to note here that he does not consider this claim a crucial component of his intended explanation (personal communication), and indeed I note at the end of sect. 3.4 that arguably he does not need to make this claim at all. Nevertheless, it is a natural

### 3.4. The justification for induction-friendliness. Douven (2021, 19) writes,

There is strong inductive evidence that induction is highly successful, which requires that our world satisfy certain uniformity conditions. Our reliance on induction here is entirely justified in light of Schurz' results.

Thus from the empirical observation that induction has been highly successful, we inductively infer that induction is indeed a *good* method, which requires a certain induction-friendliness of the world. The inductive step, this suggestion goes, is warranted by Schurz's meta-inductive justification of induction. Let us see how, exactly, we might use Schurz's argument to try and justify this step.

First, we might try to infer directly from Schurz's argument that we have reason to think induction (that is, object-induction) is a good method. (Doven seems to suggest such a reading of Schurz's argument at times, when he writes that Schurz has shown induction to be *optimal*.) But as I pointed out in sect. 2.2, Schurz's argument does not give us such a justification for inductive method. It shows, at best, that we have reason to follow inductive method *for now*, which is different from the universal conclusion that induction is a *good* method, that we may then connect to a universal conclusion of the induction-friendliness of the world. (In particular, there is no claim that inductive method is *optimal*, at least as good as any other method.)

What Douven appears to have in mind in the above passage, however, is something more like the following reasoning. Schurz's argument at least tells us that we have good reason to follow induction *now*, at this point in history. But do we then not have good reason to *now* make the above inductive inference? That is, if we have good reason to follow induction now, then it seems we have good reason to now make the inductive inference from the success of induction in the past to the universal (past and future) success of induction: to the conclusion that induction is, after all, a *good* method.

Note that this is actually a *meta*-inductive inference, an inference about methods. But it does not really matter how exactly one embeds this in the framework of Schurz's argument: as an object-inductive inference as above (where now the "objects" are actually methods and "object-induction" is justified in virtue of an optimal virtual "meta-inductivist" that is actually a meta-meta-inductivist) or as a meta-inductive inference (that is directly justified in virtue of its optimality). Namely, the second limitation discussed in sect. 2.2 now becomes relevant, and this limitation applies in both cases. This limitation is that the optimality argument does not give a justification for an object- or meta-inductivist's *universal* inductions, inferences to *universal* conclusions. But the conclusion that the world is induction-friendly is such a universal conclusion: the only way, it seems to me, to understand this conclusion in the assumed framework is as a statement about infinite outcome sequences. In sum, Schurz's argument cannot provide a justification for the conclusion that the world is induction-friendly.

This point is really a reiteration of the earlier observation in sect. 3.1 that Schurz's argument cannot explain the success of inductive method. If the argument *could* justify a belief in the induction-friendliness of the world, then we would have

---

response to give, and, I think, in itself interesting that Schurz's proposed justification cannot actually justify this.

such an explanation for free (induction works well *because* the world is induction-friendly). But it cannot give such an explanation, and Douven, in misdiagnosing the reach of Schurz’s argument, aims at an explanation more limited than that.

Indeed, one might argue that for Douven’s more limited aim, it is actually not *needed* to have a justification for the induction-friendliness of the world. For his proposed explanation of the great success of induction, this argument would go, what we model are our inductive methods *in the past*, and we *may* at least assume, in virtue of the past success of inductive methods, that the world *was* induction-friendly. Granted that the simulations are sufficiently realistic in other respects, we would then have a plausible account of how social and evolutionary factors could have facilitated our convergence on (very) successful inductive methods. Then not only does Douven’s proposal not answer the question left by Schurz’s argument: Douven himself does not actually need Schurz’s argument at all.

#### 4. CONCLUSION

I argued that Douven’s simulations do not answer the explanatory question left open by Schurz’s argument, and that, in turn, Schurz’s argument cannot justify the assumption of induction-friendliness in Douven’s simulations.

