Convergence to the Truth

Hanti Lin

University of California, Davis
ika@ucdavis.edu

To Appear In:

Sylvan, K., Sosa, E, Dancy, J. & Steup, M. (eds.) (Forthcoming).
The Blackwell Companion to Epistemology, 3rd Edition.
Wiley Blackwell.

Abstract

This article reviews and develops an epistemological tradition in philosophy of science, called convergentism, which holds that inference methods should be assessed in terms of their abilities to converge to the truth. This tradition is compared with three competing ones: (1) explanationism, which holds that theory choice should be guided by a theory's overall balance of explanatory virtues, such as simplicity and fit with data; (2) instrumentalism, according to which scientific inference should be driven by the goal of obtaining useful models, rather than true theories; (3) Bayesianism, which features a shift of focus from all-or-nothing beliefs to degrees of belief.

1 Introduction

The epistemology of scientific inference has a rich history. According to the *explanationist* tradition, theory choice should be guided by a theory's overall balance of explanatory virtues, such as simplicity, fit with data, and/or unification (Russell 1912). The *instrumentalist* tradition urges, instead, that scientific inference should be driven by the goal of obtaining useful models, rather than true theories or even approximately true ones (Duhem 1906). A third tradition is *Bayesianism*, which features a shift of focus from all-or-nothing beliefs to degrees of belief (Bayes 1763). It may be fair to say that these traditions are the big three in contemporary epistemology of scientific inference.

There is, in fact, a fourth tradition. I am tempted to call it *convergentism*, although it does not yet have a widely recognized name, as this tradition is nearly lost in contemporary philosophy despite its prominence in statistics and machine learning. The central idea, traceable to Peirce (1878), is that the concept of *convergence to the truth* should play a significant role in evaluating inference methods. This idea was further developed by Reichenbach (1938) and Putnam (1965), together with more recent contributors from statistics, machine learning, and formal epistemology. That is the story I will unfold below. Toward the end, the convergentist tradition will be briefly compared with the big three—you can expect to see not just competition, but also cooperation.

2 Peirce on Enumerative Induction

Peirce imagined a Greek tackling a certain empirical problem—testing the hypothesis that the tide would never cease to rise every half-day:

> [The Greek] had seen the tide rise just often enough to suggest to him that it would rise every half-day forever, and had proposed then to make observations to test this hypothesis, had done so, and finding the predictions successful, had provisionally accepted

> the theory that the tide would never cease to rise every half-day, ... (CP 7.215)

But what justifies the Greek's acceptance of the inductive hypothesis? Peirce's answer is that the inference method in use, enumerative induction, meets a nice standard:

> The only justification for this would be that it is the result of a method that, persisted in, must eventually correct any error that it leads us into. (CP 7.215)

This evaluative standard requires a *guarantee of eventual correction of errors*. Some important Peircean elements may not be immediately apparent from those quotes. Let me make them explicit.

A key Peircean element is, in a sense, *internalist*. That is, when inference methods are evaluated, the kind of evaluation in question can, in principle, be carried out from a *first-person* perspective, by the very agent tackling the empirical problem in question, such as the Greek in Peirce's example. Indeed, Peirce was not interested in an inference method that happens to converge to the truth in the actual world; he employed the modality 'must' to set an evaluative standard. By 'must', he had in mind a guarantee from the agent's first-person perspective, one that quantifies over the possible worlds compatible with the background assumptions or beliefs that the agent *does not doubt* when pursuing an empirical problem. This internalist stance is central to Peirce's objection to Cartesian skepticism (CP 5.438-52) and Peirce's praise of self-controlled, rational evaluation of one's own acts, thoughts, and reasoning (CP 1.591-611, 5.333-7).

Another Peircean element is nicely captured by a slogan from James (1896): *Believe truth! Shun error!* The idea is that an inference method should be evaluated on the basis of its connection to the correction of error or, better yet, the attainment of truth:

> The ... warrant for [induction] is that this method, persistently applied to the problem, must in the long run produce a convergence (though irregular) to the truth. (CP 2.775)

Thus, Peirce's view is a combination rarely seen in today's epistemology: that inference methods should be evaluated in an *internalist* way that makes explicit their connections to *truth finding* (cf. [*add cross references*]).

