

A Universal Method of Scientific Inquiry

DANIEL N. OSHERSON
IDIAP, C.P. 609, CH-1920 Martigny, Switzerland

OSHERSON@IDIAP.CH

MICHAEL STOB
Calvin College

SCOTT WEINSTEIN
University of Pennsylvania

Abstract. A paradigm of scientific discovery is defined within a first-order logical framework. Within this paradigm, the concept of “successful scientific inquiry” is formalized and investigated. Among other results, it is shown that a simple method of scientific inquiry is universal in the sense that it leads to success on every problem for which success is in principle possible.

Keywords. Computational learning theory, first-order logic, machine inductive inference, scientific methodology

1. Introduction

The present paper advances a formal paradigm of scientific investigation, defines a specific method of inquiry within it, and then demonstrates that the method is universal in the following sense. Given any scientific situation that can be represented by the paradigm, if there is some method that leads to scientific success in that situation, then our method does so also. In addition, the method is machine simulable.

The method in question is a rough implementation of the following advice to scientists.

- (1) Suppose that you know theory T to be true and that you are in possession of data-set σ . Then:
 - a) Believe T and σ .
 - b) Believe any single claim φ of the form $\forall \bar{x} \exists \bar{y} A$ (with A quantifier-free) that does not contradict T and σ .
 - c) Believe the logical consequences of your beliefs.
 - d) Believe nothing else.
 - e) Announce as a hypothesis the first sentence θ (in some fixed, well-ordering of sentences) such that:
 - i. you believe θ ;
 - ii. θ has a suitably complex or otherwise interesting form;
 - iii. θ does not follow logically from T and σ .

If you follow the advice then (e)ii will prevent you from announcing insipid hypotheses, and (e)iii will prevent you from announcing hypotheses that do not rest on genuinely inductive

inference. Of course, these latter virtues also accrue to the advice: "Advance no hypotheses at all!" So the merit of the advice depends on the fact that it underlies a universal method of inquiry, as will be demonstrated.

Before such a demonstration can be attempted, several concepts must be clarified, a task that leads naturally to the definition of a formal paradigm of empirical inquiry. Such a paradigm serves as an abstract model of genuine scientific investigation, including a representation of background theories T , of the states of Nature consistent with T , of the data σ available to scientists, and of scientific success. Within such a context, (1) can be translated into a specific method of scientific investigation, and the associated universality claim takes on precise mathematical meaning. Needless to say, the interest of such a claim is bounded by the fidelity to actual empirical inquiry of the underlying formal paradigm. Although there can be no definitive judgment about such fidelity, we note that variants of our paradigm have already been the focus of studies devoted to diverse aspects of scientific investigation (see Gaifman, Osherson & Weinstein, 1990; Osherson & Weinstein, 1990; Glymour, 1985; Osherson, Stob & Weinstein, 1988; Glymour & Kelly, 1989; and Shapiro, 1981).

The present discussion is organized as follows. Section 2 introduces and motivates the paradigm, and Section 3 reformulates (1) as a method of inquiry. In Section 4 we state a theorem which establishes that using the method is a guarantee for scientific success in any situation where success is possible and some corollaries of this theorem. Proofs of the theorem and corollaries may be found in Osherson, Stob and Weinstein (1990). Section 5 contains a brief discussion of another paradigm of scientific inquiry. The final section is devoted to concluding remarks.

2. A paradigm of scientific inquiry

2.1. Preliminaries

Our paradigm is embedded in a first-order logical context. We fix for the remainder of the discussion a countable, first-order language \mathcal{L} with identity. The sentences of \mathcal{L} are denoted by \mathcal{L}_{sen} , the formulas by \mathcal{L}_{form} . $\varphi \in \mathcal{L}_{form}$ is called *basic* just in case φ is an atomic formula or the negation of such. The set of all basic formulas is denoted BAS. The set $\{0, 1, 2, \dots\}$ of natural numbers is denoted by N . We rely on the following convention regarding implication between sets of open formulas, namely: given $X, Y \subseteq \mathcal{L}_{form}$, $X \models Y$ iff for all structures \mathcal{S} and all assignments h to the variables of \mathcal{L} , if $\mathcal{S} \models X[h]$ then $\mathcal{S} \models Y[h]$.

