A Framework for Pragmatic Reliability

Isaac Davis*†

I propose a framework for pragmatic reliability in-the-limit criteria, extending the epistemic reliability framework. I identify some common scientific contexts that complicate the application or interpretation of epistemic reliability criteria, drawing heavily from economics for illustrative examples. I then propose an extension of the standard framework, where inquiry is constrained by both epistemic and nonepistemic factors. This provides analogous notions of pragmatic underdetermination and pragmatic reliability with respect to a particular goal, as well as a principled method for extracting solvable problems from unsolvable ones.

1. Introduction. It is well established that full verification of hypotheses is rarely attainable from finite data. Thus, as finite beings with finite life spans, we must settle for something weaker. Historically, a standard approach in philosophy of science is to interpret verification as an entailment relation between hypothesis and data. Under this view, solving the problem of induction requires a precise notion of "confirmation," which quantifies the degree to which a hypothesis is justified by a given body of evidence (e.g., Carnap 1945; Hempel 1945; Putnam 1963). This provides straightforward, diachronic norms for hypothesis selection: choose the hypothesis "best confirmed" by the evidence. This *confirmation approach* can be formalized in many different ways, most recently as Bayesian updating—a rigorous framework for rationally updating beliefs in light of new evidence.

While the confirmation approach provides robust tools for belief revision, philosophers of science have struggled to provide epistemic justification for confirmation-based criteria. An alternative approach is to understand verification and refutation as success criteria for hypothesis selection methods,

Received January 2019; revised August 2019.

*To contact the author, please write to: Department of Philosophy, Carnegie Mellon University; e-mail: isaacd@andrew.cmu.edu.

†I am very grateful to David Danks and Kevin Kelly and the referees to this journal for their helpful feedback on previous drafts of this work.

Philosophy of Science, 87 (October 2020) pp. 704–726. 0031-8248/2020/8704-0008$10.00

https://doi.org/10.1086/709755 Published online by Cambridge University Press

rather than entailment relations between hypothesis and data (Kelly 1996). Rather than seeking a numerical, nonlogical score for a hypothesis given a body of evidence, the reliability theorist seeks a deductive guarantee of success for a hypothesis selection method on a data stream. This *epistemic reliability* approach can be formalized in the language of formal learning theory (and, more recently, topology; Genin and Kelly 2017) and provides a precise, mathematical characterization of underdetermination in scientific inquiry. This also allows us to evaluate scientific practices (e.g., a preference for "simpler" theories) in terms of their in-the-limit truth-seeking behavior. Even when a hypothesis cannot be deductively verified in the short run, we may still obtain deductive guarantees that a method will succeed in the limit (for a spectrum of definitions of "success"). This view is sometimes identified with the *feasibility contextualism* principle, which we can summarize as follows: given an inductive inference problem, identify your epistemic constraints, then succeed as well as possible (Kelly 2014).

This article proposes an extension to the epistemic reliability framework in which inquiry is constrained by both epistemic (e.g., the structure of the available data) and nonepistemic (e.g., research goals, environmental constraints) factors. I refer to this as "pragmatic reliability," as it defines the reliability of a method with respect to an environmentally constrained, goal-directed inference problem. To better illustrate and motivate these distinctions, consider the following toy example: Suppose an urn contains an unknown but finite number of marbles of two colors, red and blue. Each hypothesis about the initial contents of the urn corresponds to an ordered pair of integers, $U = (x, y)$. Suppose we observe a data sequence generated as follows: in each time step, we observe a red marble (R), a blue marble (B), or no marble (N) drawn from the urn. If the urn is empty, we observe N with probability 1. If the urn is nonempty, we observe N with probability 1/2, or a single marble (selected uniformly at random from those remaining) with probability 1/2. Because a nonempty urn always has a 1/2 chance of producing a null observation, it is never possible to conclude with deductive certainty that the urn is empty, no matter how many N's we observe in a row. Thus, the problem of inferring the initial contents of the urn is underdetermined by the data: any finite number of observations is insufficient to deductively infer the correct answer.

A confirmation theorist addresses the underdetermination in this problem by assessing the degree to which a hypothesis $U = (x, y)$ is confirmed or justified by a sequence of observations $D = d_1, d_2, \ldots$. A Bayesian confirmation theorist, for example, would solve this problem by positing some prior distribution $P(U)$ over possible initial contents of the urn and computing the "posterior probability" of U given evidence D using Bayes's rule: $P(U|D) = P(D|U)P(U)/P(D)$. Here, $P(D|U)$ is the probability of observing D, given that U is true, which is determined by the sampling probabilities above. The quantity $P(U|D)$ therefore reflects the degree to which U is

https://doi.org/10.1086/709755 Published online by Cambridge University Press

confirmed by the evidence D, given the theorist's background assumptions and beliefs.

An epistemic reliability theorist draws justification for a hypothesis from deductively guaranteed convergence properties of the method used to select that hypothesis. A method M is a map from evidential states D to hypotheses U, and a method is said to succeed in the limit so long as it is guaranteed to output an incorrect hypothesis only finitely many times (this notion is defined more explicitly in sec. 2). For example, consider the method M of simply counting the number of red and blue marbles observed thus far. At any step, we cannot be deductively certain that M outputs the correct answer. However, because the true urn contains only a finite number of marbles, eventually we will observe only N's. Since this method only updates its hypothesis when a non-N is observed, it is deductively guaranteed to eventually stabilize to the correct urn after finitely many observations. Note that this guarantee is completely independent of the sampling probabilities that generate the data: so long as each marble is eventually drawn from the urn, M is deductively guaranteed to converge in the limit to the correct hypothesis.

In the pragmatic context, we assume that, in addition to the learner's epistemic constraints, the learner faces some environmentally constrained goal-satisfaction problem. Suppose, for example, that the learner's environment contains four levers, corresponding to the four following possibilities: (1) U (initially) contains no red and no blue marbles, (2) U contains some red and no blue marbles, (3) U contains no red and some blue marbles, (4) U contains some red and some blue marbles. Pulling the correct lever yields a reward, while pulling an incorrect lever yields nothing. Given this setup, the following inference method is *pragmatically reliable*: $M'(D) = (I_R(D), I_B(D))$, where $I_R(D)$, $I_B(D)$ respectively return 1 if and only if (iff) D contains at least one red or blue marble, and 0 otherwise. In this case, M' is very likely to converge in the limit to an incorrect hypothesis: if, for example, the true urn is (3, 0), M' will converge to (1, 0). However, with respect the learner's environment, the hypotheses (3, 0) and (1, 0) both imply the same optimal action. Thus, even though M' will often fail to identify the correct urn, it is deductively guaranteed to converge in the limit to a hypothesis that induces the best possible action, given the learner's goal and constraints. This is the notion of pragmatic reliability, which I define formally in section 4.

The remainder of this article is organized as follows: section 2 reviews epistemic reliability as formalized in formal learning theory. I consider how it represents a scientific problem, how it characterizes underdetermination, and how it can usefully inform or justify scientific practice. Section 3 identifies certain features common in many areas of scientific research that make it difficult or impossible to use epistemic reliability criteria. I will not attempt to exhaustively characterize all problematic cases, nor will I argue that these features are unique to any disciplines in particular. The features I consider,

https://doi.org/10.1086/709755 Published online by Cambridge University Press

however, are especially pronounced in the "human sciences,"[1] and I draw heavily on economics for illustrative examples. I identify two main factors that motivate the need for a pragmatic reliability framework. First, the level of global underdetermination that often arises in these contexts is beyond the scope of traditional epistemic reliability theory. Second, there are many contexts in which the theoretical and metaphysical assumptions underlying epistemic reliability are too strong to capture how many scientists and philosophers of science conceive the nature and purpose of their research.