Douven seeks to give an explanation of how we have come to adopt (very) successful inductive methods, given that there are such methods to begin with. Whether we can judge this explanation to be convincing, depends on whether we judge plausible his modeling of inductive method, the social component, and the evolutionary process. But however the verdict, the simulations cannot explain why inductive method is successful in the first place, the question left open by Schurz.

The success of particular explications of inductive method, here variations of Carnap’s  $\lambda$ -rules, is inextricably bound up with an environment that makes it so. In order to explain the former, it will obviously not do to simply stipulate the latter—but what noncircular reason can we give for thinking that our world constitutes an environment that is friendly to inductive method? We saw that Schurz’s argument does not give us this. Indeed, we may wonder whether it makes sense at all to ask for such a thing: to put our hopes in some grand yet vague principle of “uniformity” or “induction-friendliness” of the world (Salmon, 1953; Sober, 1988).

As argued by Okasha (2001) and more recently, in the context of machine learning theory, by Sterkenburg and Grünwald (2020), we are analogously led to wonder whether “inductive method” can be plausibly explicated as some set of purely data-driven rules. Part of the attention then shifts to the choice of “local” models that inductive procedures must take as additional input; but so does the challenge of explanation: why have these choices been successful?

This local perspective does cast doubt on the plausibility of Douven’s modeling of inductive method, and as such already on his intended, more limited, explanation. Nevertheless, the leading idea that social and evolutionary factors must be part of such an explanation remains plausible; and it only appears natural to try and further connect this to the development of the local models underlying inductive inferences. Perhaps such work could even bring progress in the larger explanatory question. As it stands, however, the important problem of explaining why our inductive practices are successful at all remains wide open.

## APPENDIX

This appendix derives formal performance expectations for the agents in Douven’s simulations, extending his own remarks about “what we should expect . . . in the statistical model underlying our simulations” (2021, 17), and providing the basis to the observation in the main text about the simulations’ inherent induction-friendliness.

For both the dogmatist and the randomizer the expected cumulative squared error after each update (or *cumulative risk*), as a function of bias  $\tau$ , is 100 times

$$\mathbf{E}_{p \sim \mathcal{U}(0,1)} [(p - \tau)^2] = \tau^2 - \tau + 1/3,$$

while that of a  $\lambda$ -method (see Carnap, 1952, 67) is

$$\begin{aligned} \mathbf{E}_{X^{100} \sim \tau} \left[ \sum_{n=1}^{100} (c_\lambda(X^n) - \tau)^2 \right] &= \sum_{n=1}^{100} \mathbf{E}_{X^n \sim \tau} [(c_\lambda(X^n) - \tau)^2] \\ &= \sum_{n=1}^{100} \left( \frac{n\tau(1 - \tau) + (1/2 - \tau)^2 \lambda^2}{(\lambda + n)^2} \right). \end{aligned}$$

Plotting this we clearly see the difference in the agents’ risks (figs. 1, 2). More interestingly, in the simulations it is also the case that  $\tau$  itself is uniformly randomly generated. Taking this into account, the cumulative risk of a dogmatist/randomizer equals 100 times

$$\mathbf{E}_{\tau \sim \mathcal{U}(0,1)} [\tau^2 - \tau + 1/3] = 1/3 - 1/2 + 1/3 = 1/6,$$

or about 16.67, while that of a  $\lambda$ -rule equals

$$\begin{aligned} \mathbf{E}_{\tau \sim \mathcal{U}(0,1)} \left[ \sum_{n=1}^{100} \left( \frac{n\tau(1 - \tau) + (1/2 - \tau)^2 \lambda^2}{(\lambda + n)^2} \right) \right] &= \sum_{n=1}^{100} \frac{\int_0^1 n\tau(1 - \tau) + (1/2 - \tau)^2 \lambda^2 d\tau}{(\lambda + n)^2} \\ &= \sum_{n=1}^{100} \frac{1/12 \lambda^2 + n/6}{(\lambda + n)^2}. \end{aligned}$$