To clarify: It is often said that Peirce defines truth as whatever the scientific method converges to; however, if that definition were correct, it would trivialize Peirce's use of convergence to the truth in epistemology. There is strong textual evidence, though, that Peirce actually embraces a realist account of truth instead (Hookway 2000: chapter 2). Anyway, my focus will be on Peirce's epistemology, separate from his view on the nature of truth.

The emphasis on the long run, however, raises an obvious concern: the long run might be too long. Or, in Carnap's (1945) terms, even if a normative constraint is correctly placed on the long run, it is unhelpful because it says nothing about what we actually care about: norms governing the short run. Peirce's followers have developed a systematic reply to Carnap, to which I now turn.

3 The Long Run and the Short Run

For concreteness, I will walk you through a case study on a more precisely defined empirical problem, specified by three components:

(i) The *competing hypotheses* are 'Yes, all ravens are black' and 'No, not all are'.

(ii) Pieces of *evidence* are obtained by collecting ravens and observing their colors one by one.

(iii) The *background assumption* is that either all ravens are black or a counterexample would be observed sooner or later if the inquiry were to unfold indefinitely.

Call this *the raven problem*. The point I want to make can be equally illustrated with a different empirical problem that alters any one of the three elements (i)-(iii), such as weakening the background assumption (Lin 2022); but then the

mathematics involved would be much more complex. So, for simplicity, let me continue with the raven problem, which can be represented by the tree in figure 1. The inquiry starts at the bottom, the root of the tree. Moving upward to

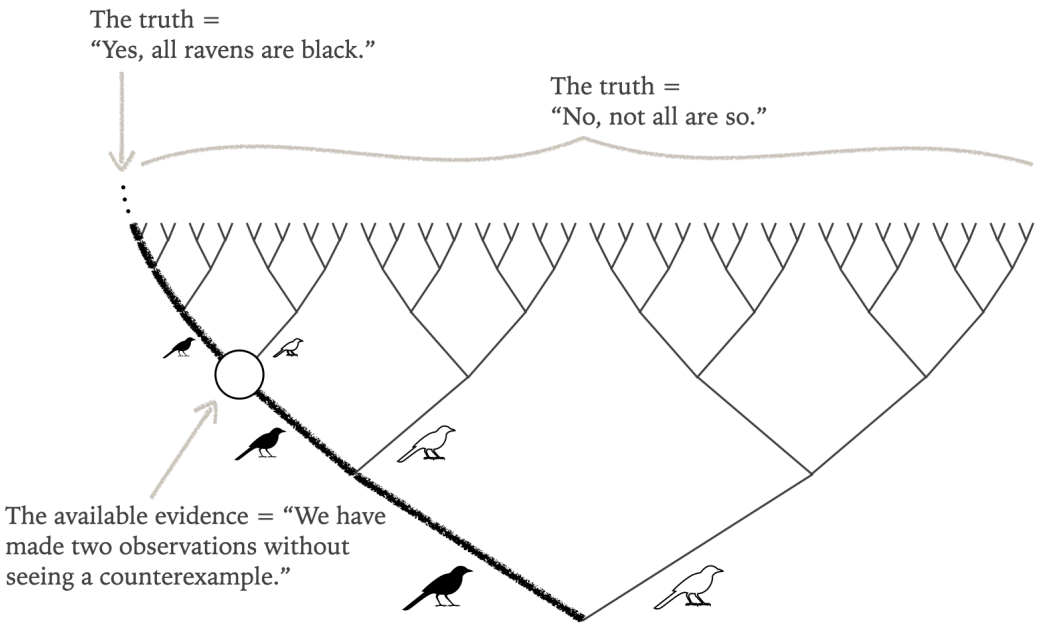

Figure 1: The raven problem represented by a tree

the right signifies observing a nonblack raven (i.e., a counterexample); moving upward to the left, a black raven (or anything other than a counterexample). So, each node represents a possible body of evidence. Each branch represents a possible world, with the tip marked by the hypothesis true in that world. To clarify: although every branch is depicted as an infinite sequence, it does not represent a world in which the agent is immortal and *will* observe an infinite number of ravens. For example, the branch that always grows to the left only represents a world in which all ravens are black, and thus every raven observed *would* be black *if* the inquiry were to extend indefinitely. Some possible worlds are not depicted at all, as they are ruled out by the background assumption of the raven problem.