$\varphi \in \mathcal{L}_{sen}$ is *existential* if φ has the form $\exists \bar{x} A \bar{x}$, where A is quantifier-free. In the sequel we shall be interested in the set of existential sentences satisfied by a given structure for \mathcal{L} . Thus we define:

(2) DEFINITION: Let structure \mathcal{S} be given. The set

$$\{\varphi \in \mathcal{L}_{sen} \mid \mathcal{S} \models \varphi \text{ and } \varphi \text{ is existential}\}$$

is denoted by $Th_3(\mathcal{S})$.

2.2. *States of Nature*

Within our paradigm, a possible state of Nature is interpreted as a countable structure that interprets \mathcal{L} . It is possible to extend our paradigm to structures with uncountable domains (see Osherson, Stob and Weinstein (1991)) for such extension in an analogous setting). For simplicity such extension is not developed here. Henceforth, by “structure” is meant countable structure that interprets \mathcal{L} .

Nature is conceived as arbitrarily choosing one structure from a predefined class of possibilities; the chosen structure is the “actualized” possibility, unknown to the scientist at first. The task of the scientist is to discover some interesting property of Nature’s choice. For this purpose, Nature provides hints in the form of basic formulas satisfiable in the structure she has chosen. The scientist must convert these hints into a stable, true and interesting hypothesis about the unknown structure. To make this clear, we rely on the following terminology.

By a *complete assignment* to a structure \mathcal{S} is meant any mapping of the (countable) set of variables of \mathcal{L} onto $|\mathcal{S}|$. Thus, a complete assignment to \mathcal{S} provides every member of $|\mathcal{S}|$ with at least one temporary name. Given a theory $T \subseteq \mathcal{L}_{sen}$, the class of structures that satisfy T is denoted $MOD(T)$. The class of structures from which Nature makes her choice is conceived as having the form $MOD(T)$ for some $T \subseteq \mathcal{L}_{sen}$. Such a class may be interpreted as all the possible realities consistent with existing scientific knowledge, namely, the knowledge embodied in the background theory T . The contents of this class may be communicated to the scientist via T .

2.3. *The data made available to scientists*

Nature provides information about her chosen structure \mathcal{S} by placing the scientist in an environment governed by \mathcal{S} . This idea is formalized as follows. An *environment* is any ω -sequence over BAS. The set of formulas appearing in an environment e is denoted by $range(e)$. We conceive of $A\bar{x} \in range(e)$ [respectively, $\neg A\bar{x} \in range(e)$] as a message from Nature of the form: “The objects assigned temporary names \bar{x} fall [do not fall] into the set that interprets A .” The following definition specifies the sense in which a structure governs an environment.

- (3) DEFINITION: Let environment e , structure \mathcal{S} , and complete assignment g to \mathcal{S} be given. e is for \mathcal{S} via g just in case $range(e) = \{\beta \in BAS \mid \mathcal{S} \models \beta[g]\}$. e is for \mathcal{S} just in case e is for \mathcal{S} via some complete assignment.

To illustrate, suppose that the following environment e is for structure \mathcal{S} .

$$Px_2 \quad \neg Qx_3x_2 \quad x_4 = x_5 \quad Px_4 \quad \dots$$

Then e may be construed as the following, endless message about \mathcal{S} .

The object given temporary name x_2 belongs to P^S . The object with temporary name x_3 is such that the pair x_3, x_2 belongs to the complement of Q^S . The objects given temporary names x_4 and x_5 are identical. Object x_4 (and hence object x_5) belongs to $P^S \dots$

An environment for a structure \mathcal{S} provides a clear picture of \mathcal{S} . This is the content of the following lemma, proved in Osherson and Weinstein (1986).

- (4) LEMMA: Let environment e and structures \mathcal{S} and \mathcal{U} be given. If e is for both \mathcal{S} and \mathcal{U} then \mathcal{S} and \mathcal{U} are isomorphic.

Given $i \in N$ and environment e , the finite initial segment of length i in e is denoted by \bar{e}_i . The next two lemmas are easy consequences of the Compactness Theorem and will be used later. Their formulation is aided by the following notation.

