Scientific Discovery from the Perspective of Hypothesis Acceptance

Eric Martin†‡
University of New South Wales

Daniel Osherson
Rice University

A model of inductive inquiry is defined within the context of first-order logic. The model conceives of inquiry as a game between Nature and a scientist. To begin the game, a nonlogical vocabulary is agreed upon by the two players, along with a partition of a class of countable structures for that vocabulary. Next, Nature secretly chooses one structure ("the real world") from some cell of the partition. She then presents the scientist with a sequence of facts about the chosen structure. With each new datum the scientist announces a guess about the cell to which the chosen structure belongs. To succeed in his or her inquiry, the scientist's successive conjectures must be correct all but finitely often, that is, the conjectures must converge in the limit to the correct cell. Different kinds of scientists can be investigated within this framework. At opposite ends of the spectrum are dumb scientists that rely on the strategy of "induction by enumeration," and smart scientists that rely on an operator of belief revision. We report some results about the scope and limits of these two inductive strategies.

1. Introduction. Contemporary perspectives on scientific method separate into two broad schools distinguished by how each perceives the investment of a rational agent in a given theory. For the Probabilist School, investments are degrees-of-belief, distributed across all the theories in play. In

†Martin: School of Computer Science and Engineering, The University of New South Wales, Sydney, NSW 2052, Australia, e-mail: emartin@cse.unsw.edu.au; Osherson: Psychology Dept., MS-25, Rice University, P.O. Box 1892, Houston TX 77005–1892, e-mail: osherson@rice.edu

‡We thank Richard Grandy and Oliver Schulte for comments on an earlier draft. Supported by the Australian Research Council Grant A49803051 to Martin and NSF IIS-9978135 to Osherson.

Copyright 2002 by the Philosophy of Science Association. All rights reserved.
contrast, advocates of Formal Learning Theory conceive of investment in more absolute terms, and designate it as "acceptance." Acceptance represents deliberate selection of a particular hypothesis as the one to elaborate, test, and defend at a particular moment of investigation. Such selection is provisional and can be modified in light of new observations. For analysis of the relations among acceptance, belief and intellectual commitment, see Cohen (1992). Comparison of the probabilist school with Formal Learning Theory is available in Earman (1992) and in Kelly (1994). Instead of pursuing the comparison, the goal of the present paper is to convey a sense of the acceptance-based perspective on scientific discovery. We begin by describing a guessing game that embodies fundamental features of the approach. The game is played between two players. Call them "Nature" and "the scientist." The following pieces are used.

1. A countable, first-order language $L$, with identity and variables $v_0$, $v_1$, $v_2$, etc. (We denote the $i$th variable by $v_i$, and similarly for other indexes needed below.)

2. A collection $C$ of countable structures that interpret $L$.

3. A partition of the collection $C$ into cells.

To illustrate, let the nonlogical vocabulary of $L$ be limited to a single unary predicate, say, $H$. Let the countable structures in $C$ be all those in which the extension of $H$ is finite. And let the partition be based on the cardinality of $H$'s extension. In one cell, the extension has cardinality 0; in another, it has cardinality 1, and so forth (there is no cell for infinite cardinality).

Nature prepares herself for the game as follows. First she chooses one structure $S$ from some cell of the partition. Pursuing our illustration, she may choose the structure consisting of the natural numbers, with $H$ interpreted as the set $\{5, 7, 9\}$. She then maps the variables of $L$ onto the domain of $S$, thereby using the variables as temporary names for all the objects in the countable universe of $S$. Nature is allowed to choose any surjective mapping she pleases. For example, she might map both $v(2i)$ and $v(2i + 1)$ to the number $i$. Nature's choice of structure and variable-mapping will be called Nature's interpretation.

