The Review of Symbolic Logic
Volume 15, Number 2, June 2022

MODES OF CONVERGENCE TO THE TRUTH: STEPS TOWARD A BETTER EPISTEMOLOGY OF INDUCTION

HANTI LIN
Philosophy Department, University of California, Davis

Abstract. Evaluative studies of inductive inferences have been pursued extensively with mathematical rigor in many disciplines, such as statistics, econometrics, computer science, and formal epistemology. Attempts have been made in those disciplines to justify many different kinds of inductive inferences, to varying extents. But somehow those disciplines have said almost nothing to justify a most familiar kind of induction, an example of which is this: “We’ve seen this many ravens and they all are black, so all ravens are black.” This is enumerative induction in its *full* strength. For it does not settle with a weaker conclusion (such as “the ravens observed in the future will all be black”); nor does it proceed with any additional premise (such as the statistical IID assumption). The goal of this paper is to take some initial steps toward a justification for the full version of enumerative induction, against counterinduction, and against the skeptical policy. The idea is to explore various epistemic ideals, mathematically defined as different modes of convergence to the truth, and look for one that is weak enough to be achievable and strong enough to justify a norm that governs both the long run and the short run. So the proposal is learning-theoretic in essence, but a Bayesian version is developed as well.

§1. Introduction. The *general* problem of induction may be taken as the problem of addressing this task: for each type of inductive inference, determine whether we can justify it with explicit reasons or arguments and, if so, identify the extent to which we can do that. Under the general problem there are several subproblems. There is, for example, the subproblem of how it is possible to reliably infer causal relations solely from observational data without experimentation—a problem that has attracted many scientists and philosophers.[1] And there is the more general subproblem of whether it is possible to escape Hume’s dilemma—a dilemma that aims to undermine any justification of any kind of inductive inference.[2] This paper addresses another subproblem of induction, one that should be familiar but is somehow seldom addressed.

Here is the background. Evaluative studies of inductive inferences are pursued with mathematical rigor in many disciplines, such as formal epistemology, statistics,

Received: September 6, 2021.
2020 *Mathematics Subject Classification*: 03B48, 62F15.
Key words and phrases: learning theory, enumerative induction, probability and inductive logic, Bayesian inference.

[1] For book-length treatments of this problem, see [18, 31, 40].

[2] See Hume ([17, Section IV]) for his formulation of the dilemma, and Reichenbach ([36, Section 38]) for an influential version of the dilemma. A classic list of attempted solutions is provided in Salmon (1966, ch. 2); for an updated list, see the survey by Henderson [14].

doi:10.1017/S1755020321000605

econometrics, and computer science. But somewhat curiously, they all have said little about a very familiar kind of inductive inference, of which an instance is this:

> We have observed this many ravens and they all are black.
> So, all ravens are black.

This is a version of enumerative induction, which may be called the *full* version. Other versions weaken the conclusion or strengthen the premise. To be sure, much work has been done for enumerative induction, but the attention has been mostly directed to the less-than-full versions. For example, sometimes the conclusion is weakened to "the ravens *observed in the future* will all be black," as is often the case in learning theory[3] and Bayesian confirmation theory.[4] Sometimes the inference is weakened with an additional premise such as "if there are nonblack ravens then we will observe one sooner or later," as is often the case in learning theory.[5] Sometimes the inference is weakened with an additional premise typical in statistics, the IID assumption, which says that data are generated *i*ndependently according to an *i*dentical *d*istribution of objective chances.

All those theories have so far set aside a serious evaluative study of full enumerative induction. But why? The reason for statisticians is obvious: their primary job is to study inductive inferences under the IID assumption or the like. The reasons for formal epistemologists such as Bayesians and learning theorists seem to run deeper, as I will explain in Section 2. That will help me formulate a subproblem of induction, which I call the *Cartesian* problem of induction. Then I will propose a solution in Section 3, focusing on the philosophical ideas. The mathematical details will then be developed in Sections 4–8. My positive account is learning-theoretic in nature, but it has a Bayesian version that employs considerations about possible futures to impose a norm on the Bayesian priors, to be presented in Section 9.

§2. The Cartesian problem of induction.

Recall the following instance of full enumerative induction:

> We have observed this many ravens and they all are black.
> So, all ravens are black.

Note that, even if the ravens observed in the past, present, and future are all black, it still leaves open whether all ravens are black—the true answer might be "yes" and, unfortunately, it might be "no." This latter possibility may be called the *Cartesian*

[3] See, for example, [21], [22], and [39].

[4] See, for example, [5, 15, 16]. Their works are concerned with the (probabilistic) inference from evidence $F(a_1) \wedge \cdots \wedge F(a_n)$ to the countable conjunction $H = \bigwedge_{i=1}^{\infty} F(a_i)$, where a_i means the i-th individual (or raven) observed in a certain agent's inquiry. So the inductive conclusion H talks about, not all ravens, but all ravens that the agent will observe in the future. What about assuming, further, that i enumerates all ravens in the world—so that the agent's first observation, second observation, *ad infinitum*, will exhaust all ravens in the world? Adding this assumption amounts to switching from full enumerative induction to a weaker version, because it in effect strengthens the premise by adding that all ravens in the world are countable in number and each of them will be observed by the agent sooner or later.

[5] See, for example, [21, 22 39].

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

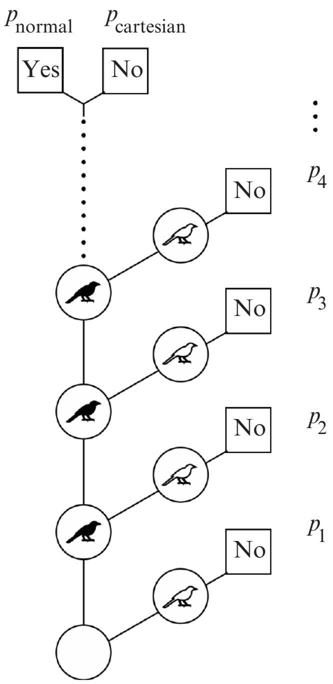

Fig. 1. A Bayesian prior over some possibilities.

scenario of induction, for it can be materialized dramatically by a Cartesian-like demon who always hides a white raven behind the inquirer. In the other possibility, the inquirer still sees only black ravens without a counterexample—and fortunately all ravens are black. This is the *normal counterpart* to the Cartesian scenario of induction. Those two scenarios, the Cartesian and the normal, are empirically indistinguishable *for the inquirer* (but perhaps not for the demon), and the hypothesis "all ravens are black" is underdetermined in that it is true in one of the two scenarios and false in the other. As we will see very soon, that causes trouble for many formal epistemologists, including both Bayesians and learning theorists.

For Bayesians, any justification of full enumerative induction requires justifying a prior probability distribution of credences that disfavors the Cartesian scenario of induction and favors its normal counterpart. To see why, consider the tree depicted in Figure 1. Branches represent possible worlds, or possible histories of inquiry. Moving straight up means observing a black raven; veering to the right means observing a nonblack raven. Let `Yes` denote the hypothesis that all ravens are black, which is true in the branch marked with '`Yes`'. Let `No` denote the hypothesis that *not* all ravens are black, which is true in the branches marked with '`No`'. The two vertical branches represent the Cartesian scenario and its normal counterpart, respectively; the Cartesian one is marked with '`No`' and its normal counterpart is marked with '`Yes`'. Now, construct a Bayesian prior by distributing probabilistic credences $p_1, p_2, \ldots, p_{\text{normal}}, p_{\text{cartesian}}$ to the branches in the way depicted in Figure 1, while making sure that those credences are nonnegative and sum to one. With that Bayesian prior, the inquirer updates her credences as follows. When she moves straight up in the tree, she observes more and

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

more black ravens and thereby rules out more and more branches that veer to the right. Then, with successive updates by conditionalization, the ratio of her posterior credence in `Yes` to that in `No` converges to the prior ratio of p_{normal} to $p_{\text{cartesian}}$. If this ratio is much greater than 1, the prior is inductive; much smaller than 1, counterinductive; close to 1, skeptical.

So, to justify the full version of enumerative induction, a Bayesian needs to provide some reason for having a prior that is inductive rather than counterinductive or skeptical. But Bayesians seldom, if ever, try to do so. Possible reasons are not hard to see. Think about objective Bayesians first, who think that a probabilistic prior is epistemically permissible just in case, roughly, it is flat or somehow flat enough to guard against a certain kind of unwarranted bias (which they specify). Objective Bayesians seem to have never mentioned the Cartesian scenario of induction and the inductive priors; see, for example, [4, 19, 43]. It might be simply because they were already too busy with justifying some less-than-full versions of enumerative induction. Or it might be because they did not find a way to justify an inductive prior as a sufficiently flat prior, for a naïve view of sufficient flatness appears to require that the prior ratio of p_{normal} to $p_{\text{cartesian}}$ be close to 1, leading to a skeptical prior.

On the other hand, subjective Bayesians (such as [6]) would say, roughly, that an inductive prior, if probabilistic, is epistemically permissible—but only because they think that any probabilistic prior is epistemically permissible.[6] So they are committed to the thesis that the counterinductive priors and the skeptical ones are epistemically permissible, too, as long as they are probabilistic. To make this commitment explicit is to invite the usual worry that "anything goes" on the subjective Bayesian account. This worry, which is all too familiar, has already arisen from many other case studies of inductive inferences. So subjective Bayesians seem to have no motivation to mention the case of full enumerative induction because this case does little to assuage (or aggravate) the familiar worry.

Learning theorists fare no better. When the adoption of an inductive principle is justified in formal learning theory, it is typically justified as a necessary condition for achieving a certain epistemic ideal: the ideal of finding the truth at least in the long run (and possibly as fast as possible) in every possible history of inquiry under consideration. This epistemic ideal is often called *identification in the limit* [12, 35]. When applied to the problem of whether all ravens are black, the ideal of identification in the limit sets an extremely high standard: finding the truth *both* in the Cartesian scenario and in its normal counterpart. But to find the truth in one of those two scenarios is to fail to do so in the other. So the epistemic standard in question is too high to be achievable and thus too demanding to serve as an evaluative standard. That is, formal learning theorists lose their evaluative standard in this case.

[6] This presentation of subjective Bayesianism is quite simplistic; see [20] for a survey of varieties of subjective Bayesianism. But the point I made here generalizes to a more precise picture of subjective Bayesianism, which is the general view that one's prior is epistemically permissible as long as it is coherent—pending an account that says what coherence amounts to. The coherence of a prior is generally taken to require at least that the prior be probabilistic. Does that exhausts what coherence requires? Yes, according to the most radical subjective Bayesians. No, according to other subjective Bayesians, who think that coherence requires more, such as so-called *regularity*, or *countable additivity*, or the *principal principle*, to name just a few that are most relevant to scientific inquiries.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

Those seem to be some of the reasons why so little has been done to justify enumerative induction in its full version. So we are still left with the problem of how we may develop an explicit argument for full enumerative induction, against counterinduction, and against the skeptical policy. I call this problem the *Cartesian problem of induction*, because of the role played by Cartesian scenarios of the sort mentioned above. The point of pursuing this problem is not to respond to every conceivable kind of inductive skeptic. The point is, rather, to push ourselves to the limit—to explore the extent to which we can justify various kinds of inductive inferences with an explicit argument.

That problem might appear to have a simple solution: just adopt the view that scientists should care about, not the colors of *all* ravens, but only the colors of the ravens that are or will actually be observed by us. This solution looks simple, but are not straightforward to defend. It presupposes a criterion of what scientists should care about, a quite stringent criterion that rules in only the objects that are or will *actually* be observed. A more lenient criterion rules in all objects that are observ*able*, whether or not they will actually be observed (e.g., [41]).[7] An even more lenient criterion rules in all objects that are physical, whether or not they are observable. It would be interesting to see how the most stringent criterion of the three might be defended in favor of the simple solution to the Cartesian problem of induction. But I believe that, before we can have an overall assessment of the competing solutions, it is important to explore and develop *at least one* positive solution for those who adopt the intermediate or lenient criterion of what scientists should care about. That is what I set out to do in this paper.

This paper aims to develop the first systematic solution to the Cartesian problem of induction. My proposal, put in a slogan, is that Bayesians go learning-theoretic and learning theorists be truly learning-theoretic. The crux is to revisit the spirit by which learning theory was created in the 1960's, give it a clear formulation, and use it to explain how normative considerations about possible futures can impose a significant constraint on the short run and even the present, including Bayesian priors.

