The Learning Power of Belief Revision

Kevin T. Kelly
Department of Philosophy
Carnegie Mellon University
kk3n@andrew.cmu.edu

Abstract

Belief revision theory aims to describe how one should change one's beliefs when they are contradicted by newly input information. The guiding principle of belief revision theory is to change one's prior beliefs as little as possible in order to maintain consistency with the new information. Learning theory focuses, instead, on *learning power*: the ability to arrive at *true* beliefs in a wide range of possible environments. The goal of this paper is to bridge the two approaches by providing a learning theoretic analysis of the learning power of belief revision methods proposed by Spohn, Boutilier, Darwiche and Pearl, and others. The results indicate that learning power depends sharply on details of the methods. Hence, learning power can provide a well-motivated constraint on the design and implementation of concrete belief revision methods.

1 Introduction

Intelligent systems act on the basis of fallible, general beliefs, such as "My car always stays where I put it" or "rough objects of a given shape and size have more air resistance than smooth ones". When general beliefs of this sort are refuted by new information, they must somehow be revised to accommodate it. Accordingly, there is increasing interdisciplinary interest in *Belief revision theory*, which aims to specify how to rationally revise refuted beliefs [1] [4] [10] [14] [32] [33] [5] [6] [36]. The official motivation for belief revision theory [7] is to minimize change or loss to one's prior belief state when new, conflicting information is incorporated.

Learning theoretic research has traditionally focused on a different consideration: *learning power*, or the ability to arrive at true beliefs in a wide range of possible environments (e.g., [29] [21] [27] [16]). Learning power is crucial for intelligent behavior since plans based upon false beliefs may fail to achieve their intended ends. Until recently, there has been little interaction between learning theory and belief revision theory, even though beliefs are revised through learning and the analogy between belief revision and scientific theory change has been recognized for over a decade [13] [34]. This paper bridges the gap between the two perspectives by analyzing the learning powers of concrete belief revision methods proposed by by Boutilier [1], Goldszmidt and Pearl [10], Darwiche and Pearl [4], and Spohn [37].[1]

Some proposed belief revision methods have maximal learning power, in the sense that they can solve every solvable learning problem, but these methods have been objected to by some belief revision theorists for altering the agent's epistemic state too much. Other methods enforcing more stringent notions of "minimal change" in the agent's initial epistemic state *restrict* learning power, in the sense that there are solvable learning problems that they cannot solve, regardless of how we attempt to optimize the agent's initial epistemic state to the nature of the learning problem at hand. This restrictiveness is manifested by a curious limitation that might be called *inductive amnesia*: the restrictive methods can predict the future — but only if they forget the past; and they can remember the past — but not if they predict the future. In other words, inductively amnestic agents who *remember* the past are doomed to repeat it!

In light of this tension between minimal epistemic change and learning power, it is interesting and relevant to isolate, for each proposed notion of minimal belief change, the "breaking point" at which the learning problem becomes sufficiently complex that the method forces inductive amnesia on the learning agent. The notion of learning problem complexity considered is the number of times the future may possibly "reverse" (e.g., the sequence 0000111000... reverses twice, once at position 4 and once at position 7). It turns out that the number of future reversals a belief revision method can handle depends rather sharply on particular details of the method. Hence, learning power can serve as an informative, well-motivated constraint on concretely proposed belief revision methods.

In this note, I first introduce the belief revision methods to be analyzed. Then I define a simple learning theoretic paradigm in which to study these methods. Next, I introduce a hierarchy of ever more complex (but nonetheless solvable) learning problems based on the number of possible future reversals in the data stream. Finally, I isolate, for each method under consideration, the least problem in the hierarchy on which it fails.

2 Iterated Belief Revision

Originally [13] [7], belief revision theory was formulated in terms of an operation $*$ that modifies a given, propositional belief state B in light of new information E to produce a revised belief state B':

$$B * E = B'.$$

[1] Earlier applications of learning theoretic analysis to belief revision theory have focused on the general axioms of belief revision theory rather than on the learning powers of concretely proposed belief revision methods [28] [27] [26] [21].