Now the game begins. Nature starts by selecting a basic formula that is true in her interpretation. (Recall that a basic formula is an atomic formula or its negation.) Following our example, she might choose any of $Hv_{10}$, $\neg Hv_5$, $Hv_{19}$, $v_4 = v_5$, $v_4 \neq v_8$, etc. In response to Nature's basic formula, the scientist designates a cell of the partition, for example, the cell consisting of all structures in which $H$'s extension has 1 member. It is then Nature's turn again. She chooses a new basic formula that is true in her interpretation. Thus, two basic formulas now stand revealed. The scientist responds to them with another designation of a partition cell. The
new cell may or may not be identical to the first one. The two players go
on like this forever. In each round, Nature reveals a new basic formula,
and the scientist responds with a conjectured cell of the partition.

Here are the rules governing play. Nature can withhold no fact forever.
That is, for every basic formula that is true in Nature’s interpretation,
there must be a round of the game in which Nature reveals it. The scientist
wins the game just in case for all but finitely many rounds, he announces
the cell from which Nature’s chosen structure was drawn. That is, the
scientist wins just in case he converges to the correct partition member.

If you are willing to conceive of the starting collection of structures as
possible worlds, then our game looks a little like science. Collections of
worlds are often called “propositions,” so the partition represents a set of
mutually exclusive propositions, one of which is true (namely, the one
from which Nature draws the real world). The scientist’s job is to discover
the true proposition by examining facts about the world. If Nature’s list
of basic formulas begins with \(-Hv3, Hv7, -Hv8, v7 \neq v8\) then the facts
can be stated informally like this:

The first object I’ve seen (denoted arbitrarily by \(v3\)) does not satisfy
\(H\). The second object does satisfy \(H\). The third object does not. The
second and third objects are distinct etc.

As in scientific settings, the available facts might never imply the correct
proposition. For example, in our illustration, at no stage of the game is
the cardinality of \(H\)’s extension revealed to the scientist; the next datum
might inflate it. But success in the game may nevertheless be possible.

Indeed, in our illustration, the scientist has a reliable strategy, one that
works no matter what Nature chooses. If at each round of the game he
guesses the cell corresponding to the smallest cardinality for \(H\) that is con-
sistent with his data, then he is guaranteed to converge to the correct cell of
the partition. So, this game can be won reliably. In contrast, suppose that
we add a cell made up of structures in which the extension of \(H\) is infinite.
Then it can be shown that no scientist is reliable for the game based on this
expanded partition. In other words, no mapping from data to cells of the
expanded partition is guaranteed to converge to the right cell.

Let’s give a name to reliable success. We say that the scientist \(solves\) a
partition if no matter what structure Nature chooses, no matter how she
maps variables onto the structure’s domain, and no matter how she lists
the basic facts, the scientist’s guesses converge to the cell from which the
structure was drawn. Thus, our first partition is \(solvable\), in the sense that
some scientist solves it. In contrast, the expanded partition is not solvable.

To describe a different kind of partition (over a new vocabulary), call
a total order \(one-sided\) if it has a least point or a greatest point but not
both. Suppose that \(L\) contains just the binary relation symbol \(R\) (along
with identity), and consider a partition of all countable structures that interpret \( R \) as a one-sided order. In one cell lie the one-sided orders with a least point. In another cell are the one-sided orders with a greatest point. There are no other cells. For future reference, let us call this example the one-sided split. This partition is solvable, as we'll see shortly. It becomes unsolvable if a third cell is added containing the two-sided orders, namely, total orders with both a least and greatest point.

The one-sided split and its extension have the particularity of being specifiable within \( L \). The class of one-sided structures with a least point (or a greatest point) can be defined by a single sentence, as can the class of two-sided orders. More generally, we call a partition first-order definable if there is a set \( T \) (possibly infinite) of sentences, and further sentences \( \theta_1 \ldots \theta_n \) such that \( T \) implies exactly one of the \( \theta_i \)'s, and cell \( i \) of the partition consists of the countable models of \( T \cup \{ \theta_i \} \). First-order definable partitions resemble scientific settings in which an underlying theory delimits the collection of possible realities, and research is oriented toward extending the theory by choice among a finite set of additional axioms. We have already seen that first-order definability is neither necessary nor sufficient for solvability. Thus, our first example—involving different finite cardinalities for the extension of the predicate \( H \)—is solvable but not first-order definable. And expanding the one-sided split by adding a cell for the two-sided orders yields a partition that is first-order definable but not solvable.