Section 4 proposes an extension to the epistemic reliability framework in which learners must act on their hypothesis so as to achieve some target goal. When a learner faces the type of underdetermination described above, the situation decomposes into two separate problems, one epistemic and one pragmatic. The epistemic problem is determining what can and cannot be reliably learned, given our available data sources. The pragmatic problem is determining what we ought to learn, and how we ought to learn it, given our nonepistemic goals and constraints. My extended framework can account for both of these problems, which provides analogous notions of "pragmatic underdetermination" and "pragmatic reliability" as well as a principled method for extracting "solvable" problems from badly underdetermined ones. This, I argue, suggests the following modification to the feasibility contextualism principle: identify your epistemic constraints, then succeed as well as necessary.

In the final section, I further explore the relation between the epistemic and pragmatic notions of reliability, by considering the case in which our research goal is prediction of future phenomena. In this case, the epistemic and pragmatic notions of underdetermination coincide in an important way. This, I suggest, allows us to understand empirical adequacy as a boundary point between epistemic (i.e., truth-seeking) and pragmatic (i.e., goal-satisfying) research problems and helps resolve a tension between realist and instrumentalist views of scientific research.

2. Epistemic Reliability

2.1. Scientific Questions and Underdetermination. Here I outline the formal framework for epistemic reliability (Kelly 1996). I define a scientific problem in terms of a set W of possible worlds and a countable set I of information states. For a world $w \in W$, I use $I(w)$ to denote the set of information states in w. Informally, an information state represents a body of evidence, and $I(w)$ denotes all evidence that we will eventually observe in w. Formally, we identify $E \in I$ with the set of possible worlds compatible with

[1] Roughly, science in which the data of interest are a product of human behavior or decision making.

https://doi.org/10.1086/709755 Published online by Cambridge University Press

that evidence. A hypothesis h is a subset of possible worlds, and a question $\mathcal{Q}$ is a partition of W into countably many answers. A hypothesis h entails an answer if it is contained entirely within that answer.

To illustrate this concretely, consider the following toy example: At the end of each day, we record whether it rained in Pittsburgh, marking 1 for rain or 0 for no rain. Each possible world w corresponds to an infinite binary sequence, and $I(w)$ is the set of finite initial subsequences of w. The hypothesis "it will rain twice in the first week" corresponds to the set of all worlds with exactly two 1's in the first seven entries. Similarly, the question "will it ever stop raining" is the partition of W into two cells: worlds with finitely many 1's and worlds with infinitely many.

This provides a formal characterization of underdetermination in scientific inquiry. For our present purposes, we distinguish between *local* and *global underdetermination*. Intuitively, a hypothesis h is locally underdetermined if any finite amount of evidence is insufficient to deductively verify or refute it. Formally, h is *locally undetermined in w* iff, for any finite evidence $E_1, \ldots, E_n \in I(w)$, there exists a world $w' \notin h$ such that $w' \in \bigcap_{i=1}^{n} E_i$ (i.e., iff any finite amount of evidence in w is consistent with some world not in h). If h is locally undetermined in every world, we drop the "in w." A locally underdetermined hypothesis is nonverifiable, although as we will see in the next section, it may be verifiable in the limit. A globally underdetermined hypothesis, however, could not be deductively verified even if we observed the entire data stream at once. Formally, h is *globally underdetermined in w* iff there exists $w' \notin h$ such that $w' \in \bigcap_{E \in I(w)} E$ (i.e., iff some world not in h is compatible with all evidence in w). This notion will be especially important in the scientific contexts we consider for section 3.

2.2. Methods and Convergence. In the formal learning theory framework, verification and refutation are defined as success criteria on *methods*. A method is a map $M : I \rightarrow \mathcal{P}(W)$ from information states to hypotheses. For a hypothesis h, we say that *M deductively verifies h* iff the following conditions hold:

1. (Infallibility) For all $E \in I(w)$, $w \in M(E)$.
2. (Convergence) $w \in h$ iff there exists $E \in I(w)$ such that $M(F) \subseteq M(E) \subseteq h$ for all $F \in I(w)$ such that $F \subseteq E$.

Intuitively, Infallibility stipulates that the method's outputs (conclusions) are deductively entailed by its inputs (evidence), and Convergence stipulates that M will converge to h (or some hypothesis entailing h) iff h is true. Similarly, *M refutes h* iff M verifies h^c, *M decides h* iff it both verifies and refutes h, and a hypothesis is verifiable/refutable/decidable iff there exists a method that verifies/refutes/decides it. Returning to our example, the

https://doi.org/10.1086/709755 Published online by Cambridge University Press

hypothesis h = “it will rain twice in the first week” is clearly decidable by the following method: output $h \cup h^c$ (the trivial hypothesis) for the first seven observations, then count the number of 1’s and output h or h_c accordingly.

Of course, Infallibility is too strict a requirement for nearly all scientific inquiry (hence the problem of induction). By dropping this condition, however, we obtain a weaker success criteria for M. Specifically, a method *verifies h in the limit* iff it satisfies condition 2, and a hypothesis is *limit verifiable* iff such a method exists for it. Thus, we can still deductively prove that a method will converge to the truth in the limit, even if we cannot deductively verify a hypothesis in the short run. To illustrate this, consider h = “it will rain only finitely many times.” This hypothesis is locally underdetermined and cannot be verified or refuted by any method: any finite evidence is consistent with a world with finitely many days of rain and a world with infinitely many. However, h is limit verifiable by the following method: if it did not rain today, output h, otherwise output h^c. In any world where h is true, there must exist a last day of rain, which means that on this last day, M will correctly stabilize to h. Conversely, in a world where h is false, M will either stabilize to h^c (if it never stops raining) or never stabilize at all. Therefore h is unverifiable but verifiable in the limit.

While we will not use them here, it is worth noting that all of these concepts have precise topological analogues, and there exists a natural hierarchy of topological properties that corresponds to the hierarchy of solvability criteria. This allows us to better analyze the solvability of a problem in terms of its topological properties. Furthermore, Genin and Kelly (2017) extend this reasoning to the statistical setting, deriving precise notions of statistical verifiability and statistical verifiability in the limit in the same topological framework.

3. Challenges for Epistemic Reliability. In this section, I identify some features common to certain scientific contexts that complicate the application or interpretation of epistemic reliability criteria. These features are not unique to any field in particular, and the goal is not to exhaustively characterize all such cases or provide a general taxonomy of which disciplines are amenable to epistemic reliability. However, the features I consider are especially prevalent in contexts in which the data under study are an output of human behavior or decision making, such as economics, cognitive science, and decision sciences, and I draw heavily on such fields—especially economics—for illustrative examples.

I identify two broadly defined kinds of challenges for epistemic reliability: technical and conceptual. On the technical side, I identify features characteristic of (although not unique to) human-based sciences that induce a level of underdetermination beyond the scope of traditional epistemic reliability analysis. On the conceptual side, I explore how hypotheses are conceived and

https://doi.org/10.1086/709755 Published online by Cambridge University Press

defined in certain research contexts, which is difficult to reconcile with the formalized notion of "hypothesis" in epistemic reliability. I conclude that, depending on one's perspective, applying epistemic reliability criteria in contexts with one or more of these features is some combination of (*a*) extremely difficult or technically impossible or (*b*) conceptually nonsensical or pointless.

3.1. Technical Challenges: Severe Global Underdetermination. Recall the motivating principle of "feasibility contextualism": identify our epistemic constraints (underdetermination) and succeed as well as possible given those constraints. If we take "success in the limit" as our baseline success criterion, we can apply this to problems that are, at worst, locally underdetermined. However, certain features characteristic of human-focused research often induce a very strong form of global underdetermination.