- (5) DEFINITION: Let σ be a finite sequence over BAS, and suppose that the free variables appearing in σ are $\bar{x} = x_1 \dots x_n$. The conjunction, in order of appearance, of the basic formulas of σ is denoted by $\wedge\sigma$. The existential closure $\exists\bar{x}\wedge\sigma$ of $\wedge\sigma$ is denoted $\exists\wedge\sigma$.
- (6) LEMMA: Let $T \subseteq \mathcal{L}_{sen}$ and structure \mathcal{S} be given. Suppose that e is an environment for \mathcal{S} . Then for all $\theta \in \mathcal{L}_{sen}$, $T \cup Th_3(\mathcal{S}) \models \theta$ iff for some $i \in N$, $T \cup \{\exists\wedge\bar{e}_i\} \models \theta$.
- (7) LEMMA: Let e be an environment for structure \mathcal{S} via complete assignment g . Let $\xi \in \mathcal{L}_{form}$ be quantifier-free. Then $\mathcal{S} \models \xi[g]$ implies that for some $i \in N$, $\wedge\bar{e}_i \models \xi$.

2.4. Scientists

Scientists are conceived as working in an environment e for a structure \mathcal{S} by examining $\bar{e}_1, \bar{e}_2, \dots$ in turn. After each such examination, the scientist announces some $\theta \in \mathcal{L}_{sen}$ to express the hypothesis that $\mathcal{S} \models \theta$. Lemma (4) ensures that no ambiguity arises about the truth of such hypotheses. To proceed formally, let SEQ be the set of all finite sequences over BAS. (Thus, $SEQ = \{\bar{e}_i \mid e \text{ is an environment and } i \in N\}$.) By a (*formal*) *scientist* is meant any function from SEQ to \mathcal{L}_{sen} . Note that scientists can be total or partial, computable or uncomputable.

2.5. Success

To be successful in a given environment, we stipulate that a scientist's successive conjectures must eventually stabilize to a sentence that is true and interesting. Stabilization is equated with "convergence," defined as follows.

- (8) DEFINITION: Let environment e , $\theta \in \mathcal{L}_{sen}$, and scientist Ψ be given. Ψ *converges* on e to θ just in case $\Psi(\bar{e}_i) = \theta$ for all but finitely many $i \in N$. In this case we write $\Psi(e) = \theta$.

We now consider the conditions under which a hypothesis θ is interesting. That θ be true may be entered as the first such condition. The other conditions are that:

- (a) θ be suitably complex; and
- (b) θ not be a logical consequence of the background theory plus the available data.

Condition (a) might require that θ exhibit a certain quantificational form such as being logically equivalent over the background theory to no Π_2 or Σ_2 sentence. Rather than attempting to stipulate once and for all the kind of complexity required, we shall leave such specification as a free parameter in our definition of "interesting hypothesis." Condition (b) is intended to disallow hypotheses that are mere logical consequences of the information available to a scientist, resting on no scientific competence of an inductive or empirical nature. Whether a hypothesis satisfies (b) thus depends on the background theory communicated to the scientist and on the structure underlying the environment in which he is working. These items also figure as parameters in our definition.

(9) DEFINITION: Let $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, structure \mathcal{S} , and $\theta \in \mathcal{L}_{sen}$ be given. θ is $\mathcal{S}, T, \mathbf{P}$ -interesting just in case:

- (a) $\mathcal{S} \models \theta$
- (b) $\theta \in \mathbf{P}$
- (c) $T \cup Th_3(\mathcal{S}) \not\models \theta$

In the definition, \mathbf{P} represents the class of potentially interesting hypotheses (having the right form or complexity), T represents the background theory communicated to the scientist, and \mathcal{S} represents Nature's choice of reality from among $MOD(T)$. By Lemma (6), clause (c) may be read: θ does not follow from T conjoined with any evidence available in an environment for \mathcal{S} .

Within our paradigm, scientific success is conceived as extending the background theory by adding a true and interesting hypothesis. Success is thus called "extension," as specified in the following definition.

(10) DEFINITION: Let $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ and scientist Ψ be given. Ψ \mathbf{P} -extends T just in case for every $\mathcal{S} \in MOD(T)$ and every environment e for \mathcal{S} there is $\theta \in \mathcal{L}_{sen}$ such that:

- (a) θ is $\mathcal{S}, T, \mathbf{P}$ -interesting,
- (b) $\Psi(e) = \theta$.

In this case, T is \mathbf{P} -extendable.

If $\mathbf{P} = \mathcal{L}_{sen}$ then the definition comes to this. Ψ extends theory T just in case for every environment e for a structure \mathcal{S} satisfying T , Ψ converges to a sentence that is true in \mathcal{S} and does not follow from $T \cup Th_3(\mathcal{S})$. If \mathbf{P} is a proper subset of \mathcal{L}_{sen} then the additional requirement is imposed that the sentence have suitable complexity or form.