An *inference method* is a function such that, whenever it receives a possible body of evidence (i.e. a node), it outputs one of the competing hypotheses or

a question mark '?' to represent judgment suspension. The task at hand is to formulate some standards to evaluate inference methods.

It would be ideal if we could have an inference method that guarantees when we would obtain the truth—a guarantee of a specific amount of evidence n that would yield the truth. This is a mode of convergence, which can be defined more precisely as follows:

> DEFINITION. An inference method M for an empirical problem P is said to achieve *uniform convergence to the truth* iff there exists an amount of evidence n such that, in every possible world compatible with the background assumption of the problem P (i.e., in every branch of the tree), M would output the truth if the number of observations were n or larger.

This mode of convergence sets an admirably high standard—an epistemic ideal that we should strive for whenever it is achievable. Unfortunately, this standard is provably too high to be met by any inference method in the raven problem.

A natural reaction is to try lowering the standard and *look for what can be achieved*. So, let's swap the two quantifiers 'there exists' and 'every' to define a weaker mode of convergence (and, for brevity, allow me to drop the relativity to empirical problems):

> DEFINITION. An inference method M is said to achieve *pointwise convergence to the truth* iff, in every branch of the tree, there exists an amount of evidence n such that M would output the truth if the number of observations were n or larger.

The idea is that the amount of evidence needed to find the truth is allowed to vary from world to world—hence the lack of uniformity, in contrast to uniform convergence as defined above. This is exactly the mode of convergence that figures in the quotes from Peirce (whether or not he had the same motivation as presented here). This lower standard is provably achievable in the raven problem. It is achieved by, for example, the method of ordinary induction

depicted at the upper left corner of figure 2, where a '`Y`' denotes an output of 'Yes, all ravens are black', and an '`N`' stands for 'No, not all are'.