As our examples suggest, it can be challenging to trace the boundary between the solvable and unsolvable partitions. Combinatorial conditions for solvability have been investigated in recent years, with progress reported in Martin and Osherson (1998).

The inductive games we have described can be modified in various ways. The restriction to countable structures can be dropped, whereas a restriction to computable scientific strategies can be imposed. The data available to the scientist can be limited to just the atomic formulas (no negations), or enriched to include universals or other kinds of formulas. Some vocabulary from the background language can be considered "theoretical," and left unrepresented in the data. The data can also be corrupted in ways that mirror the plight of real scientists (e.g., some facts can be missing; others, specious). Scientists can be made less passive, and allowed to choose the next atomic formula whose truth value will be revealed. Success can be made a graded affair, measured in terms of the probability of converging to the correct cell of a partition. Moreover, the impact of any such modification can be assessed in the context of special kinds of partitions, such as the first-order definable partitions just discussed. And the different models can be compared to the Bayesian picture, which eschews the choice of a specific hypothesis at each round of the game in favor of a distribution of credibility over all the cells.
These topics, among others, are discussed in Kelly (1996), Martin and Osherson (1998), Jain et al. (1999), and in references cited there. The results obtained help to articulate a unified perspective on deductive and inductive inquiry, starting from the concept of hypothesis acceptance. Themes expressed in Karl Popper's philosophy (e.g., Popper 1959) are thus revisited in an alternative framework.

To provide a closer look at theorizing within the acceptance-based perspective, we remain within the simple inductive paradigm presented above, and focus on a single issue, concerning the inductive power of different strategies for choosing hypotheses in response to data. (Results will be drawn from Martin and Osherson 1998, 2000, 2001.) Two approaches to hypothesis selection are contrasted. One relies on brute force, by marching through a prestored list of possibilities. This is called "induction by enumeration." The other approach relies on operations of rational belief revision, in the sense of Gardenfors (1988) and Hansson (1999).

2. Induction by Enumeration. The simplest embodiment of induction by enumeration puts all the cells of the target partition in a list, then guesses the first cell that is consistent with the data that Nature has made available so far. For example, if the partition is based on the finite cardinality of the extension of H (as in our first example), then one listing of the cells starts off with the empty extension, proceeds to the extension with one element, then two, etc. And this listing works, since guessing the first cell that is consistent with the basic formulas seen at a given stage guarantees convergence to the right answer.

If a partition has uncountably many cells, then we cannot list them in the simple way envisioned (as an ω-ordering). Hence, induction by enumeration will always fail in such a case. But this is no criticism of our method, since it is easy to see that no uncountable partition can be solved by any scientist (the range of a scientist is countable inasmuch as its domain is).

Our scheme for induction by enumeration is nonetheless a non-starter. It fails to solve the one-sided split, which (it will be recalled) consists of (a) the one-sided orders with a least point, and (b) the one-sided orders with a greatest point. This partition is solvable despite the fact that the finite amount of data available at any stage of the game rules out neither of the cells. (At each stage, all the scientist sees is a set of ordered pairs that is consistent with the existence of a least element and also with the existence of a greatest one.) Any listing of the two cells will therefore leave the scientist stranded on the first cell, unable to converge to the right answer if the data are drawn from a structure in the second. So the idea of induction by enumeration is not successfully implemented by listing the cells of the target partition.