We shortly provide examples of extendability and of nonextendability. First we discuss and set aside some trivial reasons for nonextendability.

2.6. Semi-complete theories

Suppose that $T \subseteq \mathcal{L}_{sen}$ is the background theory communicated to a scientist Ψ about the structure \mathcal{S} in which Ψ is working. If T is complete then Ψ cannot extend knowledge about \mathcal{S} , so scientific success is impossible on trivial grounds. But even if T is not complete, T may be so strong as to prevent Ψ from conjecturing interesting sentences about certain structures. The following definition formulates this idea.

- (11) DEFINITION: Let $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ be given. T is *semi-complete with respect to \mathbf{P}* just in case there is $\mathcal{S} \in \text{MOD}(T)$ such that for all $\theta \in \mathbf{P}$, either $T \cup Th_3(\mathcal{S}) \models \sigma$ or $T \cup Th_3(\mathcal{S}) \models \neg\theta$.

The reason that semi-complete theories render extension trivially impossible is this:

- (12) LEMMA: Let $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ be such that T is semi-complete with respect to \mathbf{P} . Then for some $\mathcal{S} \in \text{MOD}(T)$, $\{\theta \in \mathcal{L}_{sen} \mid \theta \text{ is } \mathcal{S}, T, \mathbf{P}\text{-interesting}\} = \emptyset$.

For example, suppose that $\mathbf{P} = \mathcal{L}_{sen}$ and that T is model complete (for the concept of model-completeness, see Chang and Keisler (1973, Section 3.1)). Then T is semi-complete with respect to \mathbf{P} . This is because of the following fact about model completeness.

- (13) Let $T \subseteq \mathcal{L}_{sen}$ be model complete. Then for every $\theta \in \mathcal{L}_{sen}$ there is existential $\varphi \in \mathcal{L}_{sen}$ such that $T \models \varphi \leftrightarrow \theta$.

None of the results of the present paper hinge on the existence of semi-complete theories.

2.7. Examples

Here is an example of extendability followed by an example of nonextendability. The examples follow as corollaries to our main result (for proofs of these corollaries see Osherson, Stob and Weinstein (1990, Section 5)).

- (14) EXAMPLE: Suppose that \mathcal{L} is limited to the binary relation symbol R , and let finite $T \subseteq \mathcal{L}_{sen}$ axiomatize the linear orders with either first or last points (or both). Assume that \mathbf{P} includes the properly Σ_2 sentences (i.e., the existential-universal sentences that are not logically equivalent to any sentence of the form $\exists \bar{x}\varphi$ or $\forall \bar{x}\varphi$ for quantifier free φ). Then some total computable scientist \mathbf{P} -extends T . Observe that for all infinite \mathcal{S} , $\mathcal{U} \in \text{MOD}(T)$, $Th_3(\mathcal{S}) = Th_3(\mathcal{U})$. So this example shows that extension is not ruled out in such cases.

(15) EXAMPLE: Let \mathcal{L} be as in Exmple (14), and let $T \subseteq \mathcal{L}_{sen}$ axiomatize the linear orders without endpoints. Then even if \mathbf{P} is taken to be all of \mathcal{L}_{sen} , T is not \mathbf{P} -extendable.

3. A universal method

The advice given in (1) inspires the design for a specific scientist, to be called “M.” M is a Turing machine with two oracles (for a discussion of such devices, see Rogers (1967, Chapter 9)). One oracle holds whatever theory T is to be extended. The other holds whatever subset \mathbf{P} of \mathcal{L}_{sen} constitutes the set of potentially interesting hypotheses. If attention is limited to cases in which both T and \mathbf{P} are recursively enumerable then M 's oracles can be replaced by two additional arguments, namely, the corresponding r.e. indices; see Corollary (19) in Section 4.

Our description of M requires two definitions.