In order to make sense of these claims, however, we must first provide some minimal characterization of the research contexts we want to address. For this purpose, economics is an informative case to consider, as economists deal both with human-generated data and mathematical models of the data-generating process. For an economic realist, a model is an attempt to (approximately) represent some real-world data-generating process, and we can at least make conceptual sense of epistemic reliability. We can very generally represent an economic research problem as follows: $\{\underline{Z}\}_{t \in T}$ denotes a data stream where T indexes the time steps of observation, and each $\underline{Z}_t$ corresponds to some panel of measurements. These measurements may be at the microlevel, denoting the behavior of individual agents (e.g., firm or consumer), or the macrolevel, denoting aggregate measures or indices (e.g., gross domestic product or unemployment). The chosen set of measurements determines our information basis I. The hypothesis space $\mathcal{H}$ is typically a model structure or class of models, posited by the economist. This, in conjunction with any theoretical assumptions, determines a parameter space, and we can equate each allowable parameterization with a possible world. The economist's challenge is to construct a method $M : I \rightarrow H$ that converges in the limit to the true model/model parameters, for some notion of convergence.

There are several sources of underdetermination common to economic research. The most common is simply the low quality of available data: economic data, particularly at the microlevel, are often biased and woefully incomplete (Friedman 1953; Windrum, Fagiolo, and Moneta 2007). Furthermore, for reasons both practical and ethical, it is often extremely difficult to generate useful experimental data in economics. Economists therefore have to rely on incomplete, often biased, observational data alone, with only marginal control over the data-collection process. This challenge is certainly not unique to economics: psychologists and cognitive scientists face similar difficulties, particularly when the variables of interest are unobservable

https://doi.org/10.1086/709755 Published online by Cambridge University Press

cognitive states on which we cannot directly intervene. The resulting difficulty of performing controlled experiments severely limits the experimenters' ability to infer causal connections between real-world variables; any two causal models that induce the same joint probability distribution over observed variables are empirically equivalent and cannot be distinguished without controlled experimentation. For some applications this may be sufficient, but many applications (e.g., designing effective policy or medical interventions) require causal knowledge beyond the scope of observational data.

A second source of underdetermination is the abundance of available models and modeling paradigms. Statisticians and econometricians are endlessly innovative in designing flexible model classes for time-series data, and the economist must often choose between two or more empirically equivalent modeling paradigms (Haavelmo 1944). In some cases, a single model class will contain multiple (sometimes infinitely many) empirically equivalent models; this occurs in econometrics, for example, when using overcomplete dictionaries to estimate complex functions (e.g., Clyde and Wolpert 2007; Belloni and Chernozhukov 2011). Such issues also arise in models that represent the unobservable inner workings of human decision processes, as is sometimes the case in behavioral micromodels and often the case in cognitive science and psychology. As it is impossible for statistical inference to decide between observationally equivalent hypotheses, global underdetermination is built into the hypothesis space itself.

A third source of underdetermination is the nature and prevalence of ceteris paribus (CP) clauses in economic hypotheses, which weaken a hypothesis of the form "all A are B" to "all else being equal, all A are B." There is some debate as to whether, and to what degree, CP clauses occur in natural sciences (e.g., Cartwright 1995; Earman and Roberts 1999), but it is widely acknowledged that most economic hypotheses contain a CP clause, either implicitly or explicitly (Hausman 1988; Cartwright 2002). In this context, we can understand a CP clause as an auxiliary hypothesis of unknown dimension; the clause specifies that our model is only true so long as certain conditions hold, but the exact conditions are left unspecified. This results in a particularly bad instance of the Duhem-Quine thesis: not only is the main hypothesis inextricably tied to a set of auxiliary hypotheses, but we cannot even specify what those auxiliary hypotheses are. A thorough study of underdetermination in economics (Sawyer, Beed, and Sankey 1997) demonstrates some technical implications of CP clauses for hypothesis testing. In time-series econometrics, for example, the modeler attempts to distinguish between "signals" that capture the causal connections of interest and "noise" that accounts for random, less systematic causes that disturb the variables of interest. In such cases, it is possible for noises to be stronger than signals, compounded by the possibility that the causal mechanisms of interest may

https://doi.org/10.1086/709755 Published online by Cambridge University Press

undergo structural changes themselves (Valente 2005). Thus, the underspecified nature of auxiliary hypotheses in economics can make precise refutations impossible, and econometricians will sometimes respond to diagnostic test failure with ad hoc adjustments like changing the estimation procedure, rather than respecifying the theory (Hendry 1980).

3.2. Conceptual Challenges: The Nature and Purpose of Scientific Hypotheses. The previous analysis assumes a certain "realist" perspective to begin with: the view that a model is an attempt to accurately represent some (often causal) aspect of the real world. Yet many economists, cognitive scientists, psychologists, and associated philosophers take a decidedly nonrealist stance on the nature of hypotheses. Even within economics, there is a wide range of "nonrealist" perspectives on the status and purpose of hypotheses, and I identify two recurring concerns among these views that complicate the interpretation of epistemic reliability criteria on a conceptual level.

The first concern relates to the notion of hypothesis specifically. In epistemic reliability, we assume a prespecified set of relevant possibilities and identify hypotheses with subsets of possibilities. This notion, however, is difficult to reconcile with how hypotheses and models are often discussed in economics.[2] Sugden (2000), for example, argues that an economic model is not a set of "real" possibilities but a counterfactual world constructed by the economist. Such an object says nothing about the world; a model supports conceptual exploration, enabling deductive inferences about what the world would look like if the assumptions in the model were true. Bridging the gap between model world and real world requires a separate inductive reasoning process, establishing the "robustness" of the model against violations or modifications of its assumptions. A less formalized, more historicist viewpoint asserts that economic hypotheses are, essentially, carefully crafted stories, which we use to explain historical events and draw lessons for the future (e.g., McCloskey 1998, 23–28). Neither conception, however, is particularly well suited to representation as subsets of a fixed set of relevant possibilities. Even if we equate individual models with relevant possibilities, the prevalence of CP clauses, noise terms of unknown dimension, and other

[2] Hausman (1992, 70–82) distinguishes between a *model* (a system of definitions specifying a set of predicates, concepts, and relations) and a *theoretical hypothesis* (a set of statements asserting that the model's assumptions are true of some part of the world). For our purposes, I intentionally lump the two terms together: what Hausman understands as models and theoretical hypotheses, I interpret as separable components of a more general hypothesis space, consisting of (*a*) the set of possible models and (*b*) the set of all ways to connect each model with some part of the real world. We may also think of this as a set of paradigms and the set of possible hypotheses within each paradigm.

https://doi.org/10.1086/709755 Published online by Cambridge University Press

underspecified simplifying assumptions make it difficult to explicitly identify a model with a set of possible worlds.

The second concern relates to how we evaluate hypotheses, independent of what kinds of objects we understand them to be. Underlying epistemic reliability is the assumption that, when faced with multiple hypotheses compatible with the same evidence, our method for selecting one should be evaluated in terms of its truth conduciveness. In the formal setup, the truth conditions for a hypothesis are built right into the framework—a hypothesis *h* is a set of possible worlds, and *h* is true iff it contains the true world. But many scientists (especially economists) take a decidedly instrumentalist stance toward their research, asserting that we cannot (or at least, have no reason to) establish firm truth conditions for a hypothesis. Rather, a hypothesis is a tool used to achieve some end—typically prediction of future data—and the only metric for evaluating a hypothesis is how well its predictions match with our observations (e.g., Friedman 1953). There are (at least) two versions of instrumentalism; one version maintains a sort of quasi realism, granting that the assumptions of a hypothesis at least refer to some part of the real world but asserting that it does not matter whether those assumptions are true. The alternative version denies that theoretical assumptions have any connection to things in the real world; in this view, the assumptions underlying a model are simply a compact way of representing the model's predictions. This latter form of instrumentalism is also common in certain areas of cognitive modeling. For example, many cognitive scientists take an "ideal Bayesian observer" approach to modeling perception, in which the goal is not to accurately model the actual causal processes underlying human cognition but to provide an "as if" explanation at the computational level (Jones and Love 2011). In either case, the instrumentalist is unconcerned with assessing the "truth" of the hypotheses, and it therefore makes little sense to evaluate the instrumentalist's methods in terms of truth conduciveness.