The next section—Section 3—presents my positive solution to the Cartesian problem of induction, with a focus on the philosophical ideas and a sketch of the supporting theorems. The mathematical details are then developed in Sections 4–8. The Bayesian version of my account is presented in Section 9. To declare the style in use: Emphasis

[7] Some clarifications are in order. The Cartesian scenario of induction and its normal counterpart are *not* empirically equivalent according to the antirealist view defended by van Fraassen [41]. For him, a theory is empirically adequate in a possible world w just in case everything it says about the observ*able* objects and their observ*able* properties or relations is true in w (1980, 12), and two theories are empirically equivalent just in case they are empirically adequate in exactly the same possible worlds. The Cartesian scenario of induction and its normal counterpart make different claims about observable objects (i.e., ravens) and their observable properties (i.e., their colors). In particular, one of those two scenarios makes it true that all ravens (as observable objects) have the blackness property (as an observable property), and the other scenario makes it false. So, those two scenarios are not empirically equivalent according to van Fraassen's definition. That said, it would be interesting to formulate and assess some variants of van Fraassen's definition under which those two scenarios become empirically equivalent. But this task has to be reserved for a paper specifically on the debate between scientific realism and antirealism.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

is indicated by *italics*, while the terms to be defined are presented in **boldface**.

§3. A solution.

An empirical problem can be understood to have three components: First, it poses a question with some potential answers as *competing hypotheses*. Second, it considers some *possible bodies of evidence* that the inquirer might receive. Third, it comes with a *presupposition* taken for granted by the inquirer. For example, we have:

> The Hard Raven Problem.
>
> (i) *Competing Hypotheses*. The inquirer asks a question, whether all ravens are black. There are two potential answers, or competing hypotheses: `Yes` and `No`.
>
> (ii) *Possible Bodies of Evidence*. She plans to collect ravens and observe their colors. A body of evidence specifies the colors of the ravens that have been observed.
>
> (iii) *Presupposition*. She takes for granted that the colors of ravens do not change in time, so the true answer to the question posed need not be indexed to any specific time. This is the *only* presupposition of the problem that she pursues.

Call this empirical problem the **hard raven problem**.

The hard raven problem can be represented by the tree in Figure 2. The observation of a black raven is represented by a datum `+`; a nonblack raven, `-`; a nonraven, `0`. A possible state of the world is represented by an entire branch, which is specified by two items: first, the infinite data stream produced in that state (represented by an infinite sequence of circular nodes); second, the hypothesis true in that state (as indicated in the box). The figure also highlights some Cartesian scenarios of induction. To be sure, each state is associated with an *infinite* data stream $(e_1, e_2, \dots, e_n, \dots)$, but that does not really represents a state in which the scientist *will* be immortal—it represents, rather, a state in which the scientist would happen to have evidence $(e_1, e_2, \dots, e_n)$ if the inquiry were to unfold up to sample size n, for any positive integer n.

A **learning method** for the hard raven problem is a mapping that sends each of the finite data sequences under consideration to one of the competing hypotheses, `Yes` or `No`, or to a question mark. Think of such a learning method as an instruction that receives data and recommends one of the three qualitative attitudes toward the general hypothesis: belief, disbelief, and suspension of judgment. (The account to be presented below has a Bayesian version that allows a learning method to output probabilistic attitudes.)

Which learning methods are the best for tackling the hard raven problem? I would like to address this issue by revisiting the root of learning theory: Putnam's [34, 35] and Gold's [11, 12] pioneering works. But before that, I would like to go all the way back to Plato's *Meno*.

3.1. Plato's Meno *revisited.*

Towards the end of his *Meno*, Plato considers an epistemic ideal:

> True opinions are a fine thing and do all sorts of good so long as they stay in their place, but they will not stay long. They run away from a man's mind; so they are not worth much until you *tether* them

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

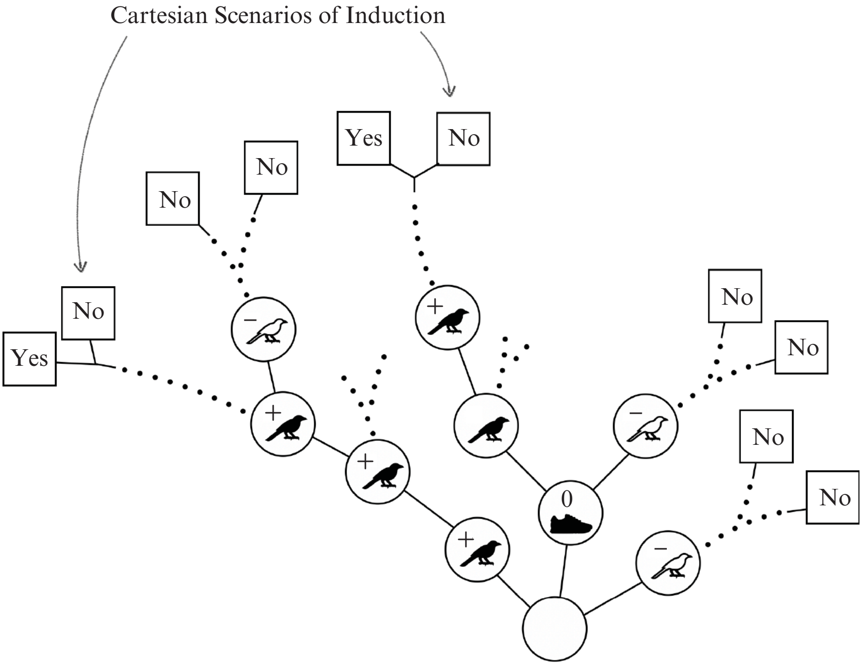

Fig. 2. A tree representation of the hard raven problem.

> by working out a reason.[8] ... Once they are *tied down*, they become knowledge, and are *stable*. That is why knowledge is something more valuable than right opinion. What distinguishes the one from the other is the *tether*. ([32, pp. 381–382], emphasis mine)

For learning theorists, the idea of *tethering to the truth* can be explicated in terms of convergence to the truth ([21, ch. 1]; [13, ch. 10]). In a state associated with an infinite data stream $(e_1, e_2, e_3, \ldots)$, a learning method M would output opinions $M(e_1), M(e_1, e_2), M(e_1, e_2, e_3), \ldots$ if the inquiry were to extend to sample sizes $1, 2, 3, \ldots$, respectively. A learning method M for an empirical problem is said to **converge to the truth** in a state/branch by sample size n if, in that state/branch, M would output the true hypothesis given sample size n and would then *always* continue to do so were the inquiry to unfold indefinitely—as if the output of this learning method were *tethered* to the true hypothesis from sample size n onwards.

The idea of tethering to the truth sounds good. But how are we to use that idea to evaluate learning methods? Learning theorists proceed with two guidelines.

3.2. Guidelines. It seems great to acquire tethering to the truth. If so, it would be even better to be guaranteed, under the presupposition of the considered empirical problem, that tethering to the truth would be acquired by a finite, fixed sample size—a single sample size that would suffice to deliver the desired tethering *uniformly* across for all states compatible with the presupposition. This is an epistemic ideal that learning

[8] Plato thinks that the process of tethering opinions to the truth is what he calls *recollection* ([32, p. 381]). Learning theorists can embrace the ideal of tethering without a commitment to Plato's theory of recollection.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

theorists call **uniform convergence to the truth**. With respect to any empirical problem, a best learning method must achieve this epistemic ideal if some learning method does. That is, this epistemic ideal is so good that it ought to be achieved whenever achievable.

Unfortunately, uniform convergence to the truth is too high an epistemic ideal to be achievable for many interesting empirical problems. Consider, for example, the **easy raven problem**, which is the same as the hard version except that it comes with a strong presupposition that rules out all Cartesian scenarios of induction. Even the easy raven problem is too hard to allow any learning method to achieve uniform convergence. It follows that, for the easy raven problem, every learning method must come to have convergence to the truth *arbitrarily late* (or even never reach convergence at all) in some states of the world under consideration. This is a mathematically inescapable result for the easy raven problem—no epistemology can help us escape it.

So, in that case, it seems inevitable to consider an epistemic ideal that allows arbitrarily slow convergence in some states under consideration. Accordingly, a learning method for an empirical problem is said to **converge to the truth everywhere** if, in every state of the world considered in that problem, this learning method converges to the truth by one or another finite sample size. This mode of convergence is often called **pointwise** convergence in mathematics; it is also called **identification in the limit** in the learning theory literature. This epistemic ideal is achievable for the easy raven problem and many other problems, studied in the branch of learning theory called *formal learning theory*.

If I am right, the above line of thought embodies two guidelines:

> GUIDELINES
>
> 1. Look for what can be achieved (in terms of tethering to the truth).
> 2. Achieve the best we can have (with respect to the empirical problem under consideration).

They reflect the spirit with which Putnam [34, 35] and Gold [11, 12] created learning theory. I take them to be the two guidelines of *learning-theoretic epistemology*. Different epistemic ideals about tethering to the truth are different modes of convergence to the truth. This raises two questions:

> MOVING PARTS
>
> 1. Which modes of convergence deserve to be taken as epistemic ideals about tethering to the truth?
> 2. As epistemic ideals, which are higher than which others?

These are the two moving parts of learning-theoretic epistemology. Although learning theorists might have a civil war over those two moving parts, they are (or should be) united by this thesis:

> THE CORE THESIS. With respect to any empirical problem, the best learning methods must at least achieve a ~~highest~~ epistemic ideal about tethering to the truth, if such a highest one exists.[9] highest achievable

This finishes my reconstruction of learning theory.

[9] For pioneers of this idea, see [21, 36, 39].

To clarify: This does *not* mean that modes of convergence to the truth exhaust all epistemic ideals that we should care about. Feel free to pursue other kinds of epistemic ideals in addition to convergence to the truth, as you see fit. This freedom is made explicit in the statement of the core thesis, as highlighted by the occurrence of 'at least'. The core thesis only imposes a constraint on the best learning methods, relative to each empirical problem. It claims that the best ones must satisfy that constraint, and says nothing about whether there exists a uniquely best one or whether each one within this constraint is as good as any other.

We will soon see that, once the two moving parts are fixed in the right way, the core thesis will shed some light on the hard raven problem.

3.3. Exploring modes of convergence. The hard raven problem is so hard that the ideal of everywhere convergence is even too high to be achievable. This is because, for any learning method, if it enjoys convergence to the truth in a Cartesian scenario of induction, it must fail to do so in its normal, empirically indistinguishable counterpart. With the impossibility to achieve everywhere convergence for the hard raven problem, the existing literature of learning theory stops there. But that seems to give up too soon. If everywhere convergence is unachievable, we should not try to achieve it. We only need to achieve the best we can have and, before that, we need to look for what can be achieved—following the two guidelines formulated above, which is what I meant when I said "be *truly* learning-theoretic."

So let's explore some modes of convergence. Everywhere convergence is only one of the many convergence criteria that concern the question of *where* convergence happens. We can also consider another question, which concerns *how* convergence happens. Have a look at Figure 3, in which various modes of convergence to the truth are arranged by two dimensions. The dimension that stretches to the upper right concerns "where." The other dimension, which stretches to the upper left, concerns "how." I introduce three modes for each of the two dimensions, so in combination there are nine modes to be considered in this paper.[10] Let me explain those modes of convergence in greater detail, focusing on their philosophical interpretations and leaving rigorous definitions to subsequent sections.

Given that no learning method can make it everywhere, let's see whether there exists a learning method that can at least make it almost everywhere. The challenge is to find a good, rigorous explication of the informal concept of "almost everywhere." I propose to proceed this way: first motivate a certain *closeness relation* between states, which defines a concept of *open neighborhoods* of a state in a state space, which then

[10] Stochastic modes of convergence (such as convergence in probability) will not be considered in this paper for two reasons. First, stochastic convergence to the truth can mean convergence in terms of chances (physical, objective probabilities), but the premise of *full* enumerative induction is too weak to be committed to the existence of chances. Second, stochastic convergence to the truth can mean that one is subjectively certain that one's posteriors in the truth will converge to full certainty. I am not sure whether this subjective mode of stochastic convergence is a necessary feature of every good Bayesian prior. But for present purposes, it suffices to note that the subjective mode of stochastic convergence is too weak to rule out all counterinductive priors for the hard raven problem: just recall Figure 1 and let $p_{\text{normal}} = 0$ and $p_{\text{cartesian}} > 0$.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

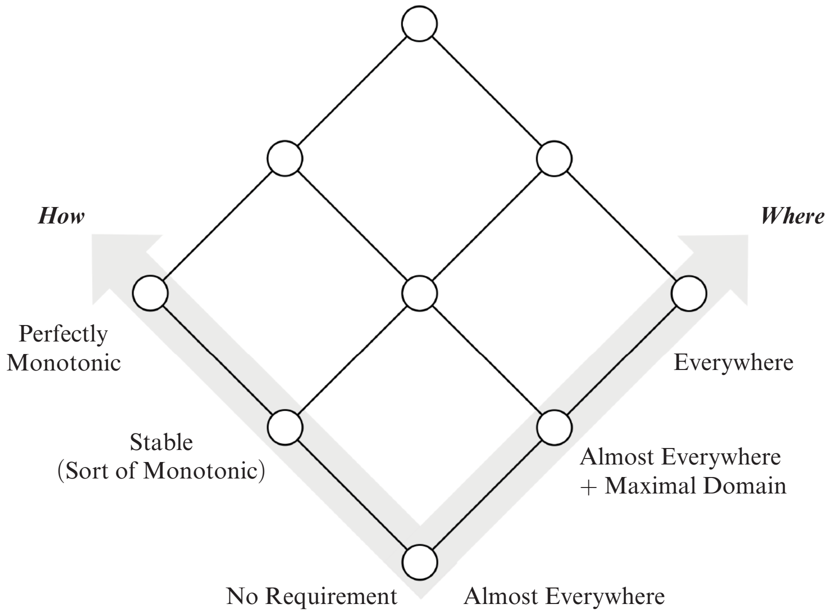

Fig. 3. Modes of convergence to the truth, arranged by two dimensions.