This formulation implies that the future updating behavior of an agent with belief state B cannot depend on the agent's updating history except insofar as that history is recorded in B. Some belief revision theorists [37] [1] [33] [10] [4] [5] [6] have responded with a weaker theory in which $*$ operates not on the agent's propositional belief state B, but on a more comprehensive *epistemic state* which in turn determines the belief state $B(\sigma)$:

$$\sigma * E = \sigma'.$$

The epistemic state may record some or all of the agent's updating history. The availability of such a record is of obvious importance in learning from experience, for it enables the learner to distinguish past observations from the consequences of refuted, past beliefs when new beliefs are being formulated.

Let W be the set of possible worlds. A *proposition* is a set of possible worlds. Informally, a belief state is a set of believed propositions, but we may identify such a state with the set of all worlds that satisfy each proposition in the set. Hence, a belief state B may be represented as a proposition.

It is usually assumed that epistemic state σ determines not only a belief state $B(\sigma)$, but also a total pre-ordering $\leq_\sigma$ on some subset $D(\sigma)$ of W. The ordering $\leq_\sigma$ is called the *implausibility* ordering induced by σ. Define $\min_\sigma(E)$ to be the set of all $\leq_\sigma$-minimal elements of $E \cap D(\sigma)$ (i.e., the most plausible states satisfying E). It is then assumed that the belief state is the set of most plausible worlds in the ordering:

$$B(\sigma) = \min_\sigma(W).$$

Most belief revision theorists agree that the updated belief state $B(\sigma * E)$ that arises when σ is updated with new information E consistent with $D(\sigma)$ should be:[2]

$$B(\sigma * E) = \min_\sigma(E).$$

There is less agreement about how to update the rest of the epistemic state. For example, Boutilier's [1] *natural* method $*_M$ performs the minimal modification of $\leq_\sigma$ consistent with the above procedure for calculating $B(\sigma * E)$. That is, worlds in $\min_\sigma(E)$ are brought to the bottom of the revised order, leaving the ordering over all other worlds unaffected. Formally, if E is consistent with $D(\sigma)$ and $\sigma' = \sigma *_M$ then

$$w \leq_{\sigma'} w' \Leftrightarrow (w \in \min_\sigma(E)) \vee (w \leq_\sigma w').$$

Another method $*_L$, presented by Spohn [37] and generalized by Nayak [33], rigidly slides all E-worlds below all non$E$-worlds.[3] In other words, the revised ranking is as much as possible like its predecessor subject to the constraint that no refuted world is

[2] Adam Grove [12] showed that any method satisfying the axioms presented in [7] can be so represented. It should be mentioned, however, that an alternative approach to revision called *updating* [15] proceeds differently.

[3] Steven Glaister [8] has recently produced a symmetry argument in favor of Nayak's rule.

more plausible than a non-refuted world. So if E is consistent with $D(\sigma)$ and $\sigma' = \sigma *_L$, then we have:

$$w \leq_{\sigma'} w' \Leftrightarrow (w \in E \wedge w' \notin E) \vee ((w \in E \Leftrightarrow w' \in E) \wedge w \leq_\sigma w').$$

Other rules are possible if one adds structure to the agent's epistemic state. For example, Spohn [37] recommends modelling the epistemic state as a (possibly partial) mapping r from possible worlds to ordinal-valued *degrees* of implausibility. If E is consistent with the domain of r, define

$$r_{\text{min}}(E) = \min\{r(w') \in \text{dom}(r) \cap E\}$$

and

$$r(w|E) = -r_{\text{min}}(E) + r(w).$$

Think of $r_{\text{min}}(E)$ as the "height" of the most plausible member of E and think of $r(w|E)$ as the "height" of w above the most plausible member of E. Then $r(w|E)$ is referred to as the *conditional implausibility* of w given E.