(16) DEFINITION: Given $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, we denote by T, \mathbf{P} -potential the collection of all pairs (φ, θ) such that:

- (a) $\varphi \in \mathcal{L}_{form}$ and $\theta \in \mathcal{L}_{sen}$
- (b) φ is universal
- (c) $\theta \in \mathbf{P}$
- (d) $T \cup \{\varphi\} \models \theta$

(17) DEFINITION: Given $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ and $\sigma \in \text{SEQ}$, we denote by T, \mathbf{P}, σ -cancelled the set of all $(\varphi, \theta) \in T, \mathbf{P}$ -potential such that either:

- (a) $T \cup \{\exists \wedge \sigma\} \models \theta$ or
- (b) $T \cup \{\wedge \sigma\} \models \neg \varphi$

M relies on two uniform methods. The first converts any oracles $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ into a T, \mathbf{P} -recursive enumeration of T, \mathbf{P} -potential. The second converts any oracles $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ and any $\sigma \in \text{SEQ}$ into a T, \mathbf{P} -recursive enumeration of T, \mathbf{P}, σ -cancelled. The compactness and completeness theorems for first-order logic guarantee the existence of such methods. Given $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, $\sigma \in \text{SEQ}$, and $n \in \mathbb{N}$, the set consisting of the first n members of M 's enumeration of T, \mathbf{P}, σ -cancelled is denoted by $[T, \mathbf{P}, \sigma$ -cancelled, $n]$.

Given $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ and $\tau \in \text{SEQ}$, $M^{T, \mathbf{P}}$ finds the first pair (φ, θ) (if such exists) in its enumeration of T, \mathbf{P} -potential such that $(\varphi, \theta) \notin \cup_{\sigma \subseteq \tau} [T, \mathbf{P}, \sigma$ -cancelled, $\text{length}(\tau)]$. If no such pair exists, then $M^{T, \mathbf{P}}(\tau)$ is undefined; otherwise, $M^{T, \mathbf{P}}(\tau) = \theta$. This concludes the description of M .

More direct translation of (1) into a mechanical procedure is obstructed by clauses (1)b, e (the set of nonconsequences of T and σ is not in general effectively enumerable). It is for this reason that M must generate the set T, \mathbf{P}, σ -cancelled, which is progressively subtracted from T, \mathbf{P} -potential. With this caveat in mind it is not hard to see that M 's behavior is in rough conformity with (1).

4. Universality of M and corollaries thereof

The following theorem establishes that M is universal in the sense of Section 1. The reader may consult Osherson, Stob and Weinstein (1990, Section 4), for a proof of this theorem.

(18) THEOREM: For all $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, if T is \mathbf{P} -extendable then $M^{T,\mathbf{P}}$ \mathbf{P} -extends T .

According to the theorem, if T is \mathbf{P} -extendable by either computable or noncomputable scientist then such extension can be achieved by the machine M outfitted with T and \mathbf{P} as oracles. From the definition of M in Section 3 it is clear that the roles played by the oracles for T and \mathbf{P} can be played equally well by r.e. indices for these sets, if such indices exist. This observation yields the following corollary.

(19) COROLLARY: There is a total computable function $f: N \times N \times \text{SEQ} \rightarrow \mathcal{L}_{sen}$ such that for all $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, if T is \mathbf{P} -extendable and if $t, p \in N$ are r.e. indices for T and \mathbf{P} respectively, then $\lambda\sigma.f(t, p, \sigma)$ \mathbf{P} -extends T .

A scientist who is successful on two environments for the same structure may nonetheless converge on them to distinct sentences. We are thus led to formulate a stronger criterion of scientific success, as follows.¹

(20) DEFINITION: Let $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ and scientist Ψ be given. Ψ uniformly \mathbf{P} -extends T just in case for every $\mathcal{S} \in \text{MOD}(T)$ there is $\mathcal{S}, T, \mathbf{P}$ -interesting $\theta \in \mathcal{L}_{sen}$ such that $\Psi(e) = \theta$ for every environment e for \mathcal{S} . In this case, T is uniformly \mathbf{P} -extendable.

It is obvious that if a theory is uniformly \mathbf{P} -extendable then it is \mathbf{P} -extendable. The converse follows from the fact—easily verified—that given any $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, $M^{T,\mathbf{P}}$ uniformly \mathbf{P} -extends T if $M^{T,\mathbf{P}}$ \mathbf{P} -extends T . Thus:

(21) COROLLARY: For all $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, if T is \mathbf{P} -extendable then T is uniformly \mathbf{P} -extendable.