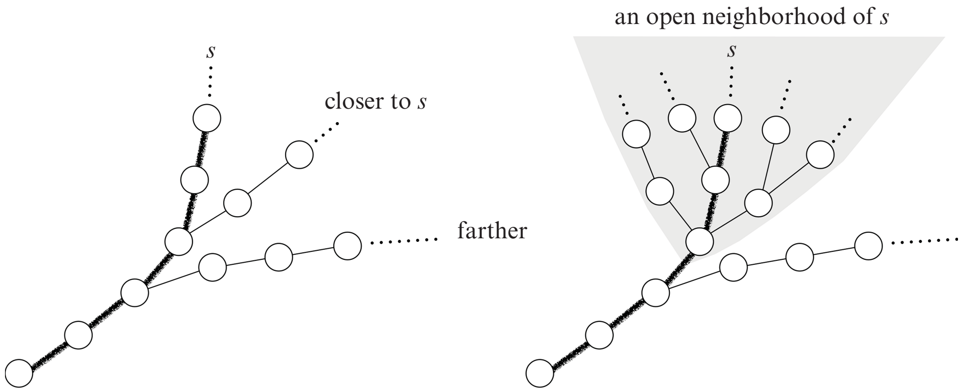

Fig. 4. The empirical topology on a space of data streams or states.

defines what it is for a region to cover *almost everywhere* in a state space. To be more specific:

> Step 1. Consider an arbitrary state s as a branch depicted on the left side of Figure 4. An alternative state is said to be **empirically closer** to s if more empirical data are needed to distinguish it from s.
>
> Step 2. Given a space of states, X, an **open neighborhood** of a state s in X is the set of states in X that are empirically close to s to at least a certain degree, as depicted on the right side of Figure 4.[11]

[11] This topology is proposed for epistemological purposes by Vickers [42] in computer science and by Kelly [21] in formal epistemology.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

> Step 3. Following the standard treatment in geometry and topology, a region within a space X is called **topologically negligible** (also called **nowhere dense**) if it can be constructed by removing an open set within *every* open neighborhood of *every* point in space X, and possibly removing more. Intuitively, we can think of a topologically negligible region as a slice of "hyper" Swiss cheese that is incredibly full of open holes. A region is said to cover **almost everywhere** in X iff it excludes only a topologically negligible region within X.

Now I can define a new epistemic ideal: Say that a learning method M for an empirical problem converges to the truth **almost everywhere** if M converges to the truth almost everywhere on the space of the states that make h true—for *every* hypothesis h considered in the problem. The idea is that we hope to make it almost everywhere—*whichever* hypothesis is true.

Here is another mode of convergence that concerns *where* convergence happens. Say that a learning method converges to the truth on a **maximal domain** if there exists no learning method that converges to the truth in the same states and in more states.

Now let's consider *how* convergence might happen. Say that a learning method achieves **perfectly monotonic** convergence to the truth if, whenever it outputs one of the competing hypotheses (rather than suspends judgment), it outputs the truth and would then continue to have the same true output were the inquiry to unfold indefinitely. That sounds like a very high ideal. A lower one is this: a learning method is said to achieve **stable** convergence to the truth if, whenever it outputs a *true* competing hypothesis, it would then continue to have the same true output were the inquiry to unfold indefinitely. What if it outputs a falsehood? Stable convergence is silent about that case, while perfectly monotonic convergence rules out that case. So "perfectly monotonic" implies "stable" but not the other way round. To be stable is to be somewhat monotonic but not necessarily perfectly so. Plato would probably love stable convergence to the truth, for it basically requires that, whenever the inquirer forms a belief in the true hypothesis, this belief is not merely a true opinion but has been "stabilized" or "tethered" to the truth, attaining the epistemic status that Plato values in *Meno*.

Finally, by 'no requirement' in Figure 3, I mean no requirement on how to converge.

The above finishes the sketch of the three modes of convergence on each of the two axes in Figure 3. So there are nine "combined" modes of convergence arranged into a two-dimensional lattice structure, in which some modes are ordered higher than some others. Here I adopt the convention of the Hasse diagram, according to which a higher node is one that we can reach by going up along edges without going down. A mode, if ordered higher, is mathematically stronger; it implies all the modes ordered lower in the lattice.

3.4. An argument for full enumerative induction. I believe that there are more modes of convergence as epistemic ideals, but the above are enough for the present purpose—for developing an explicit argument for full enumerative induction. Here you go:

A Learning-Theoretic Argument

1. (Evaluative Premise) In the lattice depicted in Figure 5, if a mode of convergence to the truth is ordered higher, it is a higher epistemic ideal. (This fixes the moving parts of learning-theoretic epistemology.)

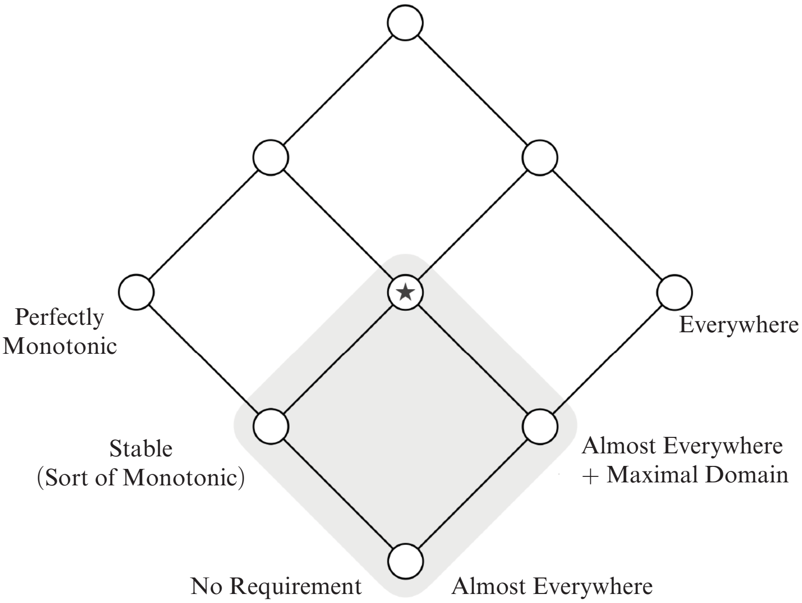

Fig. 5. Modes achievable for the hard raven problem.

2. (Evaluative Premise) For tackling any empirical problem, the best learning methods must at least achieve, of the nine epistemic ideals in the lattice, the highest achievable one—when such a highest one exists uniquely. (This implements the core thesis of learning-theoretic epistemology.)
3. (Mathematical Premise) By Theorem 7.6 of this paper, we have:
   - 3.1 For tackling the hard raven problem, the achievable modes in that lattice are exactly the four in the shaded area depicted in Figure 5.
   - 3.2 Furthermore, the starred mode—“almost everywhere” + “maximal domain” + “stable”—is achieved only by learning methods that implement full enumerative induction rather than counterinduction or the skeptical policy.
4. (Evaluative Conclusion) So, the best learning methods for tackling the hard raven problem must have at least the following properties:
   - 4.1 achieving the starred mode of convergence (by premises 1, 2, and 3.1);
   - 4.2 implementing full enumerative induction, rather than counterinduction or the skeptical policy (by premise 3.2 and the preceding clause 4.1).

This, I believe, is the first explicit argument developed in formal epistemology for *full* enumerative induction, against counterinduction, and against the skeptical policy.[12]

Now let me turn to some applications of the proposed framework.

3.5. Applications: Ockham’s razor and Bayes’ priors. One application leads to a justification of a kind of Ockham’s razor. A theorem to be presented below,

[12] I developed a very similar learning-theoretic argument to solve a problem about causal inference that parallels the Cartesian problem of induction. For details, see Lin [26], whose argument relies on a mathematical premise proved in Lin and Zhang [27].

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

Theorem 8.3, implies that, for tackling any problem, following Ockham's razor of a certain kind is necessary for achieving any mode of convergence to the truth that is strong enough to imply "almost everywhere" + "stable." This kind of Ockham's razor says roughly this: "Do not accept a competing hypothesis that is more complicated than necessary for fitting the data you have," in a sense of complexity made precise in Section 8. In light of the result just sketched, a learning-theoretic epistemologist can argue for the following evaluative thesis:

> The best learning methods for tackling a problem $\mathcal{P}$ must implement the kind of Ockham's razor just mentioned if, with respect to $\mathcal{P}$, the highest achievable mode of convergence to the truth implies "almost everywhere" + "stable."

The connection between Ockham's razor and enumerative induction is that any instance of counterinductive inference violates the kind of Ockham's razor in question, as we will see in Section 8.

Here is a second application. Bayesians can go learning-theoretic, and I propose that they do so. In fact, the results sketched above all have their Bayesian counterparts, stated in Section 9. What emerges is a view that is both Bayesian and learning-theoretic. To be more precise, consider cognitively idealized agents—those who have sharp, real-valued degrees of belief in propositions closed under certain logical connectives (such as 'and', 'or', 'not'). According to Bayesianism, those agents can be epistemically evaluated as coherent or not. Versions of Bayesianism may differ in terms of what counts as coherent—that is quite familiar. What is less familiar is that there can be a version of Bayesianism that adds the following:

> Learning-Theoretic Bayesianism. Cognitively idealized agents can also be evaluated as good or bad at tackling one or another empirical problem. The best such agents for tackling an empirical problem $\mathcal{P}$ must have at least the following properties:
>
> (i) having a coherent prior and the plan to update it by conditionalization on evidence and
>
> (ii) having a prior that, via conditionalization, achieves a highest achievable mode of convergence to the truth with respect to $\mathcal{P}$, if such a highest mode exists.

This view, which may be called *learning-theoretic Bayesianism*, seems to be a view already implicit in the minds of some Bayesians who care about designing priors that are good in terms of convergence to the truth [7, 8].[13] Clause (i) is a distinctively Bayesian thesis. Clause (ii) is a learning-theoretic thesis; in fact, it is the Bayesian version of the core thesis of learning-theoretic epistemology. The epistemological idea is still learning-theoretic; it is just that, now, the doxastic modeling in use is quantitative rather than qualitative. I believe that learning-theoretic Bayesianism deserves to be defended against other varieties of Bayesianism, but that has to be reserved for another paper. What's important for now is that learning-theoretic Bayesianism can be used, together with Theorem 9.4, to argue for the adoption of an inductive prior—rather than a counterinductive or skeptical prior—for tackling the hard raven problem.

[13] Kelly ([21, ch. 13]) can also be interpreted as a pioneer of this view.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

The above summarizes the main mathematical results and the way they are employed to solve the Cartesian problem of induction, *assuming* learning-theoretic epistemology. The rest of this paper is devoted to the mathematical details.

§4. Learning theory: The basics.

This section reviews some definitions familiar in formal learning theory.

Definition 4.1. *An (empirical)* ***problem*** *is a tuple $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ consisting of three components:*[14]

1. *a hypothesis space $\mathcal{H}$, which is a set of competing* ***hypotheses***;
2. *an evidence space $\mathcal{E}$, which is a set of finite* ***data sequences*** *$(e_1, \dots, e_n)$; and*
3. *a state space $\mathcal{S}$, which is a set of possible* ***states*** *of the world taking the form $(h, \vec{e})$, where*
   - *h, called the* ***true*** *hypothesis in this state, is an element of $\mathcal{H}$;*
   - *$\vec{e}$, called the* ***data stream*** *produced in this state, is an infinite sequence of data, written $\vec{e} = (e_1, e_2, e_3, \dots)$, whose finite initial segments $(e_1, \dots, e_n)$ are all in $\mathcal{E}$.*

The state space $\mathcal{S}$ captures the *presupposition* of the problem—it is the set of all possible ways for the inquiry to unfold indefinitely without violating the presupposition.