Using these concepts, one may define Spohn's [37] qualitative generalization of Jeffrey conditioning $*_{J,\alpha}$, which has since been explored in [10] [4]. This method lowers all E worlds until the most plausible of them are assigned implausibility degree 0, and then moves the non-E worlds until the most plausible of them is assigned implausibility α.

$$(r *_{J,\alpha} E)(w) = \begin{cases} r(w|E) & \text{if } w \in \text{ dom}(r) \cap E \\ r(w|W - E) + \alpha & \text{if } w \in \text{ dom}(r) - E \\ \uparrow & \text{otherwise.} \end{cases}$$

Darwiche and Pearl [4] have recently proposed an interesting modification $*_{R,\alpha}$ of $*_{S,\alpha}$. This rule rigidly boosts the non-E worlds up from where they currently are by a fixed ordinal α. I refer to this as the *ratchet* method, since refuted worlds always move upward by a fixed amount without backsliding.

$$(r *_{R,\alpha} E)(w) = \begin{cases} r(w|E) & \text{if } w \in \text{ dom}(r) \cap E \\ r(w) + \alpha & \text{if } w \in \text{ dom}(r) - E \\ \uparrow & \text{otherwise.} \end{cases}$$

Goldszmidt and Pearl [10] mention the procedure of boosting all non-E worlds to the fixed ordinal ω. More generally, one might consider sending all refuted worlds to a fixed ordinal α.

$$(r *_{A,\alpha} E)(w) = \begin{cases} r(w|E) & \text{if } w \in \text{ min}_{\leq_r}(E) \\ \alpha & \text{if } w \in \text{ dom}(r) - E \\ \uparrow & \text{otherwise.} \end{cases}$$

One may also reformulate the "natural" method $*_M$ in terms of ordinals:

$$(r *_{M'} E)(w) = \begin{cases} 0 & \text{if } w \in E \cap B(r(.|E)) \\ r(w) + 1 & \text{if } w \in \text{ dom}(r) - (E \cap B(r(.|E))) \\ \uparrow & \text{otherwise.} \end{cases}$$

These rules differ primarily in how much of a "boost" they apply to refuted possible worlds. The $*_M$ rule boosts refuted worlds by one step, along with many nonrefuted worlds. The lexicographic $*_L$ rule provides an infinite boost. $*_{J,\alpha}$ may provide a negative boost if α is lower than the most plausible refuted world. $*_{R,\alpha}$ sends refuted worlds up by a fixed increment α. Each time a refuted world w is inserted beneath a non-refuted world w' satisfying the inputs already received, there is some concern that w' is actually the true state, but w drops back to the bottom of the ranking before w' does. If that happens, the agent "forgets" the past data refuting w. This study concerns learning problems on which some of the rules cannot help but forget if they eventually predict the future.

3 A Simple Learning Paradigm

Many people believe that a rough ball will have more air drag than a smooth one. Suppose we invite such a subject to consider the results of an experiment. We mount both balls in a wind tunnel and measure the drag on each. We start the tunnel at a small wind velocity and raise the velocity incrementally. After each increment, we report a 0 to the subject if the drag on the smooth ball is no greater than that on the rough ball and we report a 1 otherwise. The experiment is run. The subject is smug as several 0s are presented, consistently with her current belief. But to her utter surprise, the sequence reverses and 1s continue for some time [25]. Thereafter, the sequence flips back to 0s and yields 0s thereafter. The sequential outcomes of the aerodynamical experiment look something like this:

$$e = (0000001111000\ldots).$$

Empirical outcomes of this kind are commonplace in nonlinear systems.[4]

Generalizing this example, let an *outcome stream* be an infinite sequence e of natural numbers encoding experimental outcomes, and let U be the set of all such outcome streams. An *empirical proposition* is a proposition whose truth depends only on the outcome stream. In particular, the proposition that says outcome b will be observed at position n in the outcome stream is:

$$[n, b] = \{e' \in U : e'(n) = b\}.$$

The *data stream* generated by e is therefore the sequence of propositions

$$[e] = ([0, e(0)], [1, e(1)], \ldots).$$

[4] The reversal is due to the fact that the flow past the rough ball becomes turbulent at a lower velocity than that around the smooth ball and the turbulence stays attached to the surface longer, producing a smaller wake with fewer eddies. The sequence of 1s appears when the rough ball's wake is so reduced but the smooth ball's wake is still large. After the smooth ball's flow also becomes turbulent, the rough ball's extra skin friction results in higher drag, so the sequence has 0s forever after.