4. A Framework for Pragmatic Reliability. Section 3 illustrates some obstacles to applying standard epistemic reliability criteria that commonly occur in certain scientific contexts, particularly those that involve modeling human-generated data. In this section, I propose a modified version of the epistemic reliability framework that allows us to address these obstacles. The motivating view is that, even if we cannot make sense of truth conditions, and even if we are only interested in achieving pragmatic research goals, we can and should still be concerned with the reliability of our methods. By generalizing the notions of "hypothesis" and "truth conditions," realized in the same formal learning theory framework, we obtain analogous definitions of pragmatic reliability and pragmatic underdetermination. Whether our goals are pragmatic or purely epistemic, we can use the same learning-theoretic framework to understand the achievability of those goals, given

https://doi.org/10.1086/709755 Published online by Cambridge University Press

our constraints, and identify inference methods with deductively guaranteed convergence properties.

4.1. Goal-Directed Learning and Reliability. As in the epistemic setting, we define a problem in terms of a set W of possible worlds and a set I of information states, with $I(w)$ denoting all evidence that will eventually be observed in w. Unlike the epistemic setting, the hypothesis space may be any abstract representation space and need not correspond to explicitly defined subsets of W (although we continue to use the w subscript to index the "true" world from which evidence is generated). This allows the framework to accommodate nonrealists who are uncomfortable with the metaphysical implications of a known, predefined space of possibilities. As before, we assume that evidence is observed in trials, using $M : I \rightarrow \mathcal{H}$ to denote the total evidence at the nth trial, and define a method $\mathcal{H}$ to be a map from information states to hypotheses.

By allowing $\mathcal{H}$ to be any representation space, we lose the well-defined truth conditions built into the epistemic framework. In order to evaluate a hypothesis, we must understand (a) the aims of our inquiry and (b) how we may act on our inferences so as to reach those aims. That is, in order to model goal-directed learning, we must model not only what knowledge can be acquired but how that knowledge will be put to use. To this end, we draw on the Adaptively Rational Learning framework presented in Wellen and Danks (2016). In this framework, a learner consists of an inference method M and a *decision mechanism* D, which deterministically assigns each hypothesis to an *action*. Actions can include decisions such as setting a price or choosing a medical treatment, interventions such as imposing an incentive or penalty, or inferences such as predictions about the future.

We assume that learning is motivated by some research goal G, represented as a value function $V_G(a, w)$, which assesses how well our goal is satisfied by a given action in a given world. We are intentionally vague about the nature of the goal, beyond its operationalization as a value function. Research may be motivated by data-driven goals (e.g., prediction), policy goals (e.g., optimizing interest rates, identifying effective tax interventions), or more value-laden social goals, in the vein of Dewey's (1973) pragmatist view of science. Goals may also be quantitatively evaluated (e.g., an explicit measure of forecasting accuracy) or qualitatively evaluated (e.g., replicating some qualitative, "stylized" facts about the data). We can also represent purely epistemic goals, so long as we have some way to make sense of "truth conditions" for our hypotheses.[3] Importantly, goals may be strict, requiring

[3] If each $h \in \mathcal{H}$ does correspond to a known subset of W, we can define $D(h)$ to output the subset of W to which h corresponds and define $V_G(D(h), w) = 1$ iff $w \in D(h)$. We can therefore recover the epistemic setting from the pragmatic setting.

https://doi.org/10.1086/709755 Published online by Cambridge University Press

an exact solution to some optimization problem, or approximate, only requiring a solution within some wide margin of error. While we do not consider specific examples here, sufficiently weak goals may permit short-run performance guarantees (just as sufficiently tractable epistemic problems may permit short-run convergence guarantees). Thus, our notion of goal here is very general, and the only necessary criterion is that the goal can be evaluated through some value function.

Given this notation, we define an *environmental learning problem* as $\mathcal{P} = \{W, I, \mathcal{H}, \mathcal{A}, D, V_G\}$, where $\mathcal{A}$ is the set of possible actions, D is the decision mechanism, and V_G is a goal, operationalized as a value function. For the purpose of this discussion, we assume that the decision mechanism is fixed throughout learning and deterministically outputs a best action in pursuit of the goal. For simplicity of presentation, we also assume that V_G is binary valued. With this established, we can define the following pragmatic reliability criteria:

> **Definition 1**. For a hypothesis $h \in \mathcal{H}$, a method M pragmatically verifies h in the limit under D,[4] iff for every world $w \in W$, M converges to h in w iff h induces a goal-satisfying action in w under D. Formally, this is written as
>
> $$\begin{aligned} &\forall\, w \in W((\exists\, E \in I(w))((\forall\, F \subseteq E)(M(F) = M(E) = h))) \\ &\qquad \leftrightarrow V_G(D(h), w) = 1. \end{aligned} \tag{1}$$
>
> Similarly, M pragmatically solves $\mathcal{P}$ in the limit (under D) iff for every world $w \in W$, there exists an information state $E \in I(w)$ such that
>
> 1. $M(E) = h$ for some h that satisfies $V_G(D(h), w) = 1$, and
> 2. For all information states $F \subseteq E$, $D(M(F)) = D(h)$.

Intuitively, M pragmatically verifies h so long as it only converges to h in worlds where h induces the best possible action, and M pragmatically solves $\mathcal{P}$ so long as it converges to a hypothesis that induces the best possible action in every possible world. Thus, we can extend the notion of reliability in the limit to a case in which hypotheses may not be representable as subsets of possible worlds or we cannot or have no need to make sense of "truth conduciveness." Even in such cases, we can still sensibly talk about convergence in the limit to a goal-satisfying hypothesis, given our goals and constraints.

4.2. Pragmatic Underdetermination and the Coarsening Operation. While the above definitions extend the notion of reliability beyond a strictly

[4] For notation purposes, I drop the "under D" when there is no chance for confusion.

https://doi.org/10.1086/709755 Published online by Cambridge University Press

epistemic setting, they do not address the problem of underdetermination. We reviewed the epistemic notion of underdetermination in sections 2 and 3, but when we add pragmatic constraints to the framework, we must also consider a pragmatic notion of underdetermination. In particular, pragmatic underdetermination occurs when the decision mechanism assigns two different hypotheses to the same action. That is, if two different hypotheses result in the same plan of action for achieving a particular goal, then there is no pragmatic reason to prefer one hypothesis over the other. Just as no method can limit verify a hypothesis that is globally underdetermined, no method can pragmatically limit verify a hypothesis that is pragmatically underdetermined, per definition 1. However, there is a principled way to extract a coarser but tractable problem from a pragmatically underdetermined one.

To this end, let $\mathcal{P} = \{W, I, \mathcal{H}, \mathcal{A}, D, V_G\}$ be an environmental learning problem with deterministic D, and let $\sim_D$ be a relation over $\mathcal{H}$ defined as $h_1 \sim_D h_2 \leftrightarrow D(h_1) = D(h_2)$ (following Wellen and Danks 2016). This relation induces a partition of the original hypothesis space into *pragmatic indistinguishability classes*, as any two hypotheses within the same cell result in the same action under D. We can then transform $\mathcal{P}$ into a new problem $\hat{\mathcal{P}} = \{W, I, \hat{\mathcal{H}}, \mathcal{A}, D, V_G\}$, where $\hat{\mathcal{H}}$ is the set of equivalence classes induced under $\sim_D$. Since these equivalence classes group together hypotheses that map to the same action, any two hypotheses from two distinct classes will map to distinct actions. Thus, the coarsened problem $\hat{\mathcal{P}}$ eliminates the pragmatic underdetermination faced by the original problem $\mathcal{P}$.