Definition 4.2. *A* ***learning method*** *for a problem $(\mathcal{H}, \mathcal{E}, \mathcal{S})$ is a function*

$$M : \mathcal{E} \to \mathcal{H} \cup \{\texttt{?}\},$$

where ? *represents suspension of judgment. Given each data sequence $(e_1, \dots, e_n) \in \mathcal{E}$, the output of M is written $M(e_1, \dots, e_n)$.*

Example 4.3. *The* ***hard raven problem*** *poses this question: "Are all ravens black?" This problem was already informally represented by the tree structure in Figure 2 and can now be formally defined as follows.*

- *The hypothesis space $\mathcal{H}$ is $\{\texttt{Yes}, \texttt{No}\}$, where:*
  - `Yes` *means that all ravens are black.*
  - `No` *means that not all ravens are black.*
- *The evidence space $\mathcal{E}$ consists of all finite sequences of* +, 0, *and/or* −, *where:*
  - *datum* + *denotes the observation of a black raven (a positive instance);*
  - *datum* −, *a nonblack raven (a negative instance); and*
  - *datum* 0, *a nonraven (a non-instance).*
- *The state space $\mathcal{S}$ consists of all states in any one of the following three categories*[15]:
  - (*a*) *the states $(\texttt{Yes}, \vec{e})$ in which $\vec{e}$ is an infinite* +/0 *sequence (namely, an infinite sequence whose entries are either* + *or* 0*),*

[14] Note that, although the concept of problems as defined here is enough for present purposes, it is too narrow to cover the problems that entertain statistical hypotheses.

[15] If you wish, there is a fourth category: the states $(\texttt{Yes}, \vec{e})$ in which $\vec{e}$ contains an occurrence of − (a nonblack raven). But such states are logically impossible, so they need not be considered.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

(*b*) *the states* $(\texttt{No}, \vec{e})$ *in which* $\vec{e}$ *is an infinite* `+/0` *sequence, and*
(*c*) *the states* $(\texttt{No}, \vec{e})$ *in which* $\vec{e}$ *is an infinite* `+/0/-` *sequence that contains at least one occurrence of* `-`.

The second category (*b*) *contains the states in which there are nonblack ravens but the inquirer would never observe one even if the inquiry were to extend indefinitely, so they are the* ***Cartesian scenarios of induction***.

Example 4.4. *The* ***easy raven problem*** *is basically the same as the hard raven problem except that its state space consists only of the states in categories* (*a*) *and* (*c*), *ruling out the Cartesian scenarios of induction. So the easy raven problem* presupposes *that the inquirer is not living in a Cartesian scenario of induction. It poses this question*: "Suppose that you are not living in a Cartesian scenario of induction, then are all ravens black?"

Definition 4.5. *Let M be a learning method for a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$. M is said to* ***converge to the truth*** *in a state $(h, \vec{e}) \in \mathcal{S}$ if*

$$\lim_{n\to\infty} M(e_1, \dots, e_n) = h\,,$$

namely, there exists a positive integer k such that, for each $n \geq k$, we have that $M(e_1, \dots, e_n) = h$. M is said to converge to the truth ***everywhere*** *for $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ if it converges to the truth in every state in the state space $\mathcal{S}$.*

Then it is routine to prove the following negative result:[16]

Proposition 4.6. *Although the easy raven problem has a learning method that converges to the truth everywhere, the hard raven problem does not.*[17]

§5. A topological conception of "almost everywhere".

When convergence to the truth cannot be achieved everywhere, let's see whether it can be achieved at least *almost* everywhere. This section explicates the concept of "almost everywhere" in terms of the concept of open neighborhoods, which is in turn explicated by the concept of closeness.

Consider two data streams $\vec{e}$ and $\vec{e}\,'$. Suppose that they are identical up until sample size n, that is, $e_i = e'_i$ for each $i \leq n$ but $e_{n+1} \neq e'_{n+1}$. Note that the larger n is, the later the point of departure is and the more data one needs to distinguish those two data streams. So, the larger n is, the harder it is to *empirically distinguish* those two data streams, and the closer $\vec{e}$ is to $\vec{e}\,'$—closer in the empirical sense (recall the left side of Figure 4). Accordingly, we have:

Definition 5.1. *If two data streams $\vec{e}$ and $\vec{e}\,'$ are identical up until sample size n, say that they are* ***empirically close*** *to degree n.*

[16] *Proof*. The first part is a classic, well-known result in learning theory. To prove the second part, let $\vec{e}$ be an infinite `+/0` sequence. Consider state $s = (\texttt{Yes}, \vec{e})$ and its Cartesian counterpart $s' = (\texttt{No}, \vec{e})$. Let M be an arbitrary learning method for the hard raven problem. It suffices to note that M converges to the truth in s iff M fails to do so in s' (because these two states are empirically equivalent).

[17] This impossibility result remains even if we resort to *partial* identification in the limit, a weakening of identification in the limit due to [29], which requires that, in each possible way for the inquiry to unfold indefinitely, the true hypothesis is output infinitely often and each false hypothesis is output at most finitely often.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

This allows us to define the empirical closeness between states, which can then be used to define open neighborhoods of states, as follows:

Definition 5.2. *Consider a set of states, denoted by X. Two states in X are* ***empirically close*** *to at least degree n iff their associated data streams are empirically close to at least n. Consider an arbitrary set X of states. Given a state $s \in X$ and a sample size n, we can construct the set of the states in X that are empirically close to s to at least degree n (recall the right side of Figure 4). Such a set is called a* ***basic open neighborhood*** *of state s, also called a* ***basic open set*** *in space X. The basic open sets in X form a collection called the* ***empirical topological base*** *of X.*

It is routine to verify that the empirical topological base of any state space is indeed a topological base as standardly defined.[18]

Of particular interest is the case in which the set of states in question, X, is the set of states that make a fixed hypothesis true. Consider an arbitrary hypothesis h in a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$. Let $|h|$ be the set of states that make h true; or formally:

$$|h| = \{(h', \vec{e}\,') \in \mathcal{S} : h' = h\}.$$

Let me introduce a convenient way to express the basic open neighborhoods in $|h|$. Choose a finite data sequence $(e_1, \dots, e_n)$. Construct the set of the states in $|h|$ whose data stream extends $(e_1, \dots, e_n)$. If this set is nonempty, we have a basic open neighborhood. All basic open neighborhoods in $|h|$ can be constructed that way. More formally, define

$$|(e_1, \dots, e_n)| = \{(h', \vec{e}\,') \in \mathcal{S} : \vec{e}\,' \text{ extends } (e_1, \dots, e_n)\}.$$

We can understand $|h|$ and $|(e_1, \dots, e_n)|$ as the propositions expressed by a hypothesis and a body of evidence, respectively (conditional on the presupposition of the problem).[19] It is routine to verify that the basic open neighborhoods in $|h|$ are exactly the nonempty sets taking this form: $|h| \cap |(e_1, \dots, e_n)|$, which is the set of the states that make h true and have a data stream that extends $(e_1, \dots, e_n)$.

Once we have a criterion of what counts as a basic open neighborhood in a space X, it can be used to define a topological conception of "negligible." The idea, put informally, is that a subset of a space X is said to be negligible if it can be constructed from what may be called a *hyper hole-punching procedure*:

1. Start from the entire topological space X in question.
2. For any point $x \in X$, and any open neighborhood N of x (however small N may be), punch a hole by removing a basic open set within N.
3. If you wish, remove some more points.

Examples of topologically negligible sets include straight lines in the two-dimensional Euclidean space (for a straight line can be constructed by removing an appropriate open disc from every open disc). This idea can be put formally as follows:

[18] Given a set X of points, a family $\mathcal{B}$ of subsets of X is called a **topological base** if, first, every point in X is contained in some set in $\mathcal{B}$ and, second, for any $B_1, B_2 \in \mathcal{B}$ and any point $x \in B_1 \cap B_2$, there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

[19] If you wish, the concept of problems and other learning-theoretic concepts can be defined purely in terms of propositions, as done in Baltag et al. [1] and Kelly et al. [24].

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

Definition 5.3. *Let X be a (topological) space equipped with a topological base $\mathcal{B}_X$. A* ***negligible*** *(or* ***nowhere dense****) region within X is a subset X' of X such that every nonempty basic open set in $\mathcal{B}_X$ includes a nonempty basic open set in $\mathcal{B}_X$ that is disjoint from X'. If a subset of X can be expressed as $X \setminus X'$ for some negligible region X', say that it covers* ***almost everywhere*** *in X, and that it contains* ***almost all*** *points in X.*[20]

With the above definitions, we have:

Lemma 5.4. *Consider the hard raven problem. The set of the Cartesian scenarios of induction is one of the (many) negligible regions within the space $|\mathtt{No}|$ equipped with the empirical topological base.*

Proof. Recall that a Cartesian scenario of induction is a state $(\mathtt{No}, \vec{e})$ with $\vec{e}$ being a $\mathtt{+}/\mathtt{0}$ sequence, which contains no occurrence of $\mathtt{-}$ (nonblack raven). So any set of the form $|\mathtt{No}| \cap |(e_1, \dots, e_n, \mathtt{-})|$ is disjoint from the set of the Cartesian scenarios of induction. To finish the proof, it suffices to note that each nonempty basic open set in the topological space $|\mathtt{No}|$, say $N = |\mathtt{No}| \cap |(e_1, \dots, e_n)|$, includes a nonempty basic open set, namely $N' = |\mathtt{No}| \cap |(e_1, \dots, e_n, \mathtt{-})|$, which excludes all Cartesian scenarios of induction. □

This is a lemma for proving the main result of this paper. Note that I do *not* wish to infer from "*that* is topologically negligible" to "we should ignore *that*." To infer that way is to ignore all possible states of the world, because every state forms a singleton that is topologically negligible. And it is absurd to ignore all states. A good justification for full enumerative induction has to be formulated in a more careful way, such as the learning-theoretic argument in Section 3.4.

§6. Mode of convergence: "Almost everywhere". The preceding section is purely explicatory: it provides an explication of the concept of "almost everywhere," and gives no epistemology at all. Epistemology starts from here: using the above explication of "almost everywhere" to define an epistemic ideal.

Definition 6.1. *A learning method M for a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ is said to converge to the truth* ***almost everywhere*** *if, for each hypothesis $h \in \mathcal{H}$, M converges to the truth almost everywhere on the topological space $|h|$, with respect to the empirical topology.*

I propose this definition as an *explication* of the (informal) epistemic ideal of convergence to the truth in almost all possible states. Perhaps there are other good explications of "almost all"—they should be explored. If we can develop a better explication than the present one, we should use it instead. But I tend to think that the present proposal is at least a natural starting point. If the present explication turns out to be only one of many good standards of "almost all," then we should try to see whether it is possible to meet all of those standards simultaneously.[21] The reason is learning-theoretic: let's achieve them all, if the best possible is to achieve them all.

[20] For a survey of various conceptions of "almost everywhere" developed in topology and measure theory, see [30].

[21] For example, instead of requiring convergence to happen almost everywhere on the space $|h| \subseteq \mathcal{S}$ for *each* hypothesis $h \in \mathcal{H}$, we can require convergence to happen almost everywhere on the entire state space $\mathcal{S}$, with respect to the empirical topology. So we have two definitions of almost everywhere convergence: one considers the entire state space, and the other considers each hypothesis. But we do not have to choose between those two definitions.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

The above definition makes possible a series of positive results. Here is the first one:

Proposition 6.2. *The hard raven problem has a learning method that converges to the truth almost everywhere.*

Proof. By Lemma 5.4, the topological space $|\texttt{No}|$ has the following negligible region:

$$C = \{(\texttt{No}, \vec{e}) : \vec{e} \text{ is a } \texttt{+}/\texttt{0} \text{ sequence}\},$$

which consists of the *C*artesian scenarios of induction. So it suffices to give an example of a learning method that converges to the truth in

every state in $|\texttt{Yes}|$ and
every state in $|\texttt{No}| \setminus C$.

The following learning method does the job:

> M^*: "Output hypothesis $\texttt{Yes}$ if you haven't seen a nonblack raven ($\texttt{-}$); otherwise output $\texttt{No}$."

This finishes the proof. □

It should be noted that the above is only a first step toward justifying full enumerative induction. For there remains a subproblem, which may be called the *problem of counterinduction*. Let me illustrate with an example.

Example 6.3. *Almost everywhere convergence to the truth, alone, imposes only a weak constraint, too weak to rule out certain counterinductive methods, such as this one*:

> $M^\dagger$: *"Output hypothesis* $\texttt{No}$ *if everything you have seen is a black raven* ($\texttt{+}$); *otherwise output whatever* M^* *outputs."*

This counterinductive method also converges to the truth almost everywhere for the hard raven problem. To see why, define the following pair of a Cartesian scenario and its normal counterpart:

$$s_{normal} = (\texttt{Yes}, \textit{the constant sequence of } \texttt{+}),$$
$$s_{cartesian} = (\texttt{No}, \textit{the constant sequence of } \texttt{+}).$$

It is not hard to verify that the counterinductive method defined above, $M^\dagger$, *converges to the truth in*

every state in $|\texttt{Yes}| \setminus \{s_{normal}\}$,
every state in $|\texttt{No}| \setminus (C \setminus \{s_{cartesian}\})$.

$\{s_{normal}\}$ *is negligible within* $|\texttt{Yes}|$, *because every singleton within* $|\texttt{Yes}|$ *is.* $C \setminus \{s_{cartesian}\}$ *is negligible within* $|\texttt{No}|$, *because the larger set C is* (*by Lemma* 5.4).

The above example suggests that, to rule out every counterinductive method, we need to look for additional modes of convergence.