The initial segment of e at stage k is just

$$e|k = (e(0), e(1), \ldots, e(k-1)),$$

and the corresponding sequence of data received by stage k is

$$[e|k] = ([0, e(0)], [1, e(1)], \ldots, [k-1, e(k-1)]).$$

Let $*$ be an epistemic revision operator and let σ be an epistemic state of the appropriate type. A *revision agent* is then a pair $(\sigma, *)$, which may be viewed as starting out in the *a priori* epistemic state σ and then successively updating on the propositions input thereafter. Suppose that the revision agent aims to acquire complete, true beliefs about the outcome stream e, not merely by accident, but by following a method guaranteed to lead to complete knowledge over a wide range $P \subseteq U$ of possible outcome streams. More precisely, we may say that $(\sigma, *)$ *identifies* P just in case for each $e \in P$ there is a stage k such that for each subsequent $k' \geq k, (\sigma * [e|k']) = \{e\}$. This is an adaptation of E. M. Gold's [9] concept of identification in the limit to the present situation. We will leniently count an operator as successful if there exists at least one initial epistemic state that enables it to identify P: $*$ *can identify* P just in case there exists a σ such that $(\sigma, *)$ identifies P. P is *identifiable* just in case some $(\sigma, *)$ identifies P. Then say that $*$ is *complete* if it can identify every identifiable P. Else, $*$ *restricts* learning power. The principal concern of this study is to determine whether the various iterated belief revision methods described above are complete or restrictive.

4 Learning Tail Reversals

Define the *tail reversal* operation on Boolean outcome streams as follows, where $\neg$ denotes bit reversal:[5]

$$(e' \ddagger k)(n) = \begin{cases} e'(n) & \text{if } n \leq k \\ \neg(e'(n)) & \text{otherwise.} \end{cases}$$

Then we may represent the aerodynamical outcome sequence

$$e = (0000001111000\ldots)$$

as the result of reversing the tail of the everywhere 0 sequence z in two positions:

$$e = (z \ddagger 6) \ddagger 11.$$

Since tail reversals commute and associate under composition, one may write without confusion

$$e = z \ddagger \{6, 11\}.$$

[5] This construction is similar to Nelson Goodman's construction of the "grue" predicate [11].

Define for each $n < \omega$,

$$G^n(z) = \{z \ddagger S : |S| \leq n\}.$$

Then let

$$G^\omega(z) = \bigcup_{i<\omega} G^i(z).$$

In other words, the outcome streams in $G^n(z)$ are precisely those that reverse the tail of z in at most n distinct positions and the members of $G^\omega(z)$ are the outcome streams that eventually stabilize to 0 or to 1. So the e from the aerodynamical example is in $G^2(z)$.[6] The sequence of problems $G^0(z), G^1(z), G^2(z), \ldots$ may be thought of as a crude scale of learning power, according to which stronger methods can identify problems of higher complexity.

Suppose one wishes to identify $G^\beta(z)$, where $\beta \leq \omega$. A sensible procedure would be to believe at each stage that the observed tail reversals are the only ones that will ever be observed. This procedure follows Popper's suggestion to accept the simplest hypothesis consistent with the data. This simple, Popperian procedure has three characteristic virtues with respect to the learning problem $G^\beta(z)$, where $\beta \leq \omega$:

1. it identifies $G^\beta(z)$;

2. no other method identifying $G^\beta(z)$ weakly dominates it in convergence time (i.e., finds the truth as fast on each outcome stream in $G^\beta(z)$ and faster on some such outcome stream.)

3. no other method identifying $G^\beta(z)$ has a lower worst-case bound on the number of retractions performed prior to convergence.

I do not insist that every rational agent *must* use this procedure, or even that it is a good solution to each learning problem (it isn't). It suffices that the method is trivial to compute and works well for problems of this particular sort. Hence, it would seem that any proposed method of rational belief change that *cannot* duplicate this behavior— or even identify $G^n(z)$ in the limit— is deficient, the extent of the deficiency rising as n decreases.