Evaluating the derivative w.r.t.  $\lambda$ ,

$$\sum_{n=1}^{100} \frac{n\lambda/6 - n/3}{(\lambda + n)^4} = 0,$$

we see that we must have  $n\lambda/6 - n/3 = 0$ , that is,  $\lambda = 2$ . This is the optimal value of  $\lambda$ , that comes with a cumulative risk of about 0.62 (see fig. 3). Incidentally, the choice  $\lambda = 50$  still does much better than the randomizer/dogmatist; indeed, as we let  $\lambda$  go to infinity, converging to the constant rule that is as little receptive to the data as these agents, the risk will still stay below half their risk:  $100 \cdot \mathbf{E}_{\tau \sim \mathcal{U}} [(\tau - 1/2)^2] = 100/12 \approx 8.33$ .

Squared error is only one of the two dimensions that the agents are scored on in the simulations’ evolutionary algorithm; the other is speed of convergence. Douven (ibid., 8) writes that “there is typically a trade-off to be made between speed and accuracy.” The two dimensions can indeed come apart: despite identical risk, the expected (over  $\tau \sim \mathcal{U}(0,1)$  and  $X^{100} \sim \tau$ ) convergence speed (number of updates to arrive and stay within 0.1 of  $\tau$ , with speed 101 if there is no convergence) of the randomizer is different from that of the dogmatist. Namely, the probability that uniformly randomly drawn  $\tau$  and  $p$  are 0.1-close is 0.19 (Douven, 2021, 17), so that the expected speed of the dogmatist is  $0.81 \cdot 101 = 81.81$  and that of the randomizer is  $\sum_{i=1}^{101} i \cdot 0.81 \cdot 0.19^{101-i} \approx 100.77$ . (This explains why in the first set of simulations, the dogmatists last longer than the randomizers.) In the case of the  $\lambda$ -rules, it does not seem feasible to derive an analytical expression of the expected speed, even as function of  $\tau$ . But one can write a program to calculate the latter, and plot this as in fig. 4. Comparing this to fig. 2, we see that the dimensions of accuracy and speed rather pull into the same direction. Nevertheless, approximating the areas under the curve, one can determine that the best  $\lambda$ -values now

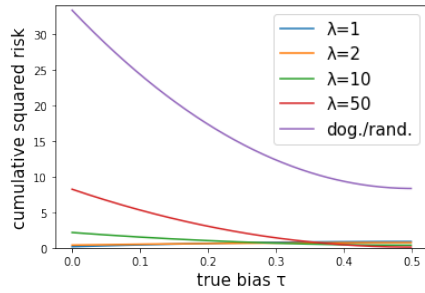


FIGURE 1. Risk of various agents.

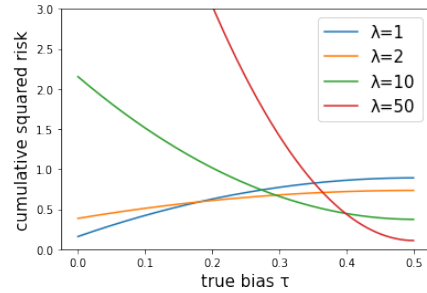


FIGURE 2. Detail of figure 1.

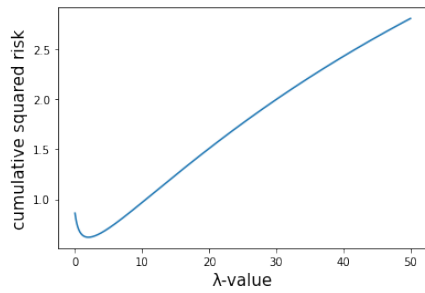
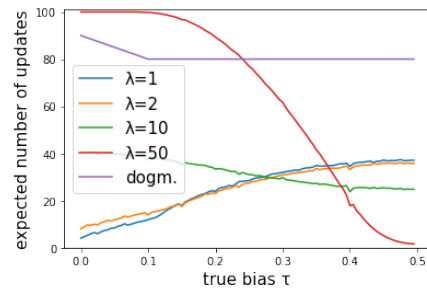
FIGURE 3. Risk of  $\lambda$ -rules for uniformly random bias  $\tau$ .