If both are good explications of "almost everywhere," then we should view them as two distinct epistemic ideals for us to strive for where possible. When we can achieve both at the same time, we should do it. It turns out that, if we were to consider both versions of almost everywhere convergence, the main theorems of this paper would remain the same. I am indebted to Alan Hájek for discussion of this idea.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

§7. Modes of convergence: "Stable" and "maximal". Before one's opinion converges to the truth, it might be false, or it might be true but to be retracted as evidence accumulates. But when one's opinion has converged to the truth, it is "tied" to the truth and will not "run away," which seems to be epistemically valuable. Hence the following definition:

DEFINITION 7.1. *A learning method M is said to* ***have converged*** *to the truth by the n-th stage of inquiry in a state $s = (h, \vec{e})$ if*

$$M(e_1, \dots, e_i) = h \quad \text{for each } i \geq n.$$

Then we can distinguish the following two epistemic ideals:

DEFINITION 7.2. *A learning method M for a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ is said to converge to the truth with* ***perfect monotonicity*** *if*

> *for any state $s = (h, \vec{e})$ and any stage n such that $M(e_1, \dots, e_n) \neq ?$, M has converged to the truth by stage n in state s.*[22]

Say that M converges to the truth with ***stability*** *if the following (weaker) condition holds:*

> *for any state $s = (h, \vec{e})$ and any stage n such that $M(e_1, \dots, e_n) = h$ (i.e., the truth in s), M has converged to the truth by stage n in state s.*[23]

Stable convergence is sort of monotonic but not necessarily perfectly so, while perfect monotonicity is quite demanding. Indeed, when a learning method for a problem achieves *everywhere* convergence to the truth with *perfect monotonicity*, it is basically what a computability theorist would call an *effective procedure* for solving that problem if we ignore whether the learning method in question is a computable function. That combination of modes, "everywhere" + "perfectly monotonic," is a great thing to have whenever achievable. But it is too demanding to be achievable for any problem that is inductive in nature, such as the hard raven problem. In fact, we have a stronger negative result:

PROPOSITION 7.3. *For the hard raven problem, it is impossible to simultaneously achieve the following two modes of convergence to the truth: "almost everywhere" and "perfectly monotonic."*

Here is another desirable mode of convergence:

[22] Everywhere convergence plus perfect monotonicity is equivalent to what's called *finite identifiability*, studied by Mukouchi [28] and Lange and Zeugman [25].

[23] The concept of stable convergence to the truth is independently developed by Genin [10], who also develops a stochastic version of it, called *progressiveness*. Stable convergence to the truth is closely related to—but significantly different from—some properties that have been studied in learning theory: Putnam's [35] and Schulte's [39] "mind-change"; Kelly and Glymour's [23] "retraction"; Kelly et al.'s [24] "opinion cycle." Those properties penalize every retraction of a hypothesis. But not every retraction is bad: the retraction of a falsehood is good. In contrast, stable convergence to the truth only penalizes the retraction of a truth, and rightly so. Closely related is the "no U-shaped learning" condition studied in Carlucci et al. [2, 3], but it penalizes the retraction of a truth only when one returns to the truth afterwards (forming a U-turn). Stable convergence to the truth penalizes the retraction of a truth whether or not there will be U-turn, and rightly so.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

DEFINITION 7.4. *A learning method M for a problem is said to converge to the truth on a* ***maximal domain*** *if there is no learning method for the same problem that converges to the truth in all states where M does and in more states.*

Then we have:

PROPOSITION 7.5. *The hard raven problem has a learning method that converges to the truth (i) almost everywhere, (ii) on a maximal domain, and (iii) with stability. Every such learning method M has the following properties*:

1. *M is* ***never counterinductive*** *in that, for any data sequence $(e_1, \ldots, e_n)$ that has not witnessed a nonblack raven, $M(e_1, \ldots, e_n) \neq$* `No`;[24]
2. *M is* ***enumeratively inductive*** *(and thus non-skeptical) in that, for any data stream $\vec{e}$ that never witnesses a nonblack raven, $M(e_1, \ldots, e_n)$ converges to* `Yes` *as $n \to \infty$.*

The idea of proof is explained in Appendix A.1, which I have tried to make as instructive as possible for those new to learning theory. The results of this section are proved in Appendix A.2.

Given Propositions 4.6, 7.3, and 7.5, the first main result follows immediately:

THEOREM 7.6 (Hard Raven Theorem). *Consider the modes of convergence to the truth arranged in the lattice in Figure 5. The four modes in the shaded area are exactly those achievable for the hard raven problem. For a learning method to achieve the strongest of those four, namely "almost everywhere" + "maximal" + "stable," a necessary condition is to implement full enumerative induction, rather than counterinduction or the skeptical policy.*

The above result is the mathematical premise of the learning-theoretic argument for full enumerative induction, as formulated above in Section 3.4. The impossibility of everywhere convergence to the truth implies that convergence to the truth has to be sacrificed in some possible states, preferably only on *some* negligible domain. Some, but *which*? The above result gives an answer: to achieve all those three modes of convergence to the truth (i)–(iii), we have no alternative but to sacrifice convergence to the truth in exactly the Cartesian scenarios of induction, at least when we are tackling the hard raven problem.

It should be noted that, like many results in learning theory, the above result is sensitive to the problem under discussion. If we turn to a problem that is even harder than the hard raven problem—too hard to allow the achievement of almost everywhere convergence—then the account I propose here will be *silent* on that problem. See Appendix A.5 for an example. Learning-theoretic epistemology provides evaluations that are sensitive to the empirical problem under discussion.

The rest of this paper is devoted to two interesting extensions of the above account: First, my argument against counterinduction can be generalized to a justification for a kind of Ockham's razor. Second, every important result in the above can be easily modified to obtain a Bayesian counterpart.

[24] It is worth mentioning that the proof of this clause (as provided in Appendix A.2) gives a stronger result: setting aside "maximal domain," the achievement of the two modes "almost everywhere" and "stable" is already sufficient for being never counterinductive.

§8. "Almost everywhere" + "stable" ⇒ "Ockham". One of the results in the above (clause 1 of Proposition 7.5) has been used to justify a norm that says when to follow this methodological principle:

> *Do not accept a counterinductive hypothesis.*

When to follow that principle? At least when tackling the hard raven problem. This section presents a strengthened result, which can be used to justify a norm that says when one ought to follow a certain version of Ockham's razor, whose informal idea can be expressed as follows:

> *Do not accept a hypothesis if it is more complicated than necessary for fitting data, where being more complicated means, roughly, having a higher capacity for fitting data.*[25]

Let me start by defining the version of Ockham's razor I have in mind.

Definition 8.1. *Let a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ be given. A data sequence $(e_1, \dots, e_n)$ and a hypothesis h are said to be **compatible** if the propositions they express have a nonempty overlap, which means that there exists a state in $\mathcal{S}$ that makes hypothesis h true and produces data sequence $(e_1, \dots, e_n)$. For each hypothesis $h \in \mathcal{H}$, let $\mathcal{E}(h)$ denote the set of data sequences in $\mathcal{E}$ that are compatible with h (so $\mathcal{E}(h)$ captures the data-fitting capacity of h). The **empirical simplicity order**, written $\prec$, is defined on $\mathcal{H}$ as follows: for all hypotheses h and $h' \in \mathcal{H}$,*

$$h \prec h' \quad \textit{iff} \quad \mathcal{E}(h) \subset \mathcal{E}(h').$$

*Or in words, h is **simpler** then h' iff h "fits" strictly less data sequences than h' does. Say that h is **no more complex** than h' if $h' \nprec h$.*

In the hard raven problem, for example, the inductive hypothesis `Yes` is simpler than the counterinductive hypothesis `No`.

The version of Ockham's razor I have in mind asks one to accept a hypothesis h only if, first, h is no more complex than necessary for fitting the available data and, second, h will continue to be accepted until it is refuted by the accumulated data. More formally, we have:

Definition 8.2. *A learning method M for a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ is said to follow **Ockham's tenacious razor** just in case it satisfies the following two conditions:*

1. *(Razor Condition) A hypothesis h is the output of M given a data sequence $(e_1, \dots, e_n)$ only if h is no more complex than any hypothesis in $\mathcal{H}$ that is compatible with $(e_1, \dots, e_n)$.*
2. *(Tenacity Condition) A hypothesis h is the output of M given a data sequence $(e_1, \dots, e_n)$ only if h continues to be the output of M given any data sequence in $\mathcal{E}$ that extends $(e_1, \dots, e_n)$ and stays compatible with h.*

[25] This informal idea of Ockham's razor is quite general and admits of different implementations. The implementation below proceeds by defining the fitting relation in terms of logical compatibility. This basically follows Popper's [33] account of Ockham's razor, according to which the easier it is to falsify a hypothesis, the simpler that hypothesis is. In the statistical context, the more standard implementation defines the complexity of a hypothesis/model by the number of adjustable parameters. See ([9, Sections 2 and 7]) for discussion of the relevant references in the statistical literature.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

In the hard raven problem, to comply with the razor condition is exactly to be never counterinductive.

Then we have the second main result, which I call the *Ockham stability theorem*:[26]

THEOREM 8.3 (Ockham Stability Theorem). *Let M be a learning method for a problem. Suppose that M converges to the truth almost everywhere. Then the following two conditions are equivalent*:

1. *M converges to the truth with stability.*
2. *M follows Ockham's tenacious razor.*

So, given almost everywhere convergence, stable convergence is the weakest normative standard that is strong enough to enforce Ockham's tenacious razor—strong enough, thanks to the 1 ⇒ 2 side; weakest, thanks to the 2 ⇒ 1 side.

The 1 ⇒ 2 side of the Ockham stability theorem has an application: it can be used to prove, as an immediate corollary, the "never be counterinductive" part of Proposition 7.5. For, when tackling the hard raven problem, to be never counterinductive is exactly to comply with the razor condition.

The result of this section is proved in Appendix A.3.

§9. Learning-theoretic Bayes. Instead of tethering one's qualitative beliefs to the truth, learning-theoretic epistemology can talk about tethering one's probabilistic degrees of belief to high accuracy. The epistemological idea is the same; it is just a change of the doxastic modeling in use. So it should not be surprising that the important results presented above all have their Bayesian counterparts, as the following shows.

DEFINITION 9.1. *Let a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ be given. Recall that $|h|$ is the set of states in $\mathcal{S}$ that make hypothesis h true, and $|(e_1, \dots, e_n)|$ is the set of states in $\mathcal{S}$ that produce data sequence $(e_1, \dots, e_n)$. Let $\mathcal{A}_\mathcal{P}$ denote the smallest σ-algebra that contains the above propositions for all $h \in \mathcal{H}$ and all $(e_1, \dots, e_n) \in \mathcal{E}$. Given a probability function $\mathbb{P}$ defined on that algebra, I will write $\mathbb{P}(h)$ as a shorthand for $\mathbb{P}(|h|)$. Similarly, I will write $\mathbb{P}(e_1, \dots, e_n)$ and $\mathbb{P}(h \mid e_1, \dots, e_n)$, where the latter stands for conditional probability as standardly defined.*[27]

DEFINITION 9.2. *A* ***probabilistic prior*** *for a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ is a probability function $\mathbb{P}$ defined on σ-algebra $\mathcal{A}_\mathcal{P}$ with $\mathbb{P}(e_1, \dots, e_n) > 0$ for each data sequence $(e_1, \dots, e_n) \in \mathcal{E}$. $\mathbb{P}$ is said to (have its posteriors)* ***converge to the truth*** *in a state $s = (h, \vec{e}) \in \mathcal{S}$ if*

$$\lim_{n \to \infty} \mathbb{P}(h \mid e_1, \dots, e_n) = 1,$$

that is, for any $\varepsilon > 0$, there exists a positive integer k such that, for each $n \geq k$, $\mathbb{P}(h \mid e_1, \dots, e_n) > 1 - \varepsilon$. $\mathbb{P}$ is said to converge to the truth ***everywhere*** *for problem $(\mathcal{H}, \mathcal{E}, \mathcal{S})$ if it converges to the truth in each state in $\mathcal{S}$. $\mathbb{P}$ is said to converge to the truth* ***almost***

[26] This theorem extends and strengthens some aspects of my earlier work with co-authors [24] in order to address the hard raven problem and the like. But this theorem also simplifies and weakens other aspects in order to highlight the core idea; in particular, the concept of Ockham's razor used here is simpler and weaker (but strong enough for present purposes).