[6] Game theorists may find the following sort of interpretation more suggestive. Fred and Jane are engaged in an indefinitely repeated prisoner's dilemma. Fred fully believes that Jane is a patsy who will cooperate no matter how often he defects. But maybe Jane simply has a veneer of civility that affords a fixed "grace period" of unconditional cooperation to new acquaintances to encourage good behavior, and punishes defections for eternity once this grace period is over (one tail reversal). Or maybe she punishes the first infraction after the grace period for a fixed time and then returns to being a patsy forever (two tail reversals). Or maybe she punishes the first infraction after the grace period for a fixed time, offers a new grace period, and punishes the next infraction for eternity (three tail reversals), etc.

5 Negative Results

Proofs of the following results are presented in [24]. Recall that if $*$ cannot identify P, then there exists no possible initial state σ such that $(\sigma, *)$ identifies P. Since an epistemic state can be an arbitrary assignment of ordinals to infinitely many infinite outcome streams, negative results in this setting are very strong. The first result concerns belief revision methods that cannot even identify the extremely simple learning problem $G^1(z)$. In other words, they are incapable of first believing in z and then upon seeing a tail reversal at position k, believing in $z \ddagger k$ therefter. This limitation is quite remarkable, given the simplicity of the learning problem.

Proposition 1 $*_{M'}, *_{J,1}, *_{A,1}$ *cannot identify* $G^1(z)$.

Moving to the case of at most two tail reversals (as in the aerodynamical example), we have

Proposition 2 $*_M, *_{A,2}, *_{R,1}$ *cannot identify* $G^2(z)$.

Darwiche and Pearl's ratchet method $*_{R,1}$ fails one level higher than the Jeffrey conditioning method $*_{J,1}$. This illustrates how learning theoretic analysis can provide sharp recommendations about subtle variations in belief revision architecture. Proceeding still higher,

Proposition 3 *for all* $n > 0, *_{A,n}$ *cannot identify* $G^n(z)$.

Say that $(r, *)$ *predictively identifies* P just in case for each $e \in P$, for all but finitely many n,

$$\emptyset \neq B(r * [e|n]) \subseteq \bigcap_{i=n}^{\infty} [i, e(i)].$$

In other words, the method is guaranteed to eventually produce only belief states that correctly predict the future (but that may fail to entail all the past observations). Say that $(r, *)$ *remembers the past* in P just in case for each $e \in P$ for each n,

$$\emptyset \neq B(r * [e|n]) \subseteq \bigcap_{i=0}^{n-1} [i, e(i)].$$

It turns out that each of the methods under consideration can predictively identify $G^\omega(z)$ (if it is not required to remember the past) and each of the methods can be made to remember the past (if it is not required to predictively identify $G^\omega(z)$).[7] Hence, each of the above, negative results is an instance of inductive amnesia.

The negative arguments are established by diagonal arguments that depend on the possibility of an *odd* number of tail reversals in the outcome stream, suggesting that some of the methods may perform better if we restrict attention to outcome streams involving

[7] Start out with the epistemic state that puts all possible worlds at the bottom level.

only even numbers of reversals. Accordingly, define the *even* tail reversal hierarchy of learning problems as follows:

$$G^n_{\text{even}}(z) = \{z \ddagger S : |S| \leq 2n \wedge |S| \text{is even}\}.$$

$$G^\omega_{\text{even}}(z) = \bigcup_{i<\omega} G^i_{\text{even}}(z).$$

The negative results concerning $*_{R,1}, *_{J,1}$ are thereby overturned, but the rest of the negative results continue to hold:

Proposition 4 *Except for those concerning $*_{R,1}, *_{J,1}$, all of the preceding negative results continue to hold when $G^n(z)$ is replaced with $G^n_{even}(z)$*

6 Positive Results

If the results were all strongly negative, they would provide little guidance for belief revision theory. The good news is that many of the methods introduced in the belief revision literature are much more powerful than one might expect. It is not too surprising that methods sending refuted possibilities to infinity are reliable, since such methods can implement the obvious, Popperian procedure (described above) of enumerating possible outcome streams and believing the first in the enumeration that is consistent with the data so far:

Proposition 5 *$*_L$, $*_{A,\omega}$, $*_{J,\omega}$, $*_{R,\omega}$ are complete identification strategies, and hence can identify $G^\omega(z)$.*

The more difficult and interesting question is whether strong learning performance can be achieved even when α is set low, so that refuted possibilities get interleaved with nonrefuted ones. This raises the possibility that refuted possibilities will work their way back down to the bottom of the ranking, in which case the fact that the possibility was refuted will have been forgotten. Clearly, this will happen for some unfortunate selections of the initial epistemic state σ. The question is whether there exist specially structured epistemic states that guarantee sufficient learning power when α is low.