The coarsening operation is neither new nor unique to the pragmatic learning paradigm. This general principle of extracting coarser but more tractable problems from finer ones is used, whether implicitly or explicitly, to justify scientific practices across a range of domains. A recent example in causation is Causal Feature Learning (Chalupka, Eberhardt, and Perona 2017), which addresses the problem of discovering high-level macrocauses from nonexperimental microvariable data. The authors endorse the view that "macrovariables should be thought of as task-specific" and propose a method for partitioning a space of microvariables into a hypothesis space of macrocauses, with respect to some set of target outputs (141). The operation underlying this framework groups together microvariables that have indistinguishable effects on the output. This is an explicit application of the coarsening operation to construct an optimal hypothesis space for identifying macrocauses of a set of effects.

A less explicit, more historical example of coarsening is seen in the first sections of Skinner (1953, 23–39), wherein he outlines and justifies a framework for behaviorism. The purpose of behaviorism, Skinner writes, is to provide a "functional analysis" of behavior, that is, an understanding of the "external variables of which behavior is a function" (35) and how we can manipulate those external variables to control that behavior. The relevant

https://doi.org/10.1086/709755 Published online by Cambridge University Press

hypothesis space is therefore a set of possible functions relating external circumstances to behavioral outputs. While Skinner does not formally specify all such functions, he characterizes a common "causal chain consisting of three links: (1) an operation performed upon the organism from without—for example, water deprivation; (2) an inner condition—for example physiological or psychic thirst; and (3) a kind of behavior—for example, drinking" (34). Under this view, it might seem that a natural characteristic of any plausible hypothesis would be separability into two functions $f = g \circ h$: one mapping external conditions to hidden internal states and one mapping internal states to actions. But, as Skinner points out, "the second link is useless in the control of behavior unless we can manipulate it" (34). That is, if g, h and g', h' are two distinct pairs of functions that, when composed, induce the same map from external inputs to behavioral outputs, then the hypotheses $g \circ h$ and $g' \circ h'$ are indistinguishable with respect to the goal of behavioral analysis. The solution, therefore, is to coarsen our hypothesis space: we group together all pairs of functions that result in the same input-output map and simply "examine the third link as a function of the first" (35). Thus, we can interpret Skinner's justification for the "behaviorist" hypothesis space as the pragmatic indistinguishability classes of some more general space of cognitive behavior functions, taken with respect to the goal of predicting and controlling behavior.

4.3. Epistemic versus Pragmatic Underdetermination. In our framework, a problem may face both pragmatic and epistemic underdetermination, and it is important to understand how these notions are distinct and how they relate to each other. Intuitively, epistemic underdetermination occurs when our epistemic constraints (i.e., the available data and possible answers under consideration) do not allow us to distinguish between two or more hypotheses. Pragmatic underdetermination occurs when our pragmatic constraints (goals, available actions, decision mechanism, etc.) do not require us to distinguish between two or more hypotheses. Whereas the traditional reliability framework deals with what we can learn from a given data source, the pragmatic extension deals with what we ought to learn. The latter, normative question is, of course, influenced by both our epistemic and pragmatic constraints: we ought not try to learn anything we cannot learn, and we ought not try to learn anything we need not learn. Thus, determining what we ought to learn requires a framework that can accommodate both kinds of constraints, which we achieve by relaxing the notions of hypothesis and truth condition and applying a coarsening operation to the hypothesis space.

To better illustrate these principles, and how they help accommodate the issues raised in section 3, we will consider some variations of the toy problem used as an example in the introduction. Recall the original problem: an urn contains some unknown number of red or blue marbles, and we observe

https://doi.org/10.1086/709755 Published online by Cambridge University Press

a sequence of draws. In each step of the sequence, we observe either a null draw N (with probability 1 if the urn is empty or probability 1/2 if the urn is nonempty) or a single marble drawn uniformly at random from those remaining in the urn. The original problem is to infer the initial contents of the urn from this sequence, and we can represent the hypothesis space $\mathcal{H}$ as the set of all ordered pairs of nonnegative integers. As we previously explained, this problem is locally but not globally underdetermined. In particular, this problem is solved in the limit by the method $M(E) = (\#r, \#b)$, where $\#r$ and $\#b$ are the number of red and blue marbles in the data sequence E, respectively.

Now consider a pragmatic extension of this problem: we are in a room with four levers, corresponding to the following four possibilities: (a) the urn is (initially) empty, (b) the urn contains some red and no blue marbles, (c) the urn contains some blue and no red marbles, and (d) the urn contains some red and some blue marbles. The hypothesis space remains unchanged, but now we must map our hypothesis to an action and choose one of the four levers. If we choose the correct lever we get a reward, otherwise we get no reward. This problem faces pragmatic underdetermination: there are multiple distinct hypotheses that map to the same action (e.g., (3, 0) and (4, 0) both map to lever b), and we therefore have no reason to distinguish between these hypotheses. We can therefore eliminate this pragmatic underdetermination by partitioning the original hypothesis space into four cells, one corresponding to each action. This allows us to define a method $\hat{M}$ that pragmatically solves the coarsened problem $\hat{\mathcal{H}}$ in the limit.

Now consider the original problem, and suppose that the data stream has been corrupted, as is often the case in sciences with human-generated data. In particular, suppose that the first draw of the sequence is obscured, so that we do not know whether the first draw was R, B, or N. In this case, the problem faces epistemic underdetermination: because we cannot observe the first draw, we cannot deduce the original contents of the urn, even if we could observe the complete data stream at once (minus the corrupted draw). This is a case in which succeeding as well as necessary surpasses what is possible. However, we can still apply the same principles to identify the strongest weakening of the original problem that is solvable in the limit. In particular, suppose that the full (corrupted) data stream contains $\#r$ red marbles and $\#b$ blue marbles. As we cannot observe the first draw, this evidence is compatible with three possibilities: $(\#r, \#b)$, $(\#r + 1, \#b)$, and $(\#r,\ \#b + 1)$. Thus, if we define a new hypothesis space by grouping hypotheses into the appropriate triples, we obtain a weaker version of the original problem that is still solvable in the limit by the method $M(E) = \{(\#r, \#b), (\#r + 1, \#b), (\#r, \#b + 1)\}$ (i.e., the method that outputs a disjunction of the three hypotheses). While this method is not guaranteed to converge to the correct singular hypothesis, it is guaranteed to converge to a set of three possibilities that

https://doi.org/10.1086/709755 Published online by Cambridge University Press

contains the correct hypothesis, which is the best possible outcome, given our epistemic constraints.

This helps illustrate the core principles of our framework and how they apply to the issues discussed in section 3. The epistemic part of the framework determines what can and cannot be reliably learned and motivates the *feasibility contextualism* principle: given an inductive learning problem, identify your epistemic constraints, then succeed as well as possible (Kelly 2014). The pragmatic setting, however, provides additional criteria for justifying inferences. In particular, the decision function and goal determine which distinctions are necessary to learn in order to satisfy the goal. Depending on the goal, then, succeeding "as well as possible" may be unnecessary, and given limited resources and finite time, learning more than necessary is often a luxury we cannot afford. We can therefore motivate our pragmatic reliability framework with a modified feasibility contextualism principle: identify your epistemic constraints, then succeed *as well as necessary*.[5]

5. Prediction and Empirical Adequacy. A common goal in scientific research, especially in human-focused sciences like economics, is the prediction of future phenomena. Of course, prediction plays an important role in any science, but many economists and philosophers of economics in particular take the primary purpose of their field to be empirical prediction (e.g., Friedman 1953). Our pragmatic reliability framework is compatible with any research goal representable as a value function, but in the case in which our goal is prediction, the pragmatic and epistemic notions of underdetermination coincide in an important way.