[27] Namely, $\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$, if $\mathbb{P}(B) \neq 0$.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

everywhere *for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ if, for each hypothesis $h \in \mathcal{H}$, $\mathbb{P}$ converges to the truth almost everywhere on the topological space $|h|$.*

Definition 9.3. *Let $\mathbb{P}$ be a probabilistic prior for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$. $\mathbb{P}$ is said to* ***have started to stably converge*** *to the truth by stage n in state $s = (h, \vec{e}) \in \mathcal{S}$ if*

1. $\mathbb{P}(h \mid e_1, \dots, e_n, \dots, e_{n+i})$ *is monotonically increasing as a function of i defined on $\mathbb{N}$.*
2. $\mathbb{P}(h \mid e_1, \dots, e_n, \dots, e_{n+i})$ *converges to 1 as $i \to \infty$.*

$\mathbb{P}$ is said to converge to the truth with ***stability*** *if, for each hypothesis $h \in \mathcal{H}$, for each state $s = (h, \vec{e}) \in \mathcal{S}$ that makes h true, and for each stage n as a positive integer, if $\mathbb{P}(h \mid e_1, \dots, e_n) > 1/2$, then $\mathbb{P}$ has started to stably converge to the truth by stage n in state s.*

In the above definition of stable convergence, you can replace the occurrence of $1/2$ by any threshold you take to stand for high probability, and do the same for the Bayesian results below. But for concreteness, let me stay with the threshold $1/2$ in this paper.

Theorem 9.4. *The hard raven problem has a probabilistic prior that converges to the truth* (i) *almost everywhere,* (ii) *on a maximal domain, and* (iii) *with stability. Every such probabilistic prior $\mathbb{P}$ has the following properties*:

1. *$\mathbb{P}$ is* ***never counterinductive*** *in that, for any data sequence $(e_1, \dots, e_n)$ that has not witnessed a nonblack raven, $\mathbb{P}(\texttt{No} \mid e_1, \dots, e_n) \leq 1/2$.*
2. *$\mathbb{P}$ is* ***enumeratively inductive*** *in that, for any data stream $\vec{e}$ that never witnesses a nonblack raven, $\mathbb{P}(\texttt{Yes} \mid e_1, \dots, e_n)$ converges to 1 as $n \to \infty$.*

Definition 9.5. *A probabilistic prior $\mathbb{P}$ for a problem $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is said to follow* ***Ockham's tenacious razor*** *just in case the following two conditions hold*:

1. (*Razor Condition*) *$\mathbb{P}(h \mid e_1, \dots, e_n) > 1/2$ only if h is no more complex than any hypothesis in $\mathcal{H}$ that is compatible with the given data sequence $(e_1, \dots, e_n)$.*
2. (*Tenacity Condition*) *$\mathbb{P}(h \mid e_1, \dots, e_n) > 1/2$ only if, for any data sequence $(e_1, \dots, e_n, \dots, e_{n+n'})$ in $\mathcal{E}$ that extends $(e_1, \dots, e_n)$ and stays compatible with h, $\mathbb{P}(h \mid e_1, \dots, e_{n+i})$ is monotonically increasing as a function of $i \in \{0, \dots, n'\}$.*

Theorem 9.6 (Ockham Stability Theorem, Bayesian Version). *Let $\mathbb{P}$ be a probabilistic prior for a problem that converges to the truth almost everywhere. Then, condition* 1 *below implies condition* 2 *below* (*but the converse does not hold*):

1. *$\mathbb{P}$ converges to the truth with stability.*
2. *$\mathbb{P}$ follows Ockham's tenacious razor.*

Although the converse does not hold,[28] we might be able to formulate a stronger version of tenacity or a weaker version of stability in Bayesian terms in order to restore the equivalence between conditions 1 and 2. But that will not be attempted here. For there is no loss in application to epistemology: to justify Ockham's razor, what is really needed is just the implication relation from the epistemic ideal expressed by 1 to the

[28] Here is the reason why the converse does not hold: the tenacity condition—as defined in the Bayesian framework—only requires the posterior probability to remain the same or go up as data accumulate, but does not require it to go up high enough to ensure convergence to 1, let alone convergence with stability.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

methodological principle expressed by 2, which shows that the latter is necessary for achieving the former. The converse does no justificatory work. Showing that Ockham's razor achieves a certain epistemic ideal does not suffice to argue that one *has to* follow Ockham's razor, for there might be other ways to achieve that ideal.

The results of this section are proved in Appendix A.4.

§10. Conclusion.

For tackling the problem of whether all ravens are black, the highest achievable epistemic ideal (among the ideals considered in this paper) is the combination of three modes of convergence to the truth: "almost everywhere" + "maximal domain" + "stable," as depicted in the lattice in Figure 5. A necessary condition for achieving that is to follow full enumerative induction rather than counterinduction or the skeptical policy. And this holds regardless whether the doxastic modeling in use is qualitative or Bayesian.

I believe that learning-theoretic epistemology is a promising approach to the evaluative studies of inductive inferences, but a defense of this philosophical claim has to be reserved for future works. The goal of this paper is, instead, to clearly identify the core thesis of learning-theoretic epistemology and use it to develop the first systematic and mathematically rigorous solution to the Cartesian problem of induction.

§A. Appendix.

A.1. The idea of proof of Proposition 7.5. Proposition 7.5 has three parts. The existential claim is easy to prove, and so is the universal claim about "be enumeratively inductive." The crucial part is the universal claim about "never be counterinductive." The reason why it is crucial is two-fold: first, it is an instructive special case of the 1 ⇒ 2 side of the Ockham stability Theorem 8.3; second, it serves as a lemma for proving the part about "be enumeratively inductive." So let me separate the crucial part for close examination:

Proposition A.1 (Never Be Counterinductive). *Let M be a learning method for the hard raven problem that converges to the truth almost everywhere with stability. Then M is never counterinductive.*

Note that this proposition does not rely on convergence on a maximal domain. It will be convenient to have the following concept:

Definition A.2. *Given a problem $(\mathcal{H}, \mathcal{E}, \mathcal{S})$, a data sequence $(e_1, \dots, e_n) \in \mathcal{E}$ is said to be* ***compatible*** *with a hypothesis $h \in \mathcal{H}$ if the propositions they express have a nonempty overlap, namely*:

$$|h| \cap |(e_1, \dots, e_n)| \neq \varnothing ,$$

which also means that $(e_1, \dots, e_n)$ can be extended into a data stream $\vec{e}$ such that $(h, \vec{e})$ is a state in $\mathcal{S}$.

The proof of the above proposition proceeds as follows. Let M be a learning method for the hard raven problem that converges to the truth almost everywhere. Suppose that M is sometimes counterinductive. That is, for some $\texttt{+}/\texttt{0}$ sequence $(e_1, \dots, e_n)$, we have that:

$$M(e_1, \dots, e_n) = \texttt{No} . \tag{1}$$

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

Since $(e_1, \ldots, e_n)$ is a `+/0` sequence, it is compatible with `Yes`. To summarize, we have had:

- M converges to the truth almost everywhere.
- $(e_1, \ldots, e_n)$ is compatible with `Yes`.

Given these two conditions, we can apply the so-called *forcing lemma* (to be stated soon) in order to "force" M to output `Yes` by extending $(e_1, \ldots, e_n)$ into another `+/0` sequence $(e_1, \ldots, e_n, \ldots, e_{n'})$. So,

$$M(e_1, \ldots, e_n, \ldots, e_{n'}) = \texttt{Yes}. \tag{2}$$

Now, choose a state s such that

$$s \in |\texttt{No}| \cap |(e_1, \ldots, e_n, \ldots, e_{n'})| . \tag{3}$$

We can always make this choice because every data sequence is compatible with `No`. By (1)–(3), we have: in state s, M outputs the truth `No` at stage n and then retracts it at stage n'. So it fails to stably converge to the truth. This finishes the proof of the part "never be counterinductive" in Proposition 7.5—as soon as the forcing lemma is stated and established.

Here is the forcing lemma:

LEMMA A.3 (**Forcing Lemma**). *Let $(\mathcal{H}, \mathcal{E}, \mathcal{S})$ be an arbitrary problem. Suppose that M is a learning method for it that converges to the truth almost everywhere, and that $(e_1, \ldots, e_n) \in \mathcal{E}$ is compatible with $h \in \mathcal{H}$. Then the above data sequence can be extended into a data sequence $(e_1, \ldots, e_n, \ldots, e_{n'}) \in \mathcal{E}$ such that:*

1. *$(e_1, \ldots, e_n, \ldots, e_{n'})$ is still compatible with h and*
2. *$M(e_1, \ldots, e_n, \ldots, e_{n'}) = h$.*

Proof. Suppose that $(e_1, \ldots, e_n)$ is compatible with h. So, $|h| \cap |(e_1, \ldots, e_n)|$ is a nonempty basic open set of topological space $|h|$. We are going to make use of the following characterization of "almost everywhere" in general topology:

> A property applies almost everywhere on a topological space (with a distinguished topological base) if, and only if, each nonempty (basic) open set U has a nonempty (basic) open subset U' such that the property applies everywhere on U'.

So, by the "only if" side and the hypothesis that M converges to the truth almost everywhere, it follows that the basic open set $|h| \cap |(e_1, \ldots, e_n)|$ has a nonempty basic open subset, $|h| \cap |(e_1, \ldots, e_n, \ldots, e_k)|$ on which M converges to the truth everywhere. Now, within this nonempty set, choose an arbitrary state $(h, \vec{e})$. So, in that state, M converges to the truth, and hence there exists a positive integer $n' \geq k$ such that M outputs the truth h at the n'-th stage along data stream $\vec{e}$. That is:

$$M(e_1, \ldots, e_n, \ldots, e_k, \ldots, e_{n'}) = h.$$

To finish the proof, it suffices to note that the above input $(e_1, \ldots, e_n, \ldots, e_k, \ldots, e_{n'})$ is compatible with h because it is an initial segment of the data stream $\vec{e}$ in state $(h, \vec{e})$. □

The above proves the "never be counterinductive" part of Proposition 7.5; now I proceed to sketch the proof of the part "be enumeratively inductive." Suppose that a learning method M converges to the truth almost everywhere with stability (and I will suppose that M converges on a maximal domain right when I really need to). Then, by the preceding result, M is never counterinductive, and hence it fails to converge to the truth in every Cartesian scenario of induction, say $(\texttt{No}, \vec{e})$, where $\vec{e}$ contains no occurrence of a nonblack raven. This failure of convergence in the Cartesian state $(\texttt{No}, \vec{e})$ opens the possibility for M to converge to the truth in its normal counterpart $(\texttt{Yes}, \vec{e})$. To turn this possibility into a reality, introduce the last supposition of the theorem, that M converges to the truth on a maximal domain. It can be shown that, in order for M to converge to the truth on a maximal domain, the domain of convergence of M has to be so expansive that it covers all states that make hypothesis `Yes` true, which implies that M is enumeratively inductive.

Finally, the existential part of Proposition 7.5 is witnessed by the method M^* constructed in Section 6, which says: "Output hypothesis `Yes` if you haven't seen a nonblack raven (`-`); otherwise output `No`."

This finishes the proof sketch of Proposition 7.5.

A.2. Proofs for Section 7: Enumerative induction.

Proof of Proposition 7.3. Suppose, for *reudctio*, that a learning method M for the hard raven problem converges to the truth everywhere with perfect monotonicity. By almost everywhere convergence, M outputs `Yes` on some data sequence $(e_1, \dots, e_n)$. But any data sequence is compatible with $|\texttt{No}|$. So choose a state $s = (\texttt{No}, \vec{e}) \in |\texttt{No}| \cap |(e_1, \dots, e_n)|$. It follows that, in that state $(\texttt{No}, \vec{e})$, M outputs a falsehood (namely, `Yes`) at some stage (namely n), which contradicts perfect monotonic convergence. □

Proof of Proposition 7.5. To establish the existential claim, it suffices to show that it is witnessed by the learning method M^* defined in Section 6: "Output hypothesis `Yes` if you haven't seen a nonblack raven (`-`); otherwise output `No`." It has been established that M^* converges to the truth almost everywhere (by Proposition 6.2). It is routine to verify that M^* converges to the truth with stability. To show that M^* has a maximal domain of convergence, note that it converges to the truth in all states in $|\texttt{Yes}|$ and in all states in $|\texttt{No}|$ except the Cartesian scenarios of induction. No learning method converges to the truth in strictly more states. For to do so is to converge to the truth in some normal state $(\texttt{Yes}, \vec{e})$ and its Cartesian counterpart $(\texttt{No}, \vec{e})$, which is impossible. This establishes that M^* has a maximal domain of convergence, and finishes the proof of the existential claim.

To establish the first part of the universal claim "never be counterinductive," it suffices to note that it has been proved (with all the details) in Appendix A.1, or that it follows immediately from the Ockham stability Theorem 8.3, to be proved in the next appendix.

To establish the second part of the universal claim "be enumeratively inductive," suppose that M is a learning method for the hard raven problem that converges to the truth almost everywhere with stability. So M is never counterinductive, thanks to the first part "never be counterinductive," which has been established. Suppose further that M converges to the truth on a maximal domain. Argue as follows that M is enumeratively inductive. Since M is never counterinductive, M fails to converge to the truth in each Cartesian scenario of induction, and hence its domain of convergence must be included in the domain of convergence of M^*, which has been proved to have a maximal domain of convergence. But M converges on a maximal domain, too,

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

so it must have the same domain of convergence as M^* does. In particular, M must converge to the truth in every state $(\texttt{Yes}, \vec{e})$ contained in $|\texttt{Yes}|$. It follows that M is enumeratively inductive, as desired. □

Theorem 7.6 follows immediately from Propositions 4.6, 7.3, and 7.5.