A more successful implementation starts by enumerating a set of for-
formulas, tailored to solve a given partition $P$. At each stage of the game, the scientist seeks the first formula $A$ in the enumeration such that the structures from $P$ satisfying both $A$ and the current data are included in a unique cell of $P$. This cell is conjectured (if there is one; otherwise, no conjecture is made). Note that formulas contradicted by the data are passed over since they yield the empty collection of structures, which is not included in a unique cell. The one-sided split again serves as illustration. In the $n$th even position of our enumeration, we place the formula $\forall y \ R \ vn \ y$, which can be read “the $n$th variable names the least point.” In the $n$th odd position of our enumeration we place $\forall \times R \times vn$, which can be read “the $n$th variable names the greatest point.” At a given stage of the game, the available data will contradict some of these formulas. For example, variable $v_{0}$ might be shown not to name a least point because the data reveal some other point below it. Since the data are finite, however, they must be consistent with some formula in our enumeration. Moreover, the first uncontradicted formula plus the present data will imply exactly one of the two cells of the one-sided split. Sooner or later, the genuine least point (or greatest point) will be named in the data, and all earlier variables will be disqualified for this role. Our scheme will then begin its convergence to the correct cell. Hence, this version of induction by enumeration solves the one-sided split.

What else can be solved this way? That depends on the character of the formulas enumerated. In the illustration above, we enumerated universal formulas with one free variable. If free variables are barred from the enumeration, then the one-sided split cannot be solved, no matter what closed formulas are employed. On the other hand, increasing the quantifier complexity of the enumerated formulas can extend the power of induction by enumeration. For example, there is a two-cell partition (not first-order definable) that can be solved by listing just two $\forall \exists$ formulas, but remains unsolved in the face of any listing of universal formulas. Universal formulas are especially adapted to solving first-order definable partitions (like the one-sided split). If such a partition is solvable then it is solved by enumerating the set of all universal formulas in arbitrary order.

On non-first-order partitions, induction by enumeration may not work, no matter what formulas are enumerated. That is, there are solvable partitions that cannot be solved via the enumeration of any set of formulas, be they universal or of greater complexity. This fact invites the thought that more could be achieved without the restriction to $\omega$-orderings of formulas. Perhaps some partition can be solved only if our formulas are arranged like successive copies of the natural numbers, or in some more complicated way. Induction by enumeration is still well defined in such a case, provided only that the more complicated listing is a well ordering.

Our hopes are dashed, however, by the existence of solvable partitions
that defy any well-ordering of any set of formulas. No matter what formulas are chosen, and no matter how they are well-ordered, our scheme will not solve these partitions—even though other kinds of scientists do solve them. Perhaps we ought not be disappointed. Induction by enumeration relies on brute force. Isn’t it reassuring that no matter how we push and pull its parameters, there are solvable partitions that escape its competence? In any event, we now consider a more old-fashioned approach to inquiry, in which hypothesis modification can be viewed as rational change in belief.

3. Inquiry via Belief Revision. Our inductive game is the same as before. To review, Nature chooses a structure from one cell of a pre-established partition over a class of countable structures. She then chooses a surjective variable-assignment and creates an enumeration of the basic diagram of the chosen structure (using variables like added constants). The scientist proceeds through the enumeration, responding to each new formula (datum) with a conjectured cell of the partition. Success consists in converging to the cell from which Nature made her choice. In the transition from induction by enumeration to belief revision, the only change is the way scientists arrive at their conjectures.

Scientists based on belief revision have two working parts. The first is a background theory, which the scientist chooses prior to encountering data. Formally, the background theory is an arbitrary set of \( L \)-formulas. The second working part is a scheme for revising the background theory under the impact of data. The scheme is called a “revision function,” which we will denote by \( + \). It maps a given theory \( X \) and a given sequence \( S \) of basic formulas (data) into another theory \( X + S \). The latter theory consists of an appropriately chosen subset of \( X \) (the fragment that survives the confrontation with fact) along with the formulas appearing in \( S \) (the facts themselves).

To fill in the picture, imagine the scientist working his way along the enumeration that Nature has provided. At each stage, the scientist views an increasing sequence \( S \) of basic formulas. In response, the scientist consults his background theory \( X \) and his revision operator \( + \). He computes \( X + S \), then asks: “Does the class of structures that satisfy the revised theory \( X + S \) fit inside a unique cell of the starting partition?” If so, then this cell is the scientist’s conjecture. In the absence of such a cell, no conjecture is made. Note that the same background theory \( X \) is used at each stage of the game. Only the data change, through accumulation. Call scientists who operate in this way revision based.