FIGURE 4. Expected speed of various agents.

(roughly) lie in the interval  $(1, 1.8)$ , rather than at 2. This is in line with the observation (ibid., 15) that the mean  $\lambda$ -value of the last generations of Carnapians lies around 1.7.

## REFERENCES

- M. Black. “Pragmatic” justifications of induction. In *Problems of Analysis. Philosophical Essays*, pages 157–190. Routledge & Kegan Paul, 1954.
- A. W. Burks. The presupposition theory of induction. *Philosophy of Science*, 20(3): 177–197, 1953.
- R. Carnap. *The Continuum of Inductive Methods*. The University of Chicago Press, 1952.
- R. Carnap and W. Stegmüller. *Induktive Logik und Wahrscheinlichkeit*. Springer, 1959.
- N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning and Games*. Cambridge University Press, 2006.
- K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computing*, 6(2): 182–197, 2002.
- I. Douven. Optimizing group learning: An evolutionary computing approach. *Artificial Intelligence*, 275:235–251, 2019.
- I. Douven. Explaining the success of induction. Forthcoming in *The British Journal for the Philosophy of Science*, 2021.
- I. Douven and S. Wenmackers. Inference to the best explanation versus Bayes’s rule in a social setting. *The British Journal for the Philosophy of Science*, 68(2):535–570, 2017.
- R. Festa. *Optimum Inductive Methods: A Study in Inductive Probability, Bayesian Statistics, and Verisimilitude*, volume 232 of *Synthese Library*. Springer, 1993.
- I. J. Good. *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. MIT Press, 1965.

- R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.
- C. Mayo-Wilson and K. J. S. Zollman. The computational philosophy: Simulation as a core philosophical method. Forthcoming in *Synthese*, 2021.
- S. Okasha. What did Hume really show about induction? *The Philosophical Quarterly*, 51(204):307–327, 2001.
- R. D. Rosenkrantz. *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*, volume 115 of *Synthese Library*. D. Reidel, 1977.
- B. Russell. *Human Knowledge: Its Scope and Limits*. Simon & Schuster, 1948.
- W. C. Salmon. The uniformity of nature. *Philosophy and Phenomenological Research*, 14(1):39–48, 1953.
- W. C. Salmon. Hans Reichenbach’s vindication of induction. *Erkenntnis*, 35:99–122, 1991.
- G. Schurz. The meta-inductivist’s winning strategy in the prediction game: A new approach to Hume’s problem. *Philosophy of Science*, 75(3):278–305, 2008.
- G. Schurz. *Hume’s Problem Solved: The Optimality of Meta-Induction*. MIT Press, 2019.
- T. Shogenji. Review of [Schurz \(2019\)](#). *International Studies in the Philosophy of Science*, 32(3-4):229–331, 2020.
- E. Sober. *Reconstructing the Past: Parsimony, Evolution, and Inference*. A Bradford Book. The MIT Press, 1988.
- T. F. Sterkenburg. The meta-inductive justification of induction. *Episteme*, 17(4):519–541, 2020.
- T. F. Sterkenburg and P. D. Grünwald. The no-free-lunch theorems of supervised learning. Available at <http://philsci-archive.pitt.edu/18505/>, 2020.

MUNICH CENTER FOR MATHEMATICAL PHILOSOPHY, LMU MUNICH  
 Email address: [tom.sterkenburg@lmu.de](mailto:tom.sterkenburg@lmu.de)