Memory is facilitated by *compressed* epistemic states. In the extreme case, an epistemic state that puts all possibilities at the bottom level makes any of the above methods into a Bayesian *tabula rasa* that always believes exactly what it has observed (so long as the data are all mutually consistent with the range of the initial epistemic state). Prediction is facilitated by *rarified* epistemic states. On this side, the extreme case is an epistemic state that places a unique world at each level (i.e. an enumeration of complete hypotheses). All of the above methods can predict $G^\omega(z)$ when started with such a state, but they may forget when refuted possibilities are "boosted" to a level lower than the truth. The challenge in obtaining the positive results is to strike a happy balance between these two extremes, so that strong inductive leaps are performed without risking memory loss. Since there are infinitely many infinite data streams to juggle, programming the initial epistemic state to optimize learning power proves to be an interesting problem. The surprising result is that the value $\alpha = 2$ suffices for $*_{J,\alpha}$ and $*_{R,\alpha}$ to identify $G^{\omega}(z)$.

Proposition 6 *$*_{J,2}, *_{R,2}$ can identify $G^{\omega}(z)$.*

In fact, $*_{R,2}$ can be shown to be a complete identification strategy (the question is currently open for $*_{J,2}$).

The positive argument for $*_{J,2}$ employs an initial state σ on which the behavior of the method $(\sigma, *_{J,2})$ is exactly that of the simple, Popperian method discussed above. That method is trivially computable in constant time even though naively simulating the rule's definition requires manipulation of infinite structures. This illustrates how even the revision of infinite epistemic states can be computationally trivial if the initial epistemic state is carefully selected.

Moving to the even tail reversal hierarchy,

Proposition 7

1. *all of the preceding positive results continue to hold when $G^n(z)$ is replaced with $G^n_{even}(z)$.*
2. *Furthermore, $*_{J,1}$, $*_{R,1}$ can identify $G^{\omega}_{even}(z)$.*

The initial epistemic state σ that makes $*_{J,1}$ and $*_{R,1}$ succeed on this problem has the following structure: $\sigma(z) = 0$ and for each data stream $e \in G^{\omega}_{even}(z)$ (i.e., each finite variant of z), $\sigma(e) =$ the number of distinct positions n such that $z(n) \neq e(n)$. The evolution of the method $*_{J,1}$ initialized on this initial epistemic state may be represented as the rigid rotation of a k-dimensional cube from vertex to vertex until the vertex labeled with the true data stream is rotated into the bottom-most position, so the rule $*_{J,1}$ may be viewed as enforcing a strong symmetry condition on the evolution of σ. It can also be shown that the result of adding just the everywhere one data stream $\neg(z)$ at any level in σ interrupts this rotation and causes the cube to lie down on an edge forever after, so the method fails to converge to the complete truth.

Further positive results include the following:

Proposition 8

1. *$*_M$ can identify $G^1_{even}(z)$ and $G^1(z)$.*
2. *$*_{A,n}$ can identify $G^n_{even}(z)$ and $G^n(z)$.*
3. *Each of the methods discussed can identify $\{z\} = G^0_{even}(z) = G^0(z)$.*

The first of these results indicates that the learning power of Boutilier's method is enhanced when non-well-ordered epistemic states are entertained, illustrating the possibility that structural constraints on epistemic states may restrict learning power. The second result matches the corresponding negative result for $*_{A,n}$. The third result is trivial: initialize any of the methods discussed with an epistemic state whose range contains only z. Now the reported positive results meet up with the negative results, providing a complete picture of the relative powers of the methods under discussion over problems of the forms $G^{\beta}(z)$ and $G^{\beta}_{\text{even}}(z)$ where $\beta \leq \omega$.