5.1. Prediction as a Research Goal. When our goal is prediction, the decision mechanism D outputs a predicted information state $D(h) \in I$.[6] Even if a hypothesis h does not explicitly correspond to a subset of W, we can identify the output of D with the set of possibilities compatible with that prediction (even if we do not explicitly know what those possibilities are).

We refer to D as *completely predictive* if, for every h, $D(h)$ outputs a complete information state for some world $w \in W$. Formally, we define the *total information in* w to be $\mathrm{Tot}(w) = \cap_{E \in I(w)} E$, and the *complete information in* w to be $\mathrm{Com}(w) = \{w' \in W | \mathrm{Tot}(w') = \mathrm{Tot}(w)\}$. Intuitively, if we have the total data in w, this means that we have all of the available

[5] And, if what is necessary surpasses what is possible, we apply these principles to identify the strongest weakened version of the problem that is solvable.

[6] We assume that D outputs an element of I because I denotes all evidence that is eventually observed, and in order to validate our predictions, we must eventually observe the true value of whatever we predicted.

https://doi.org/10.1086/709755 Published online by Cambridge University Press

evidence in w. If we have the complete data in w, this means that we have all of the available evidence in w, and we know that we have all of the available evidence in w. This distinction is subtle, but important, so it is worth briefly considering an example in which the distinction is relevant.

To this end, suppose that W is the set of all partial recursive functions from $\mathbb{N}$ to $\mathbb{N}$, and the data in each world f consist of input/output pairs $(n, f(n))$. Let f be any such function defined on every input except 1, and let f' be a function defined on every input in N that is identical to f on all inputs not equal to 1 (i.e., $f'(n) = f(n)$ for all $n \neq 1$). The *total data* in world f are the set of all input/output pairs $\mathrm{Tot}(f) = \{(n, f(n)) | n \neq 1\}$, and the total data in world f' are equal to $\mathrm{Tot}(f) \cup \{(1, f'(1)\}$. Thus, the total data in world f are compatible with world f': even if we observed every input/output pair in world f, this would not be sufficient to rule out the possibility that we are in world f'. If, however, we observe all the evidence in f and we know that this is all the available evidence, then we could rule out f', as we would know that the evidence is missing the extra input/output pair that distinguishes f' from f. Thus, this illustrates a problem in which the total data are insufficient to distinguish f from f', but the complete data are sufficient to make this distinction.

A practical example of a completely predictive mechanism is a linear regression model. In function estimation, the data consist of input/output pairs (X, Y), possible worlds are functions relating X to Y, and the complete data in any world F are the full list of input-output pairs $(X, F(X))$. For a hypothesis with parameters β, we can predict the value of Y from any value of X by computing $Y = \beta X$. Therefore, we can generate all of the input-output pairs, which constitute the complete data in some world.

With these formalizations, we can define empirical adequacy as follows: if D is completely predictive, we define the *oracle goal* G_{oracle} by the value function $V_G(D(h), w)$, which outputs 1 iff $D(h) = \mathrm{Com}(w)$. Intuitively, this specifies that under D, h is an *oracle* for w; that is, $D(h)$ predicts everything we will eventually observe in w and nothing we will not. This corresponds to van Fraassen's (1980) notion of empirical adequacy, which we can now operationalize as a research goal.

5.2. Pragmatic and Epistemic Underdetermination. When D is completely predictive, the resulting indistinguishability classes coincide in an important way with the epistemic notion of global underdetermination. In particular, for two hypotheses $h_1, h_2 \subset W$ (in the epistemic sense) we define h_1 and h_2 to be *epistemically indistinguishable* iff the problem of distinguishing between them is globally underdetermined by available data. Formally, h_1 and h_2 are epistemically indistinguishable iff for any $w \in W$,

$$h_1 \cap \mathrm{Tot}(w) \neq \varnothing \leftrightarrow h_2 \cap \mathrm{Tot}(w) \neq \varnothing. \qquad (2)$$

https://doi.org/10.1086/709755 Published online by Cambridge University Press

Recalling that h_1 and h_2 are *pragmatically indistinguishable* iff $D(h_1) = D(h_2)$, we obtain the following result:

Theorem 1. If D is completely predictive, then for any $h_1, h_2 \in \mathcal{H}$, h_1 and h_2 are pragmatically indistinguishable iff $D(h_1)$ and $D(h_2)$ are epistemically indistinguishable.[7]

Note that in the pragmatic framework, h_1 and h_2 are simply points in some representation space and may have no explicit relation to subsets of W. If D is a prediction generator, however, then $D(h)$ is an information state, which we can equate with the set of possibilities in which that information would be observed (even if we do not explicitly know what those possibilities are). Under D, the representation space $\mathcal{H}$ induces a corresponding hypothesis space of predictions $D(\mathcal{H})$. Theorem 1 shows that if D is completely predictive, the epistemic indistinguishability classes of $D(\mathcal{H})$ are exactly the pragmatic indistinguishability classes of $\mathcal{H}$, mapped under D.

A corollary of this theorem better illustrates the relation between empirical adequacy (as a pragmatic goal) and "truth seeking" (as an epistemic goal):

Corollary 1. Let $\mathcal{P} = (W, I, \mathcal{H}, D, \mathcal{A}, G_{\text{oracle}})$ be an environmental learning problem, where D is completely predictive and G_{oracle} is empirical adequacy. Let $\hat{\mathcal{H}}$ be the indistinguishability classes induced by D. Then

1. If $\mathcal{Q} = \{A_1, A_2, ...\}$ is limit solvable in every possible world (in the epistemic sense), then $\mathcal{Q}$ must be a coarsening of $D(\hat{\mathcal{H}})$.
2. $\hat{\mathcal{P}} = (W, I, \hat{\mathcal{H}}, D, \mathcal{A}, G_{\text{oracle}})$ is the hardest problem to limit solve (in the pragmatic sense). That is, converging to a solution for G_{oracle} requires at least as much data as converging to a solution for any other pragmatically solvable problem on (W, I).[8]

Intuitively, this corollary gives us a dual characterization of empirical adequacy. Under the epistemic view, empirical adequacy is the strongest problem that is limit solvable in every possible world. Under the pragmatic view, empirical adequacy is the hardest problem to pragmatically solve in the limit. To put this another way, recall the principle of feasibility contextualism (identify your epistemic constraints, then succeed as well as possible) and our pragmatic modification (identify your epistemic constraints, then succeed as well

[7] See the appendix for a proof.

[8] See the appendix for a proof.

https://doi.org/10.1086/709755 Published online by Cambridge University Press

as necessary). In this language, corollary 1 shows that, when our goal is empirical adequacy, succeeding as well as necessary is the same as succeeding as well as possible. We can therefore think of empirical adequacy as a sort of boundary point between epistemically (i.e., truth-seeking) and pragmatically (i.e., goal-satisfying) solvable problems.

6. Discussion and Future Work. As argued here and elsewhere (e.g., Kelly 1996, 2000), epistemic reliability offers a non-confirmation-based alternative to the problem of induction. Rather than characterizing a weakened entailment relation between hypothesis and evidence, epistemic reliability theorists seek deductive guarantees of in-the-limit success for hypothesis selection methods (for a spectrum of definitions of success). As we show, however, epistemic reliability is difficult to apply or make sense of in certain scientific contexts. For nonrealists and instrumentalists in such fields, hypotheses are not identified with subsets of possible worlds and lack the well-defined truth conditions that ground the epistemic notion of success. Even from a realist's perspective, there are many cases in which the level of global underdetermination goes beyond the scope of standard epistemic reliability. In many human-focused sciences, global underdetermination often results from the shoddy quality of the data, which are usually noisy, incomplete, and nonexperimental, or from overcomplete hypothesis spaces and hypotheses involving latent, unobservable entities.