Our method M has another noteworthy property. For all $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ and all environments e for any $\mathcal{S} \in \text{MOD}(T)$, if $M^{T,\mathbf{P}}$ converges on e to $\theta \in \mathcal{L}_{sen}$ then θ is $\mathcal{S}, T, \mathbf{P}$ -interesting. This is true whether or not $M^{T,\mathbf{P}}$ \mathbf{P} -extends T . Thus, even when M is incapable of \mathbf{P} -extending T , M does not misleadingly persevere on a false hypothesis in environments embraced by T . Adapting terminology from Blum and Blum (1975), scientists with this quality may be termed “reliable.”

(22) DEFINITION: Let $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$ and scientist Ψ be given. Ψ is T, \mathbf{P} -reliable just in case for every $\mathcal{S} \in \text{MOD}(T)$, every environment e for \mathcal{S} , and every $\theta \in \mathcal{L}_{sen}$, if $\Psi(e) = \theta$ then θ is $\mathcal{S}, T, \mathbf{P}$ -interesting.

Combining M 's reliability with its uniformity, we have the following.

- (23) COROLLARY: There is an oracle machine M such that for all $T, \mathbf{P} \subseteq \mathcal{L}_{sen}$, $M^{T, \mathbf{P}}$ is T, \mathbf{P} -reliable, and if T is \mathbf{P} -extendable then $M^{T, \mathbf{P}}$ uniformly \mathbf{P} -extends T .

5. Truth detection

The foregoing paradigm is but one example of a formal construal of scientific inquiry. For comparison, we will consider in this section another paradigm of scientific inquiry.

Scientific investigation is sometimes directed at determining the truth of a given conjecture on the basis of available data instead of discovering some hypothesis to explain that data. We proceed to modify the paradigm described above to model this notion of inquiry. Once again, the possible states of nature under investigation are interpreted as countable structures for a fixed, countable first-order language \mathcal{L} with identity and as before, scientists gain access to data about such structures via environments for those structures. Scientists themselves are now taken to be mappings from $\mathcal{L}_{sen} \times \text{SEQ}$ to $\{t, f\}$. Thus, a scientist Ψ may be conceived as a system that converts arbitrary $\theta \in \mathcal{L}_{sen}$ into a function $\lambda\sigma. \Psi(\theta, \sigma)$ that conjectures a truth-value for θ in whatever structure \mathcal{S} has given rise to the data σ . To be successful on θ in \mathcal{S} , Ψ must “detect” the truth-value of θ in \mathcal{S} , as specified by the following definition.

- (24) DEFINITION: Let $\theta \in \mathcal{L}_{sen}$, structure \mathcal{S} and scientist Ψ be given. Ψ detects θ in \mathcal{S} just in case for every complete assignment g to \mathcal{S} , and every environment e for \mathcal{S} via g , if $\mathcal{S} \models \theta$ then $\Psi(\theta, \bar{e}_n) = t$ for cofinitely many $n \in \mathbb{N}$, and if $\mathcal{S} \models \neg\theta$ then $\Psi(\theta, \bar{e}_n) = f$ for cofinitely many $n \in \mathbb{N}$.

Thus, we credit Ψ with detecting θ in \mathcal{S} just in case Ψ 's successive conjectures about the truth-value of θ in \mathcal{S} eventually stabilize to the correct one in response to increasingly complete information about \mathcal{S} .

- (25) DEFINITION: Let class \mathcal{K} of structures, $\theta \in \mathcal{L}_{sen}$ and scientist Ψ be given. Ψ detects θ in \mathcal{K} just in case for all $\mathcal{S} \in \mathcal{K}$, Ψ detects θ in \mathcal{S} . In this case θ is detectable in \mathcal{K} .

In Osherson, Stob and Weinstein (1991) it is shown that detection problems determined by first-order background knowledge admit of a uniform solution. This is the content of the following theorem, a proof of which may be found in Osherson, Stob and Weinstein (1991, Section 3).

- (26) THEOREM: There is an oracle machine M such that for all $\theta \in \mathcal{L}_{sen}$ and $T \subseteq \mathcal{L}_{sen}$, if θ is detectable in $\text{MOD}(T)$ then M^T detects θ in $\text{MOD}(T)$.

This theorem yields the following syntactic characterization of detectability as a corollary, a proof of which may be found in Osherson, Stob and Weinstein (1991, Section 3).

- (27) COROLLARY: Let $\theta \in \mathcal{L}_{sen}$ and $T \subseteq \mathcal{L}_{sen}$ be given. θ is detectable in $\text{MOD}(T)$ iff both θ and $\neg\theta$ are equivalent over T to existential-universal sentences.