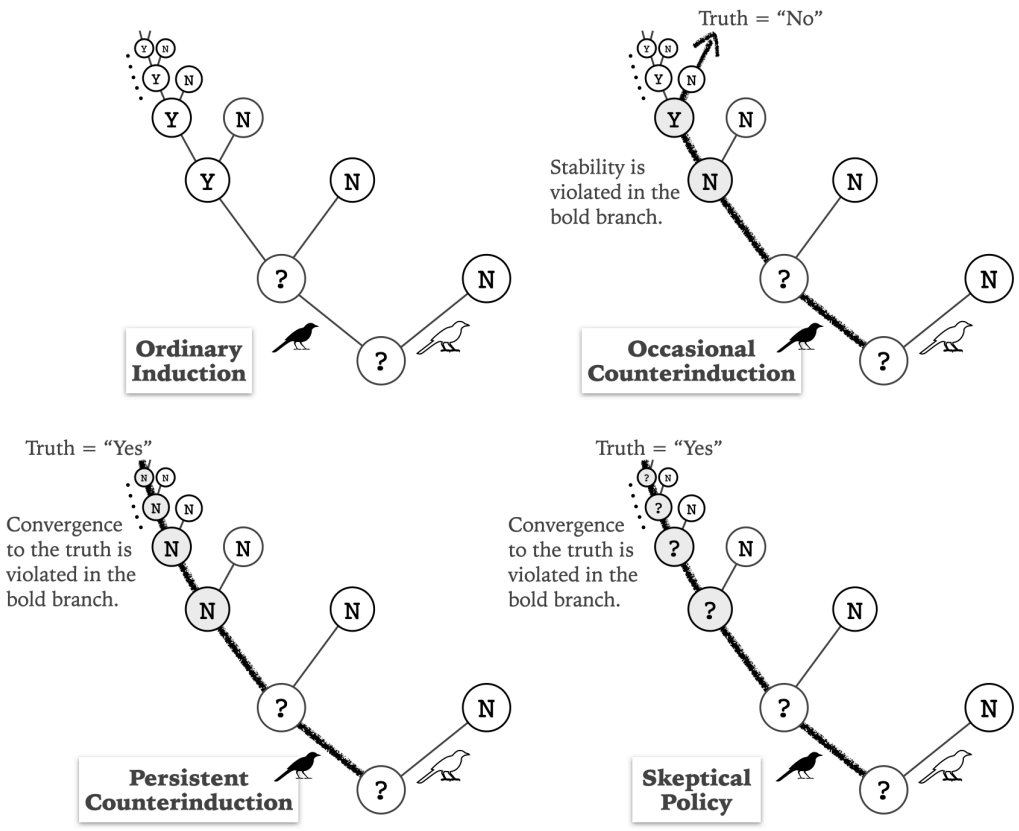

Figure 2: Four kinds of inference methods for the raven problem

Unfortunately, pointwise convergence only concerns the long run and thus imposes no constraint on the short run. It allows an inference method to engage in all sorts of erratic behavior (such as counterinduction) for the first few data points, before its eventual attainment of the truth. Indeed, it does not rule out the method of occasional counterinduction depicted in the upper right corner of figure 2. This is essentially Carnap's (1945) worry, originally formulated for Reichenbach's (1938) convergentist justification of induction, but it applies equally well to many convergentist justifications, including Peirce's own.

A strategy of reply began to emerge in the convergentist tradition since Putnam's (1965) work. The idea is simple: there is no need to settle for a

merely achievable standard such as pointwise convergence; we should, instead, *strive for the highest achievable*. So, let's try raising the bar by adding the following ideal to pointwise convergence:

> DEFINITION. An inference method M is said to achieve *stability* iff, in every branch, whenever M gets the truth, M would never let it go if the number of observations were to increase any further.

This concept simplifies Putnam's original proposal and formalizes a truth-directed ideal that Plato admires in *Meno*. Now, the combination of convergence and stability hits a sweet spot. It is weak enough to be achievable–achieved by the method of ordinary induction in the upper left corner of figure 2—yet strong enough to rule out the other three inference methods in the same figure, such as counterinductions. More generally:

> THEOREM. In the raven problem, the combined mode of convergence to the truth plus stability is weak enough to be achievable and strong enough to rule out *any* inference methods that involve application of counterinduction.

To recap: We have considered three modes of convergence to the truth. Listed as evaluative standards from high to low, they are:

- uniform convergence
- pointwise convergence with stability
- pointwise convergence

In the raven problem, the normative requirement to achieve the highest achievable standard selects the middle mode of convergence, which in turn induces a norm on the short run: never infer counterinductively in the raven problem—never, ever, including now.

A short-run rabbit is thus pulled out of a long-run hat. The trick has been mostly buried in technical works, so let me try to reveal it intuitively. Imagine that someone is deciding whether to drink and whether to drive. She reasons as follows:

I may drink, or drive, but not both.
I must drive.
So, I must not drink.

This reasoning illustrates a pattern: Once a constraint is placed jointly on two things (whether-to-drink, and whether-to-drive), a constraint imposed directly on one of the two might induce a constraint on the other. Similarly, if there is a normative constraint X placed jointly on two things—the long run and the short run—then a convergentist constraint on the long run might induce a nontrivial constraint on the short run. Such an X must then be a *diachronic* constraint, and that is the trick in reply to Carnap. The mode of stable convergence defined above is indeed diachronic, for it concerns the retention of truth as evidence accumulates. Convergentists have explored other diachronic candidates for X, to be presented below when needed.

4 A Framework for Convergentism

The above case study on the raven problem actually illustrates a framework, first adumbrated in Putnam (1965) and later developed further by Kelly (1996) and Schulte (1999) through additional case studies. A clear and general statement of the core thesis was given by Lin (2022):

> The Core Thesis. In any empirical problem, a necessary condition for an inference method to qualify as one of the best is that it achieves the highest achievable mode of convergence to the truth—pending a specification of the right hierarchy of modes of convergence as evaluative standards or epistemic ideals.

This thesis sets up what may be called the *achievabilist* framework for convergentism, for lack of a standard name. This framework encourages the exploration of modes of convergence, such as the mode of stable convergence, which is used to address Carnap's worry.