It remains to describe the kind of revision operators that equip revision based scientists. The approach that follows reformulates Hansson (1994), which in turn generalizes Alchourron and Makinson (1985). A preliminary
definition is needed. Suppose we are given a background theory \( X \) and data \( S \). A formula \( A \) of \( X \) is called \( S \)-innocent if it is a member of no \( \subseteq \)-minimal subset of \( X \) contradicted by \( S \). To spell this out, call a subset \( B \) of \( X \) \( S \)-guilty iff (i) \( B \) contradicts \( S \) (that is, \( B \) implies the negation of the conjunction of the formulas in \( S \)), and (ii) no proper subset of \( B \) contradicts \( S \). Then \( A \in X \) is \( S \)-innocent iff it is a member of no \( S \)-guilty subset of \( X \). It seems clear that every \( S \)-innocent formula in \( X \) should survive \( X \)'s confrontation with \( S \). This is the principal condition that we impose on revision based scientists. Specifically, it is required that the theory \( X + S \), resulting from the collision between background theory \( X \) and data \( S \), have the form \( Y \cup \text{range}(S) \), where \( Y \) is a subset of \( X \) that is consistent with \( S \) and includes all of \( X \)'s \( S \)-innocent members. (The symbol \( \text{range}(S) \) denotes the set of formulas that appear in \( S \).) It seems natural to require the successor theory to include \( \text{range}(S) \) because the data offered by Nature in our game are trustworthy. The successor theory must also be internally consistent, and no formula outside of \( X \) and \( S \) may appear. Call any set of formulas having these properties a potential successor to \( X \) and \( S \).

For example, let \( X = \{ \forall x (Qx \rightarrow Rx), \forall x ((Rx \& Sx) \rightarrow Tx), Sv3, \exists x \neg Qx \} \), and \( S = (Qv3, \neg Tv3) \). Then some potential successors to \( X \) and \( S \) are as follows.

1. \( \{ \forall x (Qx \rightarrow Rx), \forall x ((Rx \& Sx) \rightarrow Tx), \exists x \neg Qx, Qv3, \neg Tv3 \} \)
2. \( \{ \forall x ((Rx \& Sx) \rightarrow Tx), Sv3, \exists x \neg Qx, Qv3, \neg Tv3 \} \)
3. \( \{ \forall x ((Rx \& Sx) \rightarrow Tx), \exists x \neg Qx, Qv3, \neg Tv3 \} \)
4. \( \{ \forall x (Qx \rightarrow Rx), \exists x \neg Qx, Qv3, \neg Tv3 \} \)

In contrast, the following theories are not potential successors to \( X \) and \( S \).

1. \( \{ \forall x (Qx \rightarrow Rx), \forall x ((Rx \& Sx) \rightarrow Tx), Sv3, \exists x \neg Qx, Qv3, \neg Tv3 \} \) [inconsistent]
2. \( \{ \forall x (Qx \rightarrow Rx), \forall x ((Rx \& Sx) \rightarrow Tx), Qv3, \neg Tv3 \} \) [missing \( S \)-innocent formula]
3. \( \{ \forall x (Qx \rightarrow Rx), \forall x ((Rx \& Sx) \rightarrow Tx), Sv3, \exists x \neg Qx, \neg Tv3 \} \) [missing data]
4. \( \{ \forall x (Qx \rightarrow Rx), \forall x ((Rx \& Sx) \rightarrow Tx), \exists x \neg Qx, Qv3, \neg Tv3, \exists x \neg Rx \} \) [extraneous formula]

Any mapping \( + \) of theories \( X \) and data \( S \) into potential successors to \( X \) and \( S \) counts as a revision function.