A.3. Proofs for Section 8: Ockham's Razor. The proof of the Ockham stability theorem relies on the forcing lemma proved in Appendix A.1 and the following lemma:

LEMMA A.4. *The tenacity condition in Ockham's tenacious razor is equivalent to stable convergence to the truth.*[29]

Proof. To argue from tenacity to stability, suppose that M has the tenacity property, and that M outputs the truth h by stage n in state $s = (h, \vec{e})$. It suffices to show that M has converged to the truth by the same stage n in the same state s. Note that, for any natural number i, the data sequence $(e_1, \dots, e_{n+i})$ extends $(e_1, \dots, e_n)$ and is still compatible with h. So, by the tenacity condition, $M(e_1, \dots, e_{n+i}) = h$, for all $i \geq 0$. It follows that, by stage n in state s, M has converged to the truth, h, as desired.

Argue as follows that stability implies tenacity, by contraposition. Suppose that M violates the tenacity condition. That is, $M(e_1, \dots, e_n) = h$ and $M(e_1, \dots, e_n, \dots, e_{n'}) \neq h$, where $(e_1, \dots, e_n, \dots, e_{n'})$ is compatible with h. By that compatibility, choose a state s in the nonempty set $|h| \cap |(e_1, \dots, e_n, \dots, e_{n'})|$. It follows that, by stage n in state s, M outputs the truth h but has not converged to the truth. So M fails to converge to the truth with stability, as desired. □

Proof of Theorem 8.3. Suppose that learning method M converges to the truth almost everywhere for problem $(\mathcal{H}, \mathcal{E}, \mathcal{S})$. The side $2 \Rightarrow 1$ follows immediately from Lemma A.4. To prove the side $1 \Rightarrow 2$ by contraposition, suppose that M does not follow Ockham's tenacious razor. So M violates either the tenacity condition or the razor condition. In the former case (violation of tenacity), M fails to stably converge to the truth, thanks to Lemma A.4. It remains to discuss the latter case (violation of the razor condition), which is the only part of the proof that relies on the hypothesis of almost everywhere convergence.

Suppose that M violates the razor condition, with the goal of showing the failure of stable convergence. Since the razor condition is violated, $M(e_1, \dots, e_n) = h$, for some $(e_1, \dots, e_n) \in \mathcal{E}$ and some $h \in \mathcal{H}$, and there exists another hypothesis $h' \in \mathcal{H}$ that is simpler than h and compatible with $(e_1, \dots, e_n)$. Since $(e_1, \dots, e_n)$ is compatible with h', by the forcing Lemma A.3 and the almost everywhere convergence of M, we have: $(e_1, \dots, e_n)$ can be extended into a data sequence $(e_1, \dots, e_n, \dots, e_{n'}) \in \mathcal{E}$ such that, first, $M(e_1, \dots, e_n, \dots, e_{n'}) = h'$ and, second, $(e_1, \dots, e_n, \dots, e_{n'})$ is compatible with h'. Since $(e_1, \dots, e_n, \dots, e_{n'})$ is compatible with h' and since h' is simpler than h, it follows that $(e_1, \dots, e_n, \dots, e_{n'})$ is also compatible with the more complex hypothesis h. By that compatibility, choose a state $s \in |h| \cap |(e_1, \dots, e_n, \dots, e_{n'})|$. So, in state s, M outputs the truth h at the earlier stage n and then retracts it at or before the later stage n', because $M(e_1, \dots, e_n, \dots, e_{n'}) = h' \neq h$. It follows that M fails to stably converge to the truth. □

[29] I thank Gordon Belot for helping me identify this lemma.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

A.4. Proofs for Section 9: Learning-theoretic Bayes.

Lemma A.5 (**Forcing Lemma, Bayesian Version**). *Let $(\mathcal{H}, \mathcal{E}, \mathcal{S})$ be an arbitrary problem. Suppose that $\mathbb{P}$ is a probabilistic prior for it that converges to the truth almost everywhere, and that $(e_1, \ldots, e_n) \in \mathcal{E}$ is compatible with $h \in \mathcal{H}$. Then the above data sequence can be extended into a data sequence $(e_1, \ldots, e_n, \ldots, e_{n'}) \in \mathcal{E}$ such that:*

1. *$(e_1, \ldots, e_n, \ldots, e_{n'})$ is still compatible with h and*
2. *$\mathbb{P}(h \mid e_1, \ldots, e_n, \ldots, e_{n'}) > 1/2$.*

Proof. The proof is the same as the proof of (the qualitative version of) the forcing Lemma A.3, except that, first, the only occurrence of $M(e_1, \ldots, e_n, \ldots, e_k, \ldots, e_{n'}) = h$ is replaced by $\mathbb{P}(h \mid e_1, \ldots, e_n, \ldots, e_k, \ldots, e_{n'}) > 1/2$ and, second, each occurrence of M is replaced by $\mathbb{P}$. □

Let me prove Theorem 9.6 first before turning to Theorem 9.4.

Proof of Theorem 9.6. In fact, stability alone (without assuming almost everywhere convergence) already implies tenacity. The proof of this claim in the Bayesian setting is basically the same as the proof of the corresponding claim in the qualitative setting. To be more precise, edit the second paragraph of the proof of Lemma A.4 with the following replacements:

- First, replace $M(e_1, \ldots, e_n) = h$ and $M(e_1, \ldots, e_n, \ldots, e_{n'}) \neq h$ by $1/2 < \mathbb{P}(h \mid e_1, \ldots, e_n) > \mathbb{P}(h \mid e_1, \ldots, e_n, \ldots, e_{n'})$.
- And then replace each occurrence of M by $\mathbb{P}$.

It remains to show that almost everywhere convergence with stability implies the razor condition. The proof is basically the same as the proof of the corresponding claim in the qualitative setting. To be more precise, edit the second paragraph of the proof of Theorem 8.3 with the following replacements:

- First, replace $M(e_1, \ldots, e_n) = h$ by $\mathbb{P}(h \mid e_1, \ldots, e_n) > 1/2$.
- Then replace $M(e_1, \ldots, e_n, \ldots, e_{n'}) = h' \neq h$ by $\mathbb{P}(h' \mid e_1, \ldots, e_n, \ldots, e_{n'}) > 1/2$ and $\mathbb{P}(h \mid e_1, \ldots, e_n, \ldots, e_{n'}) < 1/2$.
- As the last step, replace each occurrence of M by $\mathbb{P}$.

This finishes the proof of the 1 ⇒ 2 side.

To prove that the converse 2 ⇒ 1 does *not* hold, construct a problem $\mathcal{P} = (\mathcal{H}, \mathcal{E}, \mathcal{S})$ as follows. Consider only the following (infinite) data streams, where m and n are arbitrary natural numbers:

$$\begin{aligned} s_\omega &= 0^\omega\,, \\ s_m &= 0^m 1^\omega\,, \\ s_{mn} &= 0^m 1^n 2^\omega\,. \end{aligned}$$

Their initial segments form the evidence space $\mathcal{E}$. The hypothesis space $\mathcal{H}$ consists of

- $h =$ "The actual sequence will not end with occurrences of 2."
- $h' =$ "It will."

The state space $\mathcal{S}$ consists of (h, s_ω), (h, s_m), and (h', s_{mn}), for all natural numbers m and n. Construct a countably additive probability function $\mathbb{P}$ that assigns the following probabilities to singletons of states:

$$\begin{aligned}\mathbb{P}\{s_\omega\} &= 0\,,\\ \mathbb{P}\{s_m\} &= \left(\frac{1}{2}\right)^{m+1} \times 60\%\,,\\ \mathbb{P}\{s_{mn}\} &= \left(\frac{1}{2}\right)^{m+1} \times 40\% \times \left(\frac{1}{2}\right)^{n+1}\,.\end{aligned}$$

Those assignments of probabilities are designed to ensure the following:

$$\begin{aligned}\mathbb{P}\{s_m\} &= \left(\frac{1}{2}\right)^{m+1} \times 60\%\,,\\ \mathbb{P}\{s_{m0}, s_{m1}, s_{m2}, \ldots\} &= \left(\frac{1}{2}\right)^{m+1} \times 40\%\,,\\ \mathbb{P}\{s_m, s_{m0}, s_{m1}, s_{m2}, \ldots\} &= \left(\frac{1}{2}\right)^{m+1}\,,\\ \sum_{m=0}^{\infty} \mathbb{P}\{s_m, s_{m0}, s_{m1}, s_{m2}, \ldots\} &= 1\,.\end{aligned}$$

It follows that, for each natural number m, we have:

$$\mathbb{P}(h \mid 0^m) = 60\%\,.$$

So $\mathbb{P}$ fails to converge to the truth H in state $s_\omega = (h, 0^\omega)$. It is routine to verify that $\mathbb{P}$ converges to the truth in all the other states. So $\mathbb{P}$ enjoys almost everywhere convergence. It is also routine to verify that $\mathbb{P}$ follows Ockham's tenacious razor. To see that stable convergence does not hold, note that, in state $s_\omega = (h, 0^\omega)$ and given evidence 0^m, $\mathbb{P}$ assigns a probability greater than $1/2$ (namely 60%) to the truth (namely h) but fails to have started to stably converge to the truth, because it even fails to converge to the truth in that state. So $\mathbb{P}$ fails to converge to the truth with stability. This finishes the proof. □

Proof of Theorem 9.4. To prove the existential claim, construct a witness $\mathbb{P}^*$ as a mixture of two other probabilistic priors:

$$\mathbb{P}^* = \frac{1}{2}\,\mathbb{P}_0 + \frac{1}{2}\,\mathbb{P}_1\,,$$

where $\mathbb{P}_0$ and $\mathbb{P}_1$ are defined as follows. Let $\mathbb{P}_0$ be the probability function generated by assuming that `Yes` is true and that observations of `+`, `0`, `-` are IID (independent and identically distributed) random variables, with equal probabilities $1/2$ for `+` and `0`, and with probability 0 for `-`. So:

$$\begin{aligned}\mathbb{P}_0(\texttt{Yes}) &= 1\,,\\ \mathbb{P}_0(e_1, \ldots, e_n) &= \begin{cases} \left(\frac{1}{2}\right)^n, & \text{if } e_i \neq \texttt{-} \text{ for each } i \leq n,\\ 0, & \text{otherwise.}\end{cases}\end{aligned}$$

Similarly, let $\mathbb{P}_1$ be the probability function generated by assuming that `No` is true and observations of `+`, `0`, `-` are IID random variables with equal probabilities $1/3$ for `+`, `0`,

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

and `-`. So:

$$\mathbb{P}_1(\texttt{No}) = 1\,,$$
$$\mathbb{P}_1(e_1, \dots, e_n) = \left(\frac{1}{3}\right)^n .$$

It suffices to show that $\mathbb{P}^*$, defined as the half-and-half mixture of $\mathbb{P}_0$ and $\mathbb{P}_1$, converges to the truth with all the three modes mentioned in the existential claim. By the construction of $\mathbb{P}^*$, we have:

$$\begin{aligned}
\mathbb{P}^*(\texttt{Yes}) &= 1/2\,,\\
\mathbb{P}^*(\texttt{No}) &= 1/2\,,\\
\mathbb{P}^*(e_1, \dots, e_n \mid \texttt{Yes}) = \mathbb{P}_0(e_1, \dots, e_n) &= \begin{cases} \left(\frac{1}{2}\right)^n, & \text{if } e_i \neq \texttt{-} \text{ for each } i \leq n,\\ 0, & \text{otherwise,}\end{cases}\\
\mathbb{P}^*(e_1, \dots, e_n \mid \texttt{No}) = \mathbb{P}_1(e_1, \dots, e_n) &= \left(\frac{1}{3}\right)^n .
\end{aligned}$$

Now, calculate conditional probability $\mathbb{P}^*(\texttt{Yes} \mid e_1, \dots, e_n)$ by plugging the above probability values into the following instance of Bayes' theorem:

$$\begin{aligned}
&\mathbb{P}^*(\texttt{Yes} \mid e_1, \dots, e_n)\\
&= \frac{\mathbb{P}^*(e_1, \dots, e_n \mid \texttt{Yes})\, \mathbb{P}^*(\texttt{Yes})}{\mathbb{P}^*(e_1, \dots, e_n \mid \texttt{Yes})\, \mathbb{P}^*(\texttt{Yes}) + \mathbb{P}^*(e_1, \dots, e_n \mid \texttt{No})\, \mathbb{P}^*(\texttt{No})} .
\end{aligned}$$

Then we have:

$$\mathbb{P}^*(\texttt{Yes} \mid e_1, \dots, e_n) = \begin{cases} \frac{1}{1+(2/3)^n}, & \text{if } e_i \neq \texttt{-} \text{ for each } i \leq n,\\ 0, & \text{otherwise,}\end{cases} \tag{4}$$

$$\mathbb{P}^*(\texttt{No} \mid e_1, \dots, e_n) = 1 - \mathbb{P}^*(\texttt{Yes} \mid e_1, \dots, e_n)\,, \tag{5}$$

$$\lim_{n\to\infty} \frac{1}{1 + (2/3)^n} = 1\,. \tag{6}$$

By the above three equations, (4)–(6), it follows that $\mathbb{P}^*$ converges to the truth in all states in $|\texttt{Yes}|$, and in all states in $|\texttt{No}|$ except the Cartesian scenarios of induction. But recall that, by Lemma 5.4, the set of the Cartesian scenarios of induction is negligible within the topological space $|\texttt{No}|$. So $\mathbb{P}^*$ converges to the truth almost everywhere. To establish that $\mathbb{P}^*$'s has a maximal domain of convergence, suppose for *reductio* that there is a probability function $\mathbb{P}$ that has a strictly more inclusive domain of convergence than $\mathbb{P}^*$ does. But $\mathbb{P}^*$ converges to the truth in all states except the Cartesian scenarios of induction. So $\mathbb{P}$ must converge to the truth in a certain normal state and in its Cartesian counterpart, which is impossible and finishes the *reductio* argument. Finally, it is routine to verify that $\mathbb{P}^*$ enjoys stable convergence. The existential claim is thus established.