This framework also features a kind of context-sensitivity: the appropriate standard for assessing inference methods should be the highest achievable,

which is sensitive to a contextual factor: the empirical problem that one tackles in one's context of inquiry. In fact, when one switches to the context of a statistical problem, the epistemic standards presented above all become unachievable, which requires convergentists to explore lower standards—weaker modes of convergence. To illustrate how that may be done, let's think about statistics, which brings us back to Peirce.

5 Peirce on Statistics

Peirce once studied a classic problem in statistics. An urn contains an unknown number of black and white balls. We ask: what is the true proportion of white balls in the urn? The three components of this problem are as follows:

(i) *Competing hypotheses* are rational numbers in the unit interval (i.e., the possible proportions).

(ii) *Evidence* is to be collected by randomly drawing balls with replacement.

(iii) The *background assumption* is that different draws are probabilistically independent.

Call this the *white ball problem*. To evaluate inference methods for this problem, Peirce proposed to employ the following mode of convergence, where, by probability, he meant physical chance:

> Definition. An inference method M for an empirical problem is said to achieve *statistical consistency* iff M has the following guarantee: (i) in any possible world compatible with the background assumption, there exists n such that M would highly probably produce a guess that gets close to the truth if the sample size were n or larger, and (ii) M has the above property for any threshold of high probability less than 1 and for any nonzero threshold of closeness.

Peirce studied this mode of convergence as an evaluative standard in an 1878 paper titled "The Probability of Induction" (CP 2.669-93). It is unclear whether Peirce influenced any statisticians of his time, but this stochastic mode of convergence was popularized by the statistician Fisher (1925: section I.3). It is now generally regarded in classical statistics as a minimum qualification for any permissible statistical methods, with different versions for various statistical problems, including estimation, hypothesis testing, and regression (i.e. curve fitting).

Peirce never explained why he employed different evaluative standards in two different empirical problems, the white ball problem and the Greek's tide problem (which is equivalent to the raven problem). From an achievabilist's hindsight, he had to employ different standards. The Greek's tide problem is easy enough to allow for a guarantee of getting exactly the true answer (at least when the amount of evidence is arbitrarily large). But this standard is too high to be achievable in the white ball problem. So, let's try lowering the bar for that problem: "getting exactly the truth" can be downgraded to "*highly probably* getting exactly the truth", which can be further downgraded to "highly probably getting *close* to the truth". This line of thought motivates statistical consistency as a lower standard, and Peirce did show how it can be achieved in the white ball problem (by using the law of large numbers).

Carnap's worry still needs to be addressed for statistical problems. The diachronic trick presented above still applies; an example will be provided in a broader context when I wrap up.

6 Closing: Comparison with the Big Three

How does the convergentist tradition fare against the big three mentioned in the introduction?

First of all, the Bayesian tradition need not be a rival to convergentism. A union has been proposed in statistics as a partial solution to Bayesians' perennial problem of the priors—the problem of identifying the correct norms to constrain the pool of the permissible distributions of prior credences (see

[*add cross references*]). A convergentist constraint on priors was proposed by the statistician Freedman (1963):

> Definition. A prior is said to achieve *Bayesian consistency* in an empirical problem iff it is guaranteed (under just the background assumption of that problem) that this prior, when guided by the diachronic rule of conditionalization, would have a high chance of its posterior credences converging to the truth among the considered hypotheses if the amount of evidence were to accumulate indefinitely—for any threshold of high chance less than 1.

Freedman did not recognize that this is another implementation of the diachronic trick in reply to Carnap: a combination of long-run convergence with a diachronic constraint, which in this case is conditionalization. The resulting constraint on the priors—and hence on the short run—turns out to be surprisingly strong in some interesting empirical problems, stronger than what traditional Bayesians have to offer. This constraint implies a particular version of Ockham's razor in statistical problems of curve-fitting (Diaconis & Freedman 1998); it also implies another version of Ockham's razor in problems of testing deterministic hypotheses (Lin 2022). It remains to be seen whether convergentist Bayesianism is better than more traditional varieties of Bayesianism.