When a learner faces this level of underdetermination, the situation decomposes into two problems, one pragmatic and one purely epistemic. The epistemic problem is determining what can be reliably learned from the available data stream. The pragmatic problem is determining what we ought to learn, and how we ought to learn it, given our pragmatic goals and constraints. By extending the epistemic framework with nonepistemic goals and actions, we obtain a framework in which both problems can be addressed simultaneously. This provides analogous notions of pragmatic reliability and pragmatic underdetermination and a principled method for extracting solvable problems from badly underdetermined ones.

Finally, our framework helps clarify the relation between purely epistemic goals (i.e., identifying the true hypothesis) and data-driven pragmatic research goals (i.e., prediction of future phenomena). By operationalizing empirical adequacy as a research goal, we can characterize it as both the strongest problem that can be solved in the limit (in the epistemic sense) as well as the hardest problem to solve in the limit (in the pragmatic sense). This dual characterization of empirical adequacy suggests two things: first, we can think of empirical adequacy as a sort of boundary point between pragmatic and purely epistemic research goals. Second, it points to empirical adequacy as a default learning goal (in the absence of any more specific pragmatic goal) whenever an inductive inference problem faces severe

https://doi.org/10.1086/709755 Published online by Cambridge University Press

global underdetermination. Because an empirically adequate hypothesis correctly predicts all possible future observations, converging to an empirically adequate hypothesis allows us to “fast forward” the data stream and use those predicted future data to identify pragmatically correct hypotheses for any “easier” pragmatic goal. We can more intuitively summarize this with the following normative principle: if the total data do not entail the correct answer, the next best solution is to identify an answer that correctly entails the total data. This helps justify the instrumentalist viewpoint, particularly common in fields like economics, that the goal of scientific research just is empirical prediction. While this principle does not justify the instrumentalist viewpoint in all contexts, it does justify prediction as a default research goal in any context that faces the level of global underdetermination we describe above.

In future work, it will be important to consider the role of resource constraints in scientific inquiry. All real-world research is bounded by resource constraints (funding, time, computational power, etc.), and realistically modeling reliable, environment-specific inquiry requires accounting for such constraints. There are several ways in which resource constraints figure into our framework. First, applying the coarsening operator requires computational resources, and in some cases, computing the exact pragmatic equivalence classes may be intractable. Thus, when faced with a globally underdetermined inference problem, there is an inherent trade-off between the level of redundancy or overcompleteness of a hypothesis space and the computational cost of applying the coarsening operator. In such cases, we must better understand the computational cost of coarsening our hypothesis space (i.e., eliminating empirically redundant hypotheses) and the benefits or drawbacks of an overcomplete hypothesis space. In some applications (e.g., function learning with neural networks), a somewhat overcomplete hypothesis space is desirable, as it enables faster and more robust convergence (Lewicki and Sejnowski 2000). Thus, applying pragmatic reliability criteria requires context-specific analysis of this trade-off.

Another critical point to address is the relevance of in-the-limit guarantees for pragmatic decision makers. While it may be reassuring to know that a method is guaranteed to eventually output the correct answer, such a method may not tell us how to act in the short run. Thus, we will often need something stronger than a convergence guarantee for a method to be useful for pragmatic inference problems. In practice, scientists often appeal to some sort of simplicity principle (e.g., Bayesian simplicity, minimum description length) when forced to choose between multiple hypotheses compatible with the same evidence. In the traditional epistemic reliability framework, simplicity criteria are interpreted as stronger convergence properties on methods, defined in terms of the number of worst-case-scenario revisions that a method may require before converging to the truth. Genin and Kelly (2015) provide precise formulations of several such criteria, demonstrating that

https://doi.org/10.1086/709755 Published online by Cambridge University Press

we can often derive bounds on the number of errors a method can make, and define the "simplest" method to be the one that is susceptible to the fewest forcible errors. This provides more direct guidelines and justification for scientists with short-term concerns.

In the pragmatic setting, however, it may be insufficient to define simplicity in terms of the total number of revisions a method can make in a worst case scenario. This is especially true when we consider inference under resource constraints, as different revisions may impose different costs. That is, the resource cost of transitioning from hypothesis h_1 to h_2 may be different from the cost of transitioning from h_1 to h_3, so it is not sufficient to assume that all revisions impose the same cost, as in the epistemic framework. In function learning with neural networks, for example, the cost of transitioning from a network h_1 to h_2 (with the same topology and slightly different weights) may be very different from the cost of transitioning from h_1 to h_3 (say, a network with a different topology and different weights). Thus, the pragmatic reliability theorist's notion of simplicity may be understood in terms of minimizing the total long-term cost of all revisions required to achieve the goal, rather than the raw number of revisions required to arrive at the truth.

Appendix

Lemma A1. For any $w, w' \in W$, $w' \in \mathrm{Tot}(w) \Rightarrow \mathrm{Tot}(w') \subseteq \mathrm{Tot}(w)$.

Proof. By definition, $\mathrm{Tot}(w) = \cap_{E\in I(w)}E$, so $w' \in \mathrm{Tot}(w)$ implies $w' \in E$ for all $E \in I(w)$. Furthermore, $I(w') = \{E \in I | w' \in E\}$, which implies $I(w) \subseteq I(w')$. Since $\mathrm{Tot}(w') = \cap_{E\in I(w')}E$, we obtain $\mathrm{Tot}(w') \subseteq \mathrm{Tot}(w)$. QED

Proof of Theorem 1. The first implication (if h_1 and h_2 are pragmatically indistinguishable, then $D(h_1)$ and $D(h_2)$ are epistemically indistinguishable) is trivial: if h_1 and h_2 are pragmatically indistinguishable, then $D(h_1) = D(h_2)$, so they contain the same possible worlds and are therefore epistemically indistinguishable. Conversely, suppose $D(h_1)$ and $D(h_2)$ are epistemically indistinguishable. By definition, there must exist $w_1, w_2 \in W$ such that $D(h_1) = \mathrm{Com}(w_1)$ and $D(h_2) = \mathrm{Com}(w_2)$. Pick any $w \in D(h_1)$. By definition, we must have $\mathrm{Tot}(w) = \mathrm{Tot}(w_1)$. Furthermore, since $w \in \mathrm{Tot}(w)$, we have $\mathrm{Tot}(w) \cap D(h_1) \neq \varnothing$, and by epistemic indistinguishability we must have $\mathrm{Tot}(w) \cap D(h_2) \neq \varnothing$. So there must exist $w' \in \mathrm{Tot}(w)$ such that $w' \in D(h_2)$, and by the definition of $D(h_2)$, we must also have $\mathrm{Tot}(w') = \mathrm{Tot}(w_2)$. Finally, since $w' \in \mathrm{Tot}(w)$, we must have $\mathrm{Tot}(w') \subseteq \mathrm{Tot}(w)$ by lemma A1. Putting this together yields

$$\mathrm{Tot}(w_2) = \mathrm{Tot}(w') \subseteq \mathrm{Tot}(w) = \mathrm{Tot}(w_1) \Rightarrow \mathrm{Tot}(w_2) \subseteq \mathrm{Tot}(w_1).$$

https://doi.org/10.1086/709755 Published online by Cambridge University Press

Similarly, if we repeat the above steps for an arbitrary $w \in D(h_2)$, we obtain $\mathrm{Tot}(w_1) \subseteq \mathrm{Tot}(w_2)$. Thus, we get $\mathrm{Tot}(w_1) = \mathrm{Tot}(w_2) \Rightarrow D(h_1) = \mathrm{Com}(w_1) = \mathrm{Com}(w_2) = D(h_2)$, so h_1 and h_2 are pragmatically indistinguishable. QED