The proof of the universal claim is basically the same as the proof of the corresponding claim in the qualitative setting. To be more precise, part 1 "never be counterinductive" follows immediately from the Bayesian version of the Ockham stability Theorem 9.6. To prove part 2 "be enumeratively inductive," it suffices to edit

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

the last paragraph of the proof of Proposition 7.5 with the following replacements:

- First, replace M by $\mathbb{P}$.
- Second, replace 'learning method' by 'probabilistic prior'.
- As the last step, replace M^* by $\mathbb{P}^*$, which is the probabilistic prior constructed above for proving the existential claim.

This finishes the proof. □

A.5. The very hard raven problem. Here is an example of a problem that is even harder than the hard raven problem:

EXAMPLE A.6. *The **very hard raven problem** poses the following joint questions*:

> *Are all ravens black? If not, will all the ravens observed in the future be black?*

So there are three potential answers:

> `Yes`: *"Yes, all ravens are black."*
> `NoYes`: *"No, not all ravens are black; yes, all ravens to be observed will be black."*
> `NoNo`, *which denies the above two hypotheses.*

`NoYes` *is a Cartesian skeptical hypothesis, a hypothesis that is akin to (but not as terrible as) the proposition that one is a brain in a vat. Hypotheses* `Yes` *and* `NoYes` *are empirically equivalent—they are compatible with exactly the same data sequences.*

PROPOSITION A.7. *For the very hard raven problem, it is impossible to achieve almost everywhere convergence to the truth.*

Proof. Suppose for *reductio* that some learning method M converges to the truth almost everywhere for the very hard raven problem. By almost everywhere convergence on the space $|\texttt{Yes}|$, there exists a $+/0$ sequence $(e_1, \dots, e_n)$ such that M converges to the truth everywhere on $|\texttt{Yes}| \cap |(e_1, \dots, e_n)|$. By almost everywhere convergence on the space $|\texttt{NoYes}|$, $(e_1, \dots, e_n)$ can be extended to some $+/0$ sequence $(e_1, \dots, e_n, \dots, e'_n)$ such that M converges to the truth everywhere on $|\texttt{NoYes}| \cap |(e_1, \dots, e_n, \dots, e'_n)|$. Choose an (infinite) data stream $\vec{e} \in |(e_1, \dots, e_n, \dots, e'_n)| \subseteq |(e_1, \dots, e_n)|$. So,

$$(\texttt{Yes}, \vec{e}) \in |\texttt{Yes}| \cap |(e_1, \dots, e_n)| \,,$$
$$(\texttt{NoYes}, \vec{e}) \in |\texttt{NoYes}| \cap |(e_1, \dots, e_n, \dots, e'_n)| \,.$$

It follows that M converges to the truth in both state $(\texttt{Yes}, \vec{e})$ and state $(\texttt{NoYes}, \vec{e})$, which is impossible. □

That is a negative result, but I believe that a learning-theoretic epistemologist does not need to apologize for that. Although a serious defense of this view has to be reserved for another paper, let me briefly sketch why I think so. The very hard raven problem embodies not just the philosophical problem of responding to the inductive skeptic, but also the problem of responding to the Cartesian external world skeptic, as highlighted by the empirical equivalence between these two hypotheses on the table: `Yes` and `NoYes`. Learning-theoretic epistemology as an epistemology of inductive inference is not, and has never been, designed to respond to the Cartesian external world skeptic. We may conjoin it with a good, independent reply to the Cartesian external world skeptic, if there is one.

What a learning-theoretic epistemologist can do, and should do, is to insist that when an inquirer tackles the hard raven problem *rather than* the very hard one, she ought to be inductive rather than counterinductive or skeptical, thanks to the learning-theoretic argument formulated in Section 3.4. Indeed, the core thesis of learning-theoretic epistemology, as formulated in Section 3.2, concerns the best learning methods *for* tackling this or that problem; it does not talk about such things as the best learning methods *simpliciter*. So learning-theoretic epistemology makes normative recommendations of this form: "*If* one tackles such and such a problem, one ought to follow a learning method having such and such properties." Such a normative recommendation is *sensitive* to the problem pursued by the inquirer.

Acknowledgements. This paper would have been impossible without Kevin T. Kelly. I thank him for the innumerable discussions that inspired many ideas in this paper. I am indebted to the following people for their detailed and helpful comments on earlier versions of this paper: Jeremy Strasser, Francesca Zaffora Blando, Alan Hájek, Julia Staffel, Ted Shear, Jason Konek, Ben Levenstein, Kenny Easwaran, Gordon Belot, Konstantin Genin, I-Sen Chen, Conor Mayo-Wilson, Liam Bright, G. J. Mattey, and two anonymous referees. I am also indebted to the following people for discussions: Ray Briggs, Tina Rulli, David Copp, Harvey Lederman, and Adam Elga. This paper was presented at the 2018 Formal Epistemology Workshop (FEW) and then discussed in Jeremy Strasser's session at the 2019 FEW; I thank the participants for discussions. This paper is supported by Fenrong Liu and her project with the Major Program of the National Social Science Foundation of China (No. 17ZDA026).

BIBLIOGRAPHY

[1] Baltag, A., Gierasimczuk, N., & Smets, S. (2015). On the solvability of inductive problems: A study in epistemic topology. In Ramanujam, R., editor. *Proceedings of the* 15*th Conference on Theoretical Aspects of Rationality and Knowledge* (*TARK-2015*). ILLC Prepublication Series PP-2015-13. ACM.

[2] Carlucci, L., & Case, J. (2013). On the necessity of U-shaped learning. *Topics in Cognitive Science*, **5**(1), 56–88.

[3] Carlucci, L., Case, J., Jain, S., & Stephan, F. (2005). Non U-shaped vacillatory and team learning. In *Algorithmic Learning Theory*. Berlin–Heidelberg: Springer, pp. 241–255.

[4] Carnap, R. (1963). Replies and systematic expositions. In Schilpp, P. A., editor. *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court.

[5] ———. (1955). Statistical and Inductive Probability (leaflet). Brooklyn, NY: Galois Institute of Mathematics and Art.

[6] de Finetti, B. (1974). *Theory of Probability*. John Wiley & Sons.

[7] Diaconis, P., & Freedman, D. (1986). On the consistency of Bayes estimates. *The Annals of Statistics*, **14**(1), 1–26.

[8] ———. (1986). Rejoinder: On the consistency of Bayes estimates. *The Annals of Statistics*, **14**(1), 63–67.

[9] Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less *Ad Hoc* theories will provide more accurate predictions. *The British Journal for the Philosophy of Science*, **45**(1), 1–35.

[10] Genin, K. (2018). The Topology of Statistical Inquiry. Ph.D. Dissertation, Carnegie Mellon University.

[11] Gold, E. M. (1965). Limiting recursion. *Journal of Symbolic Logic*, **30**(1), 27–48.

[12] ———. (1967). Language identification in the limit. *Information and Control*, **10**(5), 447–474.

[13] Glymour, C. (2015). Thinking things through: *an introduction to philosophical issues and achievements*, 2nd edition. Cambridge, MA: MIT Press.

[14] Henderson, L. (2020). The problem of induction. In Zalta, E. N., editor. *Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), https://plato.stanford.edu/archives/spr2020/entries/induction-problem/.

[15] Hintikka, J. (1966). A two-dimensional continuum of inductive methods. In J. Hintikka & P. Suppes, editors. *Aspects of Inductive Logic*. Amsterdam: North-Holland.

[16] Hintikka, J., & Niiniluoto, I. (1980). An axiomatic foundation for the logic of inductive generalization. In Jeffrey, R., editor. *Studies in Inductive Logic and Probability*, vol. **2**. Berkeley and Los Angeles: University of California Press.

[17] Hume, D. (1777). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, reprinted and edited with introduction, comparative table of contents, and analytical index by Selby-Bigge, L. A. (1975), and the third edition with text revised and notes by Nidditch, P. H. Oxford: Clarendon Press.

[18] Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

[19] Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, **4**(3), 227–241.

[20] Joyce, J. M. (2011). The development of subjective Bayesianism. In Gabbay, D. M., Hartmann, S., & Woods, J., editors. *Handbook of the History of Logic*: *Inductive Logic*, vol. 10, pp. 415–475.

[21] Kelly, K. T. (1996). *The Logic of Reliable Inquiry*. Oxford University Press.

[22] ———. (2001). The logic of success. *The British Journal for the Philosophy of Science*, Special Millennium Issue **51**, 639–666.

[23] Kelly, K. T., & Glymour, C. (2004). Why probability does not capture the logic of scientific justification. In Hitchcock, C., editor. *Contemporary Debates in the Philosophy of Science*. London: Blackwell.

[24] Kelly, T. K., Genin, K., & Lin, H. (2016). Realism, rhetoric, and reliability. *Synthese*, **193**(4), 1191–1223.

[25] Lange, S., & Zeugmann, T. (1992). Types of monotonic language learning and their characterization. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. ACM.

[26] Lin, H. (2019). The hard problem of theory choice: A case study on causal inference and its faithfulness assumption. *Philosophy of Science*, **86**, 967–980.

[27] Lin, H., & Zhang, J. (2020). On learning causal structures from non-experimental data without any faithfulness assumption. *Proceedings of Machine Learning Research*, **117**, 554–582.

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press

[28] Mukouchi, Y. (1992). Characterization of finite identification. In Jantke K. P., editor. *Analogical and Inductive Inference*, Lecture Notes in Computer Science, Vol. 642. Berlin–Heidelberg: Springer.

[29] Osherson, D., Micheal, S., & Weinstein, S. (1986). *Systems that Learn*: *An Introduction to Learning Theory for Cognitive and Computer Scientists*. MIT Press.

[30] Oxtoby, J. C. (1996). *Measure and Category*: *A Survey of the Analogies Between Topological and Measure Spaces* (second edition). New York: Springer.

[31] Pearl, J. (2009). *Causality*: *Models*, *Reasoning*, *and Inference* (second edition). New York: Cambridge University Press.

[32] Plato (1961). Meno, translated by Guthrie, W. K.C. In Hamilton, E. & Cairns, H., editors. *The Collected Dialogues of Plato*: Including the Letters. Princeton: Princeton University Press, pp. 353–384.

[33] Popper, C. (1959). *The Logic of Scientific Discovery*. London: Hutchinson & Co.

[34] Putnam, H. (1963). Degree of confirmation and inductive logic. In Schilpp, P. A., editor. *The Philosophy of Rudolf Carnap*. La Salle, IL: Open Court.

[35] ———. (1965). Trial and error predicates and a solution to a problem of Mostowski. *Journal of Symbolic Logic*, **30**(1), 49–57.

[36] Reichenbach, H. (1938). *Experience and Prediction*: *An Analysis of the Foundation and the Structure of Knowledge*. Chicago: University of Chicago Press.

[37] Salmon, W. C. (1967). *The Foundations of Scientific Inference*. Pittsburgh, PA: University of Pittsburgh Press.

[38] Savage, L. J. (1972). *The Foundations of Statistics* (second edition). New York: Dover Publications, Inc.

[39] Schulte, O. (1999). Means-ends epistemology. *British Journal for the Philosophy of Science*, **79**(1), 1–32.

[40] Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation*, *Prediction*, *and Search* (second edition). Cambridge, MA: The MIT Press.

[41] van Fraassen, B. (1980). *The Scientific Image*. New York: Oxford University Press.

[42] Vickers, S. (1989). *Topology via Logic*. Cambridge: Cambridge University Press.

[43] Williamson, J. (2010). *In defense of objective Bayesianism*. Oxford: Oxford University Press.

PHILOSOPHY DEPARTMENT
UNIVERSITY OF CALIFORNIA, DAVIS
DAVIS, CA 95616, USA
E-mail: ika@ucdavis.edu

https://doi.org/10.1017/S1755020321000605 Published online by Cambridge University Press