7 Discussion

If the various methods of iterated belief revision are thought of as notions of "minimal change" in the given epistemic state subject to the requirement of consistently incorporating new information then the results just reviewed indicate how stricter notions of "minimal change" (i.e., coherence) can compromise learning power (i.e., reliability). Methods $*_{M'}, *_{A,1}, *_{J,1}$ fail at level 1 of the tail reversal hierarchy. Methods $*_{M}, *_{A,2}, *_{R,1}$ fail at level 2. Method $*_{A,n}$ fails at level n. But $*_{L}$, $*_{A,\omega}$, $*_{J,2}$, $*_{R,2}$ succeed at each level.

These results also illustrate how learning theoretic analysis can yield strong recommendations in the design of concrete belief revision methods. Consider the question of setting the "boost parameter" in the methods $*_{S,\alpha}$ and $*_{R,\alpha}$. Prima facie, it seems a matter of small consequence whether $\alpha = 1$ or $\alpha = 2$, but this tiny cost in minimizing epistemic change is occasioned by an *infinite leap* in learning power.[8]

Also, $*_{M'}$ cannot identify $G^{1}(z)$, so the assumption that implausibility degrees are well-ordered weakens the learning power of Boutilier's "natural" method. This illustrates how learning theoretic analysis can provide a critique of structural assumptions about the nature of the epistemic state relative to other structural assumptions about how revision should proceed.

The results also illustrate how the analysis of belief revision methods can enrich learning theoretic analysis. Themes such as inductive amnesia, the duality between prediction and memory, detailed algebraic analysis of the hypothesis enumeration, and the importance of even vs. odd numbers of tail reversals do not arise naturally in learning theory alone. They arise only in light of the particular methodological constraints proposed by belief revision theorists.

Many questions remain. Is $*_{J,2}$ complete? What about problems that are not subsets of $G^{\omega}(z)$? How about problems that do not require identification of empirically complete theories? Or paradigms in which information about successive outcomes can arrive in any order (the negative arguments would be unaffected)? Is it possible to solve for the set of all initial states for which a method $*$ solves a given problem (this would correspond to a kind of "transcendental deduction" of belief revision theory)? These questions are both interesting and nontrivial.

[8] Those who view α as an indication of the epistemic force of the data themselves may view $\alpha = 2$ a critical level of epistemic force at which learning power experiences an infinite increase.

References

[1] Craig Boutilier (1993) "Revision Sequences and Nested Conditionals", in *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, pp. 519-525.

[2] Horacio Arlo-Costa (1997) "Belief Revision Conditionals: Basic Iterated Systems", *Annals of Pure and Applied Logic*, in press.

[3] Cristina Bicchieri (1988) "Strategic Behavior and Counterfactuals", *Synthese* 76: 135-169.

[4] Adnan Darwiche and Judea Pearl (1997) "On the Logic of Iterated Belief Revision" *Artificial Intelligence* 89:1-29.

[5] Nir Friedman and Joseph Halpern (1995) "A Knowledge-Based Framework for Belief Change. Part I: Foundations", in *Proceedings of the Theoretical Apsects of Reasoning about Knowledge* R. Fagin, ed.

[6] Nir Friedman and Joseph Halpern (1995) "A Knowledge-Based Framework for Belief Change. Part II: Principles of Knowledge Representation and Reasoning", in *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference* J. Doyle, E. Sandewall, and D. Porasso, eds., San Francisco: Morgan Kaufmann. eds.

[7] Peter Gärdenfors (1988) *Knowledge in Flux*, Cambridge: M.I.T. Press.

[8] Steven Murray Glaister (1997) "Symmetry and Belief Revision", unpublished manuscript.

[9] E. Mark Gold (1967) "Language Identification in the Limit", *Information and Control* 10:302-320.

[10] Moises Goldszmidt and Judea Pearl (1994) "Qualitative Probabilities for Default Reasoning, Belief Revision, and Causal Modeling", *Artificial Intelligence* 84:57-112.

[11] Nelson Goodman (1983) *Fact, Fiction and Forecast*, Cambridge: Harvard University Press.

[12] Adam Grove (1988) "Two Modellings for Theory Change", *Journal of Philosohical Logic* 17: 157-170.

[13] William Harper (1978) "Conceptual Change, Incommensurability and Special Relativity Kinematics", *Acta Philosophica Fennica* 30: 430-459.