Since our conditions are weak, the class of revision functions is broad enough to include some pathological members. An attractive additional constraint is to require that the successor theory be a \( \subseteq \)-maximal choice among potential candidates. That is, the successor theory must not be properly included in any other potential successor. A revision function of this narrower kind is called maxchoice, and implements the conservative
policy of inflicting minimal change on existing theories. For a yet narrower kind of revision function, consider the set \( Y \) in the successor \( Y \cup \operatorname{range}(S) \) to \( X \) and \( S \). In addition to the other conditions on \( Y \), we can require it to be the first candidate in a preestablished total ordering of all sets of formulas (the ordering is thought of as coding a priori preferences over theories). Revision functions that meet this additional constraint are called \textit{stringent}. It can be proved that stringent revision functions exist, and that some of them are maxichoice. Stringent revision satisfies many criteria of rationality. It manifests transitive and connected preferences among theories, and revises beliefs so as to minimize change and maximize acceptability.

Now we can consider the inductive powers of revision based scientists. What partitions do they solve? As can be expected, the scientist's prospects for success depend on his background theory. For example, if the theory consists entirely of sentences (no free variables), then no revision function can be used to solve the one-sided split (consisting of the total orders with either greatest or least point, but not both). Background theories without free variables get no traction because they are unable to formulate hypotheses about which particular object (e.g., the one labeled by variable \( v5 \) in the data) is greatest or least. With this defect corrected, it is easy to specify a revision based scientist that solves the one-sided split.

Indeed, a stronger result has been demonstrated. Let \( T \) be an arbitrary set of sentences, and let \( P \) be any partition of the collection of \( T \)'s countable models. If \( P \) is solvable (by any kind of scientist) then there is a background theory \( X \) with the following properties.

1. \( X \) is a consistent extension of \( T \), and
2. every revision based scientist using \( X \) as background theory solves \( P \).

It is also possible to specify a single, stringent revision function \(+\) that is "universal" in the following sense. For every solvable partition of the countable models of a theory \( T \) there is a consistent extension \( X \) of \( T \) such that the revision based scientist based on \( X \) and \(+\) solves the partition. This result is comforting to the thesis that successful science can be carried out via a rational process of theory selection.

The situation does not look so rosy, however, once we introduce considerations of inductive efficiency. Let \( P \) be a solvable partition. Recall that after Nature chooses a structure from some cell of \( P \), she goes on to choose a variable-mapping for that structure, and a listing of all the basic formulas made true in her interpretation. The structure is thus associated with uncountably many potential streams of data. The scientist is required to succeed on all of them, eventually converging to the correct guess about the cell from which Nature drew her choice. On any given data stream for a structure in \( P \), the successful scientist reaches an earliest stage after which
all of his conjectures are accurate. Call this the convergence point for the scientist and the data stream (relative to $P$). Now consider another successful scientist whose convergence points never come later than those for the first scientist, and sometimes come earlier. Then this second scientist may be said to dominate the first, in the sense of being a strictly more efficient inductive agent for $P$. Dominance is only a partial order on the class of scientists that solve $P$. We may nonetheless consider a scientist to be efficient for $P$ (in a weak sense) if it solves $P$ and is dominated by no other scientist that solves $P$.

It can be shown that every solvable partition of the countable models of a theory $T$ is solved efficiently by a revision based scientist using a stringent revision function. So stringent revision is consistent with efficient induction. But stringency is no guarantee of efficiency. Indeed, there are stringent revision functions that cannot be used to efficiently solve certain solvable first-order definable partitions, no matter what background theory is employed. These revision based scientists are dominated by other scientists who solve the same partitions strictly faster. It seems irrational to employ a dominated method of hypothesis choice, at least for creatures with finite lives. So it is natural to seek some additional property of revision functions (beyond stringency) that is intuitively rational and guarantees efficiency. At present, we have little idea what this property might be.

Questions about the scope and efficiency of rational inquiry become more complex when attention is limited to subclasses of scientists with characteristics that we recognize in humans. For example, the function from data to hypotheses that is implemented by real scientists is likely to be computable. Also, real scientists do not carry along the same background theory throughout their careers, modifying it over and over with accumulating data. Rather, the theory tested against the data at stage $n+1$ is often the one emerging from the encounter with the data of stage $n$. These issues and others have begun to be addressed in the literature devoted to acceptance-based induction.

REFERENCES