Proof of Corollary 1. Define $w \sim w'$ iff for all $w'' \in W$, $\mathrm{Tot}(w) \cap \mathrm{Tot}(w'') \neq \varnothing \leftrightarrow \mathrm{Tot}(w') \cap \mathrm{Tot}(w'') \neq \varnothing$. Then we can recover $D(\hat{\mathcal{H}})$ as the set of equivalence classes of W under this equivalence relation. Suppose $\mathcal{Q} = \{A_1, A_2, \ldots\}$ is solvable in the limit in every possible world. Then each> A_i must be a limit-verifiable hypothesis, so there must exist M such that, for all $w \in W$, M converges to A_i in w iff $w \in A_i$. Suppose that for some $w \in A_i$, there exists $w' \neq w$ such that $w' \sim w$, and $w' \notin A_i$. Since A_i is limit verifiable and $w \in A_i$, there must exist $E \in I(w)$ such that $M(E) \vDash A_i$, and for all $\varnothing \neq F \subseteq E, M(F) = M(E)$. Let $E \in I(w)$ be any information state satisfying these criteria. Then $\mathrm{Tot}(w) \subseteq E$. However, if w and w' are epistemically indistinguishable, there must exist a nonempty $E' \subseteq \mathrm{Tot}(w) \cap \mathrm{Tot}(w')$, which will eventually be observed in w'. Then $E' \subseteq \mathrm{Tot}(w) \subseteq E$, so by the second condition, we must have $M(E') = M(E) \vDash A_i$. Therefore M will incorrectly verify A_i in w'. Thus, if A_i is limit verifiable, then for any $w \in A_i$ and any $w' \sim w$, we must have $w' \in A_i$ as well. Thus, for every $w \in A_i$, A_i must contain the epistemic indistinguishability class of w. Since these classes are exactly the cells of $D(\hat{\mathcal{H}})$, each A_i must be a union of cells in $D(\hat{\mathcal{H}})$, so $\mathcal{Q}$ must be a coarsening of $D(\hat{\mathcal{H}})$.

Let M be a method that pragmatically solves $\mathcal{P}$ in the limit. Let D', $\mathcal{A}'$, G' be any other pragmatically limit-solvable problem on the same $(W, I, \mathcal{H})$, which is solved by some M'. For any $w \in W$, suppose $V_G(w, D(h)) = 1$. Then $D(h) = \mathrm{Com}(w)$. Since M' solves G' in the limit, there must exist some $E \in I(w)$ such that $V_{G'}(D'(M'(E)), w) = 1$, and for all $F \subseteq E, M'(F) = M'(E)$. Since $D(h) = \mathrm{Com}(w)$, $D(h) \subseteq E$, so $M'(D(h)) = M'(E)$. Thus, if h is a solution for G_{oracle}, and M' solves G' in the limit under D', then $D'(M'(D(h)))$ is a solution for G'. Thus, G_{oracle} requires at least as much data to converge to a solution as any other pragmatically solvable problem on $(W, I, \mathcal{H})$. QED

REFERENCES

Belloni, A., and V. Chernozhukov. 2011. “High Dimensional Sparse Econometric Models: An Introduction.” In *Inverse Problems and High-Dimensional Estimation*, 121–56. Berlin: Springer.

Carnap, R. 1945. “On Inductive Logic.” *Philosophy of Science* 12 (2): 72–97.

Cartwright, N. 1995. “‘Ceteris Paribus’ Laws and Socio-Economic Machines.” *Monist* 78 (3): 276–94.

———. 2002. “In Favor of Laws That Are Not Ceteris Paribus after All.” In *Ceteris Paribus Laws*, ed. J. Earman, C. Glymour, and S. Mitchell, 149–63. Dordrecht: Springer.

Chalupka, K., F. Eberhardt, and P. Perona. 2017. “Causal Feature Learning: An Overview.” *Behaviormetrika* 44 (1): 137–64.

Clyde, M. A., and R. L. Wolpert. 2007. “Nonparametric Function Estimation Using Overcomplete Dictionaries.” *Bayesian Statistics* 8:91–114.

Dewey, J. 1973. *The Philosophy of John Dewey*, ed. J. J. McDermott. New York: Putnam.

Earman, J., and J. Roberts. 1999. “Ceteris Paribus, There Is No Problem of Provisos.” *Synthese* 118 (3): 439–78.

Friedman, M. 1953. “The Methodology of Positive Economics.” In *Essays in Positive Economics*. Chicago: University of Chicago Press.

Genin, K., and K. Kelly. 2015. “Theory Choice, Theory Change, and Inductive Truth-Conduciveness.” Presented at the 15th Conference on Theoretical Aspects of Rationality and Knowledge, Carnegie Mellon University, June 4–6.

———. 2017. “The Topology of Statistical Verifiability.” arXiv.org, Cornell University. https://arxiv.org/abs/1707.09378.

Haavelmo, T. 1944. “The Probability Approach in Econometrics.” *Econometrica* 12:1–115.

Hausman, D. M. 1988. “Ceteris Paribus Clauses and Causality in Economics.” In *PSA 1988: Proceedings of the 1988 Biennial Meeting of the Philosophy of Science Association*, vol. 2. East Lansing, MI: Philosophy of Science Association.

———. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Hempel, C. G. 1945. “Studies in the Logic of Confirmation.” Pt. 1. *Mind* 54 (213): 1–26.

Hendry, D. F. 1980. “Econometrics—Alchemy or Science?” *Economica* 47:387–406.

Jones, M., and B. C. Love. 2011. “Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition.” *Behavioral and Brain Sciences* 34 (4): 169–88.

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

———. 2000. “The Logic of Success.” In *Philosophy of Science Today*, ed. P. Clark and K. Hawley, 11–38. Oxford: Clarendon.

———. 2014. “A Computational Learning Semantics for Inductive Empirical Knowledge.” In *Johan van Benthem on Logic and Information Dynamics*, ed. A. Baltag and S. Smets, 289–337. Cham: Springer.

Lewicki, M. S., and T. J. Sejnowski. 2000. “Learning Overcomplete Representations.” *Neural Computation* 12 (2): 337–65.

McCloskey, D. N. 1998. *The Rhetoric of Economics*. Madison: University of Wisconsin Press.

Putnam, H. 1963. “Degree of Confirmation and Inductive Logic.” In *The Philosophy of Rudolf Carnap*, ed. P. A. Schilpp, 761–83. La Salle, IL: Open Court.

Sawyer, K., C. Beed, and H. Sankey. 1997. “Underdetermination in Economics: The Duhem-Quine Thesis.” *Economics and Philosophy* 13 (1): 1–23.

Skinner, B. F. 1953. *Science and Human Behavior*. New York: Simon & Schuster.

Sugden, R. 2000. “Credible Worlds: The Status of Theoretical Models in Economics.” *Journal of Economic Methodology* 7 (1): 1–31.

Valente, M. 2005. “Qualitative Simulation Modelling.” Presented at the Fourth European Meeting on Applied Evolutionary Economics, Utrecht University, February.

van Fraassen, B. C. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Wellen, S., and D. Danks. 2016. “Adaptively Rational Learning.” *Minds and Machines* 26 (1–2): 87–102.

Windrum, P., G. Fagiolo, and A. Moneta. 2007. “Empirical Validation of Agent-Based Models: Alternatives and Prospects.” *Journal of Artificial Societies and Social Simulation* 10 (2): 1–8.

https://doi.org/10.1086/709755 Published online by Cambridge University Press