Theorem (26) suggests the following question: is uniform detection of the truth-value of first-order sentences possible with respect to background knowledge expressed in languages stronger than first-order logic, or does first-order logic enjoy a special status for inductive inference? In Osherson, Stob and Weinstein (1991), we give a precise sense to this question in the context of abstract model theory. We show that first-order logic can in fact be characterized uniquely in terms of the uniform solvability of the detection problems which are expressible in its language.

6. Conclusion

The results above suggest that issues in knowledge representation can be fruitfully pursued in connection with a model-theoretic approach to learning. Consider a sophisticated database, DB. Part of the knowledge stored in DB may consist of well-confirmed statements that axiomatize some class of models. To augment its knowledge, DB may rely exclusively on external sources, or DB might launch an investigation of its own aided by an automated system of scientific inquiry. In the latter case, it would be wise for DB to consider its prospects for success. Is there any guarantee that DB could succeed in an arbitrary model of its current axiom set? If DB restricts its attention to the “minimal models” of its axioms by adopting some version of the “closed world hypothesis” might success in its empirical inquiry then be guaranteed? Suppose that DB’s discovery routine requires knowledge about the truth of some sentence that does not follow from available data, but does receive support from some nonmonotone rule of inference. To what extent is the reliability of DB’s routine compromised by supplying it with information of this sort? Such questions, and others like them, are crucial to the confidence that DB may place in the results of its empirical investigation. It would, therefore, be useful to equip DB in advance with the mathematical means necessary to assess the feasibility of its contemplated empirical inquiry.

The answers to these feasibility questions depend, in part, on the kind of axioms which DB takes as its scientific starting point, for example, whether the axioms are couched in a first-order language, whether they are of special syntactic form, whether attention is restricted to finite models, etc. The answers also depend on the data available to DB and the criterion of success to which it aspires. Progress in answering these questions would in all likelihood be facilitated by deploying the considerable understanding of logical theory that has developed over the last century. This knowledge figures prominently in contemporary theoretical studies of knowledge representation. Perhaps it can be deployed, as well, in computational learning theory, and form a bridge between the two disciplines.

Acknowledgments

Research support was provided by the Office of Naval Research under contract No. N00014-89-J-1725 to Osherson and Weinstein, by a Siemens Corporation grant to Osherson, and by NSF grant DMS-88-00030 to Stob.

Notes

1. In Blum and Blum (1975) an analogous concept (called "order independence") is defined in a different inductive setting.

References

- Blum, L. & Blum, M. (1975). Toward a mathematical theory of inductive inference. *Information & Control*, 28, 125–155.
- Chang, C.C., & Keisler, H.J. (1973). *Model theory*. North Holland.
- Gaifman, H., Osherson, D. & Weinstein, S. (1990). A reason for theoretical terms. *Erkenntnis*, 32, 149–159.
- Glymour, C. (1985). Inductive inference in the limit. *Erkenntnis*, 2, 23–31.
- Glymour, C. & Kelly, K. (1989). On converging to the truth and nothing but the truth. *Philosophy of Science*, 56.
- Osherson, D. & Weinstein, S. (1986). Identification in the limit of first-order structures. *Journal of Philosophical Logic*, 15, 55–81.
- Osherson, D. & Weinstein, S. (1989). Paradigms of truth-detection. *Journal of Philosophical Logic*, 18, 1–42.
- Osherson, D. & Weinstein, S. (1990). On advancing simple hypotheses. *Philosophy of Science*, 57, 266–277.
- Osherson, D., Stob, M. & Weinstein, S. (1988). Mechanical learners pay a price for Bayesianism. *Journal of Symbolic Logic* 53, 1245–1251.
- Osherson, D., Stob, M. & Weinstein, S. (1990). A mechanical method of successful scientific inquiry. In M. Fulk & J. Case (Eds.). *Proceedings of the Third Annual Workshop on Computational Learning Theory*. Morgan-Kaufmann.
- Osherson, D., Stob, M. & Weinstein, S. (1991). A universal inductive inference machine. *Journal of Symbolic Logic*, 56, 661–672.
- Rogers, H. (1967). *Theory of recursive functions and effective computability*. McGraw-Hill.
- Shapiro, E. (1981). An algorithm that infers theories from facts. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence, IJCAI*.