The explanationist tradition can benefit from convergentism, too. According to explanationism, theory choice should be based on a theory's overall balance of explanatory virtues, so it should be based on Ockham's razor if simplicity is among these virtues. But which version of Ockham's razor? That is, which particular trade-off between simplicity and other virtues such as fit with data, under which conception of simplicity, and under which conception of fit? Explanationists often think that the choice is to be justified by intuition alone (Swinburne 1997). However, in complex empirical problems, working scientists often find that they lack the intuition needed to justify one version of Ockham's razor over another—and this is where convergentists can assist. The previous paragraph already referred to two examples of convergentist justifications of versions of Ockham's razor, in curve-fitting problems

and in problems of testing deterministic hypotheses. Let me add one more example: in problems of testing statistical hypotheses, Genin (2018) justifies another version of Ockham's razor by combining a mode of convergence with a stochastic version of stability (which he calls *progressiveness*, a guarantee that the chance of finding the truth would never drop too much if the sample size were increased by any finite amount).

Now, let's turn to the instrumentalist tradition. Due to their emphasis on pursuit of usefulness instead of truth, instrumentalists might appear to have to dismiss the significance of convergence to the *truth*. However, this is merely an appearance, I submit. The usefulness of a scientific model is often taken to include at least predictive accuracy, but predictive accuracy is a contingent matter, depending on what the actual world is like. Thus, choosing a useful model is no trivial task. The foundation of machine learning, known as *learning theory*, addresses this with certain modes of convergence: modes in which a learning algorithm or inference method is guaranteed to converge to the most useful of the predictive devices under consideration, where usefulness is identified with the chance of accurate prediction (Shalev-Shwartz et al. 2014). To be sure, this does not sound like convergence to the truth. But let's be careful: convergence to the most useful one is still convergence to a certain truth, the true answer to this practical question: *Which one is the most useful?* Instrumentalists can find a home in machine learning—a convergentist home.

The comparison just made is quite rudimentary due to the lack of an established literature. Much work is still needed to determine how convergentism competes with the other three traditions, as well as to explore opportunities for their cooperation.

References

Bayes, T. (1763) "An Essay towards Solving a Problem in the Doctrine of Chances", *Philosophical Transactions of the Royal Society*: 330-418. Reprinted with Biographical Note by Barnard, G.A., in (1958) *Biometrika* 45: 293-315.

Carnap, R. (1945) "On Inductive Logic", *Philosophy of Science*, 12(2): 72-97.

Diaconis, P. & Freedman, D. (1998) "Consistency of Bayes Estimates for Nonparametric Regression: Normal Theory", *Bernoulli*, 4(4): 411-444.

Duhem, P. (1906/1954) *The Aim and Structure of Physical Theory*, Princeton University Press.

Fisher, R.A. (1925) *Statistical Methods for Research Workers*, Oliver & Boyd.

Freedman, D. (1963) "On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case", *The Annals of Mathematical Statistics*, 34(4): 1386-1403.

Genin, K. (2018) *The Topology of Statistical Inquiry*, PhD Dissertation, Carnegie Mellon University.

Hookway, C. (2000) "Truth, Rationality, and Pragmatism: Themes from Peirce", Oxford University Press.

James, W. (1896) "The Will to Believe", *The New World*, 5: 327-347.

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

Lin, H. (2022) "Modes of Convergence to the Truth: Steps toward a Better Epistemology of Induction", *The Review of Symbolic Logic*, 15(2), 277-310.

Peirce, C.S. (1994) *Collected Papers of Charles Sanders Peirce* (Volumes I-VIII), InteLex Corp.

Putnam, H. (1965) "Trial and Error Predicates and a Solution to a Problem of Mostowski", *The Journal of Symbolic Logic*, 30(1): 49-57.

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

Russell, B. (1912) *The Problems of Philosophy*, Williams and Norgate.

Schulte, O. (1999) "Means-Ends Epistemology", *The British Journal for the Philosophy of Science* 79(1), 1-32.

Shalev-Shwartz, S. & Ben-David, S. (2014) *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press.

Swinburne, R. (1997) *Simplicity as Evidence of Truth*, Marquette University Press.