[14] Hirofumi Katsuno and Alberto O. Mendelzon (1991) "Propositional knowledge base revision and minimal change", *Journal of Artificial Intelligence* 52: 263-294.

[15] Hirofumi Katsuno and Alberto O. Mendelzon (1991) "On the difference between updating a knowledge base and revising it", in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pp. 387-394.

[16] Michael J. Kearns and Umesh V. Vazirani (1994) *An introduction to Computational Learning Theory*, Cambridge: M.I.T. Press.

[17] Kevin T. Kelly (1988) "Theory Discovery and the Hypothesis Language", *Proceedins of the 5th International Conference on Machine Learning, ed. J. Laird. San Mateo: Morgan Kaufmann.*

[18] (1989) Kevin T. Kelly "Induction from the General to the More General", in *Proceedings of the 2nd Annual Workshop on Computational Learning Theory*, ed. R. Rivest, D. Haussler, and M. Warmuth. San Mateo: Morgan Kaufmann.

[19] (1989) Kevin T. Kelly and Clark Glymour "Convergence to the Truth and Nothing but the Truth", *Philosophy of Science* 56: 185-220.

[20] (1990) Kevin T. Kelly and Clark Glymour "Theory Discovery from Data with Mixed Quantifiers", *Journal of Philosophical Logic* 19:1-33.

[21] Kevin T. Kelly (1996) *The Logic of Reliable Inquiry.* New York: Oxford University Press.

[22] Kevin T. Kelly and Oliver Schulte (1995) "The Computable Testability of Theories making Uncomputable Predictions", *Erkenntnis* 43: 29-66.

[23] Kevin T. Kelly, Oliver Schulte, and Vincent Hendricks (1996) "Reliable Belief Revision", in *Proceedings of the 10th International Congress of Logic, Methodology, and Philosophy of Science*, ed. Maria Luisa Dalla Chiara.

[24] Kevin T. Kelly (1998) "Iterated Belief Revision, Reliability, and Inductive Amnesia", Carnegie Mellon Philosophy Department Tech Report CMU Phil-88.

[25] C. A. Marchaj (1991) *Aero-hydrodynamics of Sailing*, Camden: International Marine Press.

[26] Eric Martin and Daniel Osherson (1995). "Scientific Discovery via Rational Hypothesis Revision", in *Proceedings of the 10th International Congress of Logic, Methodology, and Philosophy of Science*, ed. Maria Luisa Dalla Chiara.

[27] Eric Martin and Daniel Osherson (1996). "Scientific Discovery Based on Belief Revision", *Journal of Symbolic Logic*, in press.

[28] Eric Martin and Daniel Osherson (1996). *Elements of Scientific Inquiry*, manuscript.

[29] Daniel Oshersion, Michael Stob, and Scott Weinstein (1986) *Systems that Learn.* Cambridge: MIT Press.

[30] Daniel Osherson and Scott Weinstein (1986) "Identification in the Limit of First Order Structures", *Journal of Philosophical Logic* 15: 55-81.

[31] Daniel Osherson and Scott Weinstein (1989) "Paradigms of Truth Detection", *Journal of Philosophical Logic* 18: 1-41.

[32] Hans Rott (1992) "Preferential Belief Change Using Generalized Epistemic Entrenchment", *Journal of Logic, Language and Information* 1: 45-72.

[33] Abhaya C. Nayak (1994) "Iterated Belief Change Based on Epistemic Entrenchment", *Erkenntnis* 41: 353–390.

[34] David Poole (1988) "A Logical Framework for Default Reasoning", *Artificial Intelligence* 36: 27-47.

[35] Karl R. Popper (1968), *The Logic of Scientific Discovery*, New York: Harper.

[36] Dov Samet (1996) "Hypothetical Knowledge and Games with Perfect Information", *Games and Economic Behavior* 17: 230-251.

[37] Wolfgang Spohn (1988) "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States", *Causation in Decision, Belief Change, and Statistics, II*, ed. Brian Skyrms and William L. Harper, Dordrecht: Kluwer.

[38] Wolfgang Spohn (1990) "A General Non-probabilisistic Theory of Inductive Reasoning", *Uncertainty in Artificial Intelligence* 4: 149-159.